TOWARDS A SYNTHESIS OF CONTINENTAL SHELF DYNAMICS AND EUKARYOTIC PHYTOPLANKTON EVOLUTION

by

MATTHEW JOHN OLIVER

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Oceanography

written under the direction of

Oscar Schofield and Paul Falkowski

and approved by

____

____

____

____

____

New Brunswick, New Jersey

July 2006

Abstract of the Dissertation

TOWARDS A SYNTHESIS OF CONTINENTAL SHELF DYNAMICS AND EUKARYOTIC PHYTOPLANKTON EVOLUTION

By MATTHEW JOHN OLIVER

Dissertation Directors:
Oscar Schofield and Paul Falkowski

In the modern ocean, eukaryotic phytoplankton are central to global carbon cycling and general food web structuring. Their rise to dominance began in the Mid-Triassic with the rifting of Pangea and the flooding of continental margins. They radiated on these margins, where weathered nutrients from the continents were abundant. Turbulence, frequency of nutrient pulses, nutrient composition and concentration have been implicated as selective agents driving phytoplankton evolution, however, it has been difficult to demonstrate the relative importance of these selective agents on modern continental shelves for two reasons; (i) the inherent physical forcing dynamic of a continental shelf is not well sampled using traditional oceanographic techniques, and (ii) incomplete understanding of how evolutionary mechanisms operate on similar timescales as the physical forcing dynamic.

Because of these limitations, an empirical synthesis has yet to be achieved between the evolutionary history of phytoplankton and what we observe in the contemporary ocean. The goal of this work is two-fold; (i) to objectively elucidate the mosaic pattern of continental shelves into their statistically significant water masses and (ii) illustrate how these water masses influence evolutionary processes in diatoms.

I demonstrate, for the first time, that optical and hydrographic based time series data from *in-situ* profilers and from multiple satellites can be objectively delineated into statistically meaningful water masses that are verified by changes in current structure, salinity and macronutrients on continental shelves. Furthermore, because this approach is time-resolved, I show that the highly structured continental shelf environment is characterized by punctuated change. I use this concept of punctuated change on continental shelves to test the hypothesis that phytoplankton genomes are capable of evolving in a punctuated manner through the expression of retrotransposons. I show that retrotransposon expression is induced through nutrient limitation, and represent a punctuated evolutionary event. Finally, I examine how processes such as retrotransposon activity affect the rate of genome evolution eukaryotic phytoplankton and other eukaryotic organisms.

Acknowledgements

I thank my advisors Oscar Schofield and Paul Falkowski; they always encouraged me to tackle big problems and pushed me to be scientifically diverse. I also thank my committee members Dmitri Petrov and Kay Bidle for their consistent guidance and insightful comments. I thank my fellow graduate students and lab members. I believe interacting with two different labs, The Coastal Ocean Observation Lab and the Environmental Biophysics and Molecular Ecology Lab was a great benefit to me and has shaped my thinking and the direction of my future research. I would like to thank fellow lab members Josh Kohut, Zoe Finkel, Jen Bosch, Mike Crowley, Tuo Shi, Alex Kahl, Sherrie Whittaker, Yael Helman, Dan Tchernov, Assaf Vardi, Daniel Grzebyk, SangHoon Lee, Kevin Wyman, Courtney Kohut, and Yana Zeltser. Andrew Irwin, David Ackerly, Mark Moline and Ted Garland provided very helpful discussion and technical resources. I would like to acknowledge NSF and ONR for financial support.

I would also like to thank Josh Kohut, Christian Andrews, Tom Truman and Stan Lu and all of Outreach Red Bank for encouraging words that helped me finish. Finally I would like to thank my family, especially my wife and daughter, Korin and Samantha for unwavering support and dedication throughout my many years in school.

Table of Contents

Abstract of the Dissertation iii

Acknowlegements ....**Error! Bookmark not defined.**

List of Figures viii

List of Tables xvii

1.0. Chapter 1. Introduction 1

2.0. Chapter 2. Inversion of spectral absorption in the optically complex coastal waters of the Mid-Atlantic Bight; development, validation and coherence to hydrography........ 9

2.1. Abstract 9

2.2. Introduction 9

2.3. Methods 12

2.4. Results and Discussion 19

2.5. Conclusions 34

3.0. Chapter 3. Bioinformatic Approaches for Objective Detection of Water Masses on Continental Shelves 35

3.1. Abstract 35

3.2. Introduction 36

3.3. Methods 39

3.4. Results 52

3.5. Discussion 63

3.6. Conclusions 73

4.0. Chapter 4. Density and Nitrogen Related Patterns of Copia-Like Retrotransposon Transcription in the Diatom *Phaeodactylum tricornutum* 74

4.1.Abstract 74

4.2. Introduction 74

4.3. Methods 78

4.4. Results 81

4.5. Discussion 86

5.0 Chapter 5. The Mode and Tempo of Genome Size Evolution in Eukaryotes 90

5.1. Abstract 90

5.2. Introduction 90

5.3. Results and Discussion 92

5.4. Methods 105

6.0 Summary and Conclusions 108

Appendix I. 110

Appendix II 114

References 115

Curriculum Vita 136

List of Figures

Figure 2.1. Input spectra used to invert the *in situ* absorption values measured by the ac-9 using the OSI model. Phytoplankton spectra are averages of high-light and low-light adapted phytoplankton from Johnsen et al., 1994. Phytoplankton group one represents chlorophyll$_{a\text{-}c}$ containing classes of Bacillariophyceae, Dinophyceae and Prymnesiophyceae. Phytoplankton group two represents the phycobilin containing class Cryptophyceae. Phytoplankton group 3 represents the chlorophyll$_{a\text{-}b}$ containing classes of Chlorophyceae, Prasinophyceae and Eugelnophyceae. CDOM and detritus spectra are idealized exponential functions……………………………………………………………..17

Figure 2.2. The relationship between chlorophyll *a* fluorescence measured with a HOBI Labs hydroscat-6 and the estimated phytoplankton weight during the summers of 2000 and 2001. The phytoplankton weight was inverted from ac-9 data using the OSI$_c$ approach. The R^2 for summer 2000 and 2001 are 0.54 and 0.61, respectively…………21

Figure 2.3. Comparison of measured and (A) modeled particulate, (B) phytoplankton, (C) detritus, and (D) CDOM absorption using the minimal constraint OSI method. The measured data represents absorption spectra made for either filter pads or dissolved organics on discrete samples. The modeled data represents the predicted absorption spectra from the inverted from the ac-9 data. Data were pooled for both years and was linearly regressed at each wavelength providing both the slope (dark line with gray area) and R^2 (dark circles) for each wavelength. The gray shadow around the slope represents the standard deviation……………………………………………………………………23

Figure 2.4. Comparison of measured and modeled (A) particulate, (B) phytoplankton, (C) detritus, and (D) CDOM absorption using the OSI_c method. The measured data represents absorption spectra determined from filter pads or discrete samples for colored dissolved organics. Data were pooled for both years and were linearly regressed against measured absorption spectra at each wavelength providing both the slope (dark line with gray area) and R^2 (dark circles) for each wavelength. The gray shadow around the slope represents the standard deviation…………………………………………………………...24

Figure 2.5. Comparison of CDOM spectral absorption estimated by the OSI_c and measured with a flow through Breve-buster (Kirkpatrick et al. 2003) and on a discrete sample……………………………………………………………………………………27

Figure 2.6. Comparison of the amount of phytoplankton absorption and measured chlorophyll a associated with the three major spectral classes of phytoplankton during summers 2000 and 2001. The amount of chlorophyll a associated with each spectral class of phytoplankton was calculated via CHEMtax using the accessory pigment data measured via high performance liquid chromatography. A) Relationship between measured chlorophyll a to the OSI_m and OSI_c procedures. B) The absorption of chlorophyll c containing algae estimated with the OSI_m and the chlorophyll a associated with chromophytic algae. C) The absorption of phycobilin containing algae estimated with the OSI_m and the chlorophyll a associated with phycobilin containing algae. D) The absorption of chlorophyll c containing algae estimated with the OSI_c and the chlorophyll a associated with chromophytic algae. E) The absorption of phycobilin containing algae estimated with the OSI_c and the chlorophyll a associated with phycobilin containing algae. F) The absorption of chlorophyll b containing algae estimated with the OSI_c and the chlorophyll a associated with chlorophytic containing algae………………………..29

Figure 2.7. Time series of *in situ* data taken by the profilers during the experiment. Panel A shows density structure with water mass boundaries (white) defined by cluster analysis (see text). Panel B is the ratio of scattered and backward scattered light. Panel C is chlorophyll fluorescence measured by the optical profiler. Panel D is the OSI derived calibrated relative phytoplankton abundance. Optical and biological parameters have similar patterns as the hydrographic structure..…………………………………………31

Figure 2.8. Time series of inverted *in situ* absorption data taken by the optical profiler during the experiment. Panel A is the relative abundance of CDOM, while Panel B is the exponential slope of the CDOM curve. Panel C is the relative abundance of detritus while Panel D is the exponential slope of the detritus curve. Derived optical properties show distinct characteristics of the hydrographic structure during the experiment……………32

Figure 3.1. Temperature and reflectance maps on 7/14, 7/21, 7/31 and 8/02 2002 in this analysis. A warm core ring is evident on 8/02 as a nearshore optically dominated water mass formed nearshore. The white line is the coastline and the black indicates land or cloud……………………………………………………………………………………..40

Figure 3.2. Flow diagram of this analysis. This analysis assimilates sea surface temperature as well as two remote sensing channels for all four days. The data is standardized according to the mean and variance of the combined four day data set to make them comparable. Water types for each day are detected using four clustering algorithms, ACL, AWL, K-means and C-means. These results are combined into a Figure of Merit, where an average slope function (ASF) and threshold of acceptable flatness

(TAF) is computed. These two predictors give a range of reasonable water types. For each solution for each day, the boundaries are plotted and coincident boundaries are the most prevalent indicating similar structures found by different clustering algorithms. This indicates that the boundaries associated with this water type indicate a prevalent water mass. Finally, the predictor space distance is measured between each data point, to determine how different the water is on either side of each boundary. High values indicate a very strong boundary between water masses……………………………….42

Figure 3.3. Figure of Merit (*FOM*), Average Slope Function (*ASF*) and Threshold of Acceptable Flatness (*TAF*) calculation for each of the four days with the results of each of the clustering algorithms. A large *FOM* indicates that the variance within each cluster is comparatively large and that the cluster centroid is a generally poor predictor of the other data points within each cluster. A small *FOM* indicates that the cluster centroid better predicts the other members of its cluster, and that the variance with in the cluster is comparatively small. *ASF* is the average percent change of the four clustering algorithms compared to the maximum *FOM*. *TAF* was defined when the average change in the *FOM* was less than 1% for more than three clusters…………………………………………...48

Figure 3.4. Wind record from the RUMFS field station and Hudson River flow recorded at Waterford NY during the study time period. From 7/14 to 8/02 there were three upwelling favorable events that may have sustained phytoplankton growth near-shore. The elevated stream flow during this particular year recorded at Waterford NY may have initiated the formation of a Hudson River derived water mass during the four day study period. It has been reported that water outflow from this area takes 40 days to reach the Southern New Jersey shore (Yankovski and Garvine 1998)……………………………54

Figure 3.5. High frequency boundary locations as calculated from equation 5. The contrast indicates how often a particular pixel was designated as a boundary. The most frequent boundaries represent water types that are easily separable in predictor space. Boundaries become more distinct from 7/14 to 8/02……………………………………56

Figure 3.6. The boundary frequency calculated by equation 5 related to the total number of boundaries drawn. The days with more disorganized boundaries (7/14 and 7/21) have less low and high frequency boundaries and more medium frequency boundaries. This causes the disorganized look on these days and indicates that the clustering algorithms had a difficult time coming to similar solutions. Days 7/31 and 8/02 had more low frequency and high frequency boundaries and low medium frequency boundaries indicating that the clustering algorithms were in agreement more often and that water types were consistently distinguished……………………………………………………57

Figure 3.7. The gradient defined by equations 6, 7 and 8. The gradients are a relative measure of how different adjacent water masses are. Because no two adjacent pixels are equal, the gradient is never zero. The background gradient value for this study is approximately 10, with a standard deviation of approximately 10. Gradient values larger than 20 in this study are considered to be significant. Stronger gradients were evident in days 7/31 and 8/02. This indicates that the water types on either side of the boundary are markedly different. However, strong gradients are not necessarily coincident with high or medium frequency boundaries because two water types may be readily distinguishable in predictor space but still be relatively close to one another………………………………58

Figure 3.8. Boundary gradients overlaid with surface current fields with the surface current spatial mean subtracted for visual clarity. Areas with larger gradients are coincident with convergent and divergent areas indicating that local current structure accounts for the gradient locations. However, not all convergent areas had gradients associated with them…………………………………………………………………..60

Figure 3.9. Results of the simulated drifter experiment. The predicted location of 48 drifters on 8/02 based on the initial position of the 7/31 boundary by assimilating the CODAR measured surface currents generally approximates the location and shape of the boundary on 8/02. This indicates that the apparent movement of the boundary can be generally attributed to local advective processes. Also, this indicates that water masses in this area can be tracked effectively……………………………………………………61

Figure 3.10. Independent test of boundary analysis with salinity as an independent predictor during the LaTTE 2005 experiment. Black areas indicate areas of cloud contamination or land. Boundary analysis predicted a large bulge in the bight apex as well as a coastally trapped river plume along the New Jersey shore. Salinity transects by the R/V Hatteras confirm these boundaries…………………………………………...64

Figure 3.11. Independent test of boundary analysis with nitrate as an independent predictor during the LaTTE 2005 experiment. Black areas indicate areas of cloud contamination or land. This analysis indicates that nitrate levels in the "bulge" are very high while on the other side of the bulge boundary they are typical of open ocean values…………………………………………………………………………………65

Figure 3.12. Independent test of boundary analysis with silicate as an independent predictor during the LaTTE 2005 experiment. Black areas indicate areas of cloud contamination or land. This analysis indicates that nitrate levels in the "bulge" are very high while on the other side of the bulge boundary they are typical of open ocean values.

In addition, there are relatively high values of silicate present in an old river plume that is being advected offshore……………………………………………………………….66

Figure 4.1. Map of Copia Element in *P. tricorntnum*. The element is flanked by two 163 base pair regions and is in excess of 100 copies per genome……………………………77

Figure 4.2. Cell density, F_v/F_m, and F_m/cell for the control (A) and for treatment 1, 2, 3 (B, C, D respectively – Table 1). Grey areas indicate time points where cultures were exposed to a zero nitrogen condition……………………………………………………82

Figure 4.3. Increase in CLR transcription over the time course of the experiment. Shaded areas indicate when cells were exposed to a 0 nitrogen condition. Expression at the end of the experiment is significantly higher than at the beginning of the experiment………....84

Figure 4.4. Model II regression between fold change in CLR expression relative to histone (A) and relative to TATA (B) verses fold change in cell density. Blue dots are the control, pink dots are treatment 1, green dots are treatment 2 and gray dots are treatment 3. Regression lines were significant at the p<0.01 level and were not significantly different from each other between panel A and panel B……………………………...85

Figure 4.5. Fold change in CLR expression normalized to fold change in cell density Control expression was subtracted out and designated as a zero-line to show the change of CLR expression relative to the control after the effect of cell density is removed…...87

Figure 5.1. Maximum likelihood tree based on 18s sequences built using PHYML. Taxonomic groups highlighted in bold were analyzed for genome size evolution. Accession numbers of the 18 S rDNA sequences used in this analysis are given. Inset: Alternative eukaryotic tree based on 31 orthologs that was used to verify the general trend of genome size evolution inferred from the 18 S rDNA tree (Ciccarelli et al. 2006)……………………………………………………………………………..94

Figure 5.2. A) A tree-wise analysis of the nodal estimated genome size and the calculated contrast at each node from the 18 S rDNA tree (black dots) and the 31-ortholog tree (red dots). Estimations from both trees indicate that as genome size increases, the rate of evolution of genome size increases (shown on Log_{10} axes for plotting purposes). B) Distribution of the median absolute contrast and the median genome size of the 20 traditionally recognized taxonomic groups from the 18 S rDNA tree. Bars represent 95% bootstrapped confidence intervals. Again, a clear positive relationship between genome size and the rate of genome size evolution is evident (shown on Log_{10} axes for plotting purposes)…………………………………………97

Figure 5.3. A) Distribution of the absolute value of the standardized contrasts from the 18 S rDNA tree showing a strong deviation from the ½ normal distribution expected from a phenotypic trait under Brownian evolution. A strong deviation would be expected for a trait under proportional evolution. B) Quantile distribution of the absolute value of the standardized contrasts. These contrasts do not show a near linear relationship to the positive quantile standard deviates, indicating a strong deviation from a ½ normal distribution, which is expected for a trait under proportional evolution…………………98

Figure 5.4. A-priori Log_{10} transformation of genome size removes the proportional effect of genome size on the rate of genome size evolution so that neither, A) a tree-wise analysis of the nodal estimated genome size and the calculated contrast at each node, nor B), the distribution of the median absolute contrast and the median genome size of 20 traditionally recognized taxonomic groups show a significant correlation. Bars represent

95% bootstrapped confidence intervals. As in Figure 3A, red dots represent estimations from the 31ortholog tree and black dots represent estimations from the 18 S rDNA tree……………………………………………………………………………………..99

Figure 5.5. A) Distribution of the absolute value of the standardized contrasts calculated from Log_{10} transformed genome size and the 18 S rDNA tree. This calculation shows approximately a ½ normal distribution expected from a phenotypic trait under Brownian evolution. B) Quantile distribution of the absolute value of the standardized contrasts calculated from Log_{10} transformed genome size data. These contrasts show a near linear relationship to the positive quantile standard deviates, indicating the expected ½ normal distribution of the contrasts for a phenotypic trait under Brownian evolution…………100

Figure 5.6. Distribution of genome sizes used in this analysis in A) linear space and B) logarithmic space exhibit a log-normal distribution as predicted by proportional evolution integrated over long time periods……………………………………………………103

List of Tables

Table 3.1. Description of the four types of clustering algorithms used…………………44

Table 4.1. Nitrogen concentrations added to the F/2 growth media……………………78

Table 4.2. Optimized primer conditions…………………………………………………81

Table 5.1. Number of species in each group analyzed from the 18 S rDNA……………96

CHAPTER 1

1.0. Introduction

In the modern ocean, eukaryotic phytoplankton play a critical role in ocean biogeochemistry by participating in primary production, and exporting carbon to the deep sea through the biological pump (Dugdale and Goering 1967; Eppley and Peterson 1979), and the structuring of marine food webs (Ryther 1969; Azam et al. 1983). Given their importance, much work has focused on their geological (Katz et al. 2004), fossil (Finkel et al. 2005; Young et al. 2005), and molecular records (Kooistra and Medlin 1996; de Vargas et al. 2002) as well as their modern ecology and physiology (Falkowski and Raven 1997). Developing a synthetic view of their evolutionary history and their modern ecology is inherently difficult because it requires integrating processes operating on vastly different time and space scales. Nevertheless, great advances have been made with respect to this task *in-silico* (Tozzi et al. 2004; Fennel et al. 2005). Recent advances in ocean observational technology, genomics, and bioinformatics have opened the door to directly observe the evolution of eukaryotic phytoplankton (Rynearson and Armbrust 2000). Phytoplankton bloom in large numbers and have short generation times, thus are good candidates for directly observing evolutionary change. However, recent advances in ocean observational technology, genomics, and bioinformatics have opened the door to directly observe the evolution of eukaryotic phytoplankton. The task of directly observing the evolution of phytoplankton is also confounded by the inherent physical dynamics of the ocean. Phytoplankton populations respond quickly to a variety of episodic forcing events in the ocean environment, including aeolian dust storms, hurricanes, eddys, upwelling and river outflow. However, because phytoplankton populations are continually advected and mixed by these forcing events, tracking specific populations is exceedingly difficult. In the work presented below, I identify and address two particular issues related to the larger task of empirically synthesizing the evolutionary history and modern ecology of eukaryotic phytoplankton; i) the necessity to objectively determine the environmental structure of the marine environment (Chapters 2-3), ii) the necessity to identify observable agents of evolution that are modulated by the inherent structure of the marine environment and to demonstrate how they are integral to the overall evolutionary pattern of eukaryotic phytoplankton (Chapters 4-5).

1.1. The link between phytoplankton evolutionary history and the structure of the marine environment.

The necessity to objectively determine the environmental structure of eukaryotic phytoplankton emerges as a by-product of their evolutionary history. Eukaryotic phytoplankton are a polyphyletic photosynthetic group of organisms that, based on large subunit ribosomal sequence analyses, can be associated with at least five different deeply branching eukaryotic clades (Bhattacharya and Medlin 1995; Baldauf 2003; Falkowski et al. 2004). The broad scale phylogenetic diversity is also reflected in their physiology, nutrient quotas (Quigg et al. 2003), and life histories. However, despite their phylogenetic diversity, they are functionally united because they all require essentially the same inorganic nutrients for photosynthesis and for growth. In the ocean, understanding the distribution of phytoplankton species is inextricably linked to the distribution of these inorganic nutrients because they ultimately limit their growth (Eppley 1981; Falkowski and Raven 1997). Traditionally, ecologists have observed the various physiological characteristics of phytoplankton in a nutrient limiting environment and have considered the continual coexistence of radically different species competing for the same nutrient pool to be paradoxical (Hutchinson 1961), as the principle of competitive exclusion predicts that under equilibrium conditions in a relatively unstructured environment, only the species which is best adapted for nutrient uptake would remain (Hardin 1960). Of course, the efficacy of exclusion principle on phytoplankton populations is relevant only if the premise of an unstructured equilibrium environment is true. In the ocean, the turbulent cascade of energy down to millimeter scales structures the nutrient regimes in the ocean in such a way that it prohibits equilibrium conditions (Kolmogorov 1941; Denman and Gargett 1983). Therefore, the apparent paradox of the plankton is alleviated by demonstrating that the premise on which the exclusion principle operates does not accurately describe the ocean environment. This solution has been demonstrated in numerical models (Siegal 1998; Tozzi et al. 2004). It should be emphasized however, that while these models give formal solutions to the paradox of the plankton, they also clearly demonstrate that phytoplankton species coexistence and persist as the direct result of the inherent structure of the particular environment in which they are found. Therefore, understanding both phytoplankton ecology, as well as why evolutionary disparate groups of phytoplankton are able to coexist requires detailed information about the structure of their environment (Margalef 1961).

On geologic time scales, changes in the structure of the marine environment appears to have had major influences on the evolutionary tempo of the three major modern groups of marine eukaryotic phytoplankton (i.e. Diatoms, Coccolithophores and Dinoflagellates). In the late Triassic, the supercontinent Pangea began to rift as part of the opening of the most recent Wilson cycle (Wilson 1966). This break-up, combined with the rise in sea level flooded continental shelves and increased the total amount of coastline thus increasing the input of weathered inorganic nutrients from the continents to the newly flooded continental margins. It is during the Mesozoic when the fossil record shows that diatoms, coccolithophores and dinoflagellates radiate in parallel to the total area of flooded continental margins until the late Cretaceous suggesting that continental shelves act as engines of phytoplankton evolution (Falkowski et al. 2004). When sea level began to fall in the late Cretaceous, the diversity of coccolithophores and dinoflagellates also reduced; however, the diversity of diatoms underwent an unprecedented radiation. During this same time period, the latitudinal thermal gradient increased dramatically, thus increasing the total amount of wind-driven turbulent energy to the ocean and increased weathering of the continents resulting in a dynamic environment dominated by short, but intense nutrient pulses (Katz et al. 2004). It is in response to this highly dynamic, environment that diatoms diversify morphologically (Finkel et al. 2005) and rose to ecological and biogeochemical dominance in the phytoplankton that continues into the modern era.

While the exact oceanic environment in which diatoms radiated cannot be determined from the fossil record, diatoms have two characteristics that point to continental shelves. The first characteristic is a large intracellular storage vacuole that allows diatoms to uptake nutrients in excess of their immediate cellular requirement when nutrient concentrations are high, thus allowing them to survive on stored nutrients in a highly dynamic, non-equilibrium environment (Raven 1997; Tozzi et al. 2004). A second characteristic is their absolute requirement for silicic acids to construct their ornate frustules. Silicic acids can only be supplied by the weathering of continental rock, thus effectively tethering the majority of diatoms to regions of the ocean with significant riverine input (Falkowski et al. 2004). Highly dynamic, nutrient pulsed environments rich in silica are chief characteristics of continental shelves in mid- and high-latitudes. Therefore, it appears as if for all three major eukaryotic phytoplankton groups, the inherent dynamic structure of continental shelves have had a direct role in determining their modern ecology and distribution.

While the continental shelves represent only about 8% of the surface area of the world's oceans, they serve an important ecological function, as they are the buffer between the terrestrial environment and the open ocean (Biscaye et al. 1994; Falkowski et al. 1994). The physical forcing of the continental shelf environment results from a complex mixture of terrestrial and open ocean phenomena including storms, river outflow, upwelling and eddies which operate on the time scales of days to weeks. Furthermore, there is strong evidence that spatially and temporally dynamic episodic events drive shelf ecosystem dynamics (Malone et al. 1983).

In chapter 2, I develop and validate an inversion procedure that is capable of deconvoluting *in-situ* multi-spectral measurements of inherent optical properties. I validate my approach by comparing our derived parameters to four traditional measurement techniques (chlorophyll fluorometry, filter pad absorption, high pressure liquid chromatography, and spectral CDOM spectroscopy). Furthermore, I demonstrate that these inverted parameters are effective at discerning major water mass structures in the continental shelf environment. In chapter 3, I take advantage of the quasi-conservative nature of optical properties on the continental shelf and apply it to the Mid-Atlantic Bight. We make use of a shelf-wide observing system that synoptically captures the dynamical nature of a continental shelf and objectively elucidate the inherent structure of the continental shelf environment, which are validated by *in-situ* current, salinity and nutrient measurements and demonstrate that the environmental structure of the continental shelf can be synoptically captured by satellite.

1.2. Evolution in response to continental shelf environment

Direct observation of evolution in action is difficult in many organisms because their life histories often preclude the experimental conditions to make such measurements. Traditionally, biologists circumvent this problem by inferring evolutionary processes through comparative studies. However, if an organism has a relatively fast growth rate and is able to be cultured in large quantities, the probability of detecting mechanisms evolutionary change in response to some environmental condition increases dramatically (Elena and Lenski 2003). These studies usually include tracking point mutations and gene frequency through time of microbial populations. Such studies are done less frequently in phytoplankton, due to a lack of detailed sequence information is necessary to track rates of evolution. However, sequencing of ITS regions in phytoplankton have made it possible to track evolutionary dynamics for some phytoplankton populations (Rynearson and Armbrust 2000; Iglesias-Rodriguez et al. 2002). The phytoplankton genomes that have been sequenced (*Thalassiosira pseudonana* and *Phaeodactylum tricornutum*) hint at potential mechanisms of genome evolution. Both of these genomes contain transposable elements at appear to be active based on sequence analysis (Armbrust et al. 2004; Allen 2005) and that encode reverse transcriptase (retrotransposons).

Retrotransposons are especially interesting when considering the evolution of genomes because of their ability to self-replicate. Retrotransposons are able to produce copies of themselves which generally insert themselves nearby the original element through an RNA intermediate. Because of their self-replicating activity, retrotransposons can act as a population, thus having the potential to significantly modulate an organism's genomic structure. There are many examples of active retrotransposons, most of which have come from eukaryotic plant lineages (Grandbastien 1998). The first example of an active LTR retrotransposon was found in experimental tissue cultures of the rice *Oryza sativa* (Hirochika et al. 1996). This study showed that the stress of tissue culture increased the copy number of the Tos17 LTR retrotransposon significantly over a 16 month period. A more direct study of the OARE-1 a Ty1-copia LTR retrotransposon in the Oat species *Avenia sativa* showed that these retrotransposons were also activated by stress (Kimura et al. 2001). Activity of this retrotransposon was induced by plant wounding, exposure to UV light, and by the addition of jasmonic and salicylic acid. Increases of Ty retroelements has also been observed in *S. cerevisiae* in long term, continuous cultures (Wilke et al. 1992). In addition, dramatic activation of the Tnt1A retrotransposon in tobacco is induced in response to wounding (Grandbastien 1998). While direct evidence of retrotransposon activation in the natural environment has not been explicitly recorded, the natural distribution of the closely related BARE-1 LTR retrotransposon in natural environments suggest they are also active in natural populations. It has also been shown that showed that there was a sharp change in the distribution patterns of the BARE-1 element in wild barley (*Hordeum spontaneum*) in response to microclimate habitats (Kalendar et al. 2000). Populations of barley on adjacent north-facing and south-facing slopes of a canyon had large differences in the copy number of BARE-1. The close proximity of the populations suggested that gene flow between them was possible, however the population on the south-facing slope had a much larger copy number of the BARE-1 element. The increase of copy number was related to the harsher, more stressful environment associated with inhabiting a south-facing slope. Taken together with the laboratory studies mentioned earlier, there is mounting evidence that genome restructuring through the activities of retrotransposons in response to stress is not a rare occurrence, but an active evolutionary response to local environments. Therefore, given the punctuated evolutionary nature of retrotransposons, it is possible that we could observe this evolutionary event in response to the stresses associated with the dynamic structure of the shelf in eukaryotic phytoplankton. In chapter 4, I demonstrate that a retrotransposon in a diatom is activated under a nitrate stress condition, which is common to the continental shelf. Furthermore, in chapter 5 I demonstrate that the activity of retrotransposons and other mechanisms of insertion and deletion drive the overall pattern of eukaryotic genome evolution.

CHAPTER 2

2.0. Inversion of spectral absorption in the optically complex coastal waters of the Mid-Atlantic Bight; development, validation and coherence to hydrography.

2.1. Abstract

Recent advances in hydrologic optics offer the potential for quantitative maps of inherent optical properties, which can be inverted into optically active constituents such as CDOM, phytoplankton and detritus. During summer experiments in the Mid-Atlantic Bight (MAB) a procedure to invert bulk absorption measurements from off-the-shelf technology was developed. The inversion provides optical concentration estimates of phytoplankton, colored dissolved organic matter (CDOM), and detritus, all of which appear to be useful proxies for water mass and environmental structure. Inversion estimates were validated against chlorophyll fluorescence, filter pad absorption, and phytoplankton pigment measurements. The inversion could account for up to 90% of the observed variance in particulates, CDOM, and detritus. Robust estimates for phytoplankton community composition could be achieved but required constraints on the inversion that phytoplankton dominate the red light absorption. Estimates for the composition, as indicated by spectral slopes, for CDOM and detritus were not as robust. However, all of the inverted signals showed strong coherence to the spatial-temporal hydrographic structure of the coastal ocean, indicating that this approach shows great promise in developing optical proxies for environmental structure.

2.2. Introduction

Traditional ‘‘conservative’’ parameters (e.g., temperature and salinity) have been used to track water masses for nearly a century but developing additional parameters from chemical signatures to extend water type identification and water mass analysis into multidimensional space is of great utility (Tomczak 1999). Such a capability would improve adaptive sampling strategies (Robinson and Glenn 1999), allowing researchers to study how water masses evolve.

It has been suggested that traditional water mass markers might be complemented with standard biological measurements such as chlorophyll (Tomczak 1999) as an additional dimension would improve resolving water types in parameter space. Chlorophyll is a logical choice for this additional discrimination dimension, as it the preeminent proxy for phytoplankton abundance and can be estimated relatively easily by satellites and *in-situ* sensors. Fluorometry is a powerful *in-situ* mapping approach; however, variability in the fluorescence quantum yields requires local calibration data for deriving any quantitative estimate. This is difficult as changes in the fluorescence quantum yield reflect sensitivities to both the incident spectral irradiance and overall phytoplankton physiology (Kiefer 1973; Cullen 1982; Falkowski and Kiefer 1985), both of which can change on the timescale of hours to days (Falkowski and Raven 1997).

Another potential variable might be colored dissolved organic matter (CDOM), which has been used successfully to calibrate mass transport (Aarup et al. 1996) (Højerslev et al. 1996). Most coastal systems reflect the optical contributions of numerous in-water constituents (water, phytoplankton, CDOM, detritus, and sediment). This optical complexity compromises the accuracy of the satellite derived products (Kirk 1994; Mobley 1994); however, this complex matrix of materials provides a potential library of parameters that might be effective for discriminating water types if methods could be developed that provide reliable estimates of the optically significant constituents present.

Significant effort over the last decade has focused on measuring the spectral dependency of the in situ inherent optical properties (IOPs). The reliability in situ instrumentation that can measure the spectral IOPs is increasing (Pegau et al. 1995) (Chang and Dickey 1999; Schofield et al. 1999; Twardowski et al. 1999; Boss et al. 2001). A major advantage of these parameters is that they can be inverted to provide weights for optically active components (e.g., water, colored dissolved organic matter (CDOM), phytoplankton, detritus, sediment, etc.). These optical weights are proportional to component concentration thus making them useful for elucidating the fine scale structure of the marine environment that can change significantly on hourly time scales (Roesler and Perry 1995; Chang and Dickey 1999; Schofield et al. 1999; Gallegos and Neale 2002). These inversion techniques are often based on estimating the total absorption using generalized spectral absorption shapes for one or more of the individual absorbing components or using absorption ratios of different wavelengths that vary in a predictable way according to the components present. While promising in theory, the accuracy of these inverted measurements have not been systematically assessed over the wide optical gradients present in the nearshore coastal ocean. Furthermore, the performance of inversion techniques that do not require any ‘‘optimized’’ local *in-situ* data to derive the generalized shapes has yet to be assessed. Ideally, minimal ‘‘local’’ tuning should be applied to these inversion techniques as this would allow for a ‘‘global’’ mechanistic approach, which is particularly important in complex coastal waters where local empirical relationships are likely to be extremely variable.

In this manuscript, we assess the application of a simple optical inversion method using off-the-shelf oceanographic equipment for deriving optical parameters and assess the utility of the derived optical parameters to differentiate water types in the coastal ocean. Our goal is to assess how much information can be inverted from absorption data given a fixed number of wavelengths, which can then be used to determine fine scale environmental structure of the coastal environment.

2.3. Methods

2.3.1. Field Data

The field efforts were conducted at the Long-term Ecosystem Observatory (LEO) off the central coast of New Jersey (Glenn et al. 1998; Glenn et al. 2000; Schofield et al. 2002) during the ONR-sponsored Hyperspectral Coastal Ocean Dynamics Experiments (HyCODE) and the Coastal Ocean Modeling and Observation Program (COMOP). The LEO system is a highly instrumented 30 by 30 km research site that represents a coupled model/observation system where real-time data and model forecasts are provided to optimize field sampling. For bio-optical research, one advantage of the field site is that it ranges from very turbid estuarine waters to relatively clear offshore waters within the 30 km research box. These optical gradients reflect the variable contributions of many optically active constituents such as phytoplankton, sediments, CDOM, and detritus.

The standard shipboard transects consisted of several 15–25 km cross-shelf transects. Specific transect lines and the locations of the stations were determined by the real-time data from ocean forecast models, ships, and satellites focused on characterizing coastal upwelling dynamics (Schofield et al. 2002). At each station, vertical profiles of optical and physical data were collected using an integrated bio-optical package. The bio-optical package consisted of a WET Labs Inc. absorption/attenuation meter (ac-9), a Falmouth CTD, a profiling Satlantic spectral radiometer, and a HOBI Labs backscatter sensor (Hydroscat-6). The measurements of the inherent optical properties used in this study were collected using the standard nine wavelengths (412, 440, 488, 510, 555, 630, 650, 676, and 715 nm) of the WET Labs Inc. ac-9. At each station, the instrument was lowered to depth to remove air bubbles and the instrument was allowed to equilibrate with ambient temperature before data were collected. Only data from the upcasts were utilized. Data were averaged into 0.25 m depth bins for all subsequent analyses. The instruments were factory calibrated prior to the field season. Manufacturer recommended protocols (http://www.WETLabsInc.com/otherinfo/ugftp.htm) were used to track instrument calibration throughout the field season. This included clean water, temperature, and salinity calibrations. Whenever possible daily water calibrations were conducted; however, sampling schedules did not always allow for a daily calibration. Under these circumstances the most recent water calibration was used. It should be noted that this period without a calibration was at most three days.

A CDOM absorption mapping system was installed on the ship (Kirkpatrick et al. 2003), which consisted of a liquid waveguide capillary cell (LWCC, World Precision Instruments, Inc.) coupled to a fiber-optic spectrometer (S2000, Ocean Optics, Inc.) and a fiber-optic xenon flash lamp (PS-2, Ocean Optics, Inc.). Water was pumped by miniature peristaltic pump (P625, Instech Laboratories, Inc.) through size fractionation and cross-flow filters (MicroKros, Spectrum Laboratories, Inc.) and then through the LWCC for optical density spectra measurements. A continuous underway water supply was provided by tapping the flow through the ship’s fire suppression system.

At each ship occupied station, water was collected with Niskin bottles from both surface and bottom waters. Aliquots were filtered, under low vacuum (<10 cm Hg), through GF/F (Whatman) glass fiber filters to concentrate the particles for pigment and absorption determinations. Filters were placed into snap top vials and quick frozen in liquid nitrogen. Samples were stored at –80ºC until later analysis. Filters were analyzed for photosynthetic and photoprotective pigment complements were determined using high-performance liquid chromatography (HPLC) according to standard procedures (Wright et al. 1991). Filter pad absorption was measured on a laboratory spectrophotometer and spectra were corrected for the path length amplification factor (Roesler 1998). Detrital absorption was determined by methanol extraction of particulate material (Kishino et al. 1985). The detrital absorption was subtracted from particulate absorption to provide an estimate of phytoplankton absorption. For discrete CDOM spectra, water was filtered through a 0.2 micron Nucleopore filter, and measured on a spectrophotometer using a 5 cm long path length cuvette.

A second *in-situ* data set was collected using a two bottom-mounted nodes with profiling instrument packages. These nodes (Node B and Optical Profiler) were located approximately 4 km offshore in 13 m of water at 39°27.41 N, 74°14.75 W and collected data from Julian Day 202-215, 2000. Data measured by these profilers streamed directly to the Rutgers University Marine Field Station (RUMFS) in real time via an electro-optical cable, where it was processed and visualized. Node B included a Sea-Bird CTD mounted with a WET Labs chlorophyll fluorometer, which sampled at 2 Hz and was profiled at a vertical rate of 2 cm s^{-1} at regular intervals. The Optical Profiler included a WET Labs nine wavelength absorption/attenuation meter (ac-9) (412, 440, 488, 510, 532, 555, 650, 676 and 715 nm), which sampled at 8 Hz, and a two wavelength backscatter/fluorometer HOBI Labs HydroScat-2 (470 nm and 676 nm) that sampled at 2 Hz. The Optical profiler also profiled at a rate of 2 cm s^{-1}.

2.3.2. Inversion of *in situ* Absorption Data

The optical signature inversion method (OSI) uses measured spectral absorption data collected from the ac-9 to calculate optical weight specific absorption coefficients or material present in the water column. The OSI model calculates optical weight specific coefficients (w_i) and exponential slopes (s, r) using a nonlinear, constrained least-squares regression according to

$$a_{total}(\lambda) = w_1 a_{Phyto1}(\lambda) + w_2 a_{Phyto2}(\lambda) + w_3 a_{Phyto3}(\lambda) + w_4 a_{CDOM}(\lambda, s) + w_5 a_{Detritus}(\lambda, r) \quad (2.1)$$

where $a_{total}(\lambda)$ is the total spectral absorption measured with the ac-9 (note that ac-9 provides an absorption that has already subtracted the contribution due to water), $a_{Phyto1}(\lambda)$, $a_{Phyto2}(\lambda)$, and $a_{Phyto3}(\lambda)$ are generalized spectral absorption of chlorophyll a-c, phycobilin, and chlorophyll a-b containing phytoplankton, respectively, and $a_{CDOM}(\lambda, s)$ and $a_{Detritus}(\lambda, r)$ are the spectral absorption of CDOM and detritus (Figure 2.1). The CDOM absorption (and detritus absorption) can be described as an idealized curve as a function of wavelength and exponential slope (Kalle 1966; Bricaud et al. 1981; Green and Blough 1994),

$$a_{CDOM}(\lambda) = a_{CDOM} \exp[-s \bullet (\lambda - 412nm)] \quad (2.2)$$

The exponential s parameter (unitless) is dependent on the composition of the CDOM present and is highly variable (Carder et al. 1989; Roesler et al. 1989). Therefore it was necessary to allow the CDOM and detritus exponential slopes to vary to achieve reasonable estimates. The initial exponential slopes of CDOM were set to 0.010. The detritus exponential slope was initially set to 0.008. The slopes of the detrital curves (r) are lower (Kirk 1994) and detritus is described by equation (3),

$$a_{Detritus}(\lambda) = a_{Detritus} \exp[-r \bullet (\lambda - 412nm)] \qquad (2.3)$$

The values of w_1, w_2, w_3, w_4 and w_5 are non-spectrally dependent scalar coefficients of these input spectra. We used fixed absorption spectra measured on laboratory cultures in order to ensure that inversion was completely independent from any spectral curves encountered in the field. Spectral phytoplankton curves were of averages of high-light- and low-light-acclimated phytoplankton spectra that were normalized to absorption at 676 nm. The spectral library used was taken from (Johnsen et al. 1994) (Figure 2.1). While not optimal, we believe it was reasonable since the first-order determinant of spectral optical signals reflects the overall concentration of material rather than spectral characteristics of the materials present (Barnard et al. 1998).

We used two different inversion approaches, one with more constraints than the other. The minimal constraint OSI (OSIm) only required that all solutions be positive (equation (4)), that CDOM and detritus absorption weights were equal in the red wavelengths of light (equation (5)), and that the CDOM exponential slope is steeper than the detrital slope (equation (6)). The assumption that the CDOM and detritus absorption is equal is

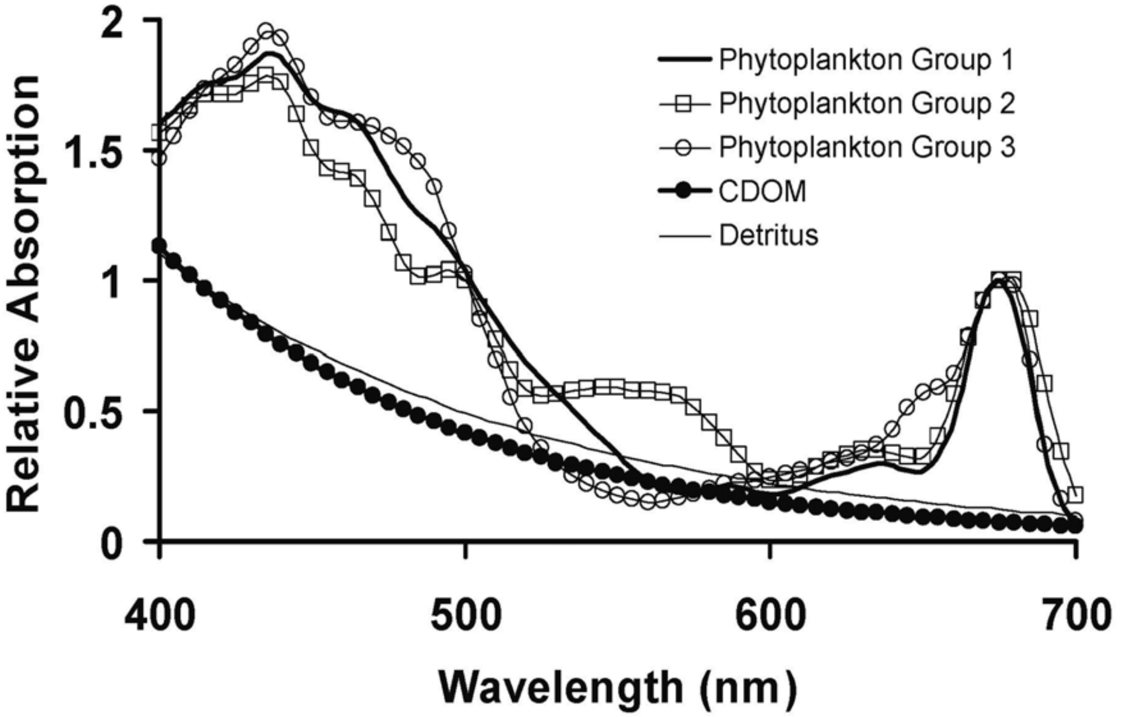


Figure 2.1. Input spectra used to invert the *in situ* absorption values measured by the ac-9 using the OSI model. Phytoplankton spectra are averages of high-light and low-light adapted phytoplankton from Johnsen et al., 1994. Phytoplankton group one represents chlorophyll$_{a\text{-}c}$ containing classes of Bacillariophyceae, Dinophyceae and Prymnesiophyceae. Phytoplankton group two represents the phycobilin containing class Cryptophyceae. Phytoplankton group 3 represents the chlorophyll$_{a\text{-}b}$ containing classes of Chlorophyceae, Prasinophyceae and Eugelnophyceae. CDOM and detritus spectra are idealized exponential functions.

artificial but we know that CDOM and detritus absorption is very low in the red wavelengths as both are exponentially decreasing curves. There are two artificial ways to ensure that the CDOM and detritus absorption do not dominate the red light absorption where phytoplankton absorption dominates. One method, more commonly used, is to set the magnitude of CDOM and detritus red absorption to a fixed low value. The second method is to anchor both curves to each other in the red, which allows the exponential slopes and amplitudes to be determined largely by the blue wavelength absorption. This second method allows the red light absorption of CDOM and detritus to be variable. The second OSI method (OSIc) added constraints so that phytoplankton absorption dominated in the red wavelengths (equations (2.7) and (2.8)), and that minor phytoplankton communities in these waters (here chlorophytes) were never dominant (equations (9) and (10)). Specifically, the constraints on the OSI optimizations were:

$$w_1 a_{Phyto1}(\lambda), w_2 a_{Phyto2}(\lambda), w_3 a_{Phyto3}(\lambda), w_4 a_{CDOM}(\lambda, s), w_5 a_{Detritus}(\lambda, r) \geq 0 \quad (2.4)$$

$$w_1 a_{Phyto1}(\lambda) \geq w_3 a_{Phyto3}(\lambda) \quad (2.5)$$

$$w_2 a_{Phyto2}(\lambda) \geq w_3 a_{Phyto3}(\lambda) \quad (2.6)$$

$$w_1 a_{Phyto1}(650nm) \geq w_4 a_{CDOM}(676nm, s) + w_5 a_{Detritus}(676nm, r) \geq 0 \quad (2.7)$$

$$w_2 a_{Phyto2}(650nm) \geq w_4 a_{CDOM}(676nm, s) + w_5 a_{Detritus}(676nm, r) \geq 0 \quad (2.8)$$

$$w_4 a_{CDOM}(676nm, s) = w_5 a_{Detritus}(676nm, r) \quad (2.9)$$

$$s \geq r \quad (2.10)$$

These assumptions were based on 5 years of experience in coastal New Jersey waters spanning both the nearshore and offshore. It should be noted that the OSIm and OSIc inversion methods using the same assumptions have been successfully used in both the oligotrophic Gulf of Mexico and the southern basin of Lake Michigan (unpublished data). When the OSI method (OSIm and OSIc) did not converge on a solution, the data were omitted from the later analysis (<5% of the total New Jersey data set). This generally reflected noise in the ac-9 data most often in near surface waters, presumably related to air bubbles, which interfered with natural inflections in the absorption curve. Of all the constraints, the requirement that green algae were always less abundant than chlorophyll c and phycobilin containing algae was admittedly the most artificial. However, running the inversion without the constraint greatly compromised the efficacy of the inversion for the overall phytoplankton absorption spectra. Phytoplankton pigment concentrations from discrete measurements during this study also confirmed that this assumption was valid as it was also confirmed a background population of green algae was detectable (Moline et al. 2004). To assess the stability of the OSI, random noise was introduced into the ac-9 spectra. For this analysis we added +0.005 m^{-1} to the data randomly across all wavelengths. Results indicated that there was no spectral bias and the quantitative impact was less than 1%.

2.4. Results and Discussion

2.4.1. Verification of the Derived Optical Products

OSI-derived particulate, detrital, and phytoplankton loads were compared to three independent data sets that included stimulated chlorophyll fluorescence, phytoplankton filter pad absorption measurements, and HPLC phytoplankton pigment concentrations. All three data comparisons suggested that the OSI method provided reasonable estimates of particulate, phytoplankton, detritus, and CDOM optical weights.

2.4.2. Fluorescence

The ac-9-derived phytoplankton absorption and stimulated *in situ* chlorophyll fluorescence were positively and linearly related to each other (Figure 2.2.). The derived phytoplankton absorption was significantly correlated ($p < 0.05$) with fluorescence and could explain 54% and 61% of the variance in the summers 2000 and 2001, respectively. The linear relationship between the fluorescence and derived phytoplankton weight was notable given that the majority of the data were collected in Case 2 waters where phytoplankton are not necessarily the dominant optical signal. In addition a significant proportion of the phytoplankton communities were probably light saturated for photosynthesis, thus fluorescence quenching was also significant and contributed to the variance in the correlation between the OSI-derived phytoplankton optical weights and chlorophyll fluorescence measurements. Xanthophyll pigment cycling (Demmig-Adams 1990; Owens et al. 1993) and photoinhibition (Prasil et al. 1992; Critchley 1994; Nickelsen and Rochaix 1994) often results in almost a 100% change in fluorescence quantum efficiency (Falkowski and Kiefer 1985; Kroon 1994). The variable fluorescence quantum yield compromises the accuracy of using fluorescence to estimate chlorophyll a biomass, the OSI phytoplankton estimates may be more desirable than the commonly used chlorophyll fluorometer because it is not subject to physiological variability. IOP sensors are now becoming operationally viable for the wider oceanographic community and inverted optical data will improve our ability for making quantitative biomass maps.

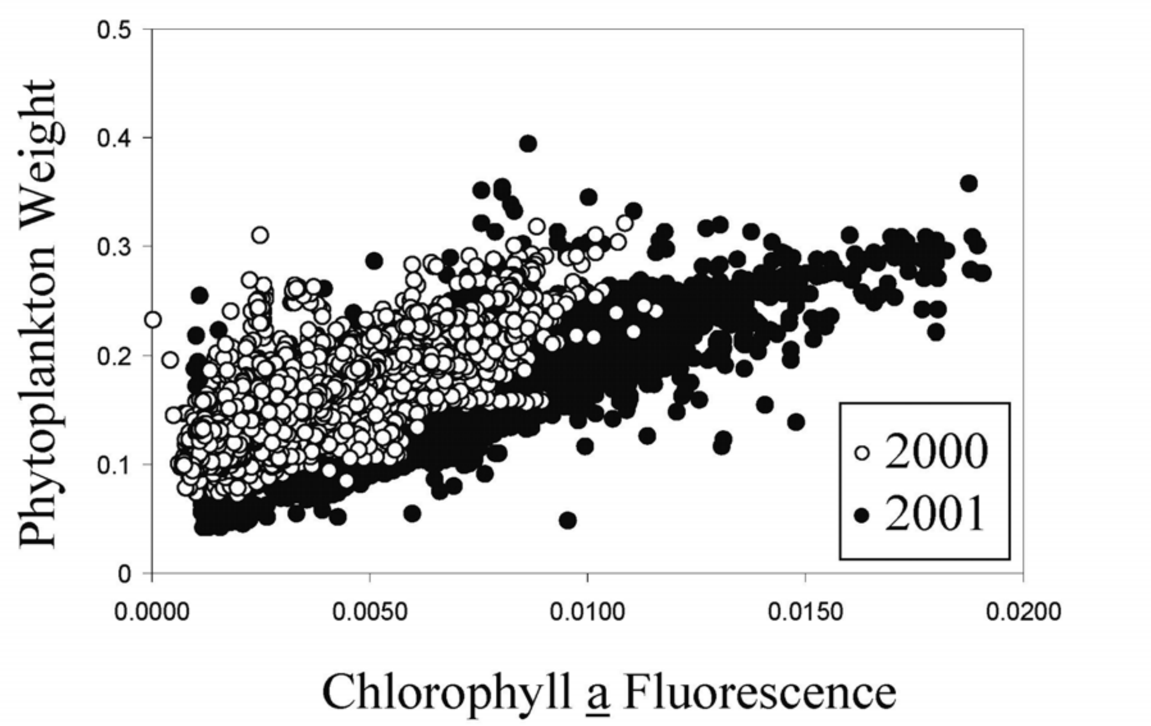


Figure 2.2. The relationship between chlorophyll *a* fluorescence measured with a HOBI Labs hydroscat-6 and the estimated phytoplankton weight during the summers of 2000 and 2001. The phytoplankton weight was inverted from ac-9 data using the OSI_c approach. The R^2 for summer 2000 and 2001 are 0.54 and 0.61, respectively.

2.4.3. Filter Pad Absorption

The OSI results were compared to 240 filter pad samples that spanned the period of the field effort (Figures 2.3 and 2.4). For OSIm, quantitative agreement with discrete samples for particulates was good with the R^2 ranging from 0.8 to 0.5 for wavelengths lower than 680 nm except in the wavelengths associated with carotenoid and phycobilin absorption (530 to 600 nm) (Figure 2.3 A). The R^2 dropped to 0.3 for wavelengths greater than 680 nm. The average slope between the measured and predicted absorption ranged 1.2 to 0.5. Given the variance within the correlations, the average slope was rarely significantly different than one. The OSIm could account for 70% of the variance in measured phytoplankton spectra except for the wavelengths associated with phycobilin absorption at wavelengths spanning from 530 nm to 600 nm (Figure 2.3 B). The R^2 dropped for wavelengths greater than 680 nm. Average slopes between measured and predicted phytoplankton absorption were insignificantly different from one except in the low blue wavelengths (<415 nm) of light where phytoplankton absorption were overestimated by as much 20% (Figure 2.3 B). OSI-derived detritus spectra could account for 70% of the variance in the absorption in the blue wavelengths of light (Figure 2.3 C), but the R^2 dropped off at the higher wavelengths because of the low signal to noise ratio associated with the exponential decline in detrital absorption with increasing wavelength. The CDOM absorption was significantly overestimated. In the blue wavelengths, this over estimate was 35% but increased to a factor of 2 in the green orange wavelengths of light (Figure 2.3 D); however, the OSIm-derived CDOM absorption could account 88% of the variance in the measured CDOM absorption in the

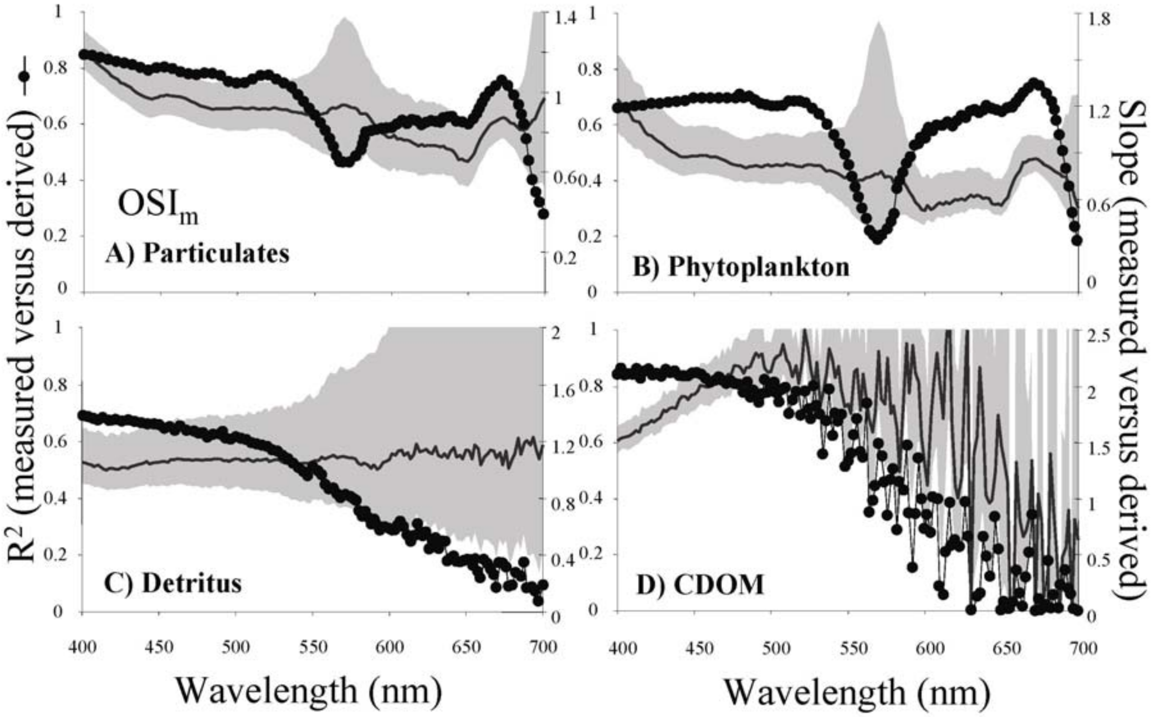


Figure 2.3. Comparison of measured and (A) modeled particulate, (B) phytoplankton, (C) detritus, and (D) CDOM absorption using the minimal constraint OSI method. The measured data represents absorption spectra made for either filter pads or dissolved organics on discrete samples. The modeled data represents the predicted absorption spectra from the inverted from the ac-9 data. Data were pooled for both years and was linearly regressed at each wavelength providing both the slope (dark line with gray area) and R^2 (dark circles) for each wavelength. The gray shadow around the slope represents the standard deviation.

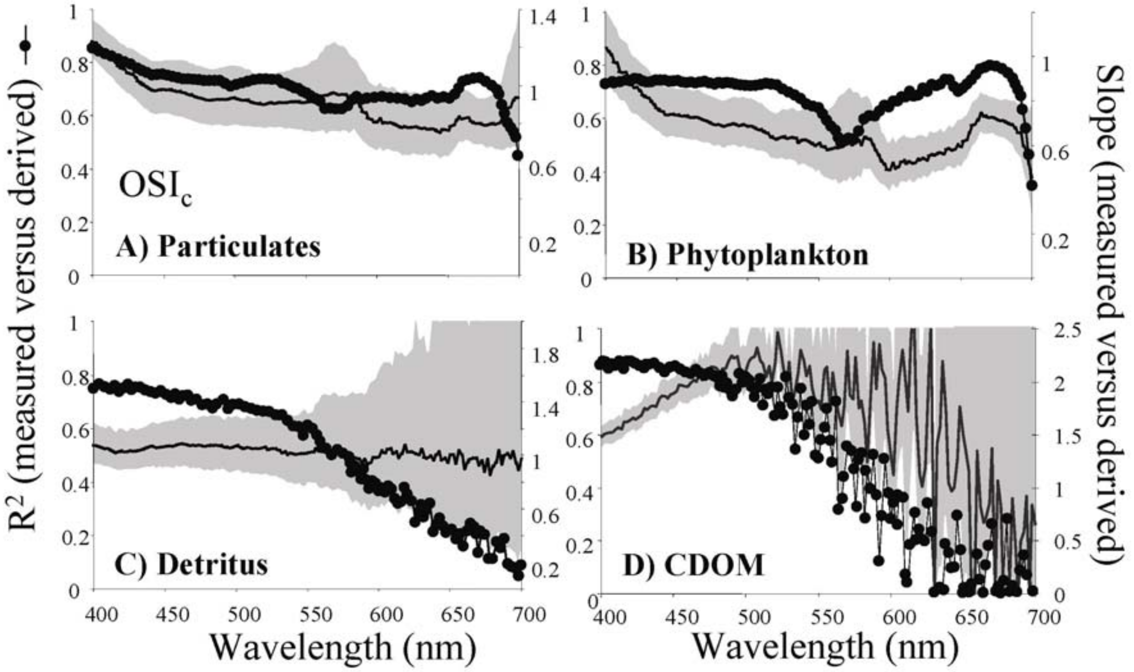


Figure 2.4. Comparison of measured and modeled (A) particulate, (B) phytoplankton, (C) detritus, and (D) CDOM absorption using the OSI$_c$ method. The measured data represents absorption spectra determined from filter pads or discrete samples for colored dissolved organics. Data were pooled for both years and were linearly regressed against measured absorption spectra at each wavelength providing both the slope (dark line with gray area) and R^2 (dark circles) for each wavelength. The gray shadow around the slope represents the standard deviation.

blue wavelengths of light. Similar to the detritus the R^2 decreased with increasing wavelength because of decreasing signal to noise. For the OSIc the quantitative agreement between measured and modeled particulate spectra was good, with the R^2 ranging from 0.5 to 0.9 with low values associated with wavelengths greater than 680 nm (Figure 2.4 A). The slope between the measured and modeled particulate spectra ranged from 1.2 to 0.8 depending on wavelength (Figure 2.4 A) and the variance was reduced from the OSIm approach especially in the wavelengths associated with phycobilin and carotenoid absorption (530–580 nm). For the majority of the wavelengths, the deviations of the average slope from 1 were rarely statistically significant (Figure 2.4 A). Quantitative agreement decreased in the red wavelengths of light where signal was low. Results indicate that the derived particulate spectra could be quantitatively derived from the ac-9 with minor spectral biases despite that only idealized spectral absorption shapes were used. Like the particulate spectra, the agreement for the OSI and measured phytoplankton spectra were good (Figure 2.4 B). The largest mismatches were in the blue wavelengths of light but 70–80% of the observed variance in the measured spectra were described by the OSIc and as with the particulate spectra the errors were lower compared to the OSIm method. Quantitative estimates for detritus were good (Figure 2.4 C), but accuracy dropped in the higher wavelengths because of low detrital absorption. The OSIc method showed no improvements over the OSIm method for predicting CDOM absorption, with the overall CDOM absorption being overestimated significantly (Figure 2.4 D). For both OSIm and OSIc the spectral slope of the CDOM was underestimated especially when discrete samples indicated high slopes (Figure 2.5).

In our approach, the input spectra for the OSI were based on laboratory data (Johnsen et al. 1994) and theoretical curves, so the OSI methods could undoubtedly be improved by customizing the input spectra for any particular location. Our goal, however, was to assess what could be derived using no local input data. The relative particulate spectra derived by the OSIc approach overestimated absorption at wavelengths of peak phytoplankton absorption (420–540 and 660–680 nm). This is consistent with the well-documented package effect, where absorption spectra are ‘‘flattened’’ when pigment packaged within a cell (Morel and Bricaud 1986). The package effect is greatest in the wavelengths of maximal absorption and increases with increasing cell size and cellular concentration of pigment. In higher-chlorophyll waters nearshore, water samples revealed high populations of large net diatoms (Moline et al. 2004), which are greatly affected by the pigment package effect (Bricaud et al. 1995). Filter pad measurements support the hypothesis that phytoplankton were highly packaged as the specific absorption at 676 nm was consistently lower (0.017 m^2 mg chl a^{-1}) than in low-chlorophyll offshore waters where populations were dominated by picoplankton. Therefore as the majority of the data collected represented nearshore stations, the package effect could account for much of spectral differences in the derived spectra. The pigment package effect has a proportionally larger impact on the wavelengths of maximum phytoplankton absorption.

The spectral mismatch resulting from the highly peaked phytoplankton contributed to the overestimated CDOM absorption. This is associated with the high pigment absorption in the blue and red wavelengths of light. Given the OSI requirement that phytoplankton dominate the red absorption peak, a flatter phytoplankton input

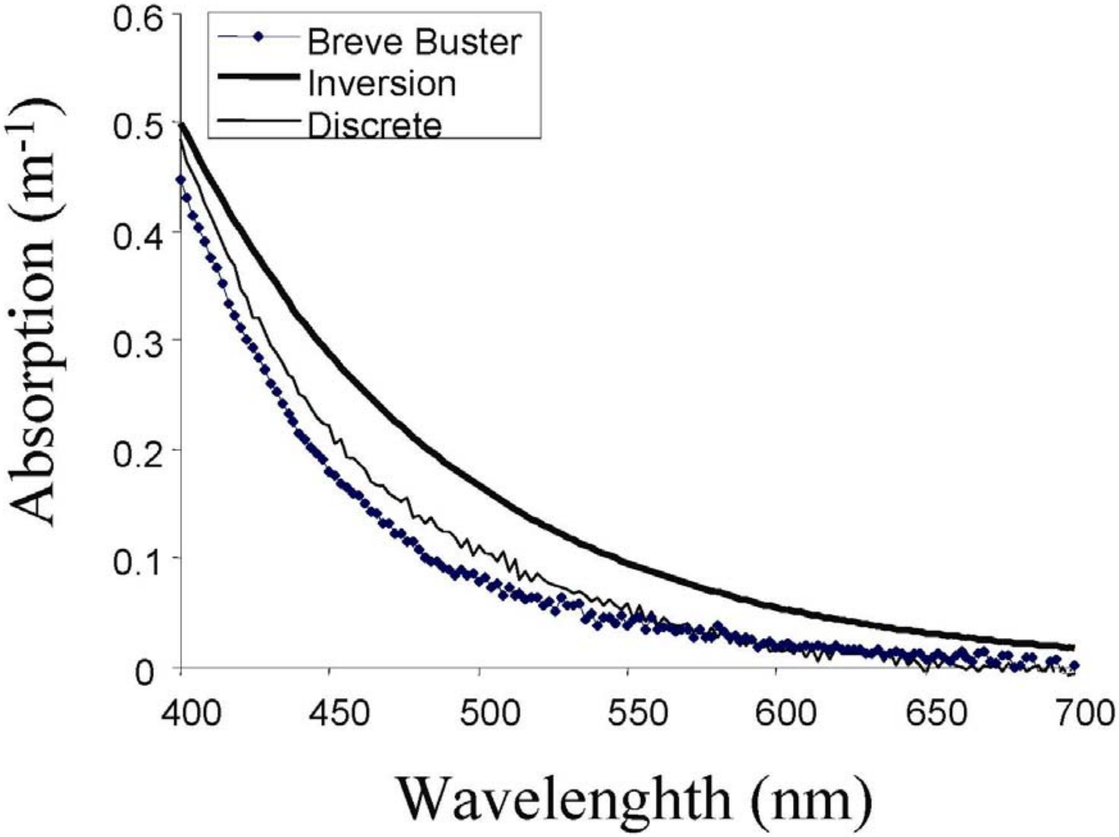


Figure 2.5. Comparison of CDOM spectral absorption estimated by the OSI_c and measured with a flow through Breve-buster (Kirkpatrick et al. 2003) and on a discrete sample.

absorption spectra would result in 1) lower the overall CDOM and detrital estimates in the blue wavelengths and 2) steeper estimates in the CDOM exponential slope. This would argue that laboratory spectra should not be used in the inversion of field data; however, it was the highly peaked pigment shoulders that allowed phytoplankton community composition to be determined. Until more wavelengths are available to allow researchers to characterize both composition and pigment-packaging, researchers will be forced to prioritize their needs. This shortcoming will improve in the coming years as the community is actively developing *in-situ* hyperspectral sensors.

2.4.4. Phytoplankton Pigments

To further assess the phytoplankton absorption inversion estimates, we examined how well the presence of the three spectral classes of phytoplankton could be determined (Figure 2.6). Using accessory pigment data and the ChemTax program (Mackey et al. 1996; Mackey et al. 1998) we estimated the proportion of total chlorophyll a associated with the three major spectral classes of phytoplankton. The inverted phytoplankton estimates from the OSIc method were significantly correlated ($p < 0.00$) with the ChemTax estimates of chlorophyll c and phycobilin-containing phytoplankton (Figure 2.6 A). There was no success in predicting the distribution of chlorophyll b, but this is consistent with the independent findings that they were a rare component of the phytoplankton community at LEO and thus had insignificant contributions to the optical signals (Moline et al. 2004). The OSIm approach had no success in predicting the phytoplankton community composition.

Overall results from the OSI show that currently available off-the-shelf technology can provide reasonable estimates of the major optical constituents (CDOM,

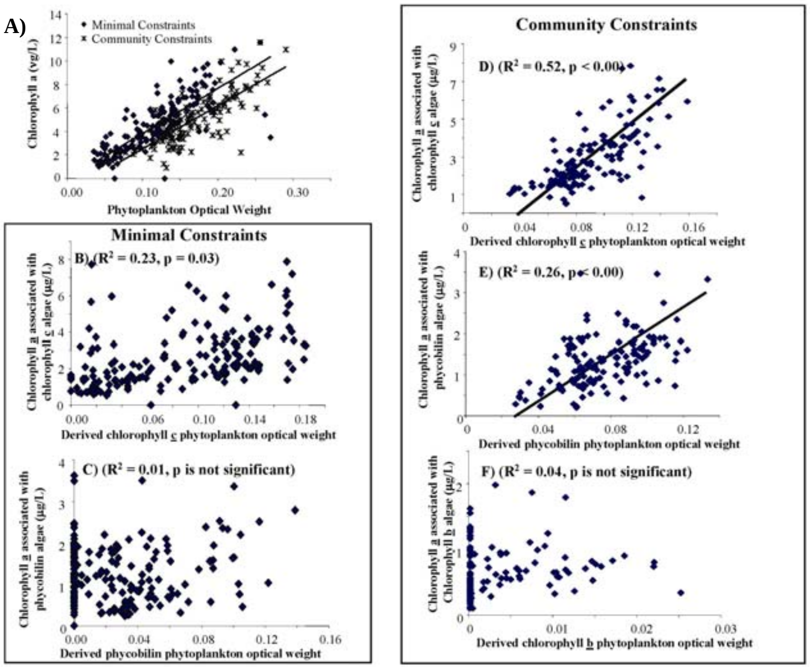


Figure 2.6. Comparison of the amount of phytoplankton absorption and measured chlorophyll a associated with the three major spectral classes of phytoplankton during summers 2000 and 2001. The amount of chlorophyll a associated with each spectral class of phytoplankton was calculated via CHEMtax using the accessory pigment data measured via high performance liquid chromatography. A) Relationship between measured chlorophyll a to the OSI_m and OSI_c procedures. B) The absorption of chlorophyll c containing algae estimated with the OSI_m and the chlorophyll a associated with chromophytic algae. C) The absorption of phycobilin containing algae estimated with the OSI_m and the chlorophyll a associated with phycobilin containing algae. D) The absorption of chlorophyll c containing algae estimated with the OSI_c and the chlorophyll a associated with chromophytic algae. E) The absorption of phycobilin containing algae estimated with the OSI_c and the chlorophyll a associated with phycobilin containing algae. F) The absorption of chlorophyll b containing algae estimated with the OSI_c and the chlorophyll a associated with chlorophytic containing algae.

detritus, particles, and phytoplankton) in Case 2 waters. While the precise phytoplankton community composition was difficult to delineate, the most dramatic gradients in phytoplankton composition could be described given constraints that maximized the phytoplankton absorption in the red wavelengths of light. Improving this and similar inversion methods will require spectral resolutions greater than nine wavelengths, ideally, at the wavelengths associated with phytoplankton accessory pigments (Jeffrey et al. 1997). Increased spectral resolution would also allow a variety of spectral pattern recognition methods to be applied (Millie et al. 1997; Schofield et al. 1999) (Kirkpatrick et al. 2000; Millie et al. 2002), which will increase our ability to discriminate the major spectral classes of phytoplankton and even specific phytoplankton taxa. Many of these approaches require spectral resolutions of 2–3 nm (Roelke et al. 1999) so developing hyperspectral instrumentation will be key to improving optical discrimination techniques for coastal waters. Despite shortcomings, this approach appears very promising in describing the major absorbing components at LEO.

2.4.4. Spatial and temporal coherence of derived optical parameters to hydrography

The high-resolution time series of hydrographic and optical data provided by the profilers allowed us to examine the spatial and temporal coherence of the derived optical parameters to the general hydrography of the coastal ocean. During the time period of their deployment, the density record of these profiles (Figure 2.7 A) indicated there where three major water column states which had radically different water column structures; i) a coastal upwelling event on days 203-207, ii) a fresh water river plume, presumably from the Hudson River (Johnson et al. 2003) on days 210-215, and

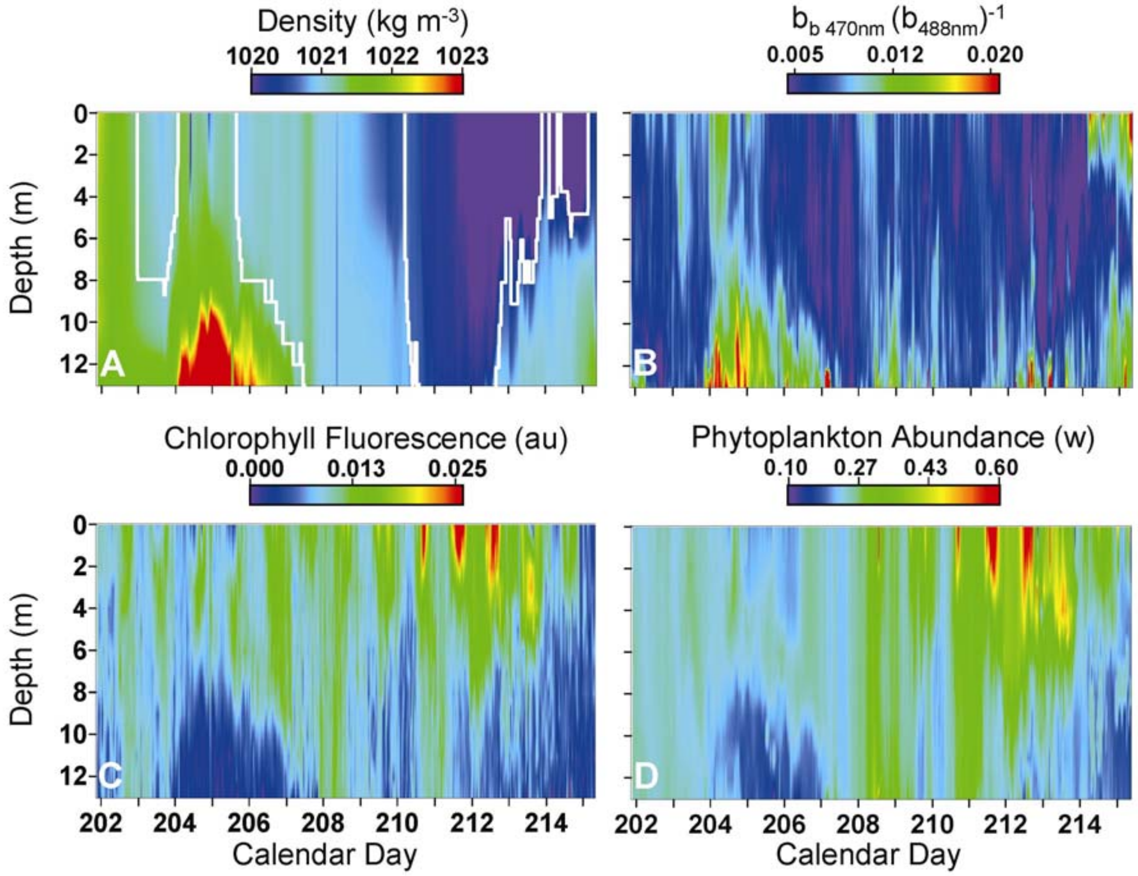


Figure 2.7. Time series of *in situ* data taken by the profilers during the experiment. Panel A shows density structure with water mass boundaries (white) defined by cluster analysis (see text). Panel B is the ratio of scattered and backward scattered light. Panel C is chlorophyll fluorescence measured by the optical profiler. Panel D is the OSI derived calibrated relative phytoplankton abundance. Optical and biological parameters have similar patterns as the hydrographic structure.

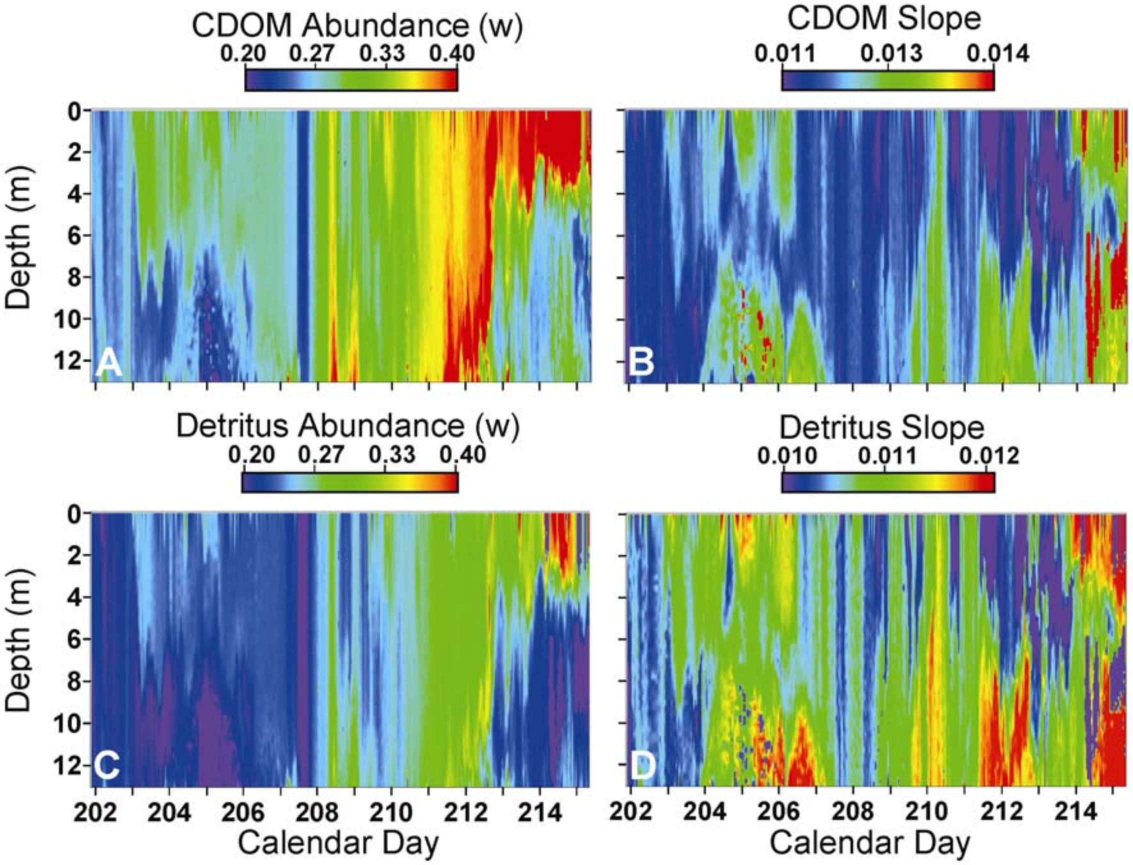


Figure 2.8. Time series of inverted *in situ* absorption data taken by the optical profiler during the experiment. Panel A is the relative abundance of CDOM, while Panel B is the exponential slope of the CDOM curve. Panel C is the relative abundance of detritus while Panel D is the exponential slope of the detritus curve. Derived optical properties show distinct characteristics of the hydrographic structure during the experiment.

iii) a well mixed regime on days 207-210. These regimes were confirmed as being statistically different from each other by a MANOVA analysis of the paired salinity and temperature records from the profilers (Pillai Trace approx. F = 2988.747, p = 0.000). The density, optical backscattering, and chlorophyll fluorescence (Figure 2.7 A-C) are all independent estimates of water column structure to which the absorption derived optical parameters could be compared. (Figure 2.7 D, 2.8 A-D).

The spatial-temporal pattern of derived phytoplankton abundance matches that of chlorophyll fluorescence very closely, while still showing the overall patterns related to the passage of the three major water masses. In addition, the backscattering ratio generally indicates that the particles associated with the river plume are larger than particles during the upwelling event. This suggests that smaller, non-absorbing particles are associated with the upwelled water as opposed to the large absorbing particles (phytoplankton) in the river plume. CDOM and detritus abundance, are, as expected, significant absorbing components of the river plume. In addition, the spectral slope of the CDOM signature is generally lower in the river plume as compared to that of the upwelled water. In coastal regions this is a typical characteristic of terrestrial derived vs. marine derived CDOM. Therefore, despite some of the weak wavelength specific correlations between the derived optical parameters and discrete samples, the derived products show the same spatial and temporal patterns as the density, backscattering and chlorophyll fluorescence indicating that the derived products are effective in capturing and characterizing the even the fine scale environmental variability.

2.5. Conclusions

Inversion of absorption data measured using off the shelf technology is possible and shows great promise. We validated our inversion procedure with other optical environmetnal optical measurements (chlorophyll fluoresence, CDOM absorption), discrete filter pad absorption, and with HPLC analysis and showed that these and similar inversion approaches can be applied to optically complex Case 2 waters. Furthermore, the spatial and temporal distribution of the derived parameters is coherent with other independent optical and hydrographic parameters. Therefore, we feel this approach has great potential utility to extend tradional water mass and environmental structure analysis into multidimensional space (Tomczak 1999) in complex coastal waters. This will provide the marine ecologist a key technology to map specific phytoplankton taxa over ecologically relevant spatial temporal scales.

Chapter 3

3.0. Bioinformatic Approaches for Objective Detection of Water Masses on Continental Shelves

3.1. Abstract

As part of the 2001 Hyper Spectral Coupled Ocean Dynamics Experiment (HyCODE) and the 2005 Lagrangian Transport and Transformation Experiment (LaTTE), sea surface temperature and ocean color satellite imagery were collected for the continental shelf of the Mid-Atlantic Bight. The imagery collected in 2001 was used to develop a water mass analysis and classification scheme that objectively describes the locations of water masses and their boundary conditions. This technique combines multivariate cluster analysis with a newly developed genetic expression algorithm to objectively determine the number of water types in the region based on ocean color and sea surface temperature measurements. Then, through boundary analysis of the water types identified, the boundaries of the major water types were mapped and the differences between them are quantified using predictor space distances. We then independently validate this approach during the 2005 LaTTE experiment with salinity and inorganic nutrient measurements. Results suggest that this approach can track the development and transport of water masses and that the boundaries of these water masses often constitute represent large changes in inorganic nutrient concentrations. Because the analysis combines the information of multiple predictors to describe water masses it is an effective tool in detecting water masses not readily recognizable with temperature or chlorophyll alone.

3.2. Introduction

Water mass analysis is an active area of research because of their potential utility for describing large scale ocean circulation (Warren 1983), assessing the impact of river plumes (Højerslev et al. 1996), understanding basin scale biogeochemistry (Broecker et al. 1985). Water masses are classically defined as waters with common formation and origin having similar conservative properties such as temperature and salinity. However, it should be noted that this conservative requirement means that for temperature and salinity to remain conservative within a mass of water, the water mass cannot be in contact with the surface ocean or its source region. The introduction of the T-S diagram was the first quantitative approach to defining water masses based on their conservative properties and has been a mainstay in the oceanographic community (Hellend-Hansen 1916). Since that time, oceanographers have used chemical isotopes to further study the circulation of water masses in the ocean interior (Broecker and Peng 1982). In the surface ocean where temperature and salinity are not considered conservative, injections of dyes and SF_6 have been successfully used to track the circulation and subduction of surface features because the presence of SF_6 can be considered conservative compared to some of the short-time scale process in the surface ocean (Upstill-Goddard et al. 1991); however, this type of research is costly and can effectively cover only relatively small space scales. To assess the impact of broad scale surface features, the key is to develop proxies that change over larger time scales than the processes being studied.

To a certain degree, optical oceanographers have addressed the issues of water mass identification in the surface ocean by classifying them based on their optical properties. Efforts by (Jerlov 1968) classified waters into nine water types. These water types were further analyzed by (Morel and Prieur 1977) and classified into the widely accepted Case 1 and Case 2 waters. These classifications have been an extremely useful tool. Water types are different than water masses in that water types occupy only similar predictor space while water masses occupy similar predictor and physical space (Tomczak 1999). A major objective over the last few decades has focused on understanding global and basin scale circulation, which operate over time scales of years to thousands of years. Therefore these processes require tracers that are relatively conservative over the same time scales (i.e. salinity). However, if the time scale of interest in detecting and tracking near surface water masses is on the order of hours to days as it often is in coastal regions, optical predictors potentially provide additional dimensions of discrimination to traditional temperature and salinity analysis. This type of optical approach has been demonstrated by tracking river influence containing anthropogenic pollutants (Højerslev et al. 1996). In addition to tracking anthropogenic pollutants, the identification of frontal regions between water masses has been used to identify important areas of mixing and biological activity (Claustre et al. 1994).

Although simple in concept, the inclusion of optics as a water mass tag presents a problem in determining the uniqueness of a water mass. Because water mass classification has traditionally relied upon hydrographic predictors only, there exists an intuitive sense, based on a century of experience, for defining significant differences in temperature and salinity predictors before discriminating between water masses. While these discriminations are inherently subjective, the inclusion of optical predictors only confounds the already subjective interpretation. This problem is not unique to oceanography, but a fundamental problem for any scientific field that assigns categories or identifiers to a known data continuum. Therefore, if optical predictors are to be used effectively in water mass analysis and identification, an objective mathematical construct is needed to for proper quantitative discrimination of water masses based on the similarity of water types (Martin-Trayovski and Sosik 2003).

One branch of science that has had to develop means to overcome the problems associated with assigning categories to a known continuum is the field of evolutionary and molecular biology. These problems manifest themselves in a variety of ways such as uncertainties in phylogenetic trees, species determination (Hey 2001; Wu 2001; Noor 2002), annotations of genomes (Meeks et al. 2001) and the expression of genes (Yeung et al. 2001). This problem has become more complex with technological breakthroughs such as DNA microarrays and automatic sequencers, and through necessity, the rapidly advancing field of bioinformatics has endeavored to produce several objective mathematical constructs to transform a data continuum into meaningful categories. This manuscript applies techniques developed by the bioinformatics field and adapts them for the use of objective water mass analysis and classification in a coastal region. We present a mathematical construct of a water mass classification method and apply it to the Mid-Atlantic Bight using optical and temperature parameters measured by satellite.

3.3. Methods

3.3.1. HyCODE

During the 2001 HyCODE experiment at the Long-term Ecosystem Observatory (LEO) off southern New Jersey, daily SeaWiFS and AVHRR passes were collected with a L-Band data acquisition system at approximately 1 km resolution over an area defined at 38.50°N – 41.50°N latitude and 76.00°W – 71.00°W longitude (Figure 3.1). These satellites were used as an adaptive sampling tool during the experiment so that data of the relevant hydrographic features in the region could be collected. Pixels from the single daily SeaWiFS pass were matched to the least cloud covered AVHRR pass using latitude and longitude. Morning AVHRR passes were used to avoid the effects of diurnal solar heating. Cloud removal was accomplished by adjusting the cloud coefficient in the MCSST algorithm. SeaWiFS data were processed using the DAAC algorithm. For this study, matched satellite passes from July 14, July 21, July 31, and August 2 2001 were chosen because of relatively little cloud cover. Each composite matrix of SeaWiFS and AVHRR imagery had between 75,000 and 105,000 cloud free pixels. Each composite matrix was sub-sampled at 6 km resolution for the analysis to increase computational speed, and to match the resolution of the surface current measurements in the region. These data were analyzed in a multi-step process that identifies predominant water mass boundaries and the gradients between water masses (Figure 3.2).

3.3.1.1. Data and Standardization

The data used from the composite matrix of AVHRR and SeaWifs in this study were sea surface temperature (SST ºC), remote sensing reflectance measured at 490 nm ($R_{rs(490)}$) and at 555 nm ($R_{rs(555)}$) (Figure 3.1). Remote sensing reflectance is a quasi-

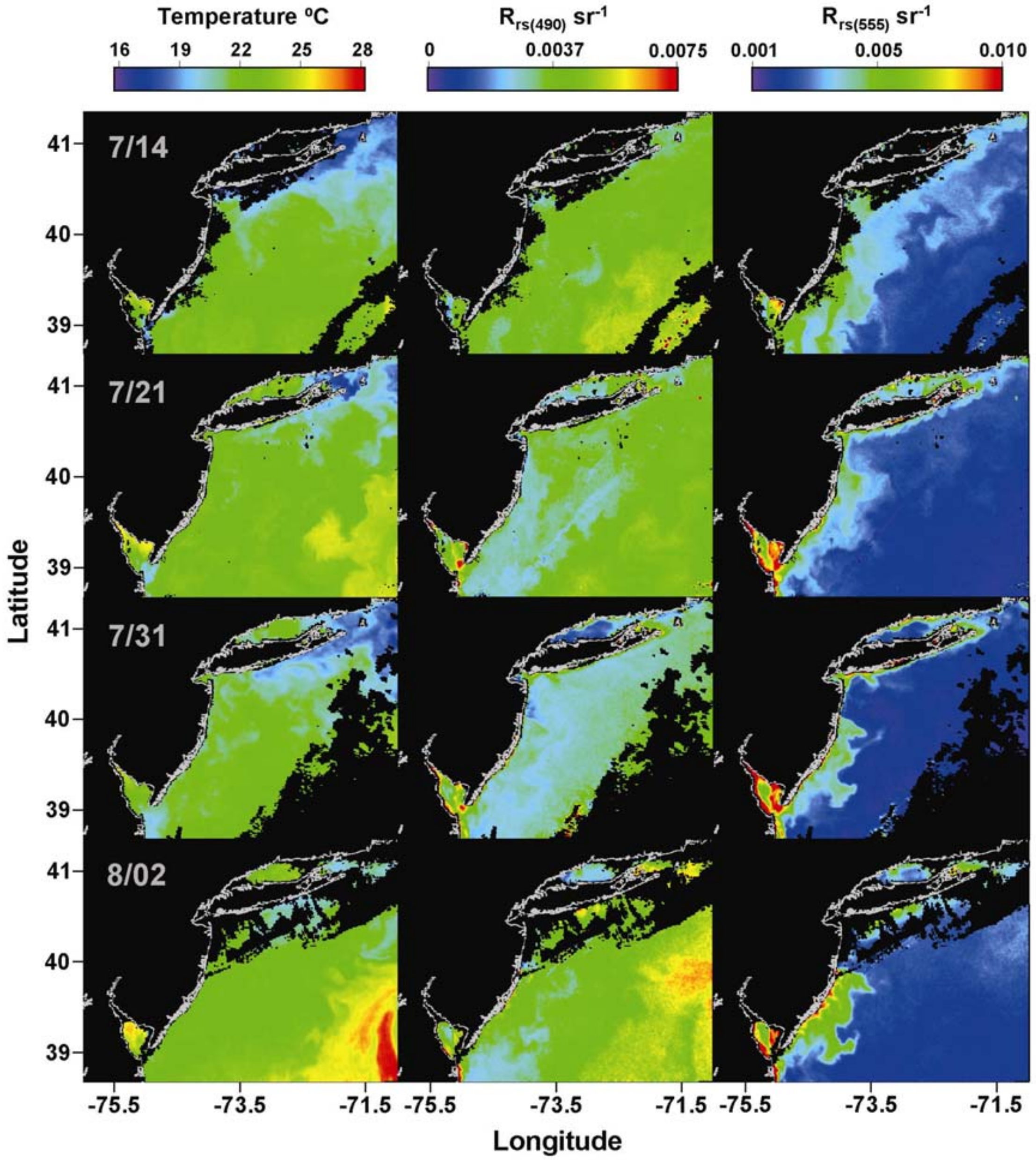


Figure 3.1. Temperature and reflectance maps on 7/14, 7/21, 7/31 and 8/02 2002 in this analysis. A warm core ring is evident on 8/02 as a nearshore optically dominated water mass formed nearshore. The white line is the coastline and the black indicates land or cloud.

inherent optical property defined as the ratio of upwelling radiance (W m^{-2} sr^{-1}) to downwelling irradiance (W m^{-2}) and has units of sr^{-1}. These data were chosen for two reasons. First, they are used in chlorophyll and primary productivity estimations. Secondly, a principal components analysis using the correlation matrix on the combined four-day data set including SST and remote sensing reflectance at 412 nm, 443 nm, 490 nm, 510 nm, 555 nm and 670 nm indicated that three linear combinations described 96.6% of the variance of the data. SST, $R_{rs(490)}$ and $R_{rs(555)}$ were the largest contributors to these linear combinations. This suggests that the majority of the waters in this analysis are Case 1 and that the other remote sensing reflecting channels are highly correlated and would not add much discrimination power. Note however, the methods described in this paper are not limited to three predictors or these specific satellite products; however in this region they represented the most useful data. Work in other areas may require some similar preliminary analysis. SST, $R_{rs(490)}$ and $R_{rs(555)}$ were standardized for this analysis by subtracting their respective means and dividing by their respective standard deviations from the combined data from the four days. This process weighted each predictor equally for any potential water mass present.

3.3.1.2. Clustering Algorithms

Four different clustering algorithms were used simultaneously in this analysis (Table 3.1). These algorithms were two agglomerative or hierarchical clustering algorithms, a K-means and a fuzzy C-means algorithm (Quackenbush 2001). From the sub-sampled data set, each pixel (observation) was projected into three dimensional standardized predictor space. The agglomerative clustering algorithms grouped observations in three dimensions according to their Euclidian distance in standardized

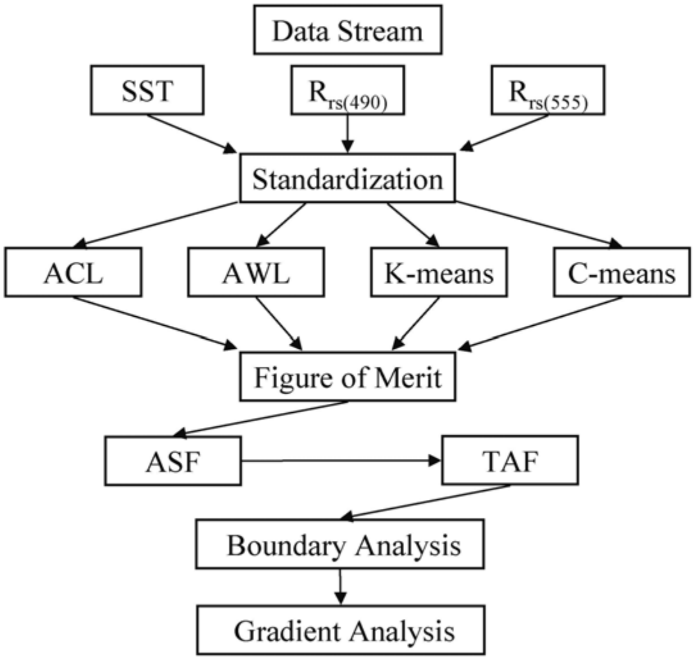


Figure 3.2. Flow diagram of this analysis. This analysis assimilates sea surface temperature as well as two remote sensing channels for all four days. The data is standardized according to the mean and variance of the combined four day data set to make them comparable. Water types for each day are detected using four clustering algorithms, ACL, AWL, K-means and C-means. These results are combined into a Figure of Merit, where an average slope function (ASF) and threshold of acceptable flatness (TAF) is computed. These two predictors give a range of reasonable water types. For each solution for each day, the boundaries are plotted and coincident boundaries are the most prevalent indicating similar structures found by different clustering algorithms. This indicates that the boundaries associated with this water type indicate a prevalent water mass. Finally, the predictor space distance is measured between each data point, to determine how different the water is on either side of each boundary. High values indicate a very strong boundary between water masses.

predictor space. The agglomerative clustering types grouped standardized predictor data hierarchically from *n* to 2 clusters from closest to furthest in predictor space where *n* is the number of observations. The difference between how the two agglomerative clustering algorithms treated the data is based on how the data was grouped in predictor space. The first agglomerative clustering type grouped data according to complete linkage (i.e. Agglomerative Complete Linkage or ACL), which determined that two clusters of data ought to be joined to a single cluster based on the maximum distance between cluster edges. The second agglomerative method grouped data according to Ward's linkage (i.e. Agglomerative Ward's Linkage or AWL) (Ward 1963). This method calculated the total sum of squared deviations from the cluster means, and joins clusters to minimize the increase of the total sum of squares deviation. The K-means clustering algorithm is a divisive clustering algorithm, which requires a user-specified cluster number. This algorithm initialized cluster centers randomly and grouped data until the within-cluster sum of squares is minimized for the number of clusters specified (Hartigan and Wong 1979). The fuzzy C-means clustering algorithm is similar to the K-means clustering algorithm except that through the use of fuzzy logic and sequential competitive learning, observations are clustered (Chung and Lee 1992).

While there are dozens of clustering schemes, these particular algorithms were chosen based on performance from the literature. (Yeung et al. 2001) observed that on real data, using agglomerative clustering with single linkage (clusters joined into a single cluster based on the minimum distance between clusters) did not produce sensible clusters of data. Rather, the K-means clustering algorithm performed very well. The ACL algorithm has been cited as very useful in producing tightly grouped clusters

Table 3.1. Description of the four types of clustering algorithms used.

Clustering Algorithm	Description
Agglomerative Complete Linkage (ACL)	Data are hierarchically grouped from *n* to 2 clusters. Data are grouped from closest to furthest based on Euclidian distance in predictor space. The distance between clusters is measured based on the maximum distance between cluster edges in predictor space.
Agglomerative Ward's Linkage (AWL)	Data are hierarchically grouped from *n* to 2 clusters. Data are grouped at each step to minimize the variance of the clusters.
K-Means	Data are divided from 1 to *k* clusters where *k* is the number of clusters requested by the user. To form *k* clusters, *k* cluster centers are randomly initialized in predictor space. Data are then assimilated into cluster centers as to minimize the within cluster sum of squares.
Fuzzy C-Means	Similar to K-means, except this algorithm clusters initial cluster centroids through competitive learning.

(Quackenbush 2001). In our opinion this is a good feature for water type identification because there is an emphasis in grouping only the most similar data. The choice of the AWL algorithm was related to previous work done by (Oliver et al. 2004), in which a priori knowledge of the number of water masses present fit well with the results of the AWL algorithm. The fuzzy C-means clustering algorithm was chosen based on the results of (Chung and Lee 1992), which showed that the competitive learning done by the fuzzy C-means algorithm produced sensible clusters.

3.3.1.3. Figure of Merit

A major difficulty in cluster analysis is determining how many clusters (or water types in this case) should be used to describe a data set as each observation could theoretically represent its own cluster. Therefore a means to analyze this structure objectively was required to identify water types in predictor space. With the advent of rapid gene sequencing and gene expression chips, the field of bioinformatics has endeavored to produce and continues to refine several algorithms that analyze gene and expression data in order to find patterns of gene expression that are linked to a variety of factors. (Yeung et al. 2001) developed and validated one such method which essentially computes the RMS deviation between individual observations and the mean of the cluster they belong too for a given algorithm. This statistic is called the Figure of Merit (*FOM*). Although this algorithm was designed to calculate the difference between expression vectors of genes, here it is used to analyze the inherent structure of clusters in predictor space detected by the clustering algorithms. In this case, "gene" expression vectors were standardized values of SST, $R_{rs(490)}$ and $R_{rs(555)}$ at each pixel. The *FOM* statistic was used to analyze the inherent structure defined by the clustering algorithms. The equation used in this study to calculate the *FOM* was:

$$FOM(c,k) = \sqrt{\frac{1}{n}\sum_{i=1}^{3}\sum_{j=2}^{k}\sum_{l=1}^{m_j}\left(\overline{a}_{ij} - a_{ijl}\right)^2} \tag{3.1}$$

where c is one of the four clustering algorithms, n is the total number of observations, i =1-3 indexes the three variables measured at each pixel, *j* is the cluster number, *k* is the number of clusters each data set was divided into, *l* is a specific observation of the total number of pixels *m* in cluster *j*, a_{ijl} is the specific standardized observation of predictor *i* in cluster *j*, and $\overline{a}_{ij}$ is the mean for each cluster. This function is essentially a measure of the variation within clusters as a function of cluster number.

Ideally, the *FOM* function will exhibit a distinct "elbow", decreasing rapidly at small *k* and much more slowly beyond a threshold *k.* This elbow represents the ideal cluster number (or number of water types in this case) for a data set because the deviation between cluster means and the individual observations in each cluster become very small. While the *FOM* statistic often show very distinct "elbows" in simulated data sets, real data sets tend to show no distinct elbow for any of the clustering algorithms (Figure 3.3, Also see Figures 1 and 3 in *(Yeung et al. 2001)*. In cases using real data, the FOM is best approximated by a power function of the number of clusters indicating that it is difficult to choose the ideal number of clusters. In this study, a threshold of acceptable flatness (TAF) of the FOM was defined by calculating the normalized average slope function ($ASF(k)$) of the FOM function at each cluster *k* for the four clustering algorithms using:

$$ASF(k) = \frac{1}{4}\sum_{c=1}^{4}\frac{FOM(c,k+1) - FOM(c,k)}{FOM_{\max}(c)} \tag{3.2}$$

where $FOM_{\max}(c)$ is the maximum FOM value for a specific cluster algorithm *c*. The TAF was defined at the smallest cluster *k* where $ASF(k)$ < 0.01 (< 1% decrease in FOM relative to the maximum FOM) for three or more consecutive clusters. Based on our own observations in which *k* was allowed to approach *n*, an $ASF(k)$ value < 0.01 indicates that the variance within each cluster no longer reduces appreciably with increasing cluster number. This established an upper bound for what we believed to be reasonable cluster numbers or water type assignments by the suite of clustering algorithms. For this study, *k* was limited to a maximum of 30 clusters, as the FOM value did not change significantly after this cluster number.

3.3.1.4. Boundary Analysis

One major difference between the clustering of a gene data set and a water mass data set is that clusters defined in a water mass data set occupy predictor space represented by standardized SST, $R_{rs(490)}$ and $R_{rs(555)}$ and physical space represented by latitude and longitude while a gene data set has no physical space representation. Water mass definitions vary slightly, so for the purposes of this analysis, our definition of a water mass is that it must occupy physical space, and water with similar properties in separate physical spaces represent different water masses. The spatial attributes of water masses provide additional useful information not generally associated with genes, and provide a useful means in delineating the physical boundaries between waters that have similar properties identified by the cluster analysis. The mapping of defined water types for any cluster number k and clustering algorithm c into physical space (this case in dimensions of latitude and longitude) defines physical boundaries between similar water types. Because each of the clustering algorithms is slightly different, the boundaries described at any specific cluster number k between water types may be different. However, it was clear that different clustering algorithms often had similar boundary solutions at different cluster numbers. This is because different water types were differentiated at slightly different cluster numbers due to differences in the clustering algorithms. Because of this a physical space representation of the clusters was used to determine which boundaries occurred most often by constructing a 2-d histogram for boundaries at $2 \le k \le TAF$. To detect the most common water mass boundaries for any cluster number, the cluster number gradient in latitude and longitude space was computed using:

$$\nabla C_{xykc} = \sqrt{\left(\frac{C_{xykc} - C_{x+\Delta x,ykc}}{\Delta x}\right)^2 + \left(\frac{C_{xykc} - C_{y+\Delta y,xkc}}{\Delta y}\right)^2} \tag{3.3}$$

where x is Longitude, y is Latitude, C_{xykc} is the cluster number assignment for k clusters for c clustering algorithm and ∇C_{xykc} is the magnitude of the cluster number

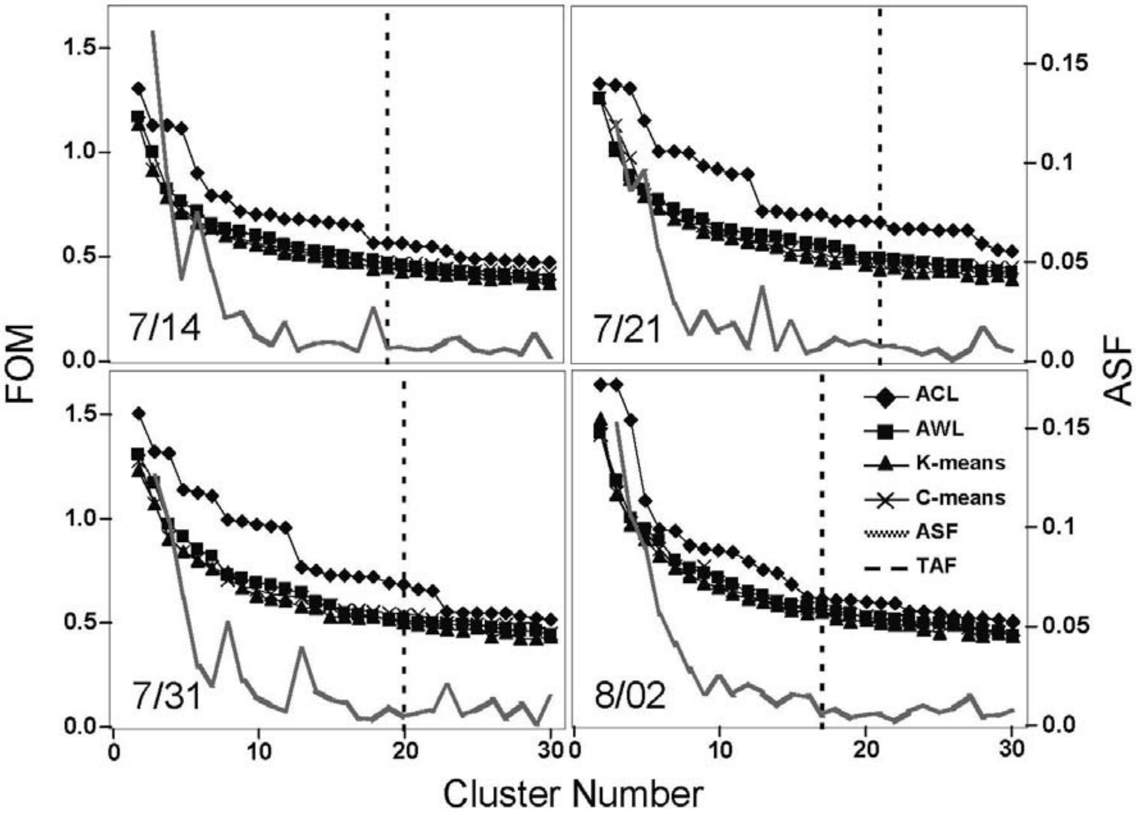


Figure 3.3. Figure of Merit (*FOM*), Average Slope Function (*ASF*) and Threshold of Acceptable Flatness (*TAF*) calculation for each of the four days with the results of each of the clustering algorithms. A large *FOM* indicates that the variance within each cluster is comparatively large and that the cluster centroid is a generally poor predictor of the other data points within each cluster. A small *FOM* indicates that the cluster centroid better predicts the other members of its cluster, and that the variance with in the cluster is comparatively small. *ASF* is the average percent change of the four clustering algorithms compared to the maximum *FOM*. *TAF* was defined when the average change in the *FOM* was less than 1% for more than three clusters.

gradient vector. Where ∇C was non-zero, it was replaced with a logical value of 1 to indicate the presence of a boundary using:

$$b_{xykc} = \begin{cases} 1 \text{ if } \nabla C_{xykc} \neq 0 \\ \\ 0 \text{ if } \nabla C_{xykc} = 0 \end{cases} \tag{3.4}$$

where b_{xyck} is the logical boundary value for a given longitude and latitude for the given cluster algorithm for k clusters. Although it is nonsensical to calculate gradients of categorical data, this method effectively detects the boundaries of the water masses. A 2-D histogram was constructed of high frequency boundaries for each of the four days using:

$$B_{xy} = \frac{\sum_{c=1}^{4}\sum_{k=2}^{TAF} b_{xyck}}{4(TAF-1)} \times 100\% \tag{3.5}$$

where B_{xy} is the frequency that a boundary (0-100%) at a given longitude and latitude. This 2-D histogram describes the most common physical boundaries between similar water types defined by the clustering algorithms. The presence of a high frequency boundary was interpreted as a boundary between separate water masses.

3.3.1.5. Gradient Analysis

In addition to determining where the major water mass boundaries are, the relative strengths of these boundaries were also estimated. Theoretically, water types could be distinctly separated in predictor space, but still be relatively close to each other in predictor space. In this case a boundary on a physical map between these water types would be drawn frequently between these distinct water types, while their differences would still be relatively minor. The purpose of the gradient analysis was to determine how different water types were in predictor space in relation to geographic space. The relative strength of the boundaries was defined as:

$$D_{x\to x+\Delta x} = \sqrt{\left(SST'_x - SST'_{x+\Delta x}\right)^2 + \left(R_{rs(490)}{'}_x - R_{rs(490)}{'}_{x+\Delta x}\right)^2 + \left(R_{rs(555)}{'}_x - R_{rs(555)}{'}_{x+\Delta x}\right)^2} \qquad (3.6)$$

$$D_{y\to y+\Delta y} = \sqrt{\left(SST'_y - SST'_{y+\Delta y}\right)^2 + \left(R_{rs(490)}{'}_y - R_{rs(490)}{'}_{y+\Delta y}\right)^2 + \left(R_{rs(555)}{'}_y - R_{rs(555)}{'}_{y+\Delta y}\right)^2} \qquad (3.7)$$

$$\nabla G(x,y) = \sqrt{\left(\frac{D_{x\to x+\Delta x}}{\Delta x}\right)^2 + \left(\frac{D_{y\to y+\Delta y}}{\Delta y}\right)^2} \qquad (3.8)$$

where SST' is standardized sea surface temperature, $R_{rs(490)}{'}$ is standardized R$_{rs(490)}$, $R_{rs(555)}{'}$ is standardized R$_{rs(555)}$, $D_{x\to x+\Delta x}$ is the standardized predictor space distance between x and $x+\Delta x$, $D_{y\to y+\Delta y}$ is the standardized predictor space distance between y and $y+\Delta y$, and $\nabla G(x,y)$ gradient in predictor space with respect to x and y. While the boundary analysis determines likely locations of water mass boundaries, $\nabla G(x,y)$ describes the strength of boundaries through simultaneous analysis of SST, R$_{rs(490)}$, and R$_{rs(555)}$.

3.3.1.6. Current Structure of the Region

Surface current maps, measured by an HF radar system, provide a dynamical context in which to evaluate the placement of water mass boundaries. The long range HF radar system used here was first deployed in 2001 (Kohut and Glenn 2003), and consists of four remote transmit/receive sites along the coast of New Jersey and a central processing site in New Brunswick, New Jersey. Using the scatter of radio waves off the ocean surface each remote site can measure the surface current component moving toward or away from the site (Barrick et al. 1977). Information from all four remote sites is then geometrically combined at the central site to provide a total vector current map.

The systems are operating at a frequency of about 5 MHz, which provides range out to 200 km offshore, a total vector grid resolution of 6 km and a surface current averaged over the upper 2.5 m of the water column. Each current map is a three hour average. For this analysis, the three hour data was averaged for the days July 21, 31, and August 2. July 14 current data was not yet available. If a particular range cell did not have at least 60% coverage over each day, the current vector in that range cell was not used in the analysis. A simple drifter experiment, which modeled 48 drifters along a boundary on 7/31, was used to determine if local advective processes could explain the changes in the boundary location during these days. This exercise attempts to predict the frontal location 51 hours later on 8/02. The current field was interpolated to the position of each drifter. The three hour average current maps were assimilated sequentially. At hourly intervals, the location of the drifter was evaluated and a new vector was assigned to the drifter. At three hour intervals a new current map was assimilated.

3.3.2. Independent Verification during LaTTE

During the 2005 LaTTE experiment in the New York Bight, we independently tested our approach by comparing surface measurements of salinity and inorganic nutrients with the water mass boundaries predicted by the cluster analysis. In addition, the Ocean Color Monitor (OCM) satellite replaced the SeaWiFS satellite as the source for ocean color. For this study, matched satellite passes from April 13, 2005 were chosen because of relatively little cloud cover. The area analyzed was between 39.50°N – 40.65°N latitude and 72.25°W – 73.2°W longitude. These data were analyzed according to the method outlined in Figure 3.2. Salinity measurements were made by flow through systems on board the research vessel R/V Hatteras. Nitrogen and silica measurements were made on an Envirotech Autolab continuous nutrient analysis system modified for 12-13 min analysis times using standard methods (Strickland and Parsons 1972). Salinity and nutrient measurements were overlaid on the boundary maps produced by the clustering algorithm to determine if the boundaries detected also represented changes in water column variables not directly used in the clustering algorithm.

3.4. Results

3.4.1. HyCODE

This study focused on a series of four composite images of SST, $R_{rs(490)}$ and $R_{rs(555)}$ from July 14 to August 2, 2001. During this period, a phytoplankton bloom developed in the northern portion of the study site and dispersed alongshore to the south (Moline et al. 2004). Offshore, part of a Gulf Stream warm core ring was observed on August 2 as it propagated from east to west (Figure 3.1). The phytoplankton bloom may have been associated by terrestrial runoff and was sustained by several upwelling events. Outflow from the Hudson River, one of the largest sources of terrestrial runoff in this region, measured at the Waterford NY site prior to the satellite passes was up to a factor of 2 larger than the 25 year mean during that time period (Figure 3.4). (Yankovski and Garvine 1998) have shown that the time lag of these outflows to reach the study area is approximately 40 days, which coincides with the time with a large outflow from the Hudson River of this study (approx. June 4). In addition, this time period had several upwelling favorable wind patterns on or around July 19, 26 and 30. These upwelling wind events are regular in this region and stimulate phytoplankton growth (Schofield et al. 2002) (Moline et al. 2004).

3.4.1.1. Evaluation of the Figure of Merit

For each of the days, FOM was calculated from k = 2 to 30 clusters for the four clustering methods (Figure 3.3). These FOM functions were generally decreasing with increasing cluster number in all cases and were similar to those found by (Yeung et al. 2001) in that no distinct "elbow" was obvious. In all FOM cases, the ACL clustering algorithm was slightly higher than the other three clustering algorithms. While not producing exactly the same FOM statistic, the AWL, K-means and C-means clustering algorithms were very similar within days. FOM curves between days were similar in shape, however they differed slightly in magnitude. The $ASF(k)$ function for these days showed the most rapid decrease occurred where $k < 10$. In addition, all of the $ASF(k)$ functions display erratic changes in value where $10 < k < 15$. For $k > 15$, the $ASF(k)$ functions in all four days flattened noticeably. The TAF value for 7/14, 7/21, 7/31, and 8/02 were 19, 20, 24 and 20 clusters respectively. These values served as the upper bound for the boundary analysis.

3.4.1.2. Location and Strengths of Common Water Mass Boundaries

The FOM analysis of the water types defined by the four clustering algorithms indicated that the "ideal" number of water types (clusters) was in the range of $2 \leq k \leq TAF$. For each c and k, k water types were defined that had boundaries described by equations 3 and 4 in physical space. Equation 5 is the frequency of these boundary observations across all c and k. A boundary frequency map (B_{xy}) was computed for each of the four days (Figure 3.5). In general, water mass boundaries become more defined from 7/14 to 8/02. The most frequent boundaries are associated with strong optical or temperature fronts. Figure 3.6 illustrates the boundary frequency differences between the

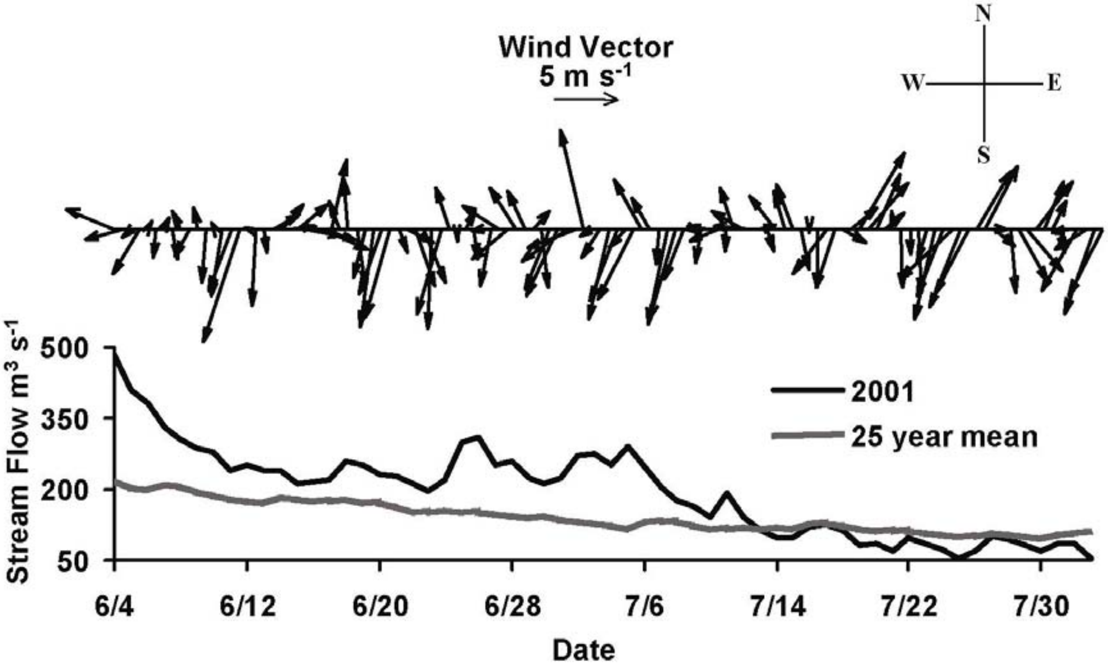


Figure 3.4. Wind record from the RUMFS field station and Hudson River flow recorded at Waterford NY during the study time period. From 7/14 to 8/02 there were three upwelling favorable events that may have sustained phytoplankton growth near-shore. The elevated stream flow during this particular year recorded at Waterford NY may have initiated the formation of a Hudson River derived water mass during the four day study period. It has been reported that water outflow from this area takes 40 days to reach the Southern New Jersey shore (Yankovski and Garvine 1998).

four days. As a function of total boundaries drawn on a map, high frequency boundaries (B_{xy} > 60%) were more spatially common on days 7/31 and 8/02 compared to 7/14 and 7/21. Also, low frequency boundaries (0% < B_{xy} < 20%) are more common on days 7/31 and 8/02 compared to 7/14 and 7/21. These two conditions cause the 7/31 and 8/02 B_{xy} maps to appear more cleanly defined. In contrast, medium frequency boundaries (20% < B_{xy} < 60%) were more common on 7/14 and 7/21 compared to 7/31 and 8/02, causing the 7/14 and 7/21 maps to appear more cluttered. On 7/21, 7/31 and 8/02, when boundaries are more distinct, the major water masses are associated with the near shore plume, shelf water, water east of the shelf break front and the warm core ring.

The objective of the cluster analysis was to describe the inherent structure and separation of water types in predictor space, which was then mapped in the form of boundaries in figure 3.5. The purpose of the gradient analysis was to determine how different water types were in predictor space in relation to geographic space. Figure 3.7 is the application of equations 6, 7, and 8 to evaluate the relative strengths of the boundaries between water masses. Because each pixel is slightly different from its neighbors, the gradient is never zero. The median value for this gradient calculation for this study is approximately 10, with a standard deviation of about 10. Therefore a strong gradient has a value in excess of 20 for this study. On 7/14 and 7/21 gradients between water masses defined in the boundary analysis are relatively weak indicating that the water types found in these days are fairly similar. In contrast, strong gradients were found associated with the nearshore optical front. These relatively strong gradients are coincident with the high frequency boundaries described in figure 3.5 indicating that these particular water types

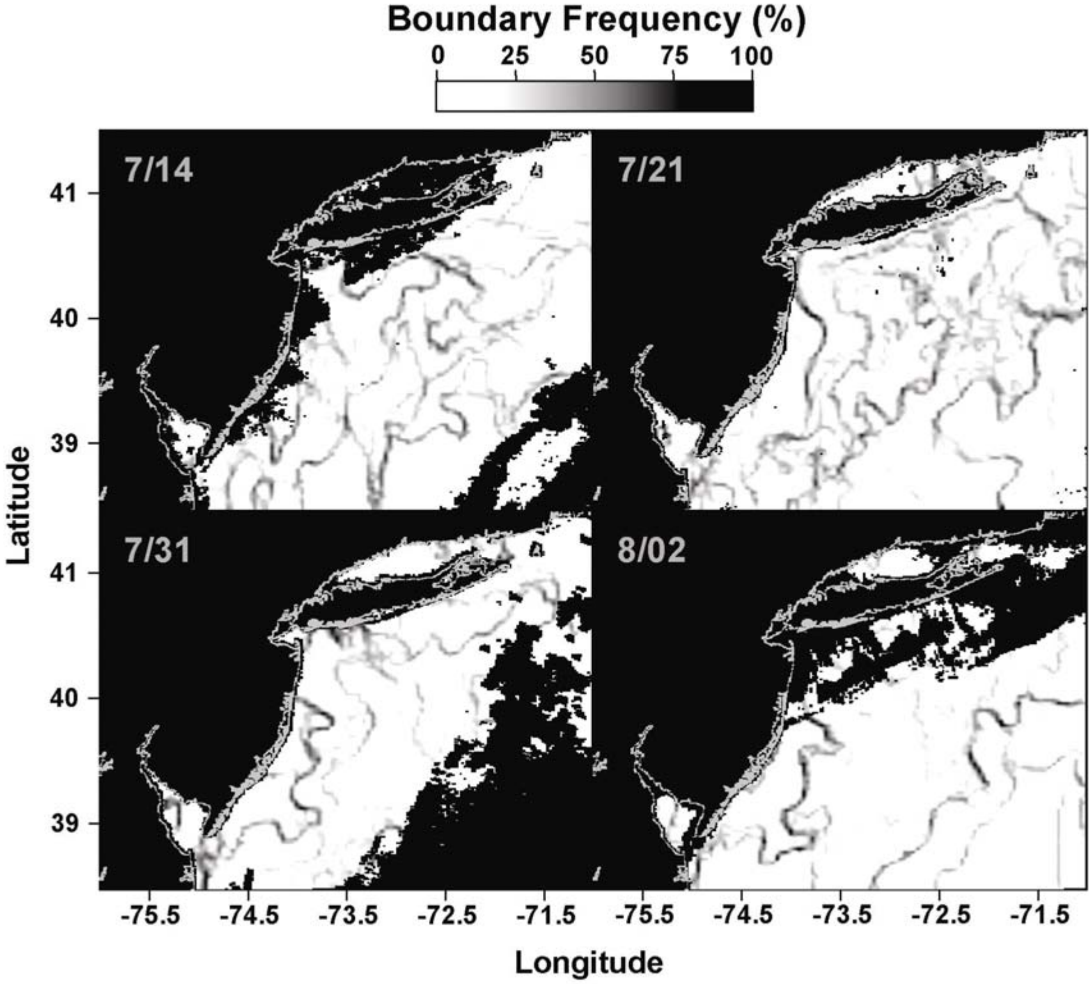


Figure 3.5. High frequency boundary locations as calculated from equation 5. The contrast indicates how often a particular pixel was designated as a boundary. The most frequent boundaries represent water types that are easily separable in predictor space. Boundaries become more distinct from 7/14 to 8/02.

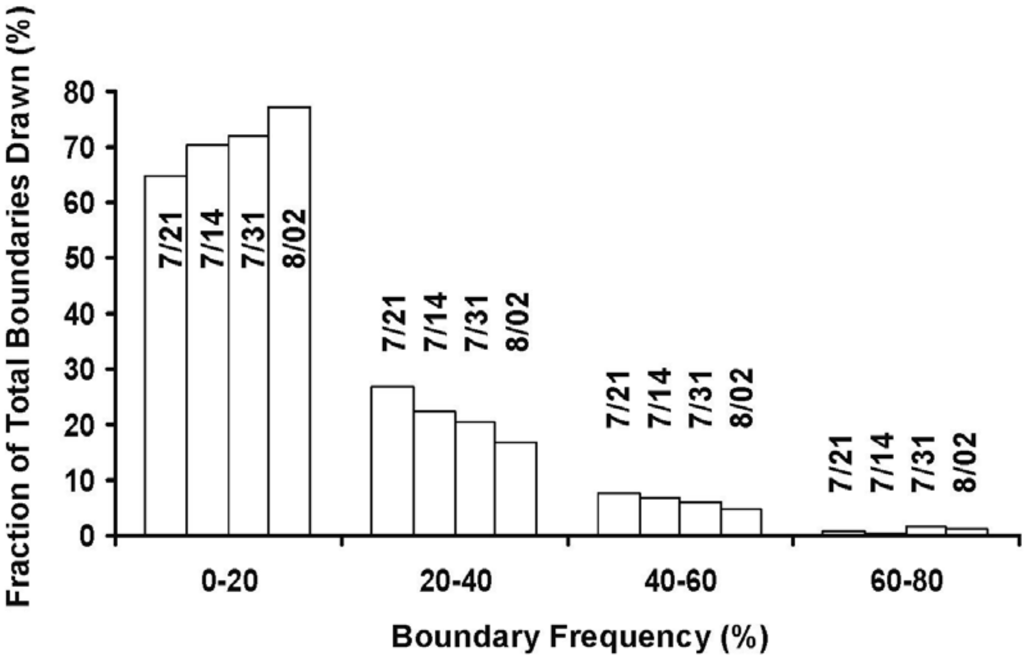


Figure 3.6. The boundary frequency calculated by equation 5 related to the total number of boundaries drawn. The days with more disorganized boundaries (7/14 and 7/21) have less low and high frequency boundaries and more medium frequency boundaries. This causes the disorganized look on these days and indicates that the clustering algorithms had a difficult time coming to similar solutions. Days 7/31 and 8/02 had more low frequency and high frequency boundaries and low medium frequency boundaries indicating that the clustering algorithms were in agreement more often and that water types were consistently distinguished.

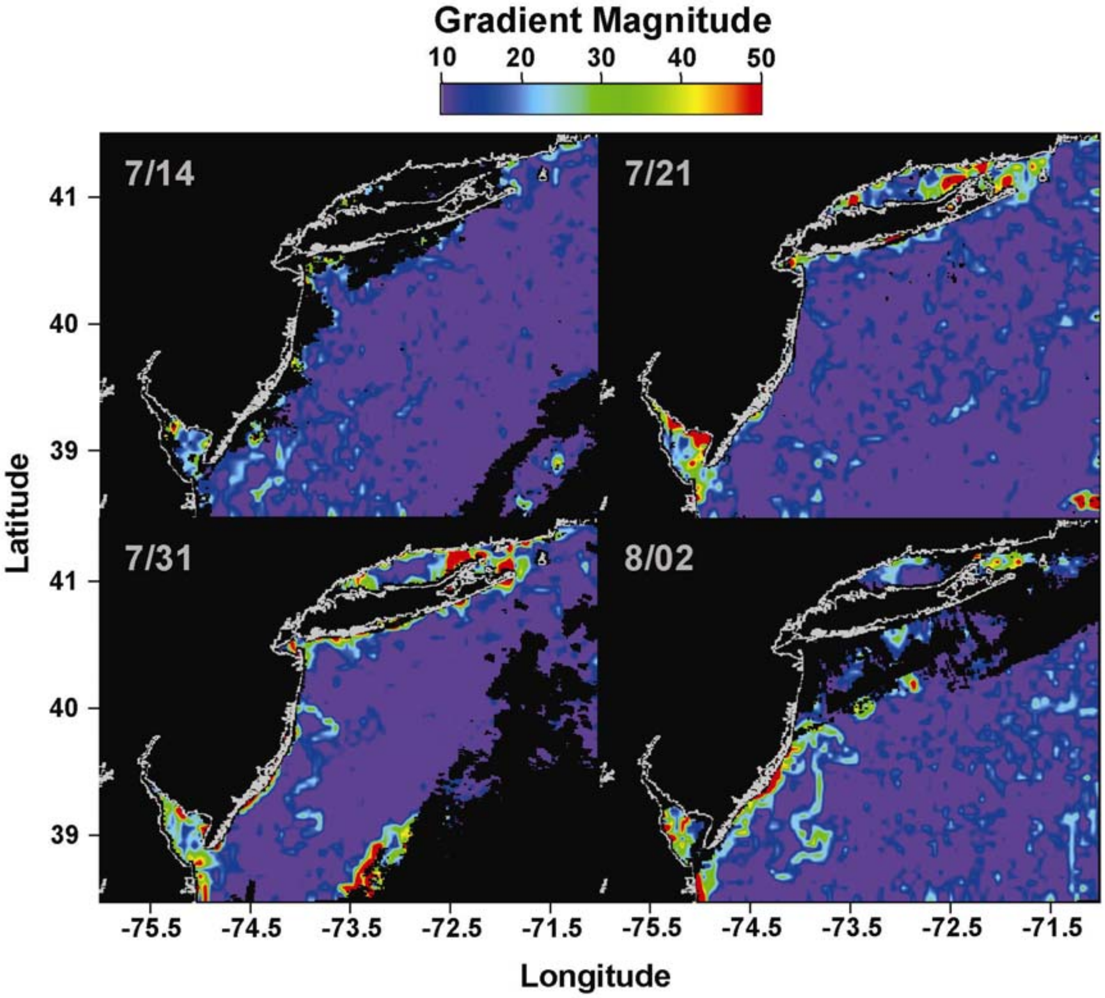


Figure 3.7. The gradient defined by equations 6, 7 and 8. The gradients are a relative measure of how different adjacent water masses are. Because no two adjacent pixels are equal, the gradient is never zero. The background gradient value for this study is approximately 10, with a standard deviation of approximately 10. Gradient values larger than 20 in this study are considered to be significant. Stronger gradients were evident in days 7/31 and 8/02. This indicates that the water types on either side of the boundary are markedly different. However, strong gradients are not necessarily coincident with high or medium frequency boundaries because two water types may be readily distinguishable in predictor space but still be relatively close to one another.

are structurally distinct and very different. In addition, strong gradients were detected near clouds which may be a result of inadequate cloud masking.

3.4.1.3. Surface Current Structure, Gradient Strengths and Boundary Locations

The seasonal mean flow in the summer time in this region is along shore toward the south (Kohut and Glenn 2003), which was generally observed in the three-hour average flow on 7/21, 7/31 and 8/02. However, the flow structure on these dates was highly variable. The current fields in figure 3.8 represent the flow field at the time of the satellite over pass with the spatial mean subtracted from it. This was done to visually enhance the fine scale current structure associated with the water mass boundary gradients. Generally speaking, gradients were associated with physical features in the flow fields such as horizontal sheer, indicating that these features were strongly influenced by advective processes. However, the strength of the gradient was not related to the strength of the horizontal sheer, nor were all horizontal sheers associated with gradients.

To determine if the apparent movement of the boundary was associated with physical advection, a simple simulated drifter experiment was performed (Figure 3.9). 48 modeled drifters were placed along the frontal boundary on 7/31 and sequentially assimilated the surface current fields in hourly time steps. The predicted position of the major boundary feature was generally in good agreement with the location of the boundary on 8/02. The predicted boundary has a more pronounced “hammer-head” appearance much like that of the boundary on 8/02. In addition the northern protrusion of the front moved southward, approximating its location on 8/02. Because the predicted position of the boundary region approximates the location of the boundary on 8/02, it

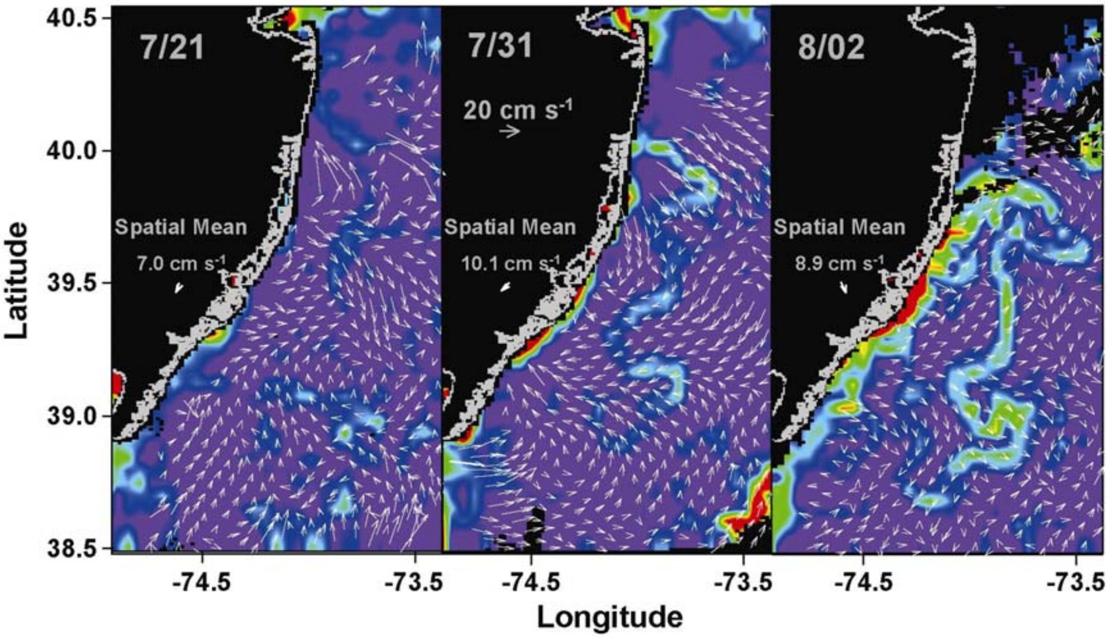


Figure 3.8. Boundary gradients overlaid with surface current fields with the surface current spatial mean subtracted for visual clarity. Areas with larger gradients are coincident with convergent and divergent areas indicating that local current structure accounts for the gradient locations. However, not all convergent areas had gradients associated with them.

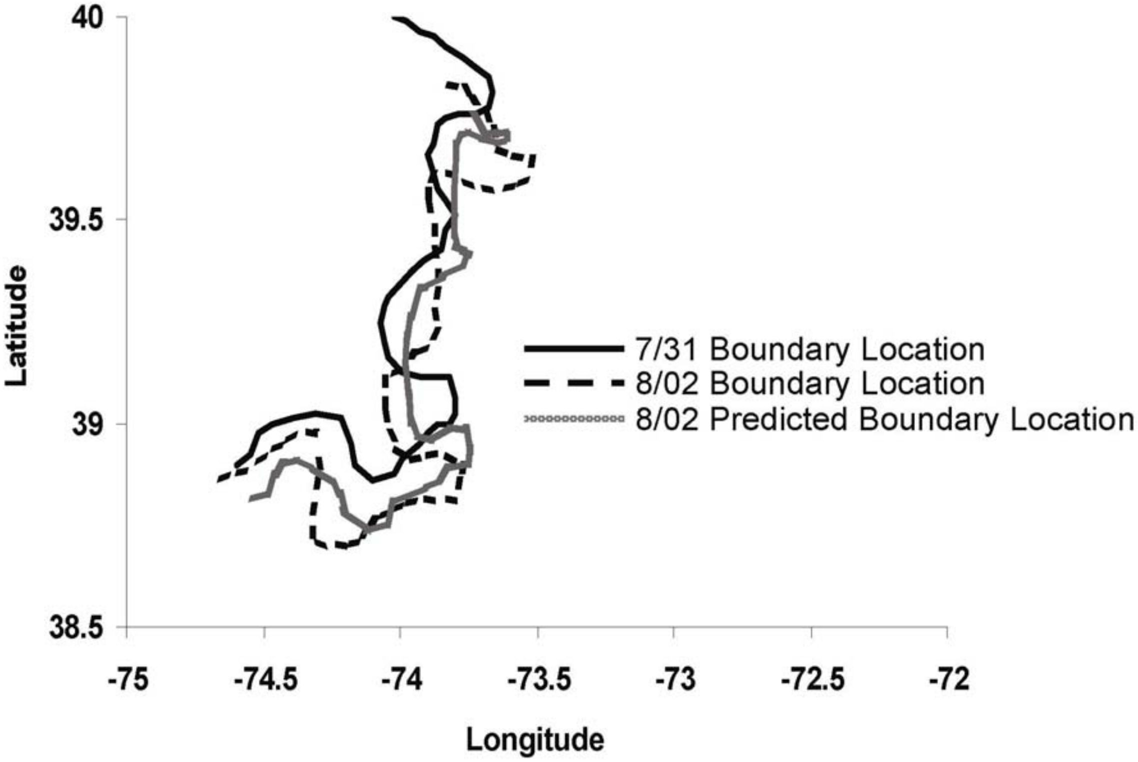


Figure 3.9. Results of the simulated drifter experiment. The predicted location of 48 drifters on 8/02 based on the initial position of the 7/31 boundary by assimilating the CODAR measured surface currents generally approximates the location and shape of the boundary on 8/02. This indicates that the apparent movement of the boundary can be generally attributed to local advective processes. Also, this indicates that water masses in this area can be tracked effectively.

suggests that local advection processes are largely responsible for changes between 7/31 and 8/02.

3.4.2. LaTTE salinity and nutrient measurements

In the days prior to 4/13/2005, the Hudson River set a 25-year record for freshwater discharge creating a complex mix of marine and fresh waters (Chant 2005). Boundary analysis of the merged OCM and AVHRR data in that day revealed highly convoluted boundaries, indicating complex interaction between the fresh and marine waters. As expected, underway salinity measurements indicate that in general water is freshest near the mouth of the Hudson River. However, salinity measurements appear to change abruptly across the water mass boundaries calculated from satellite measurements. There appears to be a fresh water “bulge” in the bight apex, as well as a coastally trapped freshwater plume along the New Jersey shore (Figure 3.10). These abrupt changes appear to be coincident with the water mass boundaries detected from space.

Nitrate and silicate measurements on this day also show large changes in concentration across water mass boundaries. Nitrate is high within the “bulge” at the bight apex, but drops to average oceanic marine levels outside the bulge indicating that this region represents a strong nutrient gradient (Figure 3.11). Silicate is also higher within the “bulge”. Interestingly, silicate is also relatively high in an offshore water mass (center of Figure 3.12). This water mass is identified as an older river plume, making its way off the shelf ((Chant 2005). Nevertheless, measurements of nitrate and silicate appear to corroborate the placement of water mass boundaries.

3.5. Discussion

AVHRR and ocean color satellite products are used to measure or infer several ocean processes. These include the tracking of the Gulf Stream (Auer 1987), the modeling of Gulf Stream rings (Glenn et al. 1990) and to estimate global ocean primary production (Behrenfeld and Falkowski 1997). New production in an ocean system has also been estimated through the combination of AVHRR and ocean color (Sathyendranath et al. 1991). To estimate new production, water types were defined intuitively, to which an idealized biomass profile was assigned. Conceivably, errors could be introduced in this type of approach if the way in which water types were defined was incorrect. (Karabashev et al. 2002) addressed the water type problem through K-means cluster analysis of SeaWiFS data, however the number of clusters chosen (k = 20) was subjective.

More recently, (Martin-Trayovski and Sosik 2003) have shown very convincingly that there exist distinct optical water types in the Mid-Atlantic Bight region, and that they can be successfully discriminated. Their study developed a feature-based classification based on remote sensing reflectance in three wave bands and used a training set of data with known water types to develop classifiers. The method was evaluated on the ability of the classifiers to properly classify pixels into the correct categories. A goodness of fit measure was used as a measure for determining how variable the water is within each water mass. This method works very well if some *a priori* knowledge about the water types or water masses present is available. The *FOM* approach builds on this technique and does not require a training set of data, nor prior knowledge of the water masses present, as it strictly looks for inherent structure in the data. Additionally the method

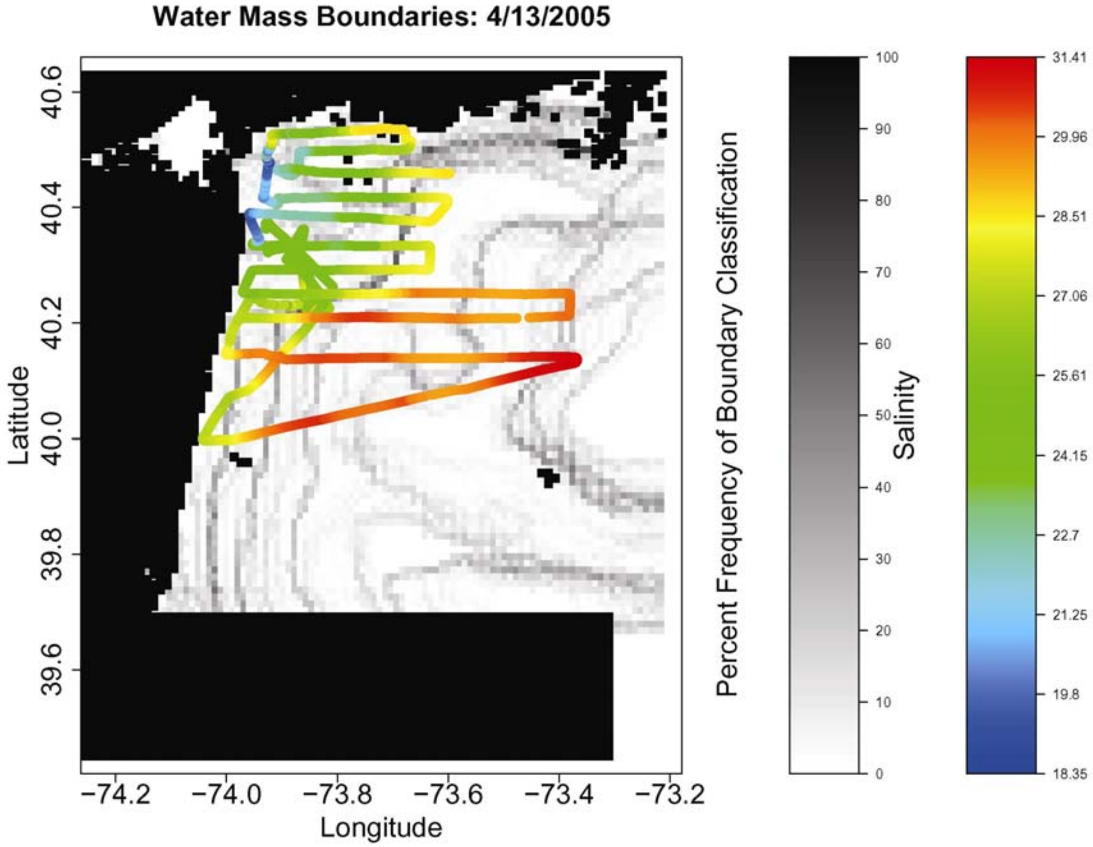


Figure 3.10. Independent test of boundary analysis with salinity as an independent predictor during the LaTTE 2005 experiment. Black areas indicate areas of cloud contamination or land. Boundary analysis predicted a large bulge in the bight apex as well as a coastally trapped river plume along the New Jersey shore. Salinity transects by the R/V Hatteras confirm these boundaries.

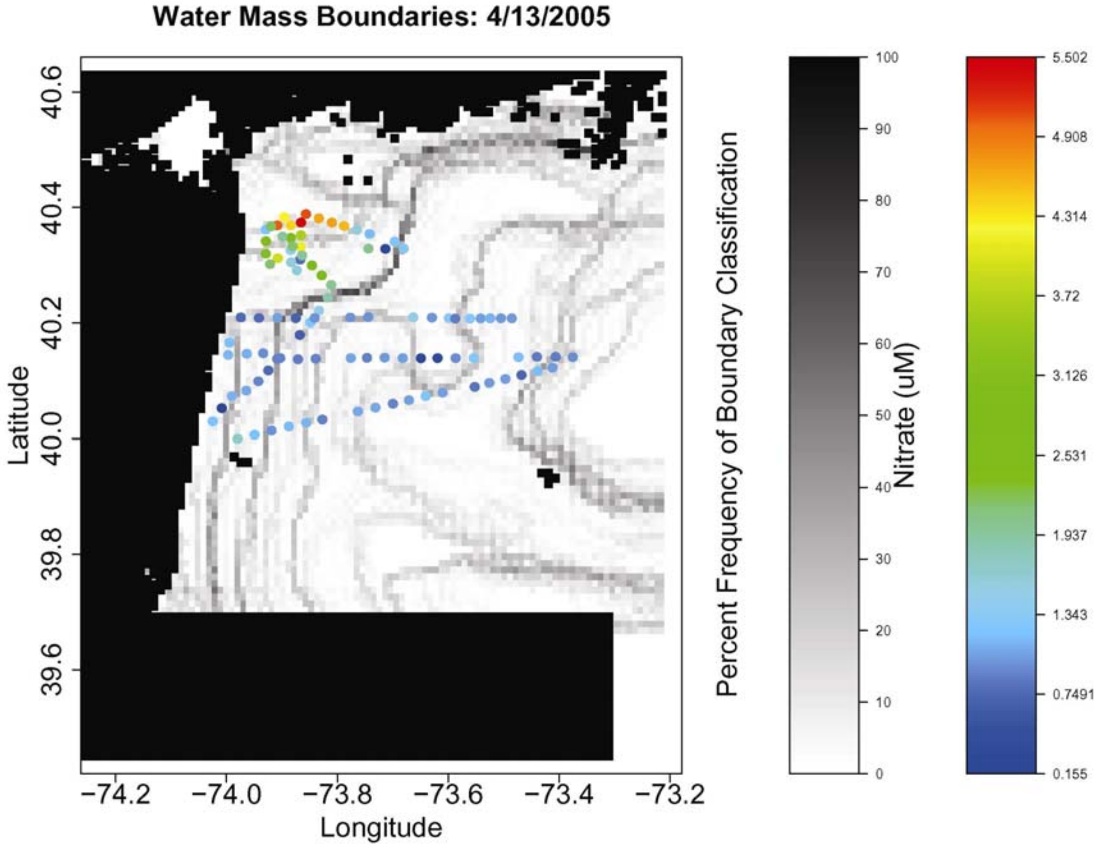
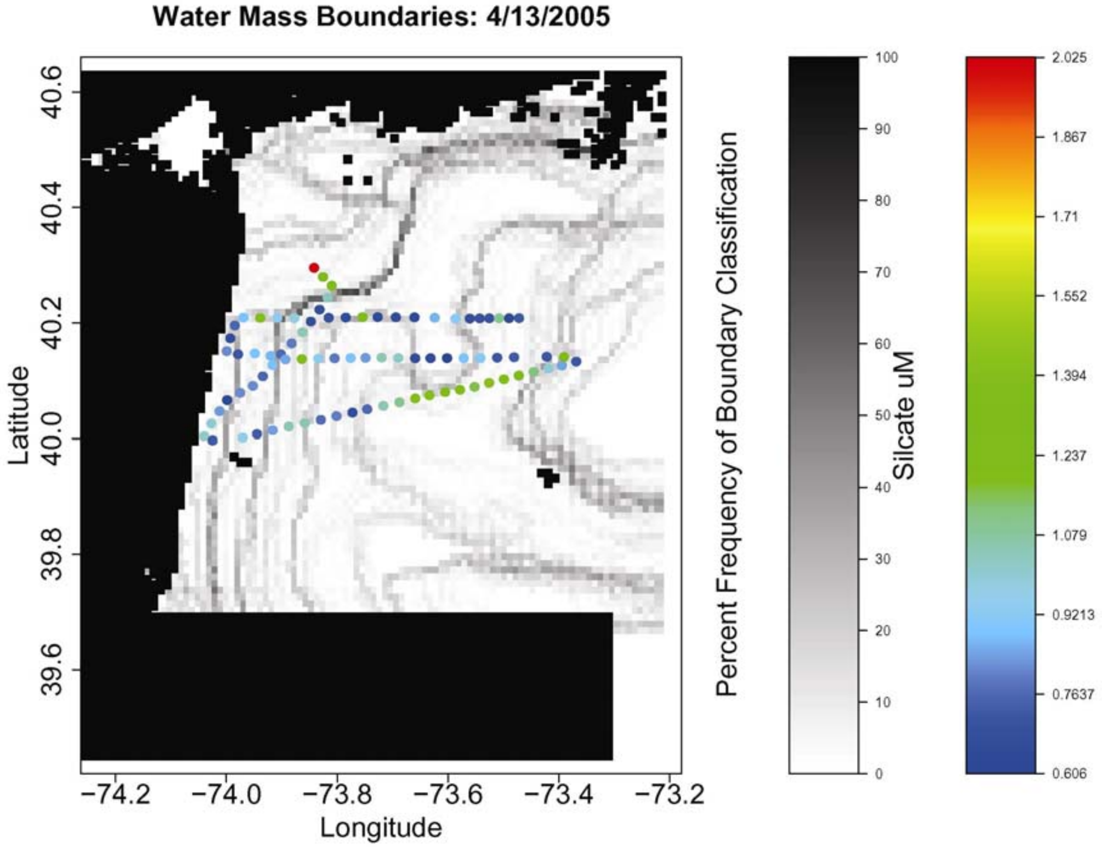


Figure 3.11. Independent test of boundary analysis with nitrate as an independent predictor during the LaTTE 2005 experiment. Black areas indicate areas of cloud contamination or land. This analysis indicates that nitrate levels in the "bulge" are very high while on the other side of the bulge boundary they are typical of open ocean values.



<u>Figure 3.12.</u> Independent test of boundary analysis with silicate as an independent predictor during the LaTTE 2005 experiment. Black areas indicate areas of cloud contamination or land. This analysis indicates that nitrate levels in the "bulge" are very high while on the other side of the bulge boundary they are typical of open ocean values. In addition, there are relatively high values of silicate present in an old river plume that is being advected offshore.

allows for the estimation of the strengths of the fronts between water types in physical space and temporal changes in boundary locations due to local advective processes. The (Martin-Trayovski and Sosik 2003) method provides a solid foundation for water mass classification from space, and complements this effort as the methods could be run in conjunction to elucidate water mass characteristics based on derived satellite products.

In general, the water masses detected in this study were a near shore plume, a water mass over the continental shelf separated by the shelf-break front, water offshore the shelf break front and a warm core ring. As for their origins, we can only speculate as satellites only detect their surface expressions. The near shore water mass is most likely from the Hudson River, but it could also be upwelled water driven by southwest winds (Glenn et al. 2004). The origin of the shelf water is from glacial melt along the southern Greenland coast that flows south to the MAB as a buoyant coastal current (Beardsley and Winant 1979) (Chapman and Beardsley 1989). Beyond the shelf break, water masses and the warm core ring reflect the Gulf Stream and or the Sargasso Sea.

This approach to water mass classification has five basic steps: i) project predictors measured for each water parcel into standardized predictor space; ii) use a suite of clustering algorithms to detect clusters in multi-dimensional predictor space data which are analogous to water types; iii) use the FOM statistic to determine a reasonable range of how many water types exist; iv) map water types into geographic space and determine the most frequent boundaries between water masses; v) evaluate the difference between water types in predictor space as a measure of the difference or gradient between defined water masses. What this analysis provides are means that validate and add mathematical rigor to intuition about the water masses present in this study. The remaining portion of the paper will discuss the factors that must be considered when interpreting the water mass boundaries and gradients calculated by this analysis.

3.5.1. Standardization of Variables

The three predictors were standardized to their respective means and standard deviations so that the variation observed in each predictor gets equal weight in this analysis. Without this standardization, temperature alone would have dominated the results because it is numerically on the order of 10^1 units while R_{rs} is numerically on the order of 10^{-3} units. However, in doing this the water mass boundaries and gradients can only be compared within the group that was standardized, in this case the four days presented here. This is an important consideration in interpreting the results of the algorithm. Large gradients and frequent boundaries surround the obvious optical load seen on days 7/31 and 8/02 in $R_{rs(555)}$ because it represented a large change in optical predictors compared to all of the data in this analysis. While this bloom is a distinct feature for those four days, if the question were whether this feature is distinct compared to a seasonal trend or annual trend, the four day data set would need to be standardized to the mean and variability of the season or year. The same principle applies for a comparison of these images to images taken in another location or in reference to larger regions. For example, for a comparison of the gradients in this image to dynamics in another coastal region, the mean and variability of both regions would have to be included for proper comparison. While this nearshore optical load may be very distinct in the context of these four days in this particular region, its distinctness seasonally or annually in this region may be different depending on the inherent mean and variability of the system.

While standardization of the variables is important for interpretation of the results, it is also important to note that standardization of the data does not guarantee that the data are normally distributed. Examining figure 3.1, one can see that the temperature and the $R_{rs(490)}$ are fairly normally distributed (i.e. the area with high values is approximately equal to area with low values, and the majority of the area is covered with mid-range values). In the case of $R_{rs(555)}$, most of the area is covered with low values and only a small area near shore is covered with high values. This means that the data have a slightly skewed distribution. Therefore, in predictor space, despite standardization of this particular data set, there is a larger range of data along the $R_{rs(555)}$ axis, thus waters with high $R_{rs(555)}$ values in this study are more easily discriminated in parameter space.

3.5.2. Predictor Space Structure, Frequent Boundaries and Gradients

The suite of clustering algorithms was used to detect the inherent structure or water types in predictor space represented in four composite data sets of SST, $R_{rs(490)}$ and $R_{rs(555)}$. For increased computational speed clusters were defined from from 2 to 30, however it is mathematically possible to define *n* water types where each observation is unique. This is the challenge associated with categorizing a known continuum of data; it is difficult to determine how different an observation of SST, $R_{rs(490)}$ and $R_{rs(555)}$ should be before it is considered a separate water type. The *FOM* statistic provides a mean to address this problem. While not providing a definitive answer as to how many water types existed in this data set, it did reduce the range of possibilities from *n* water types to 2-*TAF* water types. The geographic distribution of water types detected by the clustering algorithms between 2 and *TAF* is illustrated in figure 3.5. The significance of high frequency boundaries in this figure is that they represent consistent divisions of water types detected by more than one clustering algorithm at more than one cluster number (*k*). In essence, the four clustering algorithms vote by majority of what data in predictor space determine the dominant water types. However, because this technique uses the similarity of solutions by different clustering algorithms to determine dominate boundaries of water masses, the dissimilar solutions, which represent the low frequency boundaries in figure 3.5, represent somewhat of a "forced" result due to low signal.

While boundaries may be consistently reflecting recognizable water types in predictor space by the clustering algorithms, the frequency of boundaries is not necessarily related to the gradients separating the water masses. For example, on 7/14 several high frequency boundaries were present indicating that the clustering algorithms were finding consistent structure in predictor space indicating discrete water types. However, gradient analysis of that same day indicates that while distinct water types are present in the data set, the differences between them are relatively small. This is different than days 7/31 and 8/02 when the most frequent boundary also reflected a strong gradient. Therefore, for complete interpretation of water mass characteristics, both frequency of boundaries and gradient strengths must be considered. For example, a high frequency water mass boundary is calculated on 7/21 at approximately 40°N, 73°W which is the same frequency as the water mass boundary calculated for the nearshore "hammer-head" shape on 7/31 and 8/02 (Figure 3.5), however the gradient calculated for this boundary (Figure 3.7) is weak compared to gradients found on 7/31 and 8/02. This result indicates that the boundary on 7/21 is separating distinct water types in predictor space, however the water masses represented by these water types are not nearly as different as the water masses separated along the "hammer-head" shape on days 7/31 and

8/02. A distinct frontal region cam be inferred on 7/21 in this area, but the water masses that are meeting at this front are not as different as ones encountered elsewhere in this analysis.

3.5.3. Current Structure, Boundaries and Gradients

The measured current structure associated with the boundaries and gradients indicate that physical features in the current field such as convergent zones and horizontal sheers are generally associated with water mass boundaries. This suggests that the physical processes are driving the propagation of the frontal region, as opposed to spurious changes in the optics due to changes in biomass or SST due to solar sea surface warming. Furthermore, it has been shown that optical properties are highly related to spatial physical dynamics in this region (Oliver et al. 2004). However, it should be noted that the current resolution (6 km) averaged over three hours might be too coarse to resolve all pertinent currents that are shaping these complex fronts. The drifter simulation (Figure 3.9) from 7/31 to 8/02 shows that the positions of water mass boundaries in this study are also related largely to local advective processes. The predicted boundary location of the 7/31 boundary on 8/02 using assimilated CODAR fields is very similar to the observed boundary position on 8/02. The current magnitudes and directions are sufficient to explain not only the general location of the water mass boundary, but also how some of the specific features form such as the protrusion of the northern horn of the "hammer-head" shape. Discrepancies between the predicted location of the boundary on 8/02 and the actual location of the boundary on 8/02 may be due to local vertical sheers. The CODAR system measures the current velocity of approximately the top meter of the water column, while the boundary location is responding to the integrated depth averaged current. Despite this, these results suggest that at least over the short term in this coastal region, water masses can be identified and tracked.

3.5.4. Independent verification by salinity and inorganic nutrients

Having developed an algorithm based on a separate set of imagery, we sought to rigorously test our assertion that this type of cluster analysis can deconvolve a complex continental shelf environment into meaningful water masses. Based on the comparison of salinity, nitrate, and silicate concentrations, it appears that the boundaries delineated from space correspond to changes in these three independent variables (Figures 3.10-3.12). In some cases, the changes in these variables are slightly offset from the location of the boundary. This could be due to time of satellite pass issues, as the satellite only measures once a day while it could take a research vessel the better part of a day to cover the same area.

Presently, ocean observatories are being developed world wide and the water mass analysis presented here is an efficient way to assimilate observational data and objectively describe prevalent water types in a system as well as describe the strengths of the boundaries between them. From an operational standpoint, this can be a powerful tool in determining sampling strategies for specific experiments. Depending on the variables of interest, this type of analysis can be used when the position of water masses defined by other predictors or many predictors are more cryptic and non-intuitive. With the development of remote sensing optical inversion algorithms that detect functional groups of phytoplankton, this analysis can be used to detect clusters of communities and identify ecotones. These ecotone regions often have higher primary and secondary production leading to higher fish production (Pingree et al. 1974). In addition, this type of analysis can be used in understanding the biogeochemistry of a particular water mass and be able to track it in the context of an observing system, and to sample its populations over time.

3.6. Conclusion

The goal of this study was to determine if specific water types could be identified and mapped as distinct water masses in a coastal region using satellite data, and whether the measured surface currents, salinity and nutrient fields supported the boundaries and gradients in these maps. Because of the episodic and dynamic nature of coastal regions, optical discriminators were added to a water mass analysis to resolve water types that would not be resolved only by a single suite of parameters. To do this tools were adapted from the field of bioinformatics to constrain the number of water types in this study. Based on the boundary and gradient analysis, water types based on temperature and remote sensing reflectance could be mapped and that the relative differences between them could be estimated. Furthermore, the boundaries and gradients were generally co-located with features in the current, salinity and nutrient fields. Simulated drifter experiments show that the location of these boundaries is largely a result of local advective processes. This suggests that the predictors used in this experiment change slow enough to act as effective tracers of water masses over short time scales and that combining satellite data products is an effective method for discriminating water masses found on continental shelves.

Chapter 4

4.0. Density and Nitrogen Related Patterns of Copia-Like Retrotransposon Transcription in the Diatom *Phaeodactylum tricornutum*

4.1. Abstract

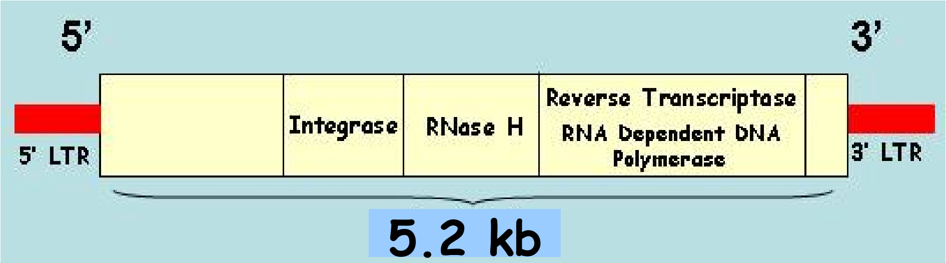
Retrotransposons are mobile genetic elements that encode the necessary machinery for their own replication. Many studies have linked their expression to environmental stress. Genome sequencing and EST libraries of the diatom *Phaeodactylum tricornutum* indicate that this organism has an active copia-like retrotransposon. In this study we test the hypothesis that this mobile element is induced by nitrogen stress. We show for the first time in a diatom that a retrotransposon is induced by nitrogen stress, but also find that the predominant pattern of transcription is related to cell density. Based on this pattern, we suggest that the expression of the copia element is also possibly related to cell surveillance and signaling.

4.2. Introduction

Retrotransposons are nearly ubiquitous among eukaryotes, often comprising a large fraction of the genome (Lynch and Conery 2003). Retrotransposons are capable of autonomous proliferation within a genome through the activity of self-encoded reverse transcriptase, but also are capable of self-deletion through illegitimate recombination (Devos et al. 2002). In addition to the presence of reverse transcriptase, many retrotransposons have group specific antigen, protease, polymerase, enveloping, and endonuclease domains, making the basic structure of a retrotransposon reminiscent of a modern retrovirus (Kazazian 2004). The activity of retrotransposons can be destructive by interrupting or rearranging crucial sequences, but also potentially beneficial by introducing novel combination of sequences. Because of this activity, these elements are thought to be major drivers of genome evolution by increasing the bulk rate of genome mutation (Kazazian 2004).

The estimated rates of retrotransposition are highly variable and depend on host of factors including transcriptional activity, efficiency of re-insertion, the relative fitness of the insertion and the effective population size, many of which are unknown factors. However, there are now many examples of retrotransposons becoming active during environmental stress, most of which have come from eukaryotic plant lineages (Grandbastien 1998). For example, a transcriptionally active retrotransposon was found in experimental tissue cultures of the rice *Oryza sativa* (Hirochika et al. 1996). This study showed that the stress of tissue culture increased the copy number of the Tos17 LTR retrotransposon significantly over a 16-month period. A more direct study of the OARE-1 a Ty1-copia LTR retrotransposon in oat (*Avenia sativa*) showed that these retrotransposons were also activated by the stress of UV light exposure, and by the addition of jasmonic and salicylic acid and plant wounding (Kimura et al. 2001). In addition, dramatic activation of the Tnt1A retrotransposon in tobacco is also induced in response to wounding (Grandbastien 1998). The natural distribution of the closely related BARE-1 LTR retrotransposon in natural environments also suggest they are active in natural populations. It has also been shown that there was a sharp change in the distribution patterns of the BARE-1 element in wild barley (*Hordeum spontaneum*) in response to microclimate habitats (Kalendar et al. 2000). Populations of barley on adjacent north-facing and south-facing slopes of a canyon had large differences in the copy number of BARE-1. The increase of copy number was related to the harsher, more stressful environment associated with inhabiting a south-facing slope. Taken together with the laboratory studies mentioned earlier, there is mounting evidence that genome restructuring through the activities of retrotransposons in response to stress is not a rare occurrence, but an active response to local environments that increases the rate of genome evolution.

Like south-facing mountain slopes, the coastal marine environment is also a stressful environment for diatoms. The physical dynamics of the coastal marine environment result in highly episodic nitrogen pulses to the coastal ocean (Malone et al. 1983). Diatoms living in this environment are often mixed in and out of large river plumes and upwelling waters rapidly, resulting in rapid changes in ambient nutrient concentrations. Diatoms respond to this sort of nutrient regime by following a "boom and bust" growth model, where diatom biomass increases rapidly over a short period of time, then dramatically crashes once nutrients run out. This phenomenon often results in anoxic or hypoxic "dead zones" in coastal areas due to bacterial respiration of large amounts of diatom biomass. However, because diatoms strip nitrogen out of the water column so efficiently, nitrogen limitation is a common and recurrent biological stress for diatoms. Therefore, diatoms might exhibit increased transcriptional activity of retrotransposons in the presence of variable nitrogen regimes. Recent sequencing and EST libraries indicate that a Copia-like retrotransposon (CLR) is transcriptionally active in the coastal marine diatom *Phaeodactylum tricornutum* (Allen 2005) (Figure 1). However, whether the transcription of this retrotransposon is induced by environmental stress is not known. In this study, we hypothesize that the transcription of the CLR in *P.*



<u>Figure 4.1.</u> Map of Copia Element in *P. tricorntnum*. The element is flanked by two 163 bp regions and is in excess of 100 copies per genome.

tricornutum is related to the stress of varying nutrient regimes. We expose batch cultures of *P. tricornutum* to rapid changes in nitrogen concentration and track the transcription of the CLR using RT-PCR. We find that the while CLR does exhibit a significantly higher transcription in low nitrogen conditions, cell concentration is a significant predictor of CLR transcription levels.

4.3. Methods

4.3.1. Culture Conditions and Treatments

Batch cultures of *P. tricornutum* were tracked for a period of 14 days (enumerated from 0-13) and exposed to various nutrient regimes (Table 4.1). All cultures were grown in a 14-10 light-dark cycle at 200 μmole photons m^{-2} s^{-1} and moderately bubbled. On day 0, one 4L culture was started and grown for a few days to provide enough biomass for the separate nitrogen treatments. This initial culture was grown in F/2 media (Guillard and Ryther 1962) and had a starting concentration of 800 μM nitrate. On day 3, the biomass was equally split into four cultures; one control culture and three cultures with various nitrate treatments. Approximately ¼ of the biomass from the initial culture was spun down and re-suspended in new respective growth media. The control cells were resuspended back into 3L of F/2 media and the three treatment cultures were placed into 3L F/2 –N media (F/2 media with zero nitrogen added). Before the transfer of these cultures, they were washed with F/2 –N media twice to remove as much nitrogen as

Table 4.1. Nitrogen concentrations added to the F/2 growth media.

Days	0-3	3-9	9-13
Control	800 μM	800 μM	800 μM
Treatment 1		0 μM	800 μM
Treatment 2		0 μM	75 μM
Treatment 3		0 μM	75 μM*

*Nitrogen was added back in the form of ammonia possible. On day 9, all four cultures were again centrifuged and re-suspended into 2L fresh media for nitrogen recovery. The control culture and treatment 1 were placed back into nitrogen replete F/2 media. Treatment 2 was placed into F/2 with 75µM nitrate and treatment 3 was placed into F/2 with 75 µM ammonia. Cells were harvested from each of these conditions approximately twice per day. 250ml of culture was spun down at 10000 rpm at 4ºC for 10 minutes. Most of the supernatant was poured off, then concentrated culture was transferred to 2ml micro tubes. Cells were spun again for 1 minute at 10000 rpm and flash frozen in liquid nitrogen before being stored at –80ºC.

4.3.2. Cell Counts

Cell density was measured using a Coulter Multisizer II. Cells were measured using a 70 µm glass orifice in triplicate. Cells were diluted between 100 and 1000 times into 0.4 µm filtered seawater and were counted immediatly. 500 µl of the diluted samples were counted and coincidence levels were less than 4%.

4.3.3. Photosynthetic Physiology

The photosynthetic physiological response to the various nitrogen treatments was measured using a Satlantic Fluorescence Induction and Relaxation (FIRe) fluorometer (Gorbunov and Falkowski 2004). The ratio of variable chlorophyll fluorescence to the maximum chlorophyll fluorescence (F_v/F_m) is the quantum efficiency of photosystem II, thus is an indicator of how efficient a photosynthetic organism can utilize light energy. Because photosystem II is the primary energy gateway for many photosynthetic organisms and is nitrogen rich, F_v/F_m has been used as a general "health" index of photosynthetic organisms (Kolber et al. 1998). In addition to measuring F_v/F_m for each of the treatments, F_m/cell was calculated by dividing chlorophyll fluorescence by cell density as a proxy for how well the nitrogen additions were incorporated as intracellular pigment.

4.3.4. RNA Extraction and QPCR

RNA was extracted from the frozen pellets using Tri Reagent® according to manufacturers recommended protocol (Ambion). The RNA was subsequently treated with Turbo DNA-free™ kit, using the most stringent DNAse treatment recommended by the manufacturer (Ambion). DNAse treated RNA was then reverse transcribed into first-strand cDNA with the SuperScript™ III First-Strand Synthesis System for RT-PCR (Invitrogen) using oligo-dT primers. Gene transcription was measured using the Brilliant® SYBR® Green QPCR Core Reagent Kit and the Stratagene MX3000P QPCR machine (Stratagene).

4.3.4.1. House Keeping Genes and Primer Optimization

The transcription of CLR was quantified relative to the transcription of two other genes, histone and the TATA box binding protein (TATA). The histone protein is integral to DNA organization and the TATA box binding protein is a basal transcription factor. Because of their critical importance to the cell, they are considered housekeeping genes and are assumed to be constitutively expressed. They have been used successfully as normalizing genes in previous *P. tricornutum* experiments (Allen 2005). Two genes were chosen to normalize transcription of the CLR to avoid bias interpretation stemming from a single normalizing gene (Thellin et al. 1999).

CLR primer sequences were 5'-GTGTTCTTGCTGCAAATGGA-3' (forward) and 5'-ATTCATCGGGGTCACCAATA-3' (reverse). They were designed to amplify a 174 bp region of the CLR reverse transcriptase domain. The primers used for histone were 5'-

AGGTCCTTCGCGACAATATC-3' (forward) and 5'-ACGGAATCACGAATGACGTT-3' (reverse) and amplified a 150 bp region. The primers used for TATA were 5'-CGGAATGCGCGTATACCAGT-3' (forward) and 5'-ACCGGAGTCAAGAGCACACAC-3' (reverse) and amplified a 180 bp region. These amplicons were cloned into plasmids, which were linearized by restriction digest. The linearized plasmids were diluted in series from 10X to 1e-7X to serve as pure target for optimization (Table 4.2). CLR, histone and TATA efficiency of amplification were 99%, 95% and 95% respectively.

Table 4.2. Optimized primer conditions.

Gene	Forward:Reverse Primer Ratio	Individual Primer Concentration
CLR	1:1	800 μM
Histone	1:1	800 μM
TATA	1:1	300 μM

4.4. Results

4.4.1. Cell Density and Photosynthetic Physiology

In general, cell density increased throughout the experiment after the inoculation and dilution events (Figure 4.2). During exposure to F/2 –N media, cells increased in number initially, but noticeably reduced their growth rates after a few days. F_v/F_m is in the control culture was >0.45 throughout the experiment, indicating that the control stayed healthy throughout the experiment. When the three treated cultures were exposed to F/2 –N media, F_v/F_m dropped quickly for all cultures, stabilizing at a value of ~0.23, indicating that these cultures were physiologically stressed (Figure 4.2 B C D). In

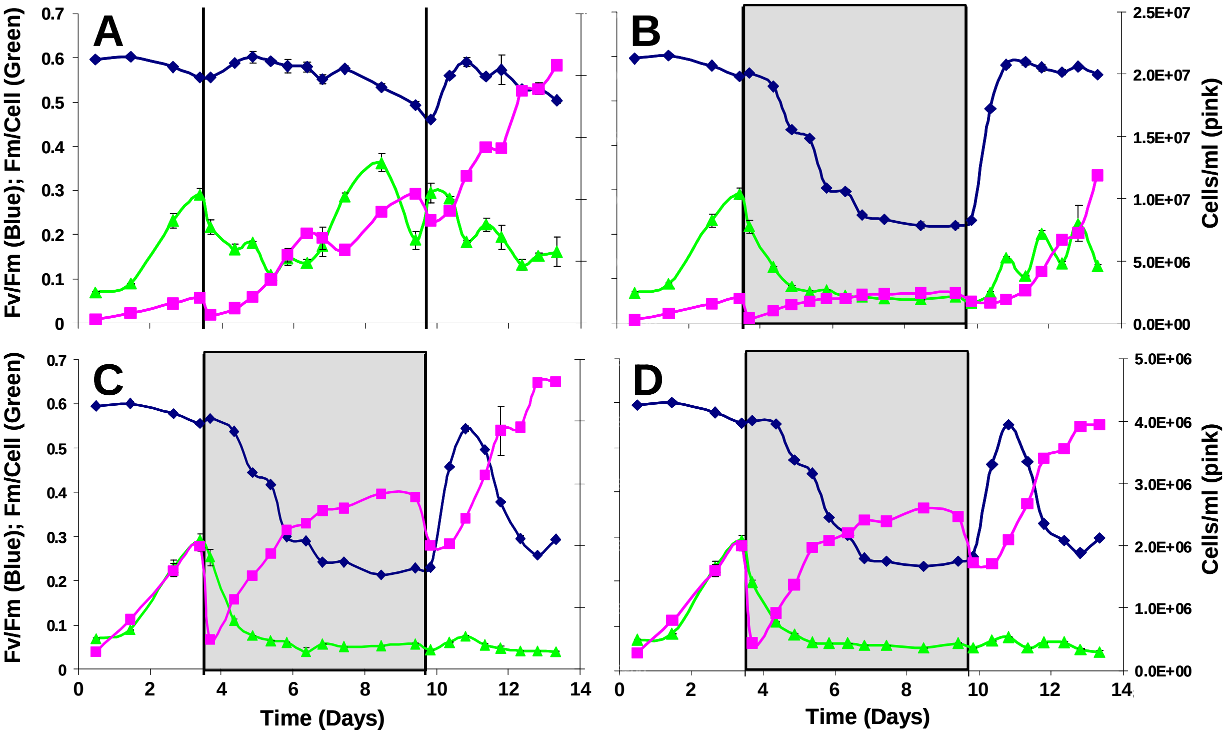


Figure 4.2. Cell density, F_v/F_m, and F_m/cell for the control (A) and for treatment 1, 2, 3 (B, C, D respectively – Table 1). Grey areas indicate time points where cultures were exposed to a zero nitrogen condition.

addition, F_m/cell decreased relative to the control indicating that the amount of chlorophyll per cell was decreasing during nitrogen stress. This is consistent with the internal harvesting of nitrogen rich chlorophyll by the cells to maintain growth. For all three treatment cultures, the addition of nitrogen back into the system resulted in a dramatic increase in F_v/F_m within 10 hours and approached control F_v/F_m levels in 23 hours. In treatment 1, F_v/F_m remained high after nitrogen addition and F_m/cell increased indicating that the nitrogen addition was resulting in a downstream increase of chlorophyll. In treatments 2 and 3, where only 75μm nitrate and 75μm ammonia were added back to the culture, F_v/F_m increased for a short time, but then dropped dramatically as the nitrogen was used by the cells. While cell number increased dramatically after the low-level nutrient addition, F_m/cell did not show a large increase indicating that the nitrogen addition was not being used to manufacture chlorophyll.

4.4.2. Relative Transcription of Copia-like Retrotransposon

Fold change in CLR transcription normalized to histone transcription (Copia/Histone) and in CLR expression normalized to the TATA box binding protein (Copia/TATA) showed a general increase throughout the experiment (Figure 4.3). Transcription of the CLR element was 3-7 fold higher at the end of the stress time period than at the beginning of the experiment. This indicated that CLR expression was not primarily related to stress associated with nitrogen starvation. Model II regressions of the fold change in Copia/Histone and Copia/TATA to fold change in cell number indicated that CLR transcription was significantly positively correlated to fold change in cell density ($P < 0.01$ for both relationships) (Figure 4.4). Furthermore, the slopes of these two regressions were not significantly different ($P < 0.01$). Notably, the experimental

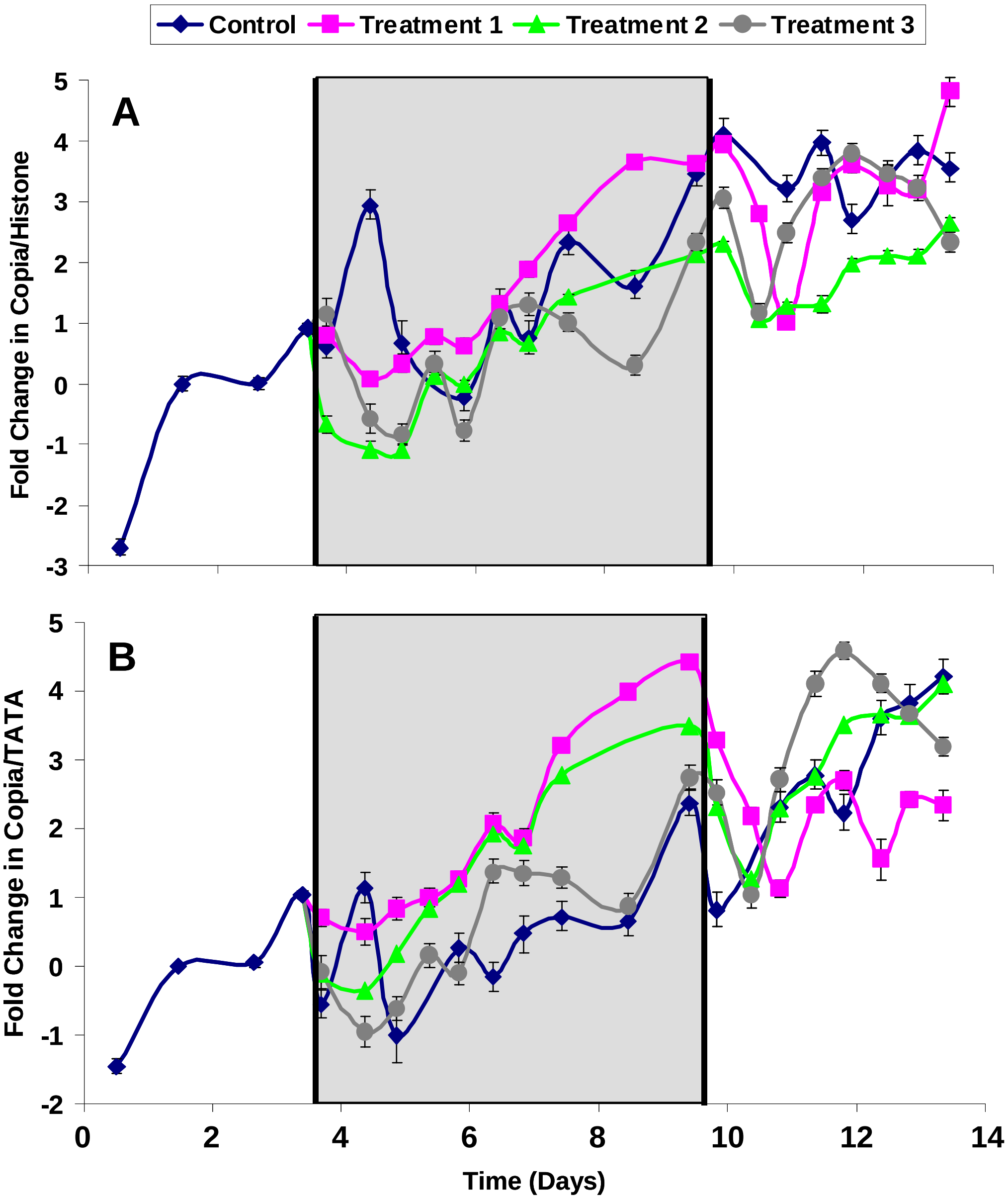


Figure 4.3. Increase in CLR transcription over the time course of the experiment. Shaded areas indicate when cells were exposed to a 0 nitrogen condition. Expression at the end of the experiment is significantly higher than at the beginning of the experiment.

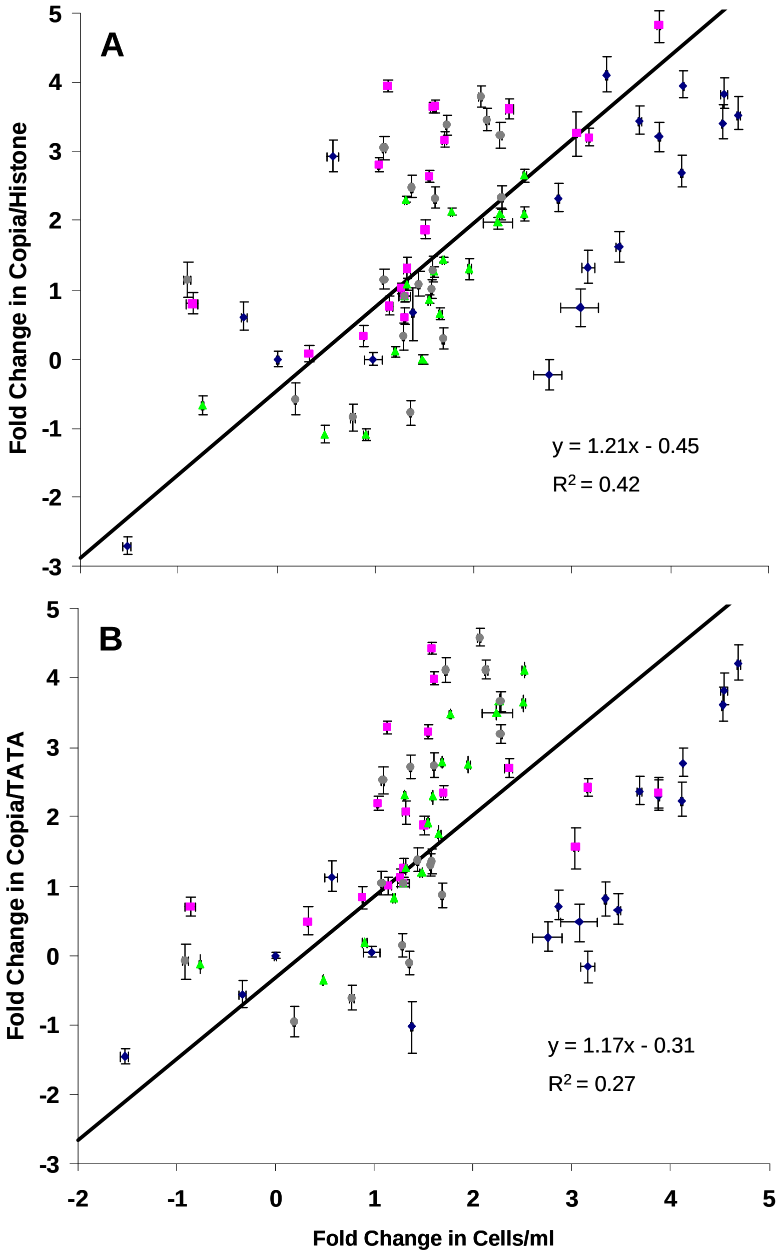


Figure 4.4. Model II regression between fold change in CLR expression relative to histone (A) and relative to TATA (B) verses fold change in cell density. Blue dots are the control, pink dots are treatment 1, green dots are treatment 2 and gray dots are treatment 3. Regression lines were significant at the $p<0.01$ level and were not significantly different from each other between panel A and panel B.

conditions did not overlay each other. The control condition was below best-fit line while the three treatment conditions were above it. This indicates that for equivalent cell densities, the cultures that were exposed to low nitrogen conditions had a higher transcription level of CLR as compared to the control culture.

To remove the density dependent effect of CLR transcription, we normalized the fold transcription of CLR to fold increase in cells/ml. Then, for plotting purposes, we subtracted the control condition from the three treatment conditions to determine if, after the density dependent effect was removed, there was significantly higher transcription of CLR during nitrogen treatments (Figure 4.5). Analysis of the time series in this manner revealed that CLR transcription was significantly higher only after cells had spent 2-6 days under nitrogen limited conditions. A large excursion in CLR transcription relative to the control was observed after the initial dilution, however, the second dilution did not show the same effect. The addition of nitrogen on day 9 appeared to alleviate the increased transcription of CLR relative to the control and initiated a general down regulation of CLR transcription through the end of the time series which converged to the expression level in the control.

4.5. Discussion

Retrotransposons have been traditionally described as genetic parasites because they encode for their own reproduction (Doolittle and Sapienza 1980; Orgel and Crick 1980). This activity makes them a potentially harmful mutagenic force within a genome. Stress induction of retrotransposons has been observed in a variety of organisms (Wessler 1996; Grandbastien 1998), leading to the idea that an organism under stress is less able to control these otherwise harmful mutations. An alternative interpretation of stress induced

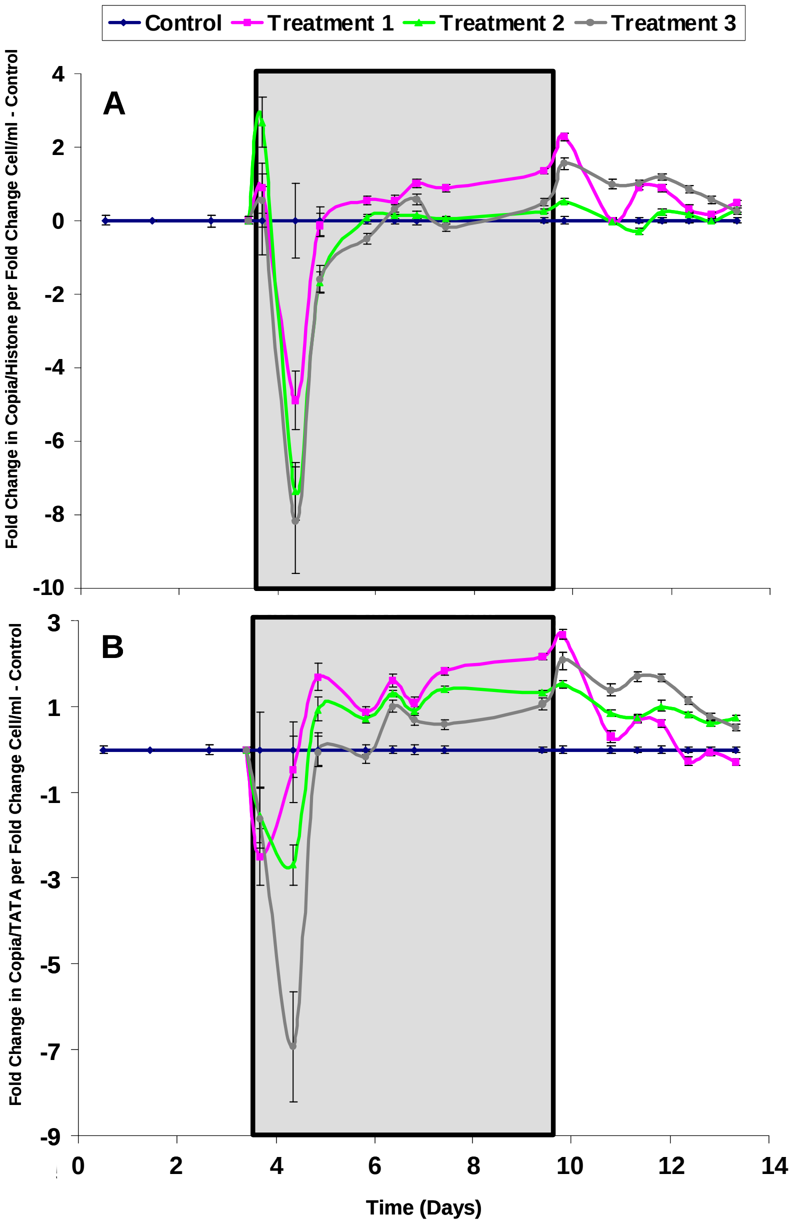


Figure 4.5. Fold change in CLR expression normalized to fold change in cell density Control expression was subtracted out and designated as a zero-line to show the change of CLR expression relative to the control after the effect of cell density is removed.

retrotransposition is that these elements are employed by the organism to increase the genetic diversity in hopes of genetically adapting to the stress environment (McClintock 1984). Because it is still unclear to what level retrotransposition affects the fitness of the organism, distinguishing between these interpretations is difficult. However, if the expression of a particular retrotransposon can be linked to some other nominal cell function or have some net genetic benefit, this would suggest that retrotransposons are in some cases beneficial (Kidwell and Lisch 1997). For example, retrotransposons sequences are often found in promoter regions of genes (Takeda et al. 1999).

In this study, we found that exposing *P. tricornutum* to nitrogen stress resulted in up to a two-fold increase in the CLR compared to a nitrogen replete culture. This represents the first demonstration of an environmentally related expression of a retrotransposon in a phytoplankton. However, the major CLR induction pattern appeared to be also significantly correlated with cell density. Density-dependent expression of a gene has been traditionally linked to cell signaling and quorum sensing (Fuqua et al. 1994; Bassler 1999). Several studies have shown that diatoms exhibit complex quorum sensing behavior that has implications for their predators (Ianora et al. 2004) as well as sensing nutrient conditions resulting in the induction of autocatalytic cell death (Vardi et al. 2006). Furthermore, it has been suggested that retrotransposons are regulated by cell signaling pathways (Labudova and Lubec 1998). Therefore, it seems possible that the density dependent transcription of CLR is in someway related to cell signaling. While this is a speculative assertion that requires further testing, we feel that the density dependent expression of CLR independent of stress conditions suggest that the function of this retrotransposon is not solely selfish, but may be employed by the cell under normal growth conditions.

Chapter 5

5.0. The Mode and Tempo of Genome Size Evolution in Eukaryotes

5.1. Abstract

Eukaryotic genome size varies over five orders of magnitude; however, the distribution is strongly skewed towards small values. Genome size does not appear to be an entirely neutral character, as it is highly correlated to a host of phenotypic traits, making it possible that the relative lack of large genomes is due to selective removal. However, these observations have not been considered with respect to the rate of genome size evolution. Here, using phylogenetic contrasts, we show that the rate of genome size evolution is proportional to genome size, with the fastest rates of evolution occurring in the largest genomes. This trend is evident across all major clades analyzed, indicating that, on long time scales, proportional change is the dominant and universal mode of genome size evolution in eukaryotes. Our results show that the proportional mode of evolution is sufficient to describe the skewed distribution of eukaryotic genome sizes in nature without invoking strong selection against large genomes.

5.2. Introduction

Genome size is a unique biological trait because it lies at the intersection of genotype and phenotype. While the size of the genome does not necessarily confer genotypic information, it might have great evolutionary significance evidenced by its large number of phenotypic correlates, including cell size (Gregory 2001), metabolic rate (Kozlowski et al. 2003) and genomic landscape (i.e., the relative number of genes, introns and mobile genetic elements) (Lynch and Conery 2003; van Nimwegen 2003). Many causal explanations have been hypothesized to account for the strong statistical correlations between these traits and genome size (Petrov et al. 2000; Gregory 2001; Petrov 2001; Gregory 2003; Cavalier-Smith 2005). While these explanations differ regarding the particular evolutionary mechanisms that ultimately determine genome size, natural selection acting on its phenotypic correlates might provide means by which the genome size distribution in nature is determined. Large eukaryotic genomes are rare in nature (Gregory 2005; Knight et al. 2005), and studies on the rates of extinction and species richness suggest that large genome size is a deleterious trait which is selectively removed from Eukarya (Vinogradov 2003; Vinogradov 2004). However, the influence of the underlying molecular mechanisms on the dynamics of genome size evolution has not been fully considered. Here, we show that the dynamics of genome size evolution necessarily leads to the comparable lack of large genomes, even in the absence of selection against them.

The rate of genome size evolution is the balance between the rates of DNA insertion and deletion (indels). Thus, fundamentally, genome size evolution is governed by the molecular mechanisms that produce indels and by the processes that lead to their fixation in populations (Petrov et al. 2000). In eukaryotes, the dominant mechanisms include unequal chromosome crossover (Smith 1976), DNA replication errors (Albertini et al. 1982; Bebenek and Kunkel 1990; Kunkel 1990), polyploidization (Soltis and Soltis 1999), and the proliferation and recombination of transposable elements (Devos et al. 2002; Kazazian 2004). These mechanisms of DNA mutation potentially have variable responses to selection pressures (or lack thereof) and, depending on the organism in which they occur, will have variable rates of fixation, reflecting the mosaic of genome size evolution (Petrov 2001). While the modes of indel production are diverse, what is essential here is that virtually all changes in the rates of indel generation and/or fixation are proportional to the initial genome size. For example, the increase in DNA resulting from polyploidy is proportional to the initial genome size, as is the probability of total insertions and deletions due to random replication errors. In addition, the probability of transposition is a function of the initial transposon copy number, as well as the number of potential target insertion sites (Zhu et al. 2003; Kazazian 2004). Therefore, we might expect the rate of genome size evolution to also reflect these underlying proportional mechanisms that alter genome size.

In this study, we estimated the rate of genome size evolution in 20 traditionally recognized eukaryotic taxonomic groups comprising 168 species, and use the concept of Brownian evolution (Bookstein 1987) and phenotypic contrasts (Felsenstein 1985) to test the hypothesis of proportional genome size evolution in eukaryotes. In our analysis, we find strong evidence of proportional evolution in eukaryotic genome size and suggest that observed genome size distribution in eukaryotes emerges necessarily from the underlying mechanics of proportional evolution.

5.3. Results and Discussion

The absolute magnitude of evolutionary change (i.e., rate of evolution) for phenotypic traits under a simple Brownian model behaves as if drawn randomly from a ½ normal variance distribution at each time step. In other words, the variance of the underlying evolutionary rate is fixed, and is not correlated with the preceding phenotype. However, a trait under proportional evolution violates the Brownian model because the mean and the variance of the underlying evolutionary rate scale with, and depend on, the preceding phenotype. Therefore, if a phenotypic trait, such as genome size, evolves primarily in a proportional manner, we would expect two clear patterns of genome size evolution to emerge. First, the absolute rate of genome size evolution should be positively correlated with genome size, while the variance of the underlying evolutionary rate should clearly deviate from the ½ normal distribution predicted under Brownian evolution. Second, if genome size data were proportionally transformed a-priori (Log_{10}), thus removing the dependency of the underlying evolutionary rate variance on the preceding phenotype, the absolute rate of genome size evolution should show no correlation to genome size. Furthermore, the proportional transformation should result in a ½ normal variance of the underlying evolutionary rate, thus approximating the simple Brownian model.

We estimated the rate of genome size evolution in eukaryotes using the phylogenetic contrast method. This method uses a local maximum likelihood estimation of a phenotypic trait (genome size) at each node in a tree based on the trait at its tips. The main tree in this analysis is based on 18 S rDNA sequences (Figure 1). A contrast is the quantitative difference between the genome sizes of the subtending branches for each node, standardized to their evolutionary distance based on the subtending branch lengths. The absolute value of this standardized contrast represents an estimate of the underlying rate of genome size evolution, based on divergence from a common ancestor as long as the mutation rate of the 18 S rDNA sequence and genome size change are not directly coupled (i.e. branch length is not correlated to genome size) (Garland Jr. 1992). In our analysis, there was no such correlation; however, we emphasize that the rates inferred in this method are relative, as the tree is clocked in the units of 18 S rDNA evolution and not in either absolute or generation time. However, even though the 18 S rDNA mutation

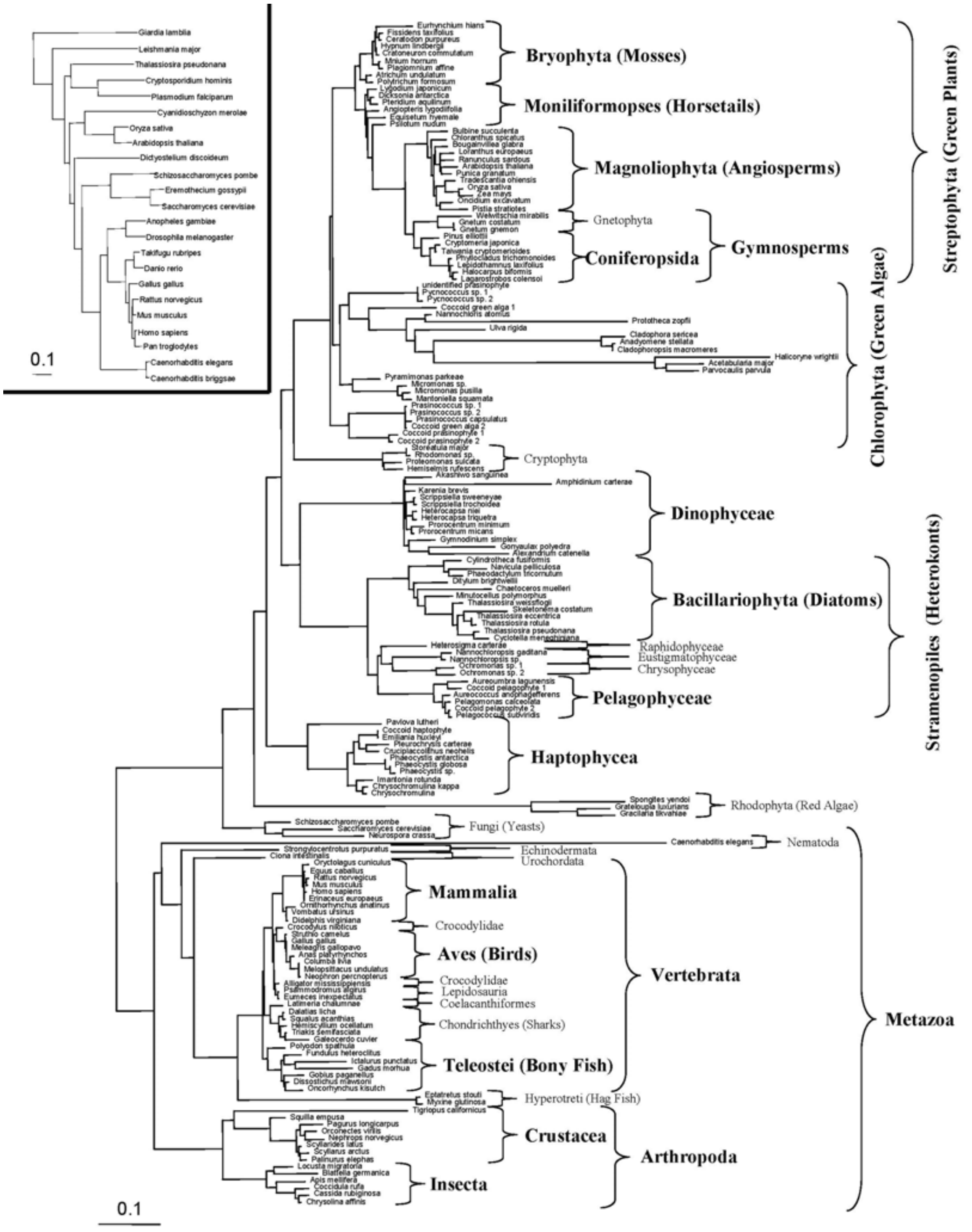


Figure 5.1. Maximum likelihood tree based on 18s sequences built using PHYML. Taxonomic groups highlighted in bold were analyzed for genome size evolution. Accession numbers of the 18 S rDNA sequences used in this analysis are given. Inset: Alternative eukaryotic tree based on 31 orthologs that was used to verify the general trend of genome size evolution inferred from the 18 S rDNA tree (Ciccarelli et al. 2006).

rates do vary among taxonomic groups, this variation likely has minimal effects on the estimation of genome size evolution because genome size varies by multiple orders of magnitude while mutation rates in the 18 S rDNA sequence vary by less than one order of magnitude. To determine empirically if the estimation of genome size evolution was significantly influenced by variable mutation rate in the 18 S rDNA gene, we also estimated the rate of genome size evolution using a smaller (23 species), published eukaryotic tree based on 31 concatenated orthologs (Ciccarelli et al. 2006).

We examined the relationship between genome size and the absolute value of standardized contrasts (i.e. absolute magnitude of the rate of divergence) in two ways. First, the maximum likelihood estimation of genome size at each node was compared to the contrast calculated at each node for the whole 18 S rDNA tree and for the 31-ortholog tree (Figure 2A). Second, the 18 S rDNA tree was divided into 20 traditionally recognized taxonomic sub-trees, from which the median transformed genome size and median contrast for each sub-tree was taken as the representative for the group (Table 1, Figure 2B). The 31-ortholog tree was not divided because of its small size. These analyses show a significant positive relationship between initial or median genome size and the rate of genome size evolution, while analyses of the distribution of the absolute value of the contrasts reveal a significant deviation from a ½ normal distribution as predicted by the Brownian model (Figure 3). In addition, the pattern of genome size evolution inferred from the 18 S rDNA tree and the 31-ortholog tree are consistent with each other, indicating that variance in the mutation rate in the 18 S rDNA tree does not significantly influence the rate estimate of genome size evolution in eukaryotes (Figure 2A).

Table 5.2. Number of species in each group analyzed from the 18 S rDNA tree.

Taxonomic Group	***N***
Streptophyta (Green Plants)	37
Bryophyta (Mosses)	9
Moniliformopses (Horse Tails)	6
Magnioliophyta (Angiosperms)	12
Gymnosperms	10
Coniferopsida	7
Chlorophyta (Green Algae)	23
Dinophyceae	12
Stramenopiles (Heterokonts)	23
Bacillariophyta (Diatoms)	12
Pelagophyceae	6
Haptophyceae	11
Metazoa	52
Vertebrata	33
Mammalia	9
Aves (Birds)	7
Teleostei (Bony Fish)	7
Arthropoda	14
Crustacea	8
Insecta	6

An alternative and more direct test of proportional genome size evolution involves an a-priori transformation of the genome size data, thus removing any proportional dependency between the rate of genome size evolution and genome size. Comparisons of Log_{10} transformed genome size and their calculated contrasts reveal no significant correlation (Figure 4), indicating that the underlying specific (i.e., proportional) rate of genome size evolution is independent of genome size. Again, estimates of the rate of the Log_{10} transformed genome size evolution are consistent between the 18 S rDNA tree and the 31-ortholog tree. The analysis of the distribution of the contrasts calculated from Log_{10} transformed genome size also approximate a ½ normal distribution, thus fitting the Brownian model of evolution quite well (Figure 5). This indicates that the dominant mode of genome size evolution is proportional, with the

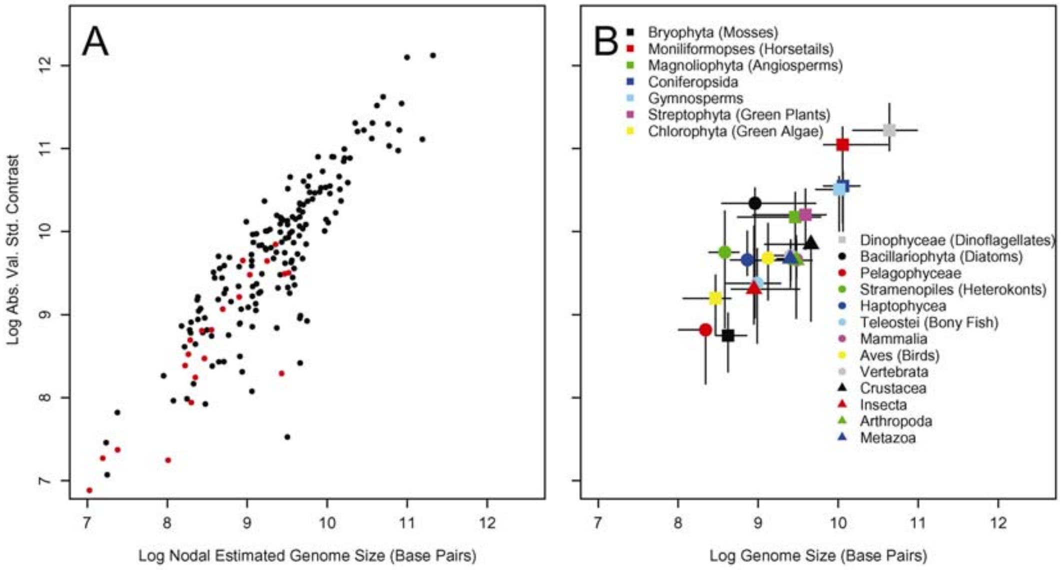


Figure 5.2. A) A tree-wise analysis of the nodal estimated genome size and the calculated contrast at each node from the 18 S rDNA tree (black dots) and the 31-ortholog tree (red dots). Estimations from both trees indicate that as genome size increases, the rate of evolution of genome size increases (shown on Log_{10} axes for plotting purposes). B) Distribution of the median absolute contrast and the median genome size of the 20 traditionally recognized taxonomic groups from the 18 S rDNA tree. Bars represent 95% bootstrapped confidence intervals. Again, a clear positive relationship between genome size and the rate of genome size evolution is evident (shown on Log_{10} axes for plotting purposes).

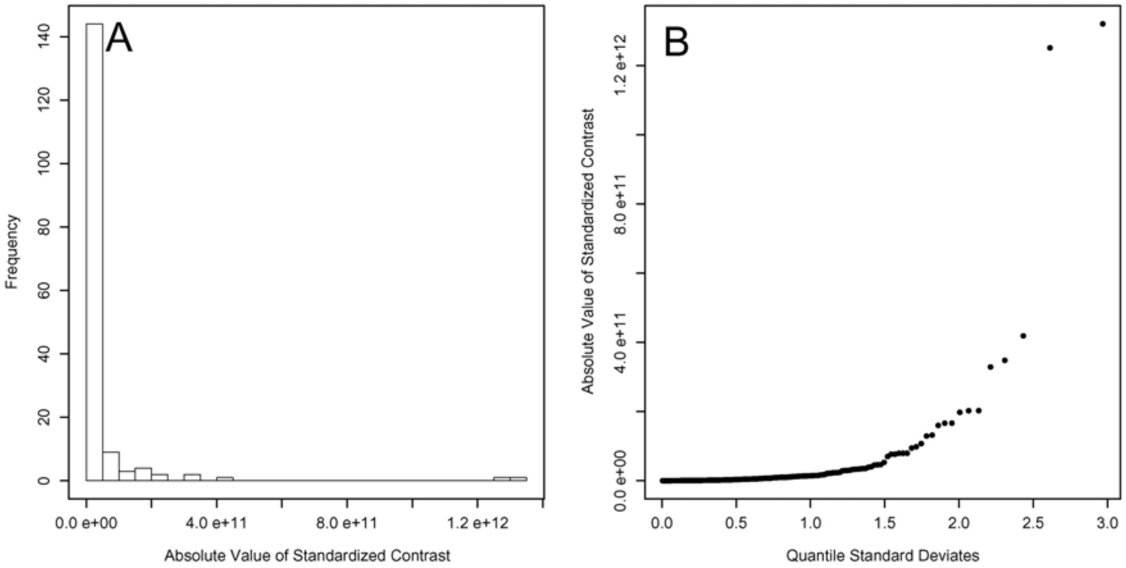


Figure 5.3. A) Distribution of the absolute value of the standardized contrasts from the 18 S rDNA tree showing a strong deviation from the ½ normal distribution expected from a phenotypic trait under Brownian evolution. A strong deviation would be expected for a trait under proportional evolution. B) Quantile distribution of the absolute value of the standardized contrasts. These contrasts do not show a near linear relationship to the positive quantile standard deviates, indicating a strong deviation from a ½ normal distribution, which is expected for a trait under proportional evolution.

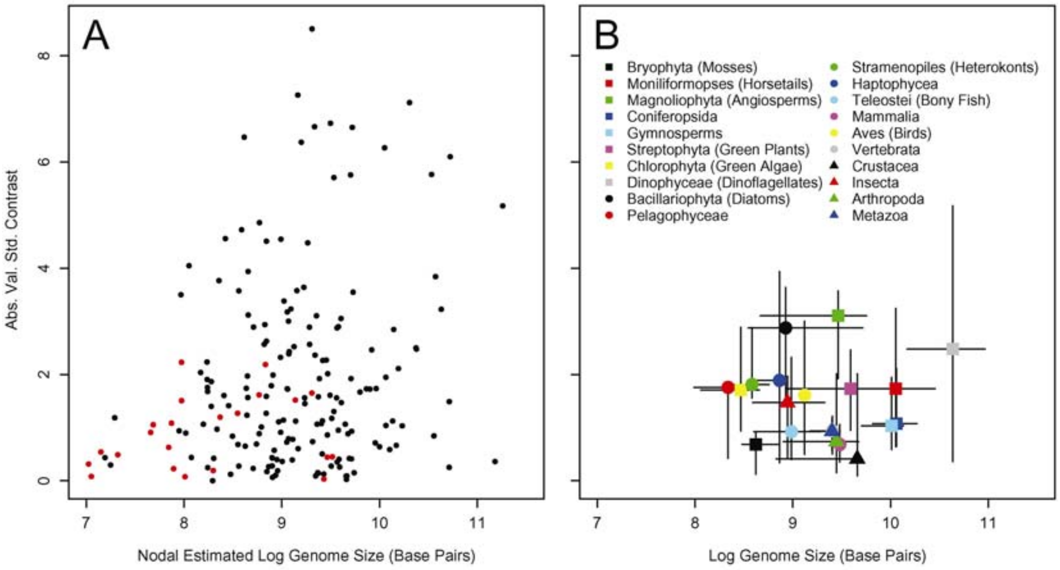


Figure 5.4. A-priori Log$_{10}$ transformation of genome size removes the proportional effect of genome size on the rate of genome size evolution so that neither, A) a tree-wise analysis of the nodal estimated genome size and the calculated contrast at each node, nor B), the distribution of the median absolute contrast and the median genome size of 20 traditionally recognized taxonomic groups show a significant correlation. Bars represent 95% bootstrapped confidence intervals. As in Figure 3A, red dots represent estimations from the 31ortholog tree and black dots represent estimations from the 18 S rDNA tree.

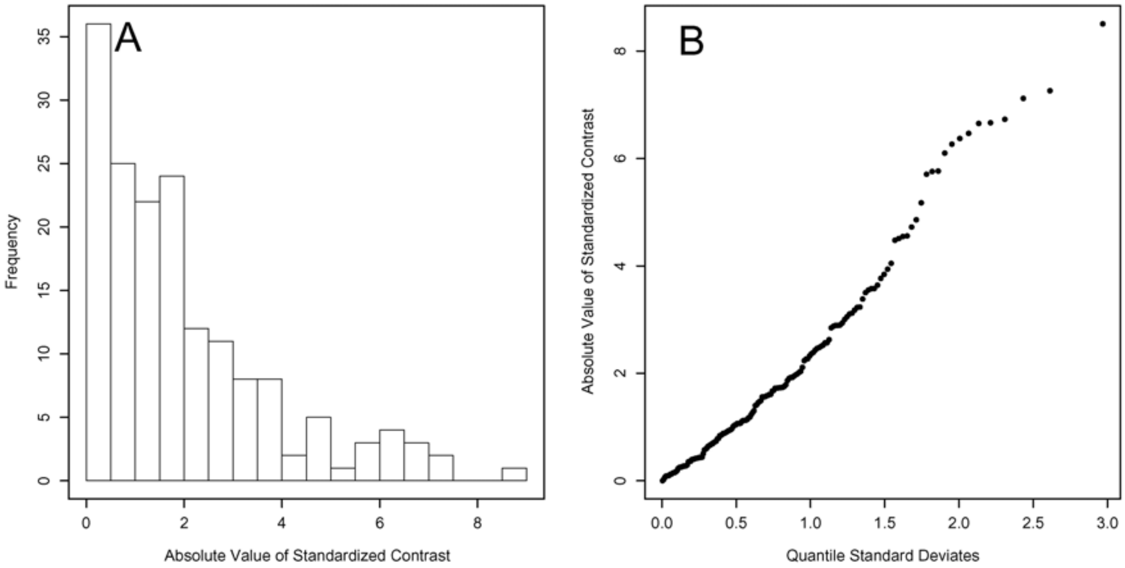


Figure 5.5. A) Distribution of the absolute value of the standardized contrasts calculated from Log_{10} transformed genome size and the 18 S rDNA tree. This calculation shows approximately a ½ normal distribution expected from a phenotypic trait under Brownian evolution. B) Quantile distribution of the absolute value of the standardized contrasts calculated from Log_{10} transformed genome size data. These contrasts show a near linear relationship to the positive quantile standard deviates, indicating the expected ½ normal distribution of the contrasts for a phenotypic trait under Brownian evolution.

tempo increasing with genome size. Hence, in eukaryotes, the larger the genome, the faster its size is evolving.

Traditionally, the paucity of large genomes in eukaryotes has been interpreted as a universal selection against this trait, yet precise descriptions of the specific selection pressures against large genomes are admittedly indirect (Vinogradov 2004; Knight et al. 2005). This is not to say that there are no real reductive selection pressures on genome size in specific instances; loss of DNA in organisms co-opted as organelles (i.e., mitochondria and plastids) and in organisms with parasitic life histories suggests that there could be strong selective forces on genome size (Cavalier-Smith 2005). However, it is not clear if these reductions are due to the inherently higher fitness of reduced genome size, or to the apparent inability of mobile elements to flourish in asexual organisms (Arkhipova and Meselson 2000; Wright and Finnegan 2001). The observed anti-correlations between environmental factors, extinction rates and genome size in some eukaryotic groups also suggest that natural selection may act against species with large genomes, possibly at higher taxonomic levels (Knight and Ackerly 2002; Vinogradov 2004; Knight et al. 2005). However, it is not clear whether such selection is necessary or sufficient to generate the observed skew in the distribution of genome sizes in eukaryotes. Here we offer an alternative explanation for the lack of large eukaryotic genomes that does not rely on selection acting on genome size.

We suggest that proportional evolution of genome size necessarily leads to the skewed distribution of genome sizes in nature. For instance, let us consider a fixed normal positive distribution, with a mean of 1 and bounded at 0, that represents all the possible changes in genome size for all genomes. For a specific genome at a given time step, either a stochastic or selective process produces a single variate draw from this distribution at each time step. However, the effect the variate draw has on genome size change is a multiple of genome size. Therefore, the effect of the variate draw at each time step on small genomes is much less than for large genomes. In short, under proportional evolution, it is more difficult for small genomes to become and stay large, and more likely for large genomes to become and stay small. Therefore, by virtue of proportional evolution, which integrates both random and selective forces, we expect far more small genomes than large genomes in eukaryotes. This corroborates the observation that eukaryote families eukaryotes characterized by large genomes tend to have a much larger range of genome sizes than families with small genomes (Hinegardner 1972), and the recent suggestion that it is difficult for small genomes to increase in size at all after a prolonged phase of genome reduction (Ciccarelli et al. 2006). The general expectation of proportional evolution is that the distribution of a trait under this mode of evolution should approximate log-normality after sufficiently long periods of time (Lewontin and Cohen 1969). Eukaryotes are thought to be between 1.45-2 billion years old (Embley and Martin 2006), thus it is a reasonable expectation that large eukaryotic genomes are rare not because of a universal selection pressure against them, but because of the underlying molecular mechanics that drive the proportional evolution of genome size. The distribution of genome sizes used in this analysis (Figure 6) support this hypothesis. Similar trends in genome size distribution have also been noted in Teleosts (Hinegardner 1972), Angiosperms (Knight et al. 2005) and Metazoans (Gregory 2005).

Despite the small sample size for each of the taxonomic groups, there appear to be some interesting trends in these specific rates of genome size evolution (Figure 4B). For

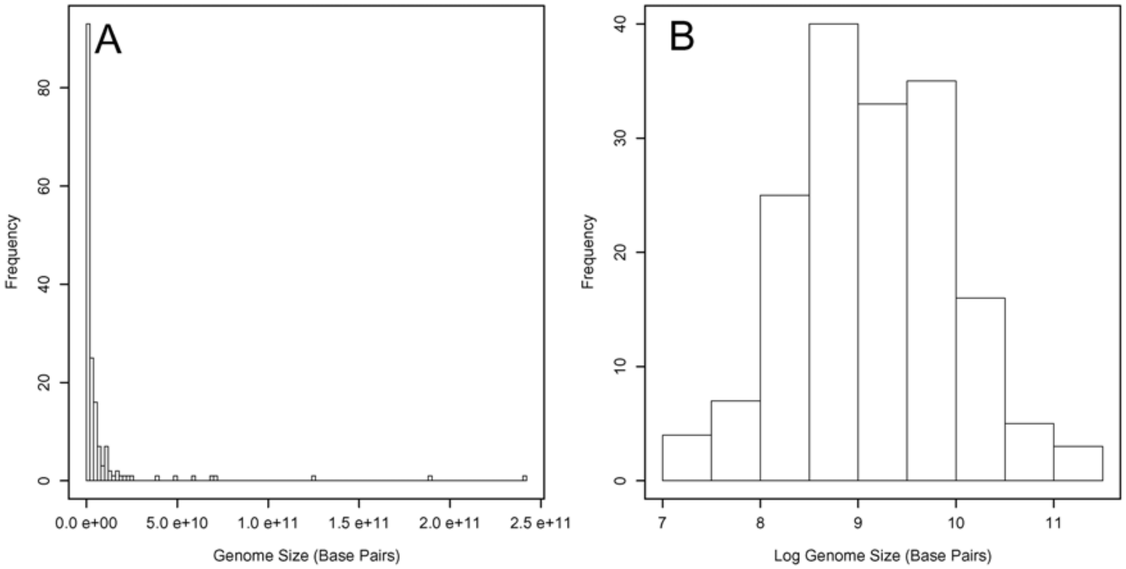


Figure 5.6. Distribution of genome sizes used in this analysis in A) linear space and B) logarithmic space exhibit a log-normal distribution as predicted by proportional evolution integrated over long time periods.

example, bird genomes have been hypothesized to evolve at a slower rate compared to other eukaryotes (Gregory 2002). However, our analysis suggests their rate of genome size evolution is not especially slow, but is actually near the expected rate, if the underlying proportionality of genome size evolution is considered. Furthermore, our analysis suggests that only Magnoliophyta and Bacillariophyta genomes evolve at statistically significantly higher specific rates than the other eukaryotic groups, possibly due to frequent polyploidy (Soltis and Soltis 1999; Chepurnov et al. 2002). It should be noted, however, that because the main effect on the rate of genome size evolution is removed via a-priori logarithmic transformation of genome size, it is also possible that inter-group differences in 18 S rDNA mutation rates influence estimates of the specific rates of genome size evolution. While the lack of statistical significance does not imply the true lack of the differences among or within species groups because of small sample size, future investigations of genome size evolution will need to take into account the dominance of the proportional mode of evolution of genome size before inferring unusually fast or slow patterns of genome size evolution.

Our results suggest the tempo of genome size evolution is positively correlated to genome size across broad eukaryotic diversity. This relationship is consistent with a proportional model of genome size change as the dominant mode of genome evolution. Furthermore, the proportional evolution of genomes provides an alternative explanation for the distribution of genome size in nature and is not reliant on a universal selection pressure against large genomes. Of the taxa examined here, none appeared to violate proportional genome size evolution; therefore, we conclude that taxa-specific selection pressures on genome size must operate within the umbrella of proportionality.

5.4. Methods

5.4.1. Tree Building and Analysis

There are two trees used in this analysis. The first is based on 18 S rDNA sequences that simultaneously allowed for broad coverage across the eukaryotic tree of life, as well as incorporated variable mutation rates in these sequences associated with various reproductive strategies and life histories. Therefore, the rates of evolution are in terms of 18 S rDNA divergence. These sequences were first automatically aligned using ClustalX and then the alignment was hand-edited. A Maximum Likelihood tree was computed using PHYML (GTR model, 1000 bootstraps). See http://atgc.lirmm.fr/phyml/. The second tree used in this analysis is a small published eukaryotic tree estimated from 31 concatenated orthologs (Ciccarelli et al. 2006). The main purpose of this tree was to determine if the inherent variation in 18 S rDNA mutation rates significantly skewed our estimation of genome size evolution. Figures 2A and 4A both indicate that the overall trend of proportional genome evolution in eukaryotes is evident from both trees.

Genome size (1C values) estimates for the 18s rDNA tree come from various literature (Shuter et al. 1983; Veldhuis et al. 1997) and web sources (Appendix 1). These sources tabulate genome sizes from other research efforts, and have those references within. Most eukaryotic genome size estimates are from only a few taxanomic groups (namely green plants and animals). The goal of this study is to look at the broad scale pattern of eukaryotic genome size evolution, therefore, not all available estimates of genome size were used, in favor of a more even distribution of species from across the eukaryotic tree. From the two largest databases of genome size, the Kew database http://www.kew.org/cval/homepage.html and the Animal Genome Size Database http://www.genomesize.com/, a random number generator was used to pick 6-10 species without replacement. Clearly, not all species or taxanomic groups could be included in this type of analysis, however we feel we achieved broad taxanomic coverage of the eukaryotic domain. Genome size estimates for the 31-ortholog tree also come from various sources (Appendix 2). This tree does have some overlap with the 18 S rDNA tree, but also includes some parasitic eukaryotes not included in the 18 S rDNA tree.

Standardized independent contrasts were calculated for the 20 taxonomic groups in Table 1 using the Analyses of Phylogenetics and Evolution Package (Paradis et al. 2004) in the statistical program R http://www.R-project.org.

5.4.2. Regression Analysis

For Figure 2A,B, the data are shown on a Log_{10} transformed axis, but the statistics were done on the linear data. For Figure 2A, a standard OLS regression of the two variables indicated a significant positive correlation ($R^2 = 0.67$, $P << 0.001$). However, the local maximum likelihood estimations of genome size at each node are not independent of each other, since the estimation of the genome size at any node depends on the distal nodes above it, therefore making a standard P value unreliable. Hence, to determine if the positive correlation was significant, we used the PDSIMUL module of the PDAP program to simulate proportional evolution of genome size (Garland Jr. et al. 1993). Parameterization of the model was based on the distribution of the genome sizes and the tree topology based on the 18 S rDNA divergence used in this analysis. Correlations computed from 1000 Monte Carlo simulations of proportional evolution of genome sizes were used to estimate the significance of the of the OLS correlation coefficient computed in Figure 2A. The correlation fell within the 95% confidence interval of the expected correlation between the nodal estimation of genome size and the absolute value of the standardized contrast ($P = 0.226$), indicating the trend in Figure 2A was not significantly different than what would be expected under proportional evolution of genome size. Non-independence of regression variables was also taken into account for Figure 2B due to the hierarchical nature of the sub groups considered. For example, Vertebrata are not independent of Metazoa because Metazoa subsumes Vertebrata. Therefore, regression analysis was done only on the medians of the mutually exclusive sub groups ($R^2 = 0.84$, $P << 0.001$). The same statistical precautions were taken for Figure 4A,B, which was based on a-priori Log_{10} transforming genome size. For Figure 4A, a standard OLS regression showed no significant relationship ($R^2 = 0.021$, $P = 0.057$). Monte Carlo simulation of proportional evolution of genome size indicated that the OLS correlation fell within the 95% confidence interval of the expected correlation between the nodal estimation of Log_{10} genome size and the absolute value of the standardized contrast ($P = 0.137$), indicating the trend in Fig 4A was not significantly different than what would be expected under proportional evolution of genome size. For Figure 4B, the median values of the mutually exclusive sub-groups showed no significant correlation ($R^2 = 0.006$, $P = 0.787$). While figures 2B and 4B affirm the overall proportional relationship between genome size and the rate of genome size evolution, we emphasize that correlation of medians should be interpreted with caution and therefore should be treated as visual heuristic companions to Figures 2A and 4A.

6.0. Summary and Conclusions

Geological records point to marine phytoplankton as major players in establishing Earth's climate. Because of this, much weight has been given to understanding the evolutionary trajectory of marine phytoplankton. Geological records and computer modeling point to turbulent continental shelves as major drivers of diatom evolution, however what the particular mechanisms of diatom evolution are on the shelf remain occluded because i) an accurate picture of continental shelf dynamics required intensive sampling and ii) whether or not a major driver of evolution could respond on the same time scales as the shelf dynamic. In this work, I attempted to tackle both of these issues to move toward a synthesis between inferences drawn from geological records and what can currently be observed in the coastal ocean.

In chapters 2-3, I take advantage of the NEOS observing system, which provided me a well-sampled coastal ocean. I integrated optical parameters as indicators of continental shelf dynamics and produced the first objective classifying scheme for the coastal ocean. In these efforts we found that optical parameters in the coastal ocean significantly mimic hydrography. Because they are quasi-conservative, they were useful in delineating particular kinds of water masses that are encountered by diatoms. This method not only elucidated the turbulent nature of the coastal ocean, but also provided an objective statistic to define the coastal ocean environment. Analysis of the continental shelf in this manner revealed that nutrient pulses to the shelf were highly episodic.

In chapters 4-5, I expose a diatom to highly episodic nitrogen regimes and determine for the first time, a copia-like retrotransposon in a diatom responds to nitrogen stress. Retrotransposons are major drivers of evolution. Therefore, the upregulation of the copia transcription indicates that there is a major evolutionary force capable of acting on the same time scales as the inherent shelf dynamic. Furthermore, I investigate what the net effect of retrotransposon and other DNA indels have on eukaryotic genomes. Comparative analysis of the rate of genome evolution in response to these indels indicate that diatoms possibly have one of the fastest evolving genomes.

Appendix I

Genome sizes and accession numbers for the sequences used for the 18 S rDNA tree in Chapter 5. When necessary, the number of base pairs was estimated from the mass of DNA using: Base Pairs = DNA Mass (pg)*0.978 X. Also, when multiple genome sizes were given from different cell types, the mean value was used.

Species	Accession Number	Genome Size (Base Pairs)	Source
Fissidens taxifolius	X95934.1	3.2E+08	Royal Botanic Gardens Kew
Ceratodon purpureus	Y08989.1	3.8E+08	Royal Botanic Gardens Kew
Eurhynchium hians	U18501.1	4.2E+08	Royal Botanic Gardens Kew
Hypnum lindbergii	AF229922.1	3.0E+08	Royal Botanic Gardens Kew
Cratoneuron commutatum	Y15482.1	3.0E+08	Royal Botanic Gardens Kew
Mnium hornum	X80985.1	8.6E+08	Royal Botanic Gardens Kew
Plagiomnium affine	AF023711.1	8.8E+08	Royal Botanic Gardens Kew
Atrichum undulatum	X85093.1	7.2E+08	Royal Botanic Gardens Kew
Polytrichum formosum	X80982.1	5.2E+08	Royal Botanic Gardens Kew
Lygodium japonicum	AB001538.1	1.1E+10	Royal Botanic Gardens Kew
Dicksonia antarctica	U18624.2	1.1E+10	Royal Botanic Gardens Kew
Pteridium aquilinum	U18628.1	6.3E+09	Royal Botanic Gardens Kew
Angiopteris lygodiifolia	D85301.1	7.0E+09	Royal Botanic Gardens Kew
Loranthus europaeus	L24153.1	8.1E+09	Royal Botanic Gardens Kew
Ranunculus sardous	L24092.1	3.2E+09	Royal Botanic Gardens Kew
Arabidopsis thaliana	X16077.1	1.3E+08	Lynch and Conery 2003
Punica granatum	U38311.1	7.1E+08	Royal Botanic Gardens Kew
Bougainvillea glabra	AF206873.1	4.0E+09	Royal Botanic Gardens Kew
Oryza sativa	AF069218.1	4.7E+08	Lynch and Conery 2003
Zea mays	AF168884.1	2.7E+09	Royal Botanic Gardens Kew
Tradescantia ohiensis	AF069213.1	1.8E+10	Royal Botanic Gardens Kew
Oncidium excavatum	U42791.1	2.1E+09	Royal Botanic Gardens Kew
Pistia stratiotes	AF168869.1	3.2E+08	Royal Botanic Gardens Kew
Chloranthus spicatus	D29787.1	3.5E+09	Royal Botanic Gardens Kew
Bulbine succulenta	AF206876.1	1.1E+10	Royal Botanic Gardens Kew
Welwitschia mirabilis	AF207059.1	7.1E+09	Royal Botanic Gardens Kew
Gnetum costatum	AY755661.1	3.9E+09	Royal Botanic Gardens Kew
Gnetum gnemon	AY755660.1	3.8E+09	Royal Botanic Gardens Kew
Cryptomeria japonica	D85304.1	1.1E+10	Royal Botanic Gardens Kew
Taiwania cryptomerioides	D38250.1	1.9E+10	Royal Botanic Gardens Kew
Phyllocladus trichomonoides	D38244.1	9.8E+09	Royal Botanic Gardens Kew
Lepidothamnus laxifolius	AF342755.1	6.6E+09	Royal Botanic Gardens Kew
Halocarpus biformis	AF342762.1	1.1E+10	Royal Botanic Gardens Kew
Lagarostrobos colensoi	AF342753.1	1.4E+10	Royal Botanic Gardens Kew
Pinus elliottii	D38245.1	2.3E+10	Royal Botanic Gardens Kew
Psilotum nudum	X81963.1	7.1E+10	Royal Botanic Gardens Kew
Equisetum hyemale	U18500.1	1.2E+10	Royal Botanic Gardens Kew
Pycnococcus sp. 1	AF122889.1	4.0E+08	Veldhuis et al 1997
Pycnococcus sp. 2	AY425305.1	1.5E+08	Veldhuis et al 1997
unidentified prasinophyte	AJ010406.1	4.3E+08	Veldhuis et al 1997
Nannochloris atomus	AB080303.1	1.2E+08	Veldhuis et al 1997

Prototheca zopfii	X63519.1	6.7E+07	Shuter et al 1983
Ulva rigida	AJ005414.1	1.5E+08	Royal Botanic Gardens Kew
Cladophora sericea	Z35320.1	2.9E+08	Royal Botanic Gardens Kew
Anadyomene stellata	AF510147.1	1.3E+09	Royal Botanic Gardens Kew
Cladophoropsis macromeres	AF510144.1	2.0E+09	Royal Botanic Gardens Kew
Acetabularia major	Z33462.1	1.2E+09	Royal Botanic Gardens Kew
Parvocaulis parvula	Z33471.1	4.4E+08	Veldhuis et al 1997
Halicoryne wrightii	AY165786.1	9.3E+08	Royal Botanic Gardens Kew
Coccoid green alga 1	U40921.1	6.2E+07	Veldhuis et al 1997
Pyramimonas parkeae	AB017124.3	6.7E+09	Veldhuis et al 1997
Micromonas pusilla	AY425320.1	1.1E+08	Veldhuis et al 1997
Mantoniella squamata	X73999.1	6.6E+08	Veldhuis et al 1997
Micromonas sp.	AJ010408.1	1.2E+08	Veldhuis et al 1997
Prasinococcus sp. 1	AF203403.1	4.5E+08	Veldhuis et al 1997
Prasinococcus capsulatus	AB058384.1	1.1E+08	Veldhuis et al 1997
Coccoid green alga 2	U40919.1	2.6E+08	Veldhuis et al 1997
Prasinococcus sp. 2	AF203401.1	5.1E+08	Veldhuis et al 1997
Coccoid prasinophyte 1	AF203402.1	7.2E+07	Veldhuis et al 1997
Coccoid prasinophyte 2	AF203399.1	9.9E+07	Veldhuis et al 1997
Storeatula major	U53130.1	5.0E+09	Veldhuis et al 1997
Rhodomonas sp.	AB183594.1	1.7E+09	Veldhuis et al 1997
Proteomonas sulcata	AJ007285.1	3.0E+09	Veldhuis et al 1997
Hemiselmis rufescens	AJ007283.1	4.8E+08	Veldhuis et al 1997
Akashiwo sanguinea	AB183672.1	6.8E+10	Shuter et al 1983
Amphidinium carterae	AF274251.1	1.1E+10	Veldhuis et al 1997
Scrippsiella sweeneyae	AF274276.1	1.4E+10	Shuter et al 1983
Scrippsiella trochoidea	AJ415515.1	1.6E+10	Shuter et al 1983
Heterocapsa niei	AF274265.1	5.8E+10	Veldhuis et al 1997
Heterocapsa triquetra	AF022198.1	2.2E+10	Veldhuis et al 1997
Prorocentrum minimum	AJ415520.1	3.9E+10	Veldhuis et al 1997
Prorocentrum micans	AJ415519.1	2.4E+11	Veldhuis et al 1997
Karena brevis	AF352822.1	4.8E+10	Shuter et al 1983
Gonyaulax polyedra	AJ415511.1	1.3E+11	Shuter et al 1983
Alexandrium catenella	AJ535392.1	1.9E+11	Veldhuis et al 1997
Gymnodinium simplex	U41086.1	5.0E+08	Shuter et al 1983
Navicula pelliculosa	AY485454.1	6.8E+07	Shuter et al 1983
Phaeodactylum tricornutum	AY485459.1	2.4E+08	Veldhuis et al 1997
Cylindrotheca fusiformis	AY485457.1	4.1E+08	Shuter et al 1983
Thalassiosira eccentrica	X85396.1	2.5E+10	Shuter et al 1983
Thalassiosira rotula	AF462059.1	5.0E+09	Shuter et al 1983
Skeletonema costatum	AY485473.1	3.2E+08	Shuter et al 1983
Thalassiosira pseudonana	AY485452.1	3.8E+08	Veldhuis et al 1997
Cyclotella meneghiniana	AJ535172.1	1.2E+09	Veldhuis et al 1997
Thalassiosira weissflogii	AY485445.1	5.5E+09	Veldhuis et al 1997
Minutocellus polymorphus	AY485478.1	1.4E+09	Veldhuis et al 1997
Chaetoceros muelleri	AY485453.1	5.7E+08	Veldhuis et al 1997
Ditylum brightwellii	AY485444.1	1.2E+10	Veldhuis et al 1997
Nannochloropsis gaditana	AF045039.1	1.6E+07	Veldhuis et al 1997
Nannochloropsis sp.	U41094.1	1.8E+07	Veldhuis et al 1997

Ochromonas sp. 1	U42382.1	4.8E+08	Veldhuis et al 1997
Ochromonas sp. 2	U42381.1	5.4E+08	Veldhuis et al 1997
Heterosigma carterae	U41650.1	5.5E+09	Veldhuis et al 1997
Aureoumbra lagunensis	U40258.1	9.8E+07	Veldhuis et al 1997
Coccoid pelagophyte 1	U40926.1	2.5E+08	Veldhuis et al 1997
Coccoid pelagophyte 2	U40927.1	1.3E+08	Veldhuis et al 1997
Pelagococcus subviridis	U14386.1	1.9E+08	Veldhuis et al 1997
Pelagomonas calceolata	U14389.1	2.5E+08	Veldhuis et al 1997
Aureococcus anophagefferens	AF117778.1	2.6E+08	Veldhuis et al 1997
Coccoid haptophyte	U40924.1	3.5E+08	Veldhuis et al 1997
Emiliania huxleyi	M87327.2	4.5E+08	Veldhuis et al 1997
Pleurochrysis carterae	AJ544120.1	3.3E+09	Veldhuis et al 1997
Cruciplaccolithus neohelis	AJ246262.1	7.9E+08	Veldhuis et al 1997
Phaeocystis globosa	AF182110.1	1.1E+09	Veldhuis et al 1997
Phaeocystis sp.	AJ278035.1	8.2E+08	Veldhuis et al 1997
Phaeocystis antarctica	X77477.1	7.3E+08	Veldhuis et al 1997
Imantonia rotunda	AJ246267.1	2.4E+08	Veldhuis et al 1997
Chrysochromulina kappa	AJ246271.1	6.0E+08	Veldhuis et al 1997
Chrysochromulina	AJ004868.1	5.9E+09	Veldhuis et al 1997
Pavlova lutheri	AF106053.1	6.7E+08	Veldhuis et al 1997
Grateloupia luxurians	U33132.1	2.0E+08	Royal Botanic Gardens Kew
Gracilaria tikvahiae	M33640.1	2.0E+08	Royal Botanic Gardens Kew
Spongites yendoi	U60948.1	1.5E+08	Royal Botanic Gardens Kew
Saccharomyces cerevisiae	AY790536.1	1.2E+07	Lynch and Conery 2003
Neurospora crassa	AY046271.1	4.3E+07	Lynch and Conery 2003
Schizosaccharomyces pombe	AY251644.1	1.4E+07	Lynch and Conery 2003
Oryctolagus cuniculus	X06778.1	3.0E+09	Animal Genome Size Data Base
Rattus norvegicus	M11188.1	2.7E+09	Lynch and Conery 2003
Mus musculus	X00686.1	2.5E+09	Lynch and Conery 2003
Equus caballus	AJ311673.1	3.1E+09	Animal Genome Size Data Base
Homo sapiens	M10098.1	2.9E+09	Lynch and Conery 2003
Erinaceus europaeus	AJ311675.1	3.5E+09	Animal Genome Size Data Base
Vombatus ursinus	AJ311678.1	3.8E+09	Animal Genome Size Data Base
Didelphis virginiana	AJ311677.1	4.0E+09	Animal Genome Size Data Base
Ornithorhynchus anatinus	AJ311679.1	2.9E+09	Animal Genome Size Data Base
Struthio camelus	AF173607.1	2.1E+09	Animal Genome Size Data Base
Meleagris gallopavo	AJ419877.1	1.5E+09	Animal Genome Size Data Base
Anas platyrhynchos	AF173614.1	1.3E+09	Animal Genome Size Data Base
Melopsittacus undulatus	AF173629.1	1.2E+09	Animal Genome Size Data Base
Neophron percnopterus	AF173633.1	1.5E+09	Animal Genome Size Data Base
Columba livia	AF173630.1	1.2E+09	Animal Genome Size Data Base
Gallus gallus	AF173612.1	1.2E+09	Animal Genome Size Data Base
Crocodylus niloticus	AJ311672.1	3.2E+09	Animal Genome Size Data Base
Alligator mississippiensis	AF173605.1	2.4E+09	Animal Genome Size Data Base
Psammodromus algirus	AY217918.1	2.1E+09	Animal Genome Size Data Base
Eumeces inexpectatus	AY217939.1	2.5E+09	Animal Genome Size Data Base
Dalatias licha	AY049827.1	8.7E+09	Animal Genome Size Data Base
Squalus acanthias	M91179.1	6.6E+09	Animal Genome Size Data Base
Triakis semifasciata	AF212180.2	4.6E+09	Animal Genome Size Data Base

Galeocerdo cuvier	AY049833.1	4.6E+09	Animal Genome Size Data Base
Hemiscyllium ocellatum	AY049835.1	5.2E+09	Animal Genome Size Data Base
Latimeria chalumnae	L11288.1	4.1E+09	Animal Genome Size Data Base
Polyodon spathula	AF188371.1	1.9E+09	Animal Genome Size Data Base
Fundulus heteroclitus	M91180.1	1.3E+09	Animal Genome Size Data Base
Ictalurus punctatus	AF021880.1	9.7E+08	Animal Genome Size Data Base
Gadus morhua	AF518205.1	6.3E+08	Animal Genome Size Data Base
Gobius paganellus	AF518189.1	4.0E+08	Animal Genome Size Data Base
Dissostichus mawsoni	AF518188.1	9.7E+08	Animal Genome Size Data Base
Oncorhynchus kisutch	AF030250.1	2.7E+09	Animal Genome Size Data Base
Eptatretus stouti	M97572.1	2.6E+09	Animal Genome Size Data Base
Myxine glutinosa	M97574.1	4.1E+09	Animal Genome Size Data Base
Ciona intestinalis	AB013017.1	1.6E+08	Lynch and Conery 2003
Strongylocentrotus purpuratus	L28056.1	8.0E+08	Lynch and Conery 2003
Caenorhabditis elegans	AY268117.1	1.0E+08	Lynch and Conery 2003
Tigriopus californicus	AF363306.1	2.4E+08	Animal Genome Size Data Base
Pagurus longicarpus	AF436018.1	4.8E+09	Animal Genome Size Data Base
Orconectes virilis	AF235965.1	4.5E+09	Animal Genome Size Data Base
Nephrops norvegicus	Y14812.1	4.7E+09	Animal Genome Size Data Base
Scyllarus arctus	AF498677.1	1.9E+09	Animal Genome Size Data Base
Palinurus elephas	AF498678.1	4.1E+09	Animal Genome Size Data Base
Scyllarides latus	AF498669.1	6.7E+09	Animal Genome Size Data Base
Squilla empusa	L81946.1	5.7E+09	Animal Genome Size Data Base
Locusta migratoria	AF370793.1	5.6E+09	Animal Genome Size Data Base
Blattella germanica	AF220573.1	1.9E+09	Animal Genome Size Data Base
Cassida rubiginosa	AY676687.1	9.7E+08	Animal Genome Size Data Base
Chrysolina affinis	AJ622062.1	8.0E+08	Animal Genome Size Data Base
Coccidula rufa	AF427603.1	7.0E+08	Animal Genome Size Data Base
Apis mellifera	AY703484.1	2.1E+08	Animal Genome Size Data Base

Appendix II

Genome sizes used for the 31-ortholog tree in Chapter 5.

Species	**Genome Size (Base Pairs)**	**Source**
Thalassiosira pseudonana	3.4E+07	JGI
Cryptosporidium hominis	9.2E+06	NCBI
Plasmodium falciparum	2.3E+07	http://gib.genes.nig.ac.jp/single/main.php?spid=Pfal_3D7
Oryza sativa	3.9E+08	http://www.nature.com/nature/journal/v436/n7052/full/nature03895.html
Arabidopsis thaliana	1.3E+08	Royal Botanic Gardens Kew
Cyanidioschyzon merolae	1.7E+07	Nature. 2004 Apr 8;428(6983):653-7.
Dictyostelium discoideum	3.4E+07	Nature. 2005 May 5; 435(7038): 43–57.
Eremothecium gossypii	9.2E+06	Science. 2004 Apr 9;304(5668):304-7. Epub 2004 Mar 4.
Saccharomyces cerevisiae	1.2E+07	http://www.ensembl.org/Saccharomyces_cerevisiae/index.html
Schizosaccharomyces pombe	1.2E+07	Nature **415**, 871-880 (21 February 2002) | doi: 10.1038/nature724
Anopheles gambiae	2.2E+08	NCBI
Drosophila melanogaster	1.8E+08	NCBI
Takifugu rubripes	3.9E+08	http://www.ensembl.org/Fugu_rubripes/index.html
Danio rerio	1.7E+09	http://www.ensembl.org/Danio_rerio/index.html
Rattus norvegicus	2.7E+09	http://www.ensembl.org/Rattus_norvegicus/index.html
Mus musculus	2.7E+09	http://www.ensembl.org/Mus_musculus/index.html
Homo sapiens	3.4E+09	http://www.ensembl.org/Homo_sapiens/index.html
Pan troglodytes	2.7E+09	http://www.ensembl.org/Pan_troglodytes/index.html
Gallus gallus	1.1E+09	http://www.ensembl.org/Gallus_gallus/index.html
Caenorhabditis elegans	1.0E+08	http://www.ensembl.org/Caenorhabditis_elegans/index.html
Caenorhabditis briggsae	1.0E+08	NCBI
Leishmania major	3.3E+08	http://www.sanger.ac.uk/Projects/L_major/
Giardia lamblia	1.2E+07	NCBI

References

Aarup, T., N. Holt and N. K. Hojerslev (1996). "Optical measurements in the North Sea-Baltic Sea transition zone. II. Water mass classification along the Jutland west coast from salinity and spectral irradiance measurements." Continental Shelf Research **16**(10): 1343-1353.

Albertini, A. M., M. Hofer, M. P. Calos and J. H. Miller (1982). "On the formation of spontaneous deletions: The importance of short sequence homologies in the generation of large deletions." Cell **29**(2): 319-328.

Allen, A. (2005). Personal Communication.

Arkhipova, I. and M. Meselson (2000). "Transposable elements in sexual and ancient asexual taxa." PNAS **97**(26): 14473-14477.

Armbrust, E. V., J. A. Berges, C. Bowler, B. R. Green, D. Martinez, N. H. Putnam, S. Zhou, A. E. Allen, K. E. Apt, M. Bechner, M. A. Brzezinski, B. K. Chaal, A. Chiovitti, A. K. Davis, M. S. Demarest, J. C. Detter, T. Glavina, D. Goodstein, M. Z. Hadi, U. Hellsten, M. Hildebrand, B. D. Jenkins, J. Jurka, V. V. Kapitonov, N. Kroger, W. W. Y. Lau, T. W. Lane, F. W. Larimer, J. C. Lippmeier, S. Lucas, M. Medina, A. Montsant, M. Obornik, M. S. Parker, B. Palenik, G. J. Pazour, P. M. Richardson, T. A. Rynearson, M. A. Saito, D. C. Schwartz, K. Thamatrakoln, K. Valentin, A. Vardi, F. P. Wilkerson and D. S. Rokhsar (2004). "The Genome of the Diatom *Thalassiosira Pseudonana*: Ecology, Evolution, and Metabolism." Science **306**(5693): 79-86.

Auer, S. J. (1987). "Five-year climatological survey of the Gulf Stream system and its associated rings." Journal of Geophysical Research **92**(11): 11709-11726.

Azam, F., T. Fenchel, J. Field, J. Gray, L. Meyer-Reil and F. Thingstad (1983). "The ecological role of water-column microbes in the sea." Marine Ecology Progress Series **10**(3): 257-263.

Baldauf, S. L. (2003). "The Deep Roots of Eukaryotes." Science **300**(5626): 1703-1706.

Barnard, A. H., W. S. Pegau and J. R. J. Zaneveld (1998). "Global relationships of the inherent optical properties of the oceans." Journal of Geophysical Research **103**: 24955-24968.

Barrick, D. E., M. W. Evans and B. L. Weber (1977). "Ocean surface currents mapped by radar." Science **198**: 138-144.

Bassler, B. (1999). "How bacteria talk to each other: regulation of gene expression by quorum sensing." Current opinion in microbiology **2**(6): 582-587.

Beardsley, R. C. and C. D. Winant (1979). "On the mean circulation in the Mid-Atlantic Bight." Journal of Physical Oceanography **9**: 612-619.

Bebenek, K. and T. A. Kunkel (1990). "Frameshift errors initiated by nucleotide misincorporation." Proceedings of the National Academy of Sciences of the United States of America **87**(13): 4946-4950.

Behrenfeld, M. J. and P. G. Falkowski (1997). "Photosynthetic rates derived from satellite-based chlorophyll concentration." Limnology and Oceanography **42**: 1-20.

Bhattacharya, D. and L. Medlin (1995). "The phylogeny of plastids: a review based on comparisons of small-subunit ribosomal RNA coding regions." Journal of Phycology **31**: 489-498.

Biscaye, P. E., C. N. Flagg and P. G. Falkowski (1994). "The shelf edge exchange experiment, Seep-II: An introduction to hypotheses, results and conclusions." Deep Sea Research II **41**(2): 231-253.

Bookstein, F. L. (1987). "Random Walk and the Existence of Evolutionary Rates." Paleobiology **13**(4): 446-464.

Boss, E., W. S. Pegau, W. D. Gardner, J. R. J. Zaneveld, A. H. Barnard, M. S. Twardowski, G. C. Chang and T. D. Dickey (2001). "Spectral particulate attenuation and particle size distribution in the bottom boundary layer of a continental shelf." Journal of Geophysical Research **106**: 509-516.

Bricaud, A., M. Babin, A. Morel and H. Claustre (1995). "Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: Analysis and parameterization." Journal of Geophysical Research **100**(13): 321-332.

Bricaud, A., A. Morel and L. Prieur (1981). "Absorption by dissolved organic matterof the sea (yellow substance) in the UV and visible domains." Limnology and Oceanography **26**: 45-53.

Broecker, W., T. Takahashi and T. Takahashi (1985). "Sources and Flow Patterns of Deep-Ocean Waters as Deduced From Potential Temperature, Salinity, and Initial Phosphate Concentration." Journal of Geophysical Research **90**(C4): 6925-6939.

Broecker, W. S. and T. H. Peng (1982). Tracers in the sea. New York, Lamont-Doherty Geological Observatory.

Carder, K. L., R. G. Steward, G. R. Harvey and P. B. Ortner (1989). "Marine humic and fulvic acids: Their effects on remote sensing of chlorophyll a." Limnology and Oceanography **34**: 68-81.

Cavalier-Smith, T. (2005). "Economy, Speed and Size Matter: Evolutionary Forces Driving Nuclear Genome Miniaturization and Expansion." Annals of Botany **95**(1): 147-175.

Chang, G. C. and T. D. Dickey (1999). "Partitioning *in situ* total spectral absorption by use of moored spectral absorption-attenuation meters." Applied Optics **38**(18): 3876-3887.

Chant, R. (2005). Personal Communication.

Chapman, D. C. and R. C. Beardsley (1989). "On the origin of shelf water in the Middle Atlantic Bight." Journal of Physical Oceanography **19**: 384-391.

Chepurnov, V. A., D. G. Mann, W. Vyverman, K. Sabbe and D. Danielidis (2002). "Sexual Reproduction, Mating System, and Protoplast Dynamics of *Seminavis* (Bacillariophyceae)." Journal Of Phycology **38**(5): 1004.

Chung, F. L. and T. Lee (1992). "Fuzzy competitive learning." Neural Networks **7**(3): 539-551.

Ciccarelli, F. D., T. Doerks, C. von Mering, C. J. Creevey, B. Snel and P. Bork (2006). "Toward Automatic Reconstruction of a Highly Resolved Tree of Life." Science **311**(5765): 1283-1287.

Claustre, H., P. Kerherve, J. C. Marty, L. Prieur, C. Videau and J. H. Hecq (1994). "Phytoplankton dynamics associated with a geostrophic front - Ecological and biogeochemical implications." Journal of Marine Research **54**(4): 711-742.

Critchley, C. (1994). D1 protein turnover: Response to photodamage or regulatory mechanism? Photoinhibition of Photosynthesis From Molecular Mechanisms to the Field. A. R. Baker and J. R. Bowyer. Oxford, BIOS Science**:** 195-201.

Cullen, J. J. (1982). "The deep chlorophyll maximum: comparing profiles of chlorophyll *a*." Canadian Journal of Fisheries and Aquatic Sciences **39**: 791-803.

de Vargas, C., M. Bonzon, N. Rees, J. Pawlowski and L. Zaninetti (2002). "A molecular approach to biodiversity and ecology in the planktonic foraminifera *Globigerinella siphonifera* (d'Orbigny)." Marine Micropaleontology **45**: 101-116.

Demmig-Adams, B. (1990). "Carotenoids and photoprotection in plants: a role for the xanthophyll zeaxanthin." Biochimica et Biophysica Acta **1020**: 1-24.

Denman, K. L. and A. E. Gargett (1983). "Time and space scales of vertical mixing and advection in the upper ocean." Limnology and Oceanography **28**: 801-815.

Devos, K. M., J. K. M. Brown and J. L. Bennetzen (2002). "Genome Size Reduction through Illegitimate Recombination Counteracts Genome Expansion in Arabidopsis." Genome Research **12**(7): 1075-1079.

Doolittle, W. F. and C. Sapienza (1980). "Selfish genes, the phenotype paradigm and genome evolution." Nature **284**: 601-603.

Dugdale, R. C. and J. J. Goering (1967). "Uptake of new and regenerated forms of nitrogen in primary productivity." Limnology and Oceanography **12**: 196-206.

Elena, S. F. and R. E. Lenski (2003). "Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation." Nature Reviews Genetics **4**: 457-469.

Embley, M. T. and W. Martin (2006). "Eukaryotic evolution, changes and challenges." Nature **440**: 623-630.

Eppley, R. W. (1981). Relationships between nutrient assimilation and growth in phytoplankton with a brief review of estimates of growth rate in the ocean.

Physiological Bases of Phytoplankton Ecology. T. Platt. Ottawa, Journal of Canadian Fisheries Research Board**:** 251-263.

Eppley, R. W. and B. J. Peterson (1979). "Particulate organic matter flux and planktonic new production in the deep ocean." Nature **282**: 677-680.

Falkowski, P. G., P. E. Biscaye and C. Sancetta (1994). "The lateral flux of biogenic particles from the eastern North American continental margin to the North Atlantic Ocean." Deep Sea Research II **41**(2): 583-602.

Falkowski, P. G., M. E. Katz, A. H. Knoll, A. Quigg, J. A. Raven, O. Schofield and F. J. R. Taylor (2004). "The Evolution of Modern Eukaryotic Phytoplankton." Science **305**(5682): 354-360.

Falkowski, P. G. and D. A. Kiefer (1985). "Chlorophyll *a* fluorescence in phytoplankton: Relationship to photosynthesis and biomass." Journal of Plankton Research **7**: 715-731.

Falkowski, P. G. and J. A. Raven (1997). Aquatic Photosynthesis. Malden, Blackwell Science.

Felsenstein, J. (1985). "Phylogenies and the Comparative Method." The American Naturalist **125**(1): 1-15.

Fennel, K., M. Follows and P. G. Falkowski (2005). "The Co-evolution of the nitrogen, carbon and oxygen cycles in the Proterozoic ocean." American Journal of Science **305**: 526-545.

Finkel, Z. V., M. E. Katz, J. D. Wright, O. M. E. Schofield and P. G. Falkowski (2005). "Climatically driven macroevolutionary patterns in the size of marine diatoms over the Cenozoic." Proceedings of the National Academy of Sciences of the United States of America **102**(25): 8927-8932.

Fuqua, W. C., S. C. Winans and E. P. Greenberg (1994). "Quorum Sensing in Bacteria: the LuxR-LuxI Family of Cell Density-Responsive Transcriptional Regulators." Journal of Bacteriology **176**(2): 269-275.

Gallegos, C. L. and P. J. Neale (2002). "Partitioning spectral absorption in case 2 waters: discrimination of dissolved and particulate components." Applied Optics **41**(21): 4220-4233.

Garland Jr., T. (1992). "Rate Tests for Phenotypic Evolution Using Phylogenetically Independent Contrasts." The American Naturalist **140**(3): 509-519.

Garland Jr., T., A. W. Dickerman, C. M. Janis and J. A. Jones (1993). "Phylogenetic Analysis of Covariance by Computer Simulation." Systematic Biology **42**(3): 265-292.

Glenn, S. M., R. A. Arnone, T. Bergmann, W. P. Bisset, M. Crowley, J. Cullen, D. Gryzmski, D. Haidvogel, J. T. Kohut, M. Moline, A,, M. J. Oliver, C. Orrico, R. Sherrell, T. Song, A. D. Weidemann, R. Chant and O. Schofield (2004). "Biogeochemical impact of summertime coastal upwelling on the New Jersey Shelf." Journal of Geophysical Research **109**(C12S02): doi:10.1029/2003JC002265.

Glenn, S. M., T. D. Dickey, W. P. Bisset and O. Schofield (2000). "Long-term real-time coastal ocean observation networks." Oceanography **13**: 24-34.

Glenn, S. M., G. Z. Forristall, P. Cornillon and G. Milkowski (1990). "Observations of Gulf Stream Ring 83-E and their interpretation using feature mode." Journal of Geophysical Research **95**: 13043-13063.

Glenn, S. M., D. Haidvogel, O. Schofield, F. Grassle, C. J. von Alt, E. R. Levine and D. C. Webb (1998). "Coastal predictive skill experiment at the LEO-15 national littoral laboratory." Sea Technology **39**(63-69).

Gorbunov, M. Y. and P. G. Falkowski (2004). Fluorescence Induction and Relaxation (FIRe) Technique and Instrumentation for Monitoring Photosynthetic Processes and Primary Production in Aquatic Ecosystems. Photosynthesis: Fundamental Aspects to Global Perspectives. A. van der Est and D. Bruce. Montreal, Allen Press.

Grandbastien, M. A. (1998). "Activation of plant retrotransposons under stress conditions." TRENDS in Plant Science **3**(5): 181-187.

Green, S. A. and N. V. Blough (1994). "Optical absorption and fluorescence properties of chromophoric dissolved organic matter in natural waters." Limnology and Oceanography **39**: 1903-1916.

Gregory, T. R. (2001). "Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma." Biological Reviews **76**: 65-101.

Gregory, T. R. (2002). "A Bird's-eye View of the C-Value Enigma: Genome Size, Cell Size, and Metabolic Rate in the Class Aves." Evolution **56**(1): 121-130.

Gregory, T. R. (2003). "Is small indel bias a determinant of genome size?" Trends in Genetics **19**(9): 485-488.

Gregory, T. R. (2005). "The C-value Enigma in Plants and Animals: A Review of Parallels and an Appeal for Partnership." Annals of Botany **95**(1): 133-146.

Guillard, R. R. L. and J. H. Ryther (1962). "Studies of marine planktonic diatoms. I. Cyclotella nana Hustedt and Detonula confervacea Cleve." Canadian Journal of Microbiology **8**: 229-239.

Hardin, G. (1960). "The competitive exclusion principle." Science **131**: 1292-1297.

Hartigan, J. A. and M. A. Wong (1979). "Algorithm AS 136: A$K$-Means Clustering Algorithm." Applied Statistics **28**(1): 100-108.

Hellend-Hansen, B. (1916). "Nogen hydrografiske metoder form." Skand. Naturf. Mote.: 357-359.

Hey, J. (2001). "The mind of the species problem." TRENDS in Ecology and Evolution **16**(7): 326-329.

Hinegardner, R. R., D. E. (1972). "Cellular DNA content and the evolution of Teleostean fishes." The American Naturalist **106**(951): 621-644.

Hirochika, H., K. Sugimoto, Y. Otsuki, H. Tsugawa and M. Kanda (1996). "Retrotransposons of rice involved in mutations induced by tissue culture." Proceedings of the National Academy of Sciences of the United States of America **93**: 7783-7788.

Højerslev, N. K., N. Holt and T. Aarup (1996). "Optical measurements in the North Sea-Baltic Sea transition zone. I. On the origin of the deep water in the Kattegat." Continental Shelf Research **16**(10): 1329-1342.

Hutchinson, G. E. (1961). "The paradox of the plankton." Amer Natur **95**: 137-145.

Ianora, A., A. Miralto, S. A. Poulet, Y. Carotenuto, I. Buttino, G. Romano, R. Casotti, G. Pohnert, T. Wichard, L. Colucci-D'Amato, G. Terrazzano and V. Smetacek (2004). "Aldeyde suppression of copepod recruitment in blooms of a ubiquitous planktonic diatom." Nature **6990**: 403-407.

Iglesias-Rodriguez, D., S. A. Garcia, R. Groben, K. J. Edwards, J. Batley, L. K. Medlin and P. K. Hayes (2002). "Polymorphic microsatellite loci in global populations of the marine coccolithophorid *Emiliania huxleyi*." Molecular Ecology Notes **2**: 495-497.

Jeffrey, S. W., R. F. C. Mantoura and S. W. Wright (1997). Phytoplankton Pigments in Oceanography. Paris, U.N. Education, Science and Cultural Organization.

Jerlov, N. G. (1968). Optical Oceanography. Amsterdam, Elsevier Oceanography.

Johnsen, G., L. Samset, L. Granskog and E. Sakshaug (1994). "In vivo absorption characteristics in 10 classes of bloom-forming phytoplankton: Taxonomic characteristics and responses to photoadaptation by means to discriminate and HPLC analysis." Marine Ecology Progress Series **105**: 149-157.

Johnson, D. R., J. Miller and O. Schofield (2003). "Dyanmics and optics of the Hudson River outflow plume." Journal of Geophysical Research **108**(C10): doi:10.1029/2002JC001485.

Kalendar, R., J. Tanskanen, S. Immonen, E. Nevo and A. H. Schulman (2000). "Genome evolution of wild barley (Hordeum spontaneum) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence." Proceedings of the National Academy of Sciences of the United States of America **97**(12): 6603-6607.

Kalle, K. (1966). "The problem of glebstoff in the sea." Oceanographic and Marine Biology Review **4**: 91-104.

Karabashev, G., M. Evdoshenko and S. Sheberstov (2002). "Penetration of coastal waters into the Eastern Mediterranean Sea using the SeaWiFS data." Oceanology Acta **25**: 31-38.

Katz, M. E., Z. Finkel, D. Grzebyk, A. H. Knoll and P. G. Falkowski (2004). "Evolutionary trajectories and and biogeochemical impacts of marine eukaryotic phytoplankton." Annual review of ecology, evolution and systematics **35**: 523-556.

Kazazian, H. H., Jr. (2004). "Mobile Elements: Drivers of Genome Evolution." Science **303**(5664): 1626-1632.

Kidwell, M. G. and D. Lisch (1997). "Transposable elements as sources of variation in animals and plants." Proceedings of the National Academy of Sciences of the United States of America **94**: 7704-7711.

Kiefer, D. A. (1973). "Chlorophyll *a* fluorescence in marine centric diatoms: responses of chloroplasts to light nutrient stress." Marine Biology **23**(39-46).

Kimura, Y., Y. Tosa, S. Shimada, R. Sogo, M. Kusaba, T. Sunaga, S. Betsuyaku, Y. Eto, H. Nakayashiki and S. Mayama (2001). "OARE-1, a Ty1-copia Retrotransposon in Oat Activated by Abiotic and Biotic Stresses." Plant Cell Physiology **42**(12): 1345-1354.

Kirk, J. T. O. (1994). Light and Photosynthesis in Aquatic Ecosystems. Cambridge, Cambridge University Press.

Kirkpatrick, G., D. F. Millie, M. Moline, A, and O. Schofield (2000). "Absorption-based discrimination of phytoplankton species in naturally mixed populations." Limnology and Oceanography **42**: 467-471.

Kirkpatrick, G. J., C. Orrico, M. J. Oliver, M. Moline, A, and O. Schofield (2003). "Continuous real-time determiniation of hyperspectral absorption of colored dissolved organic matter." Applied Optics **42**: 6564-6568.

Kishino, M., M. Takahashi, N. Okami and S. Ichimura (1985). "Estimation of the spectral absorption coefficients of phytoplankton in the sea." Bulletin of Marine Science **37**: 634-642.

Knight, C. A. and D. D. Ackerly (2002). "Variation in nuclear DNA content across environmental gradients: a quantile regression analysis." Ecology Letters **5**(1): 66-76.

Knight, C. A., N. A. Molinari and D. A. Petrov (2005). "The Large Genome Constraint Hypothesis: Evolution, Ecology and Phenotype." Annals of Botany **95**(1): 177-190.

Kohut, J. T. and S. M. Glenn (2003). "Improving HF radar surface current measurements with measured antenna beam patterns." Journal of Atmospheric and Oceanic Technology **20**: 1303-1316.

Kolber, Z. S., O. Prasil and P. G. Falkowski (1998). "Measurements of variable chlorophyll fluorescence using fast repetition rate techniques: defining methodology and experimental protocols." Biochim. Biophys. Acta **1367**: 88-106.

Kolmogorov, A. N. (1941). "Dissipation of energy in a locally isotropic turbulence." Dokl Akad Nauk SSSR **32**: 141.

Kooistra, W. H. C. F. and L. K. Medlin (1996). "Evolution of the Diatoms (Bacillariophyta)." Molecular Phylogenetics and Evolution **6**(3): 391-407.

Kozlowski, J., M. Konarzewski and A. T. Gawelczyk (2003). "Cell size as a link between noncoding DNA and metabolic rate scaling." Proceedings of the National Academy of Sciences of the United States of America **100**(24): 14080-14085.

Kroon, B. M. A. (1994). "Variability of photosystem II quantum yields and related processes in Chlorella pyrenoidsa (Chlorophyta) acclimated to an oscillating light regime simulating a mixed photic zone." Journal of Phycology **30**: 841-852.

Kunkel, T. A. (1990). "Misalignment-mediated DNA synthesis errors." Biochemistry **29**(35): 8003-8011.

Labudova, O. and G. Lubec (1998). "cAMP Upregulates the Transposable Element mys-1: A possible link between signaling and mobile DNA." Life Sciences **62**(5): 431-437.

Lewontin, R. C. and D. Cohen (1969). "On population growth in a randomly varying environment." Proceedings of the National Academy of Sciences of the United States of America **62**: 1056-1060.

Lynch, M. and J. S. Conery (2003). "The origins of genome complexity." Science **302**(5649): 1401-1404.

Mackey, M., H. Higgins, D. Mackey and D. Holdsworth (1998). "Algal class abundances in the western equatorial Pacific: Estimation from HPLC measurements of chloroplast pigments using CHEMTAX." Deep Sea Research I **45**: 1441-1468.

Mackey, M., D. Mackey, H. Higgins and S. Wright (1996). "CHEMTAX—A program for estimating class abundances from chemical markers:

Application to HPLC measurements of phytoplankton." Marine Ecology Progress Series **144**: 265-283.

Malone, T. C., P. Falkowski, T. S. Hopkins, G. T. Rowe and T. E. Whitledge (1983). "Mesoscale response of diatom populations to a wind event in the plume of the Hudson River." Deep-Sea Research **30**(2A): 149-170.

Margalef, R. (1961). "Communication of structure in planktonic populations." Limnology and Oceanography **6**(2): 124-128.

Martin-Trayovski, L. V. and H. M. Sosik (2003). "Feature-based classification of optical water types in the Northwest Atlantic based on satellite ocean color data." Journal of Geophysical Research **108**(C5): 1-19.

McClintock, B. (1984). "The significance of responses of the genome to challenge." Science **226**: 792-801.

Meeks, J. C., J. Elhai, T. Theil, M. Potts, F. Larimer, J. Lamerdin, P. Predki and R. Atlas (2001). "An overview of the genome of *Nostoc punctiforme*, a multicellular symbiotic cyanobacterium." Photosynthesis Research **70**(85-106).

Millie, D. F., O. Schofield, G. Kirkpatrick, G. Johnsen and T. J. Evens (2002). "Using absorbance and fluroscence spectra to discriminate micro algae." European Journal of Phycology **37**: 313-322.

Millie, D. F., O. Schofield, G. Kirkpatrick, G. Johnsen, P. Tester and B. T. Vinyard (1997). "Phytoplankton pigments and absorption spectra as potential 'biomarkers' for harmful algal blooms: A case study of the Florida red-tide dinoflagellate, *Gymnodinium breve*." Limnology and Oceanography **42**: 1240-1251.

Mobley, C. D. (1994). Light and Water: Radiative Transfer in Natural Waters. San Diego, Academic.

Moline, M., A,, S. M. Blackwell, R. Chant, M. J. Oliver, T. Bergmann, S. M. Glenn and O. M. E. Schofield (2004). "Episodic physical forcing and the structure of phytoplankton communities in the coastal waters of New Jersey." Journal of Geophysical Research **109**(C12S05): doi:10.1029/2003JC001985.

Morel, A. and A. Bricaud (1986). Inherent optical properties of algal cells including picoplankton: Theoretical and experimental results. Photosynthetic Picoplankton. T. Platt and W. K. W. Li, Canadian Bulletin of Fisheris and Aquatic Science. **214:** 521-559.

Morel, A. and L. Prieur (1977). "Analysis of Variations in Ocean Color." Limnology and Oceanography **22**(4): 709-722.

Nickelsen, J. and J. D. Rochaix (1994). Regulation of synthesis of D1 and D2 proteins of photosystem II. Photoinhibition of Photosynthesis From Molecular Mechanisms to the Field. N. Baker, R, and J. R. Bowyer. Oxford, BIOS Science**:** 179-190.

Noor, M. A. F. (2002). "Is the biological species concept showing its age?" TRENDS in Ecology and Evolution **17**(4): 153-154.

Oliver, M. J., O. Schofield, T. Bergmann, S. M. Glenn, C. Orrico and M. Moline (2004). "Deriving In Situ Phytoplankton Absorption for Bio-optical Productivity Models in Turbid Wate." Journal of Geophysical Research **109**(C07S11): doi:10.1029/2002JC001627.

Orgel, L. E. and F. H. C. Crick (1980). "Selfish DNA: the ultimate parasite." Nature **284**: 604-607.

Owens, T. G., A. P. Shreve and A. C. Albrecht (1993). Dynamics and mechanism of singlet energy transfer between carotenoids and chlorophylls: Light-harvesting and nonphotochemical fluorescence quenching. Research in Photosynthesis. N. Murata. Norwell, Kluwer Acadamy. **IV:** 179–186.

Paradis, E., K. Strimmer, J. Claude, G. Jobb, R. Opgen-Rhein, J. Dutheil, Y. Noel and B. Bolker (2004). "Analyses of Phylogenetics and Evolution Package." R.

Pegau, W. S., J. S. Cleveland, W. Doss, D. C. Kennedy, R. A. Maffione, J. L. Mueller, R. Stone, C. Trees, A. D. Weidemann, W. H. Wells and J. R. J. Zaneveld (1995). "A comparison of methods for the measurement of the absorption coefficient in natural waters." Journal of Geophysical Research **100**(C7): 13201-13220.

Petrov, D. A. (2001). "Evolution of genome size: new approaches to an old problem." Trends in Genetics **17**(1): 23-28.

Petrov, D. A., T. A. Sangster, J. S. Johnston, D. L. Hartl and K. L. Shaw (2000). "Evidence for DNA Loss as a Determinant of Genome Size." Science **287**(5455): 1060-1062.

Pingree, R. D., G. R. Forster and G. K. Morrison (1974). "Turbulent convergent tidal fronts." Journal of Marine Biology Association, United Kingdom **54**: 469-479.

Prasil, O., N. Adir and I. Ohad (1992). Dynamics of photosystem II: Mechanism of photoinhibtion and recovery processes. The photosystems: Stucture, Function and Molecular Biology. J. R. Barber. New York, Elsevier. **11:** 295-348.

Quackenbush, J. (2001). "Computation Analysis of Mictoarray Data." Nature **2**: 418-427.

Quigg, A., Z. V. Finkel, A. J. Irwin, Y. Rosenthal, T.-Y. Ho, J. R. Reinfelder, O. Schofield, F. Morel and P. Falkowski (2003). "The evolutionary inheritance of elemental stoichiometry in marine phytoplankton." Nature **425**: 291-294.

Raven, J. A. (1997). "The vacuole: A cost-benefit analysis." Advances in Botanical Research **25**: 59-86.

Robinson, A. R. and S. M. Glenn (1999). "Adaptive sampling for ocean forcasting." Naval Research Reviews **51**: 26-38.

Roelke, D., D. C. Kennedy and A. D. Weidemann (1999). "Use of discriminant and fourth-derivative analyses with high-resolution absorption spectra for phytoplankton research: Limitations at varied signal to noise ratio and spectral resolution." Fisheries and Aquatic Ecology **17**: 17-28.

Roesler, C. and M. J. Perry (1995). "In situ phytoplankton absorption, fluorescence emission, and particulate backscattering spectra determined from reflectance." Journal of Geophysical Research **100**: 13279-13294.

Roesler, C. S. (1998). "Theoretical and experimental approaches to improve the accuracy of particulate absorption coefficients derived from the quantitative filter technique." Limnology and Oceanography **43**: 1649-1660.

Roesler, C. S., M. J. Perry and K. L. Carder (1989). "Modeling in situ phytoplankton absorption from total absoption spectra in productive inland marine waters." Limnology and Oceanography **34**(8): 1510-1523.

Rynearson, T. A. and E. V. Armbrust (2000). "DNA fingerprintin reveals extensive gentic diversity in a field population of the centric diatom *Ditylum brightwellii*." Limnology and Oceanography **45**(6): 1329-1340.

Ryther, J. H. (1969). "Photosynthesis and fish production in the sea." Science **166**(901): 72-76.

Sathyendranath, S., T. Platt, E. P. W. Horne, W. G. Harrison, O. Ulloa, R. Outerbridge and N. Hoepffner (1991). "Estimation of new production in the ocean by compound remote sensing." Nature **353**: 129-133.

Schofield, O., T. Bergmann, W. P. Bisset, F. Grassle, D. Haidvogel, J. T. Kohut, M. Moline, A, and S. M. Glenn (2002). "Linking regional coastal observatories to provide the foundation for a national ocean observation network." Journal of Ocean Engineering **27**(2): 146-154.

Schofield, O., T. Bergmann, W. P. Bisset, F. Grassle, D. Haidvogel, J. T. Kohut, M. A. Moline and S. M. Glenn (2002). "The long term ecosystem observatory: An integrated coastal observatory." IEEE Journal of Oceanic Engineering **27**(2): 146-154.

Schofield, O., W. P. Bisset, G. J. Kirkpatrick, D. F. Millie, M. Moline and C. S. Roesler (1999). "Optical Monitoring and Forecasting Systems for Harmful Algal Blooms: Possibility or Pipe Dream?" J. Phycol. **35**(6): 1477-1496.

Shuter, B. J., J. E. Thomas, W. D. Taylor and A. M. Zimmerman (1983). "Phenotypic correlates of genomic DNA content in unicellular eukaryotes and other cells." The American Naturalist **122**: 26-44.

Siegal, D. A. (1998). "Resource competition in a discrete environment: Why are plankton distributions paradoxical." Limnology and Oceanography **43**(6): 1133-1146.

Smith, G. P. (1976). "Evolution of Repeated DNA Sequences by Unequal Crossover." Science **191**(4227): 528-535.

Soltis, D. E. and P. S. Soltis (1999). "Polyploidy: recurrent formation and genome evolution." Trends in Ecology & Evolution **14**(9): 348-352.

Strickland, J. D. H. and T. R. Parsons (1972). A Pratical Handbook of Seawater Analysis, Ottawa.

Takeda, S., K. Sugimoto, H. Otsuki and H. Hirochika (1999). "A 13-bp *cis*-regulatory element in the LTR promoter of the tobacco retrotransposon *Tto1* is involved in the responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors." The Plant Journal **18**(4): 383-393.

Thellin, O., W. Zorzi, B. Lakaye, B. De Borman, B. Coumans, G. Hennen, T. Grisar, A. Igout and E. Heinen (1999). "Housekeeping genes as internal standards: use and limits." Journal of Biotechnology **75**: 291-295.

Tomczak, M. (1999). "Some historical, theoretical and applied aspects of quantitative water mass analysis." Journal of Marine Research **57**: 275-303.

Tozzi, S., O. Schofield and P. G. Falkowski (2004). "Historical climate change and ocean turbulence as selective agents for two key phytoplankton functional groups." Marine Ecology Progress Series **274**: 123-132.

Twardowski, M. S., J. M. Sullivan, P. Donaghay and J. R. J. Zaneveld (1999). "Microscale quantification of the absorption by dissolved and particulate material in coastal waters with an ac-9." Journal of Atmospheric and Oceanic Technology **16**: 691-707.

Upstill-Goddard, R. C., A. J. Watson, J. Wood and M. I. Liddicoat (1991). "Sulphur hexafluoride and helium-3 as sea-water tracers: deployment techniques and continuous underway analysis for sulphur hexafluoride." Analytica Chimica Acta **249**(2): 555-562.

van Nimwegen, E. (2003). "Scaling laws in the functional content of genomes." Trends in Genetics **19**(9): 479-484.

Vardi, A., F. Formiggini, R. Casotti, A. D. Martino, F. o. Ribalet, A. Miralto and C. Bowler (2006). "A Stress Surveillance System Based on Calcium and Nitric Oxide in Marine Diatoms." PLoS Biology **4**(3).

Veldhuis, M. J. W., T. L. Cucci and M. E. Sieracki (1997). "Cellular DNA Content of Marine Phytoplankton using two new Fluorochromes: Taxonomic and Ecological Implications." Journal Of Phycology **33**: 527-541.

Vinogradov, A. E. (2003). "Selfish DNA is maladaptive: evidence from the plant Red List." Trends in Genetics **19**(11): 609-614.

Vinogradov, A. E. (2004). "Genome size and extinction risk in vertebrates." Proceedings of the Royal Society of London. Series B **271**: 1701-1705.

Ward, J. H. (1963). "Hierarchical grouping to optimize an objective function." Journal of the American Statistics Association **41**: 236-244.

Warren, B. A. (1983). "Why is no deep water formed in the North Pacific?" Journal of Marine Research **41**: 327-347.

Wessler, S. R. (1996). "Plant retrotransposons: Turned on by stress." Current Biology **6**(8): 959-961.

Wilke, C. M., E. Maimer and J. Adams (1992). "The population biology and evolutionary significance of Ty elements in *Saccharomyces cerevisia*." Genetica **86**.

Wilson, J. T. (1966). "Did the Atlantic close and then re-open?" Nature **211**: 676-681.

Wright, S. and D. Finnegan (2001). "Genome evolution: Sex and the transposable element." Current Biology **11**: 296-299.

Wright, S. W., S. W. Jeffrey, R. F. C. Mantoura, C. A. Llewellyn, T. Bjornland, D. Repeta and N. Welschmeyer (1991). "Improved HPLC method for the analysis of chlorophylls and carotenoids from marine phytoplankton." Marine Ecology Progress Series **77**(183-196).

Wu, C. I. (2001). "The genic view of the process of speciation." Journal of Evolutionary Biology **14**: 851-865.

Yankovski, A. E. and R. W. Garvine (1998). "Subinertial dynamics on the inner New Jersey shelf during the upwelling season." Journal of Physical Oceanography **28**: 2444-2458.

Yeung, K. Y., D. R. Haynor and W. L. Ruzzo (2001). "Validating clustering for gene expression data." Bioinformatics **17**(4): 309-318.

Young, J. R., M. Geisen and I. Probert (2005). "A review of selected aspects of coccolithophore biology with implications for paleobiodiversity estimation." Micropaleontology **51**(4): 267-288.

Zhu, Y., J. Dai, P. G. Fuerst and D. F. Voytas (2003). "Controlling integration specificity of a yeast retrotransposon." Proceedings of the National Academy of Sciences of the United States of America **100**(10): 5891-5895.

Matthew John Oliver
Institute of Marine and Coastal Sciences
Rutgers University
New Brunswick, NJ 08901
Tel: 732-932-6555 ext. 222 Fax: 732-932-4083
E-mail: oliver@imcs.rutgers.edu

Education

2001 – present	Ph.D., Biological Oceanography, Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, NJ.
1999 – 2001	M.S., Biology, Highest Honors, California Polytechnic State University, San Luis Obispo, CA.
1996 – 1999	B.S., Ecology and Systematic Biology, Summa Cum Laude, California Polytechnic State University, San Luis Obispo, CA.
1993 – 1996	A.A. Natural Sciences, Honors, Cerritos College Norwalk, CA.

Honors/Awards

2000	John David Jackmen Memorial Award for Excellence in Biology
1999	Outstanding Graduating Senior, Ecology and Systematic Biology
	John David Jackmen Memorial Award for Excellence in Biology
	E. H. Lehman Memorial Natural History Award
	Montgomery/Richards Marine Biology Scholarship
	Dean's Honor List, Winter
	Green and Gold Foundation Scholarship
	Kevin M. Wright Memorial Biological Scholarship
1998	Dean's Honor List, Fall
	Kevin M. Wright Memorial Biological Scholarship
	Dean's Honor List, Spring
	Green and Gold Foundation Scholarship
1997	GTE Mobile Net Scholar Athlete of the Year
	Burger King Scholar Athlete of the Year
	Dean's Honor List, Spring
	Dean's Honor List, Winter
1996	Dean's Honor List, Spring
	Academic Excellence, Cerritos College Foundation
1995	Gold Falcon Service Award
	Captain's Award
	Mike Merkle Memorial Award for Athletic Excellence
	Gold Falcon Service Award

Employment

2001 – present	Graduate/Teaching Assistantship, Institute of Marine and Coastal Sciences, Rutgers University
1999 – 2001	Graduate/Teaching Assistantship, Environmental Biotechnology Institute, California Polytechnic State University

Field Research Experience

2005 R.V. Oceanus – Lagrangian Transport and Transformation Experiment (LaTTE) focused on primary production, and community composition in the Hudson River plume.
2004 R.V. Cape Hatteras – Lagrangian Transport and Transformation Experiment (LaTTE) focused on evolution of optical properties, and community composition of the Hudson River plume.
2003 Norfolk Naval Base, Mine Warfare Readiness and Effectiveness Measuring (MIREM) focused on optical mine detection and radiative transfer in coastal systems.
2003 R.V. Suncoaster – Ecology and Oceanography of Harmful Algal Blooms (ECOHAB) focused on detection and diel cycles of *Karenia brevis*.
2003 R.V. Suncoaster – Monitoring and Event Response for Harmful Algal Blooms (MERHAB) focused on optical and molecular detection of *Karenia brevis.*
2001 Rutgers Tuckerton Field Station (LEO-15), Office of Naval Research Hyperspectral Coupled Ocean Dynamics Experiment (HyCODE)
2001 R.V. Walford – Coastal predictive skill experiments focused on coastal upwelling
2000 Rutgers Tuckerton Field Station (LEO-15), Office of Naval Research Hyperspectral Coupled Ocean Dynamics Experiment (HyCODE)
2000 R.V. Walford – Coastal predictive skill experiments focused on coastal upwelling
2000 R.V. Endeavor – Utilization of KSS laser lidar for assessing thermocline depth, CTD/FRRF/Bathyphotometer profiles
2000 R.V. Point Sur, Monterey Bay – Assisted operation of Bathyphotometer and Schindler trap profiles
2000 Morro Bay Estuary, Optical quantification of particulate and phytoplankton transport
2000 Morro Bay Estuary, DNA Ribotyping Analysis of Non-point Source Fecal Coliforms in conjunction with Regional Water Quality Control Board/California Department of Health
1999 Marine Laboratory Center for Coastal and Tropical Benthic Ecology; Internship
1998 T.S. Golden Bear Educational Oceanographic Cruise, Cal Poly Quarter at Sea Program

Teaching Experience

2003 Teaching Assistant, Physical Oceanography, Fall, Rutgers University
2000 Teaching Assistant, Undergraduate Marine Biology, Fall, Cal Poly
2000 Teaching Assistant, Undergraduate Computer Applications in Biology, Winter, Cal Poly
1999 Laboratory Instructor, Undergraduate Introduction to Animal Physiology, Winter, Cal Poly
1999 Laboratory Instructor, Undergraduate Introduction to Organismal Diversity, Fall, Cal Poly

Society Memberships

American Geophysical Union
AAAS

Publications Accepted or In Press

Schofield, O., Kerfoot, J., Mahoney, K., Moline, M., **Oliver, M.,** Lohrenz, S., Kirkpatrick, G. Vertical Migration of the Toxic dinoflagellate *Karenia brevis* and its Impact on Ocean Optics. Journal of Geophysical Research (accepted).

Schofield, O., Bosch, J., Glenn, S. M., Kirkpatrick, G., Kerfoot, J., Moline, M., **Oliver, M. J.**, Bissett, W. P. Harmful algal blooms in a dynamic environment: How can optics help the field-going and sample poor biologist? In Real Time Coastal Observing systems for ecosystems dynamics and harmful algal blooms. Babin, M. And Cullen, J. J. (Eds) UNESCO, Paris. (in press).

Glenn, S. M., Schofield, O., Bergmann, T., Chant, R., **Oliver, M. J**., Crowley, M., Cullen, J., Haidvogel, D., Kohut, J., Moline, M. A. 2004. Studying the Biogeochemical Impact of Summertime Upwelling Using a Coastal Ocean Observatory. Journal of Geophysical Research 109, C12S02, doi:10.1029/2003JC002265.

Schofield, O., Bergmann, T., Bissett, W. P. Moline, M. A., Orrico, C., **Oliver, M. J.** 2004. Inversion of Spectral Absorption in the Optically Complex Coastal Waters of the Mid-Atlantic Bight: Journal of Geophysical Research 109, C12S04, doi:10.1029/2003JC002071.

Glenn, S., Schofield, O., Dickey, T., Chant, R. Kohut, J., Barrier, H., Bosch, J., Bowers, L., Creed, E., Haldeman, C., Hunter, E., Kerfoot, J., Mudgal, C., **Oliver, M.** , Roarty, H., Romana, E., Crowley, M., Barrick D., and Jones C. 2004. The expanding role of ocean color and optics in the changing field of operational oceanography. Oceanography 17(2): 86-95.

Schofield, O., Arnone, R., Bissett, W. P., Dickey, T., Davis, Curt, Finkel, Z., **Oliver, M. J.**, Moline, M. A. 2004. Watercolors in the coastal zone: What can we see? Oceanography 17(2): 30-37.

Moline, M. A., Blackwell, S., Chant, R., **Oliver, M. J.**, Bergmann, T., Glenn, S., Schofield, O. Episodic physical forcing and the structure of phytoplankton communities in the coastal waters of New Jersey. 2004. Journal of Geophysical Research 110, C12S05, doi:10.1029/2003JC001985.

Oliver, M. J., Schofield, O., Bergmann, T., Glenn, S. M., Moline, M. A., Orrico, C. Deriving In Situ Phytoplankton Absorption for Bio-optical Productivity Models in Turbid Waters. 2004. Journal of Geophysical Research 109, C07S11, doi:10.1029/2002JC001627.

Oliver, M. J., Kohut, J. T., Irwin, A. J., Glenn, S. M., Schofield, O., Moline, M. A., Bissett, W. P. Bioinformatic Approaches for Objective Detection of Water Masses. 2004. Journal of Geophysical Research 109, C07S04, doi:10.1029/2003JC002072.

Moline, M.A., Arnone, R., Bergmann, T., Glenn, S., **Oliver, M. J.**, Orrico, C., Schofield, O., Tozzi, S. 2004. Variability in spectral backscatter estimated from satellites and its relation to in situ measurements in optically complex coastal waters. Journal of International Remote Sensing. 24: 1-4.

Kirkpatrick, G. J., Orrico, C., Moline, M. A., **Oliver, M. J.**, Schofield, O. 2003. Continuous hyperspectral absorption measurements of colored dissolved organic material in aquatic systems. Applied Optics 42(33): 6564-6568.

Publications Submitted

Schofield, O., **Oliver, M.,** Moline, M. A. Mixing and photoacclimation in coastal Antarctica: Impact on photosynthetic quantum yields (submitted JGR).

Oliver, M. J., Petrov, D., Ackerly, D, Schofield, O. M. Falkowski, P.G. The Mode and Tempo of Genome Size Evolution in Eukaryotes (submitted PLoS).

Grants

NASA 2006-2009, Bioinformatic mapping of ocean biogeochemical provinces, **Matthew Oliver**, Andrew Irwin, Oscar Schofield, Paul Falkowski ($491,000).

Invited Lectures

Bioinformatic Approaches for Objective Detection of Water Masses on Continental Shelves: Early Results from LaTTE 2005, Lamont-Doherty Earth Observatory, Columbia University, NY, May 5, 2005.

Evolution of Dinoflagellate and Diatom Genomes; thoroughbreds of the Eukaryotes, Mote Marine Laboratory, Sarasota, FL. June, 2005.

Contributed Abstracts

Oliver, M. J., Petrov, D., Ackerly, D., Falkowski, P., Schofield, O. The Rapid Evolution Of Diatom And Dinoflagellate Genomes. Ocean Sciences Meeting, Honolulu, Hawaii, Feb 20-24, 2006.

Bosch, J., Schofield, O., Kohut, J., Glenn, S., Gogte, **M. Oliver**, M. East Coast Plumes and Blooms: Monitoring On-Ramp Traffic to the Ocean Highway off New Jersey. Ocean Sciences Meeting, Honolulu, Hawaii, Feb 20-24, 2006.

Connolly, J., Moline, M., Knight, C., **Oliver, M.** Exploring the Evolutionary Implications of Diatom (Bacillariophyceae) Genome Size Variation. Ocean Sciences Meeting, Honolulu, Hawaii, Feb 20-24, 2006.

Frazer, T. K., Schofield, O., Moline, M. M., Glenn, S. M., Kohut, J. T., Chant, R. J., Keller, S. R., **Oliver, M. J.**, Reinfelder, J. R., Zhou, M., Chen, R. F. LaTTE 2005: Super Size Me! Ocean Sciences Meeting, Honolulu, Hawaii, Feb 20-24, 2006.

Oliver, M. J., Finkel, Z, Schofield, O. M., Falkowski, P. G., de Vargas, C. Retrotransposons in Diatom Taxa. The International Ocean Research Conference, UNESCO Headquarters, Paris, France, June 6-10, 2005.

Kohut, J., Chant, R., Glenn, S., Schofield, O., **Oliver, M. J.** Observed response of the Hudson river plume to wind forcing. The International Ocean Research Conference, UNESCO Headquarters, Paris, France, June 6-10, 2005.

Kohut, J., Bosch, J. A., **Oliver, M. J.,** Glenn, S. M. and Schofield, O. M. E. Evolution of Fronts in the Mid-Atlantic Bight (MAB): What Exit on the Ocean Highway off New Jersey? American Geophysical Union Fall Meeting, San Francisco, CA. Dec 13-17, 2004.

Oliver, M. J., Kohut, J. T., Irwin, A. J., Glenn, S. M., Schofield, O., Moline, M. A., Bissett, W. P. Bioinformatic Approaches for Objective Detection of Water Masses. Ocean Optics XVII, Fremantle, Au, Oct 25-29, 2004.

Oliver, M. J., Finkel, Z. V., Schofield, O. M. and Falkowski, P. G. A Hypothesis of Genome Structure in Marine Phytoplankton. 56th Annual Meeting of The Society of Protozoologists June 2-6, Bryant College, Smithfield, Rhode Island, 2004.

Matteson, R. S., Moline, M. A., Bellingham, J. G., Blackwell, S. M., Chavez, F. P., Haddock, S., McManus, M. A., **Oliver, M. J.**, Schofield, O. M. Distribution of Optical Constituents in Response to Episodic Upwelling in Monterey Bay ASLO/TOS Ocean Research Conference Feb 15 - 20 Honolulu, HI, 2004.

Oliver, M. J., Bergmann, T., Glenn, S., Moline, M., Orrico, C., Schofield, O. Application of Optical Inversion Model: Implications for Constituent Specific Absorption and Bio-Optical Modeling of Primary Production. Ocean Optics XVI, Santa Fe, NM. 2002.

Schofield, O., Bergmann, T., Bissett, W. P., Kirkpatrick, G., **Oliver, M. J.**, Orrico, C., Moline, M. A., Glenn, S. Inversion of the Inherent Optical Properties and Their Utility for Delineation of Water Masses in Turbid Coastal Waters. Ocean Optics XVI, Santa Fe, NM. 2002.

Moline, M. A., Bergmann, T., Bissett, W. P., Case, J., Herren, C., Mobley, C. D., **Oliver, M. J.**, Schofield, O., Sundman L. (2002). Integrating optics and biology: Estimation of bioluminescence leaving radiance from an autonomous vertical profiler. Ocean Optics XVI, Santa Fe, NM, 2002.

Oliver, M. J., Moline, M. A., Schofield, O., Bergmann, T., Glenn, S., Bisset, W. P., Bio-Optical Estimates of Phytoplankton Productivity From an Autonomous In Situ Profiler in the Coastal Waters of the Mid-Atlantic Bight. Ocean Sciences Meeting, Honolulu, Hawaii, 2002.

Kirkpatrick, GH., **Oliver, M. J**., Berg, B., Orrico, C., Moline, M. A., Lohrenz, S. E., Schofield, O. Continuous, Real-Time Determination Of Hyperspectral Absorption Of Colored Dissolved Organic Material. Ocean Sciences Meeting, Honolulu, Hawaii, 2002.

Pearson, J. A., Blackwell, S. M., Doughty, N., Moline, M. A., **Oliver, M. J.,** Orrico, C., Optical estimation of Phytoplankton and Sediment Transport in Morro Bay Estuary. Ocean Sciences Meeting, Honolulu, Hawaii, 2002.

Moline, M. A., Arnone, R., Bergmann, T., Glenn, S., **Oliver, M. J.**, Orrico, C., Schofield, O., Tozzi, S., Variability in Spectral Backscatter Estimated from Satellites and its Relation to In-Situ Measurements in Optically Complex Coastal Waters. Presented at, Oceans From Space 2000 Venice, Italy. 2000. Sponsored by the Joint European center and NASA. (Best Poster Award)

Current Research Projects

- Stress induction of retrotransposable elements in diatoms. Collaborators: Collaborators: Paul Falkowski, Oscar Schofield, Kay Bidle (Rutgers University), Dmitri Petrov (Stanford University).
- Evolution rates of cell size in Diatoms, Prymnesiophytes and Dinoflagellates: Collaborators: Colomban deVargas, Paul Falkowski (Rutgers University), Diana Nemergut (University of Colorado), Zoe Finkel (Mount Allison University)
- Global water mass identification based on SeaWifs and AVHRR data. Collaborators: Andrew Irwin (Mount Allison University), Josh Kohut, Paul Falkowski, Oscar Schofield (Rutgers University).
- Predicting internal light source depth from water leaving radiance using neural nets: Collaborators: Mark Moline (California Polytechnic State University), Wayne Slade (University of Maine), Curt Mobley (Sequoia Scientific).

Community Service

2003-present	Ecology Tour Guide, Hutchison Memorial Forest, NJ.
2005	Science Judge, Shore Bowl (NOSB, New Jersey)
2004	Science Judge, Shore Bowl (NOSB, New Jersey)
2003	Science Judge, Shore Bowl (NOSB, New Jersey)

Special Skills

Visual Basic Programming
R Programming
S+ Programming
Matlab Programming
PADI Rescue Diver
Real-Time PCR