INVESTIGATION OF THE POPULATION GENETIC STRUCTURE OF THE TOXIC DINOFLAGELLATE *KARENIA BREVIS* IN THE GULF OF MEXICO

A Dissertation

by

DARREN WAYNE HENRICHS

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Lisa Campbell
Committee Members,	John R. Gold
	James R. Manhart
	Alan E. Pepper
Head of Department,	U. Jack McMahan

December 2012

Major Subject: Biology

ABSTRACT

Karenia brevis is the major harmful bloom forming dinoflagellate in the Gulf of Mexico. The toxin produced by this dinoflagellate can cause large fish kills, marine mammal mortality, respiratory irritation, and neurotoxic shellfish poisoning in humans. Blooms can occur anywhere in the Gulf of Mexico (hereafter Gulf) but are predominantly observed off the west coast of Florida and the coast of Texas. The west coast of Florida has been hypothesized to be the origin for blooms of *K. brevis* in other regions within the Gulf based upon the frequent formation of blooms in this region. To investigate this possibility, microsatellite markers were used to determine the population-genetic structure of *K. brevis* in the Gulf of Mexico. The difficulties of culturing *K. brevis* required development and use of a single-cell PCR amplification protocol for preserved cells. Lugol's iodine-preserved bloom samples of *K. brevis* were destained with sodium thiosulfate and subjected to two rounds of PCR amplification. The destaining protocol resulted in the successful, simultaneous amplification of five microsatellite markers from single cells of *K. brevis.* A total of 18, highly polymorphic microsatellite markers are available for *K. brevis*. Each marker was amplified from 40 cultures of *K. brevis* isolated from water samples from Florida and Texas. Observed genetic diversity was high but similar to the genetic diversity observed in other phytoplankton species. No genetic divergence was detected between isolates from Florida and isolates from Texas. Single cells from a total of 38 field samples were analyzed at five microsatellite markers to determine if population-genetic structure was

present in *K. brevis* in the Gulf. Significant genetic divergence between several individual samples was detected, reflecting the high genetic diversity present within the species. Observed genetic divergence was low between blooms from the west coast of Florida and the coast of Texas and supports the hypothesis of a common origin for blooms of *K. brevis* in the Gulf of Mexico.

DEDICATION

I dedicate this work to my family for the love and support they offered throughout this endeavor. My wife, Sarah, sacrificed many late nights and weekends that permitted me to finish my work and was instrumental in providing words of encouragement when needed. I am truly blessed to have her in my life. My son, Keegan, provided the final impetus to complete my work and also has provided many entertaining moments in the past year. I am sure many more entertaining moments are to still to come. My parents, Rick and Carol, deserve a great deal of thanks for allowing me to become a biologist over the years. I cannot convey the excitement I felt when I was finally given permission to install a pond in the backyard of my childhood home. My first pond served only to nurture my lifelong fascination with water and the creatures living in it. R. C. Henrichs, we made it!

ACKNOWLEDGEMENTS

I thank my committee chair, Dr. Lisa Campbell, and my committee members, Dr. John R. Gold, Dr. James R. Manhart, and Dr. Alan E. Pepper for their guidance and support throughout the course of this research. Thanks also go to the faculty and staff in the Department of Biology for their support and teaching assistantships throughout my time at Texas A&M University. I also want to thank the ECOHAB program (NOAA) for providing funding to complete the field study. A special thanks is given to all individuals and groups, including the Fish and Wildlife Research Institute (Florida Fish and Wildlife Conservation Commission) and Texas Parks and Wildlife Department, who collected field samples from the various locations in the Gulf of Mexico.

TABLE OF CONTENTS

Page

ABSTRACT ii

DEDICATION iv

ACKNOWLEDGEMENTS v

TABLE OF CONTENTS vi

LIST OF FIGURES viii

LIST OF TABLES ix

CHAPTER I INTRODUCTION 1

- Molecular Analyses 3
- Microsatellite Markers 4
- Phytoplankton Population Studies 5
- Physical Models 7
- Objectives 8

CHAPTER II PCR AMPLIFICATION OF MICROSATELLITES FROM SINGLE CELLS OF *KARENIA BREVIS* PRESERVED IN LUGOL'S IODINE SOLUTION 10

- Introduction 10
- Materials and Methods 11
- Results and Discussion 14

CHAPTER III GENETIC DIVERSITY AMONG CLONAL ISOLATES OF *KARENIA BREVIS* 21

- Introduction 21
- Materials and Methods 23
- Results 27
- Discussion 39

Page

CHAPTER IV POPULATION-GENETIC STRUCTURE OF *KARENIA BREVIS* IN THE GULF OF MEXICO 43

Introduction.. 43
Materials and Methods 47
Results..... 50
Discussion.... 60

CHAPTER V CONCLUSIONS: GENETIC DIVERSITY AND POPULATION STRUCTURE IN *KARENIA BREVIS* WHAT HAVE WE LEARNED?..... 65

Introduction 65

REFERENCES 74

APPENDIX A 88

LIST OF FIGURES

FIGURE		Page
1.1	A single cell of *Karenia brevis*	2
2.1	PCR amplification products from microsatellite *Kbr*10	16
3.1	Principal coordinates analysis showing the distribution of cultures based on multilocus genotypes from 16 microsatellite markers and 38 cultures	35
3.2	Allele frequency distribution for microsatellite marker *Kbr*5 from cultures and field cells	40
4.1	Collection locations for samples from the Gulf of Mexico	45
4.2	Spatial autocorrelation correlogram of correlation at four distance classes	59
4.3	Results of the STRUCTURE analysis separated by year for the no admixture model with 'locprior' option invoked	60

LIST OF TABLES

TABLE		Page
2.1	Size range of alleles detected at five microsatellites among 27 clonal cultures of *Karenia brevis*	18
2.2	Cultures genotyped to obtain a size range of alleles at five microsatellite loci in *Karenia brevis*	19
3.1	Collection information for cultured isolates of *Karenia brevis*	22
3.2	Primer sequences for new microsatellite markers	28
3.3	Microsatellite allele sizes for 40 cultures of *Karenia brevis* for microsatellites *Kbr*1 - *Kbr*10	29
3.4	Microsatellite allele sizes for 40 cultures of *Karenia brevis* for microsatellites *Kbr*11 - *Kbr*19	31
3.5	Summary of microsatellite information for cultures of *Karenia brevis*	37
3.6	Summary of microsatellite information for field cells of *Karenia brevis*	38
4.1	Collection location, date, and number of cells included in genetic analyses for surface samples used in this study	46
4.2	Sample diversity statistics for five microsatellite loci	51
4.3	Diversity statistics for grouped samples	53
4.4	Pairwise values of genetic divergence between samples	55
4.5	Pairwise values of genetic divergence between groups	58

CHAPTER I

INTRODUCTION

Harmful algal blooms (HABs) have been increasing in occurrence both globally and in the Gulf of Mexico (Brand and Compton 2007; Magaña *et al*. 2003). The major bloom-forming HAB species in the Gulf of Mexico (hereafter Gulf) is *Karenia brevis*, a haploid, toxic, unarmored dinoflagellate (Fig. 1.1). *Karenia brevis* has been found throughout the Gulf (Brand and Compton 2007; Licea *et al*. 2004; Magaña *et al*. 2003) and as far north as North Carolina (Tester *et al*. 1991), with two regions, the west coast of Florida and the coastline of Texas, garnering the most attention. Blooms of *K. brevis* frequently have been associated with fish kills and respiratory problems in humans (Steidinger *et al*.1998); however, information about this species, including the health effects of *K. brevis* toxins on humans and marine mammals, several life-history stages, how blooms are initiated, sustained, and dissipated, and the relationships, if any, between blooms in the Gulf of Mexico, remains relatively unknown.

Increased knowledge in all areas of research regarding *K. brevis* will allow better management and mitigation strategies to be developed. These strategies are important because they can reduce the overall impact of *K. brevis* blooms on humans and other species. Previous work has focused on identification and quantification of *K. brevis* toxins and their effects (Cheng *et al.* 2005; Pierce *et al.* 2005; Leblond and Chapman 2002; Casper *et al*. 2007), growth rates (Brown *et al.* 2006), swimming behavior, and physiological responses (McKay *et al.* 2006; Schaeffer *et al.* 2007). The numerous

harmful impacts of *K. brevis* also motivated studies in monitoring, early detection, and forecasting of blooms (Robbins *et al*. 2006; Wynne *et al*. 2005).

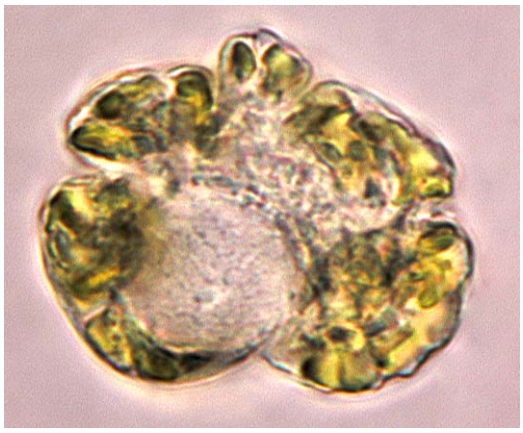

Figure 1.1 A single cell of *Karenia brevis*.

Future work in the early detection and forecasting of blooms must focus on predictive understanding of how, why, and when blooms of *K. brevis* form. This can be accomplished by incorporating information about physical processes with the genetic population structure and understanding how those processes work to move, condense, and dissipate blooms. Determining the population structure of *K. brevis* in the Gulf of Mexico could provide information applicable to many different areas of research. Many studies concerning *K. brevis* focus primarily on blooms in Florida and/or blooms in Texas (Brand and Compton 2007; Hetland and Campbell 2007; Magaña *et al*. 2003; Stumpf *et al.* 2008; Walsh *et al.* 2006). These two regions, Florida and Texas, are well sampled and, being geographically distant from each other, may provide a good estimate

of population structure for *K. brevis* in the Gulf of Mexico. If genetic differences are found between samples of *K. brevis* from Florida and Texas, then investigators looking for information about where blooms are formed could shift their focus (from the entire Gulf) to local or regional environments. On the other hand, if no genetic difference is found, it could lead investigators to look for regions with physical processes with the potential to distribute cells of *K. brevis* throughout the Gulf. This information also could allow physical models of bloom initiation and movement (Hetland and Campbell 2007; Stumpf *et al.* 2008) to be tested by providing a phylogenetic tree of local blooms and their probable sources, ultimately aiding local authorities in tracking and forecasting future bloom events and their potential impacts on coastal communities. Identifying the population structure of *K. brevis* requires a technique to identify individual blooms and molecular tools are currently available to do just that.

MOLECULAR ANALYSES

Most of the molecular work conducted on *K. brevis* has focused on determining whether *K. brevis* was present in a sample and in what concentration. The presence of *K. brevis* in a field sample could indicate that a bloom is forthcoming and/or that shell fishing in the immediate area would need to be closed. In order to establish a quick and reliable identification protocol, Gray *et al.* (2003) used real-time reverse transcription-PCR (RT-PCR) to target the *rbcL* gene and successfully detected and quantified *K. brevis* concentrations in field samples. Mikulski *et al*. (2005) developed genetic markers for *K. brevis*, using large subunit ribosomal RNA (LSU rRNA), and were able to identify *K. brevis* in samples containing different but morphologically similar species, while

Casper *et al*. (2007) developed a protocol that provided field detection and quantification of *K. brevis* through the use of a handheld, nucleic acid sequence-based amplification (NASBA) analyzer.

An important note is that these studies all focused on identifying one species from a field sample containing many, possibly related, different species. The molecular markers were species specific, but had no ability to distinguish one cell of *K. brevis* from another. Determining the genetic population structure of a species requires markers that can distinguish differences among individuals and/or populations. An attempt to identify intra-specific variation (Loret *et al*. 2002) identified physiological differences among five isolates of *K. brevis* and attempted to identify genetic differences, using sequences from the internal transcribed spacer (ITS) and 18S rRNA regions of *K. brevis*. However, the sequences obtained from the five isolates of *K. brevis* were identical for both the ITS and 18S regions (Loret *et al*. 2002). Mikulski *et al*. (2005) also noted that the LSU rRNA sequence was identical among all isolates of *K. brevis* tested. Identical sequences are ideal when identifying a species from other similar organisms but are not suitable for use when conducting a population study and/or identifying intra-specific differences. This research led to the use of microsatellites as the molecular tool of choice to identify genome-based differences among isolates of *K. brevis*.

MICROSATELLITE MARKERS

Microsatellites, also known as simple sequence repeats (SSRs), are stretches of DNA that contain tandemly repeated sequences of 1-6 base pairs (bp) in length, are presumed to be selectively neutral, and are used widely in population genetics (Selkoe

and Toonen 2006). Nine nuclear-encoded microsatellites were identified in *K. brevis* and found to be polymorphic among thirteen isolates from Florida and Texas (Renshaw *et al*. 2006). Microsatellite markers have been used successfully in several studies of marine phytoplankton, including *Alexandrium tamarense* (Nagai *et al*. 2007), *Cochlodinium polykrikoides* (Nagai *et al.* 2009), *Ditylum brightwellii* (Rynearson and Armbrust 2004), *Gymnodinium catenatum* (Bolch *et al*. 1999), *Pseudo-nitzschia multiseries* (Evans *et al*. 2004), and *Pseudo-nitzschia pungens* (Evans *et al*. 2005). These studies showed geographically separated populations of marine phytoplankton to be distinct genetically, with one exception; Evans *et al*. (2005) concluded the German North Sea supported a single unstructured population of *Pseudo-nitzschia pungens.*

PHYTOPLANKTON POPULATION STUDIES

Rynearson and Armbrust (2004), in a population study of *Ditylum brightwellii*, a marine diatom, used three microsatellite markers to identify three genetically distinct populations in two connected estuaries, Puget Sound and the Strait of Juan de Fuca, Washington, USA. Rynearson and Armbrust (2004) believed that the water currents in the estuaries did not provide enough force to maintain a constant gene flow between populations and therefore allowed them to diverge. The extent of genetic divergence was not correlated with distance or time; the two most genetically diverged populations, based on the microsatellite data, had identical 18S rDNA sequences. In another study, Nagai *et al*. (2007) used nine microsatellite markers to identify distinct populations of a marine dinoflagellate, *Alexandrium tamarense*, along the coastlines of Japan and Korea. Nagai *et al*. (2007) showed that genetic distance did correlate with geographic distance,

suggesting that tidal currents did not provide enough of a dispersal mechanism to maintain genetic homogeneity. These studies indicate that phytoplankton populations can be genetically distinguished when separated by as few as several hundred kilometers (Rynearson and Armbrust 2004) or as many as several thousand kilometers (Nagai *et al*. 2007). In a study of another red tide forming dinoflagellate, *Cochlodinium polykrikoides,* Nagai *et al.* (2009) used ten microsatellite markers and was able to show samples from the Sea of Japan were more similar, genetically, to other samples from the Sea of Japan than they were to samples taken elsewhere. In the same study, samples of *C. polykrikoides* from the Pacific coast of Japan were shown to be more similar, genetically, to other samples from the Pacific coast of Japan than to samples taken elsewhere. However, there was no correlation between genetic distance and geographic distance, and Nagai *et al*. (2009) suggested that a large genetic barrier had occurred between the populations in the Sea of Japan and populations along the Pacific coast. In contrast, Evans *et al*. (2005) used six microsatellite markers to genotype *Pseudo-nitzschia pungens*, a marine diatom, and showed that isolates from different spatial and temporal samples exhibited weak genetic divergence (only 22 of 192 F_{ST} values differed significantly from zero). However, no correction for multiple tests was conducted and the true number of significant tests is likely less. There were no apparent barriers to gene flow in the geographic area sampled, suggesting that the *P. pungens* in the North Sea along the coast of Germany is well-mixed and represents a single, large population (Evans *et al*. 2005).

One interesting note about these studies concerns the amount of diversity observed among phytoplankton. Though asexual reproduction is present in the life cycle, there were a large number of unique genotypes in each study. The three populations identified in Rynearson and Armbrust (2004) possessed different allele distributions, were composed of cells with different physiological qualities, and produced 101 unique genotypes from 105 isolates (96%). Nagai *et al*. (2007) found that, while four sample sites produced a few repeated genotypes (exact numbers are not given), six other sample sites produced no repeated genotypes among them (n=300). Evans *et al.* (2005) also identified high levels of diversity; 453 unique genotypes were identified from 464 isolates (98%). The use of microsatellite markers appears to be ideal for identifying differences among individual cells, even with a small number of loci, and can uniquely identify, and possibly link, individual populations. The ability to link local populations will allow further research into the physical processes that serve to concentrate, move, and dissipate blooms.

PHYSICAL MODELS

The extent of genetic differentiation among isolates of a species taken from different geographic locations appears largely dependent upon currents between sample locations (Evans *et al*. 2005; Nagai *et al*. 2007). Hetland and Campbell (2007) proposed a numerical model where the timing and magnitude of bloom formation along the western coast of the Gulf of Mexico can be predicted. However, their model does not predict movement or dissipation of a bloom after it has moved near shore. Stumpf *et al*. (2008) proposed a model to explain development of blooms under low nutrient

conditions in the eastern Gulf of Mexico. Although the model described mean conditions under which blooms may form, it was not able to predict bloom distributions accurately due to daily changes in wind patterns and circulation fields (Stumpf *et al*. 2008). Information about large- and small -scale population structure of *K. brevis* will permit further testing of these models and could be used in identifying physical mechanisms needed for bloom movements into certain areas.

OBJECTIVES

There are currently several possibilities concerning population structure of *K. brevis* in the Gulf of Mexico: (i) one large, source population of *K. brevis* that is occasionally and randomly dispersed into the coastal waters of Gulf states, (ii) two, or more, physically distinct (and presumably genetically distinct) populations that bloom independently of one another and are separated by a currently unknown physical barrier, or (iii) blooms in the western Gulf of Mexico are the result of blooms being moved from the eastern Gulf of Mexico (Florida to Texas).

This research project will test whether significant genetic differences exist between blooms of *K. brevis* in Florida and Texas. This will be accomplished by genotyping and characterizing isolates of *K. brevis* obtained from bloom events in Florida and Texas that have occurred over the last fifty years. Field populations of *K. brevis* will be genotyped from single cells taken from blooms occurring in waters off of Florida and Texas in 2005, 2006 and 2009. Data obtained will help to identify the genetic population structure, if any, of *K. brevis* blooms in the Gulf of Mexico and aid in the study of many other aspects of research on *K. brevis*, including monitoring, early

detection, and forecasting (and possibly mitigation) by providing information that could link small bloom populations to each other. When combined with physical parameters (i.e. wind speed, wind direction, current flow), this information could help identify the location of possible bloom sources.

CHAPTER II

PCR AMPLIFICATION OF MICROSATELLITES FROM SINGLE CELLS OF *KARENIA BREVIS* PRESERVED IN LUGOL'S IODINE SOLUTION*

INTRODUCTION

The major harmful algal bloom (HAB) species in the Gulf of Mexico is *Karenia brevis,* an unarmored dinoflagellate responsible for both fish kills and respiratory problems in humans (Steidinger *et al*. 1998). Factors influencing initiation, development, and dissipation of blooms of *K. brevis*, however, are not well understood. A more detailed understanding of genetic diversity within and among blooms is needed so that the dynamics and demography of this dinoflagellate can be studied in relation to environmental parameters.

Hypervariable, nuclear-encoded genetic markers such as microsatellites are powerful tools for assessment of population structure and have been developed for several dinoflagellate species (Nagai *et al*. 2006, 2007), including *K. brevis* (Renshaw *et al*. 2006). In these and other studies (Rynearson and Armbrust, 2004) of genetic diversity among phytoplankton species, clonal cultures were required for extraction of sufficient quantities of DNA for genotyping. Unfortunately, in contrast to other phytoplankton species that have been studied, e.g., *Alexandrium tamarense* (Nagai *et al*. 2007), *Ditylum brightwellii* (Rynearson and Armbrust 2004), and *Emiliania huxleyi* (Iglesias-Rodriguez *et al*. 2006), there are few isolates of *K. brevis* available for genetic studies (http://ccmp.bigelow.org/). An advantage in working with dinoflagellate species

*Reprinted from Springer Marine Biotechnology vol. 10, 2008, 122-127, PCR amplification of microsatellites from single cells of *Karenia brevis* preserved in Lugol's iodine solution, D. W. Henrichs, M. A. Renshaw, C. A. Santamaria, B. Richardson, J. R. Gold, L. Campbell, with kind permission from Springer Science and Business Media, Copyright 2008.

(e.g., *A. tamarense*) is the use of resting-stage cysts to establish clonal cultures. The resting stage cyst for *K. brevis*, however, has not been identified or reproducibly produced in the laboratory. Consequently, clonal cultures of *K. brevis* must be established *de novo* from individual cells isolated from a bloom, a difficult, time consuming, and challenging task because of the high mortality of isolated single cells (B. Richardson, personal observation).

Here, we describe a simple procedure for PCR (polymerase-chain-reaction) amplification of nuclear-encoded microsatellites from Lugol's iodine (LI) preserved single cells of *K. brevis*. The procedure allows microsatellite genotypes to be acquired from a large number of individual cells within a bloom. Successful PCR amplification of microsatellites from cells preserved in LI solution, the preferred preservation method for marine flagellates, has the advantages that (i) external cell morphology is preserved for identification, and (ii) genotypes can be acquired from historical and time-course samples, permitting tests of hypotheses linking genetic diversity and population structure of *K. brevis* with temporally varying physiological and ecological parameters.

MATERIALS AND METHODS

Cell preservation and isolation

A 1.3 ml aliquot of cultured cells of *K. brevis* (SP1 isolate, Loret *et al*. 2002) was placed in a 1.5 ml Eppendorf tube, stained with 50-µl of LI solution (10g I_2, 20g KI, 20ml glacial acetic acid, 200ml dH_2O), and placed in the dark at 4°C for three hours. Subsequently, 10 µl 1M sodium thiosulfate (Tittel *et al*. 2003) was added to destain cells. The tube was then gently inverted four times (LI coloration generally dissipated

immediately) and the cells were ready to isolate once the solution was devoid of color. A 200 µl aliquot of destained cells was placed on a microscope slide and individual cells isolated using a method modified from Ki *et al*. (2005); individual cells were then transferred, using a Pasteur pipet, to a PCR tube (0.2ml; VWR International, West Chester, PA) in a minimum volume (<2 µl) of sterile Optima water (Fisher Scientific, Fair Lawn, NJ). Individual PCR tubes were then observed under an Olympus SZX12 stereomicroscope to confirm presence of a single cell.

DNA extraction and amplification

PCR tubes were centrifuged (1,177 x g) for 30 sec and subjected to three cycles of freeze/thawing (-80°C for one min and 75°C for one min constituted one cycle) to lyse the cells (Sebastián and O'Ryan 2001). The lysate was then subjected to two rounds of PCR amplification. The first round was a multiplex reaction that employed five PCR primer pairs in a 20 µl reaction containing 12 µl Go*Taq* Green Master Mix (Promega, Madison, WI), 5 µl Optima water (Fisher Scientific), and 3 pmol of each forward and reverse primer. The microsatellites amplified were *Kbr*5, *Kbr*7, *Kbr*8, *Kbr*9, and *Kbr*10; details, including primer sequences, of these microsatellites be found in Renshaw *et al*. (2006). Amplification was carried out using a Bio-Rad PTC 100 thermal cycler (Bio-Rad, Hercules, CA) as follows: initial (one cycle) denaturation at 95°C for 180 sec, followed by eight cycles of denaturation at 95°C for 80 sec, annealing at 52°C for 165 sec, extension at 72°C for 80 sec, 50 cycles of denaturation at 95°C for 60 sec, annealing at 52°C for 105 sec, extension at 72°C for 60 sec, and one final extension at 72°C for 30 min. Product in each tube was diluted with 20 µl 1X Tris-EDTA (1XTE) and used as a

template for five separate reactions that used each of the five PCR primer pairs. This second round of PCR employed 10 µl reactions containing 5 µl Go*Taq* Green Master Mix, 1.4 µl Optima water, 5 pmol of fluorescently labeled forward primer, 5 pmol of reverse primer, and 2 µl of template. The fluorescent dyes employed were FAM, HEX (Invitrogen, Carlsbad, CA), and NED **(**Applied Biosystems, Foster City, CA). Amplification was carried out using a Bio-Rad PTC 100 thermal cycler as follows: initial (one cycle) denaturation at 95°C for 180 sec, annealing at 52°C for 120 sec, extension at 72°C for 80 sec, 40 cycles of denaturation at 95°C for 60 sec, annealing at 52°C for 75 sec, extension at 72°C for 60 sec, and one final extension at 72°C for 30 min. PCR products were diluted with 10 µl 1XTE and separated and visualized on a 5% polyacrylamide gel (Long Ranger Singel Pack; Cambrex Bio Science Rockland, Inc., Rockland, ME) using an ABI PRISM 377 DNA sequencer (Applied Biosystems). Gels were run for 2.5 hours at 3kV, 100W, and a laser power of 39mW. A size standard, 400HD ROX (Applied Biosystems), was loaded with each sample in order to estimate fragment sizes. All gels were analyzed using GENESCAN ANALYSIS 3.1.2® (Applied Biosystems); allele-calling was performed with GENOTYPER® software, version 2.5 (Applied Biosystems) and with STRAND 2.3.48 (UC Davis-Veterinary Genetics Lab, Davis, CA). Genotypes obtained were compared to genotypes compiled previously from cetyl trimethylammonium bromide (CTAB)-extracted DNA (after Doyle and Doyle 1990) from a pellet of cultured cells of the SP1 isolate of *K. brevis*. Comparison of genotypes obtained from single cells with those from pooled cells of a culture initiated from a single cell was to confirm that products obtained from single cells were identical

in size to products obtained from a cell pellet of the same culture and not a product of random amplification. The first round of PCR utilized extended denaturation, annealing, and extension times in order to maximize product from each cycle. The extended denaturation time ensured that all double-stranded DNA was denatured. The extended annealing time allowed primers from all five microsatellites to anneal to their target sequence. This step appeared especially critical in insuring equal amplification of each microsatellite and minimizing the chance that a single microsatellite would monopolize available resources. The longer extension time ensured complete synthesis of the new strands. The two rounds of PCR amplification were necessary to increase the copy number of each target sequence and allow template DNA from a single cell to be used in multiple reactions. Each reaction in the second round of PCR amplified a single microsatellite and included a fluorescently labeled forward primer.

RESULTS AND DISCUSSION

Initially, all five microsatellites were amplified successfully from single cells of *K. brevis* (SP1 isolate) fixed with LI solution. Out of ten trials, three microsatellites (*Kbr*7, *Kbr*8, and *Kbr*9) amplified successfully in all cases, while the two remaining microsatellites (*Kbr*5 and *Kbr*10) amplified successfully in 90% of trials. Amplification also was successful with single, LI-preserved cells (n = 16) sampled from a bloom occurring in Fulton Harbor near Rockport, Texas, in the fall of 2000 and that had been stored refrigerated at 4°C for six years. We then used the protocol to amplify the five microsatellites from single cells (n = 129) isolated from a recent bloom of *K. brevis* sampled from shorelines around Corpus Christi, Texas, during the fall of 2005. *Kbr*9

amplified successfully in 97% of trials; *Kbr*8 amplified successfully in 96% of trials; *Kbr*5 amplified successfully in 87% of trials; *Kbr*10 amplified successfully in 85% of trials; *Kbr*7 amplified successfully in 71% of trials.

PCR amplifications from single cells frequently generated multiple, extraneous bands for all five microsatellites; an example is shown in Figure 2.1 (SP1 Single Cells 4 and 5). CTAB-extracted DNA from cell pellets of the same culture did not produce multiple bands due, presumably, to a higher initial copy number of template DNA. The target-band range for each of the five microsatellites (Table 2.1) was determined based on observed, single-band genotypes (phenotypes) of CTAB-extracted DNA from 27 different cultures (Table 2.2). The extraneous bands were observed in about 40% of the amplifications from single cells and invariably fell outside of the target band range (Fig. 2.1). In addition, the target band was the brightest band in the target range and almost always the brightest band observed. Finally, in amplifications of single cells from the same culture, the same target band was observed, whereas the extraneous bands would be of different sizes.

Figure 2.1 PCR amplification products from microsatellite *Kbr*10. Samples 1-5 are from single-cells of *K. brevis* (SP1 isolate); samples 6 and 7 are from CTAB-extracted DNA (SP1 isolate); sample 8 is a negative control (sterile water added instead of template DNA); sample 9 is a negative control (nothing added in place of template DNA); and samples 10-12 are from single-cell samples of *K. brevis* isolated from an LI-preserved field sample (20051013-3) collected in 2005. Target band range is 169-181bp. SP1 allele size is 177bp. Similar results were obtained at the other four microsatellites.

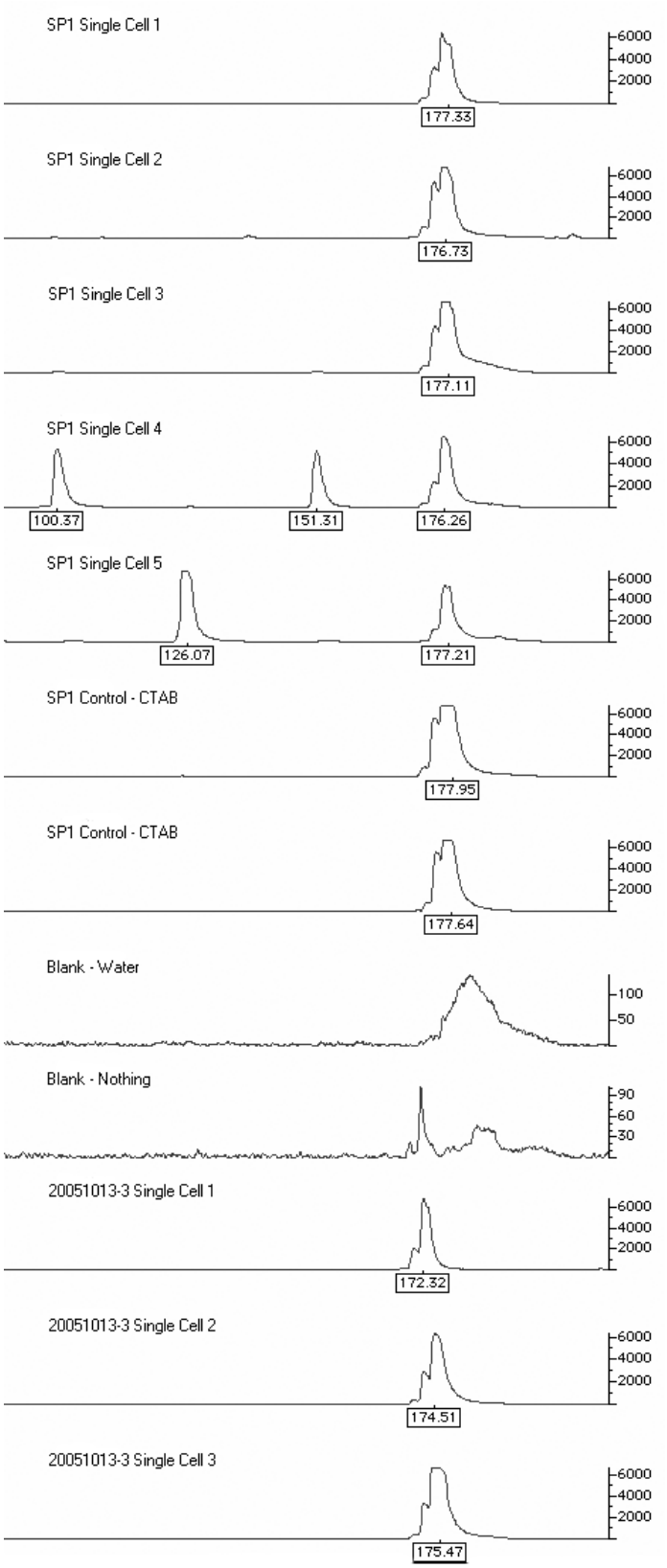
SP1 Single Cell 1
6000
4000
2000
177.33
SP1 Single Cell 2
6000
4000
2000
176.73
SP1 Single Cell 3
6000
4000
2000
177.11
SP1 Single Cell 4
6000
4000
2000
100.37
151.31
176.26
SP1 Single Cell 5
6000
4000
2000
126.07
177.21
SP1 Control - CTAB
6000
4000
2000
177.95
SP1 Control - CTAB
6000
4000
2000
177.64
Blank - Water
100
50
Blank - Nothing
90
60
30
20051013-3 Single Cell 1
6000
4000
2000
172.32
20051013-3 Single Cell 2
6000
4000
2000
174.51
20051013-3 Single Cell 3
6000
4000
2000
175.47

Table 2.1 Size range of alleles detected at five microsatellites among 27 clonal cultures of *Karenia brevis*.

Microsatellite	*Range (in base pairs)*
*Kbr*5	182 - 190
*Kbr*7	252 - 261
*Kbr*8	128 - 146
*Kbr*9	158 - 167
*Kbr*10	169 - 181

Table 2.2 Cultures genotyped to obtain a size range of alleles at five microsatellite loci in *Karenia brevis*.

Collection Number	Collection Location	Collection Date
CCFWC250	Neptune Beach, FL	October 1999
CCFWC251	Neptune Beach, FL	October 1999
CCFWC252	Neptune Beach, FL	October 1999
CCFWC253	Duck Key, FL	February 1995
CCFWC254	New Pass, FL	October 1999
CCFWC256	Charlotte, FL	May 1996
CCFWC257	Charlotte, FL	May 1996
CCFWC258	Mexico Beach, FL	June 1998
CCFWC259	Mexico Beach, FL	June 1998
CCFWC260	Mexico Beach, FL	June 1998
CCFWC261	Apalachicola Bay, FL	June 1998
CCFWC262	Apalachicola Bay, FL	June 1998
CCFWC263	Panacea, FL	May 1996
CCFWC265	Panacea, FL	May 1996
CCFWC266	South Padre Island, TX	October 1999
CCFWC267	South Padre Island, TX	October 1999
CCFWC268	John's Pass, FL	1953
CCFWC269	Corpus Christi Bay, TX	1986
CCMP2228	Sarasota, FL	August 2001
CCMP2229	Manasota Key, FL	August 2001
CCMP2281	Navarre, FL	September 1999
CCMP718	John's Pass, FL	1953
SP1	South Padre Island, TX	October 1999
SP2	South Padre Island, TX	October 1999
TSP3	South Padre Island, TX	October 1999
NTSP3	South Padre Island, TX	October 1999
NBK	Nueces Bay, TX	February 2002

Different approaches to PCR amplification of microsatellites from single cells of *K. brevis* were also evaluated: 95% ethanol preservation and pre-extraction precipitation, ethanol precipitation after destaining of LI-fixed cells, Chelex (Bio-Rad, Hercules, CA) extraction as described by Richlen and Barber (2005), CTAB extraction (Doyle and Doyle 1990), and a freeze/thaw, buffer-incubation method as described by Kai *et al*. (2006). Ethanol preservation and precipitation after destaining of LI-fixed cells yielded cells that were difficult to ascertain visually, precluding species identification and confirmation of single cells inside PCR tubes. Chelex extraction required a small volume (~10 µl) of Chelex solution to be added to the PCR tube containing the single cell. Once Chelex extraction is complete the entire supernatant (minus beads) must be transferred to

another PCR tube and used as template. This significantly reduced DNA template concentration for the initial PCR amplification and resulted in inconsistent amplification. The CTAB extraction method contained several steps that involved addition/removal of solutions to the tube containing the cellular DNA, and similar to Chelex extraction significantly reduced DNA template concentration. The freeze/thaw buffer-incubation (Kai *et al*. 2006) also involved addition of buffer, again reducing initial DNA template concentration.

We also tried the whole-genome-amplification (WGA) method, using GENOMIPHI (GE Healthcare, UK) and the phi29 polymerase (Raghunathan *et al*. 2005). Results using WGA produced gels that were difficult to score because of apparent (and extensive) non-specific amplification, presumably artifacts of background synthesis (Hutchison *et al*. 2005; Raghunathan *et al*. 2005). Successful amplifications were achieved using destained, LI-fixed cells and the lysis buffers (SDS/Proteinase K and TritonX-100/Proteinase K) as described in Kai *et al*. (2006). However, resulting gels contained numerous additional bands relative to those observed using the freeze/thaw extraction method.

The method reported here permits successful microsatellite genotyping of single cells of *K. brevis* and bypasses the need to establish cultures. The method is straightforward and relatively rapid, and it significantly reduces the amount of time needed to obtain multiple genotypes from a bloom.

CHAPTER III

GENETIC DIVERSITY AMONG CLONAL ISOLATES OF *KARENIA BREVIS*

INTRODUCTION

Factors influencing bloom dynamics of *Karenia brevis*, in particular bloom initiation, are not well understood. Prior studies have shown considerable physiological variation exists among clones of *K. brevis* but little work has been done to identify genetic variation (Loret *et al*. 2002; Brown *et al.* 2006; McKay *et al.* 2006; Errera *et al.* 2010). Because blooms of dinoflagellates result from accumulations of haploid vegetative cells that reproduce by binary fission, it might be expected that genetic diversity would be low within a bloom. In fact, in a number of bloom-forming dinoflagellates, high levels of genetic diversity have been observed based on microsatellites; (Nagai *et al*. 2007, 2009; Alpermann *et al*. 2009; Lowe *et al.* 2010; Erdner *et al*. 2011). A more detailed understanding of genetic diversity within and among blooms of *K. brevis* is needed so that the dynamics of toxic blooms of *K. brevis* can be described and links to environmental factors investigated.

Here I focus on the use of microsatellite markers to identify genetic diversity among clones of *K. brevis*. Nine microsatellite markers currently exist for *K. brevis* (Renshaw *et al.* 2006). In this study, I first developed new microsatellite markers to combine with previously identified microsatellites for *K. brevis* in order to obtain a better estimate of the genetic diversity in cultured isolates of *K. brevis.* Cultures of *K. brevis* have proven very difficult to establish; consequently, relatively few isolates are

available for study (Table 3.1). To obtain an estimate of how well the current isolates represent the field population, genotypes from single cells of *K. brevis* were compared with genotypes from cultured strains.

Table 3.1 Collection information for cultured isolates of *Karenia brevis*.

Collection number	Collection location[a]	Collection date
CCFWC121	Clam Pass, FL	July 2006
CCFWC122	Clam Pass, FL	July 2006
CCFWC123	Clam Pass, FL	July 2006
CCFWC124	Clam Pass, FL	July 2006
CCFWC125	Clam Pass, FL	July 2006
CCFWC126	5 nau. mi. W of Stump Pass, FL	July 2006
CCFWC127	5 nau. mi. W of Stump Pass, FL	July 2006
CCFWC128	5 nau. mi. W of Stump Pass, FL	July 2006
CCFWC129	Mouth of Caloosahatchee River, FL	July 2006
CCFWC130	Mouth of Caloosahatchee River, FL	July 2006
CCFWC250	Neptune Beach, FL	October 1999
CCFWC251	Neptune Beach, FL	October 1999
CCFWC252	Neptune Beach, FL	October 1999
CCFWC253	Duck Key, FL	February 1995
CCFWC254	New Pass, FL	October 1999
CCFWC256	Charlotte, FL	May 1996
CCFWC257	Charlotte, FL	May 1996
CCFWC258	Mexico Beach, FL	June 1998
CCFWC259	Mexico Beach, FL	June 1998
CCFWC260	Mexico Beach, FL	June 1998
CCFWC261	Apalachicola Bay, FL	June 1998
CCFWC262	Apalachicola Bay, FL	June 1998
CCFWC263	Panacea, FL	May 1996
CCFWC265	Panacea, FL	May 1996
CCFWC266	South Padre Island, TX	October 1999
CCFWC267	South Padre Island, TX	October 1999
CCFWC268	John's Pass, FL	1953
CCFWC269	Corpus Christi Bay, TX	1986
CCMP2228	Sarasota, FL	August 2001
CCMP2229	Manasota Key, FL	August 2001
CCMP2281	Navarre, FL	September 1999
CCMP2820	New Pass, Sarasota, FL	February 2005
CCMP718	John's Pass, FL	1953
EPA JR	Pensacola Beach, FL	1999
NBK	Nueces Bay, TX	February 2002
NOAA-1	Charlotte Harbor, FL	1999/2000
NSP3[b]	South Padre Island, TX	October 1999
SP1	South Padre Island, TX	October 1999
SP2	South Padre Island, TX	October 1999
SP3[c]	South Padre Island, TX	October 1999

[a] Approximate collection locations are given.
[b] Called NTSP3 in Henrichs *et al*. (2008).
[c] Called TSP3 in Henrichs *et al*. (2008).

The goal of the present study is to describe the genetic variation among cultured strains of *K. brevis* and to determine if genetic differences exist between isolates from different regions of the Gulf. Clonal cultures of *K. brevis* available from several different laboratories were examined to address the following questions:

1) Are geographic isolates of *K. brevis* from the northern Gulf genetically homogeneous?

2) Are strains currently in culture representative of the genetic diversity present in field populations?

3) Can microsatellite markers provide a diagnostic tool to differentiate clonal cultures?

MATERIALS AND METHODS

DNA isolation

Cell pellets for 40 clonal cultures were obtained by centrifugation (10000*g* for 15 min.) of 1.7mL Eppendorf tubes (VWR; Radnor, PA, USA), each containing approximately 1.5 mL of a dense culture. The resulting supernatant was removed and discarded. Genomic DNA from the cell pellet was extracted using the cetyl trimethylammonium bromide (CTAB) buffer extraction method described by Doyle and Doyle (1990). The 40 strains have been isolated over a span of more than 50 years; most were isolated from water samples taken from the Gulf of Mexico. Collection information for each culture can be found in Table 3.1.

Microsatellite development

Two approaches were taken to develop additional microsatellite markers. In the first approach, expressed sequence tag (EST) sequences from *K. brevis* were

downloaded from GenBank and imported into Sequencher (v4.2; Gene Codes, Ann Arbor, MI, USA). Sequences not containing a microsatellite motif were removed from the dataset. The remaining sequences were visually inspected and sequences lacking suitable flanking regions at both ends of the microsatellite motif were removed. Sequences with identical microsatellite motifs were then aligned and manually inspected to eliminate duplicate sequences.

The second approach identified microsatellite markers from genomic DNA, using the method of Renshaw *et al.* (2006). Primers for all microsatellite markers were developed using Primer3 (http://frodo.wi.mit.edu/primer3/; Rozen and Skaletsky 2000). All microsatellite markers (EST-based and those obtained from genomic DNA) were tested using DNA from four different strains of *K. brevis*; those markers producing a product consisting of a single visible band after gel electrophoresis (2% agarose) were further tested using DNA from all remaining cultures. Polymorphic microsatellites were visually identified by gel electrophoresis (4% agarose). The forward primer of each identified polymorphic microsatellite was labeled with a fluorescent label from Applied Biosystems standard dye filter set D (Applied Biosystems; Foster City, CA, USA). All polymorphic microsatellite markers were tested further using DNA from a wide variety of dinoflagellate species (*Alexandrium monilatum*: CAAE 106; *Crypthecodinium* sp.; *Scrippsiella* sp.; *Karlodinium micrum*: CCMP1974, CCMP2282, CCMP415; *Karenia mikimotoi*: C21 [isolated in 2001 from Corpus Christi Bay, Texas]; *K. papilionacea*: CAWD91; *K. bidigitata*: CAWD92; *K. selliformis*: CAWD79; *Oxyrrhis marina*;

Pfiesteria spp.: CCMP2301, CCMP2362; and *Pseudopfiesteria* sp.: CCMP2089) to confirm specificity of the primers.

Allele sizing

Each culture was genotyped by polymerase chain reaction amplification (10μL reaction) of each microsatellite, from CTAB extracted DNA. Allele sizing was conducted by running the resulting product on a 5% polyacrylamide gel (Long Ranger Singel Pack, Cambrex Bio Science Rockland, Rockland, ME, USA) and ABI Prism 377 genetic analyzer (Applied Biosystems). Gels were analyzed with Genescan 3.1.2 (Applied Biosystems) and alleles (fragment length) scored in Genotyper version 2.5 (Applied Biosystems).

Genetic analysis

Strains were placed into two groups according to the geographic location of where they were collected: Florida (FL, USA) and Texas (TX, USA). For each microsatellite, number of alleles, effective number of alleles, allelic ranges, and estimates of gene diversity (*H*; Nei 1973) were calculated using PopGene v1.32 (Yeh and Boyle 1997). Unbiased estimates of gene diversity were calculated according to Nei (1987) to account for the small sample sizes from each location. Tests for genotypic disequilibrium between pairs of loci were run in Genepop v1.2 (Raymond and Rousset 1995) and Bonferroni correction for multiple tests was applied following the method of Rice (1989).

The number of strains isolated from the waters around Florida is four times higher than the number of isolates from the coast of Texas (32 FL: 8 TX). To account for

this difference, allelic richness estimates for the cultures from Florida were calculated by rarefaction following the method of El Mousadik and Petit (1996). For the rarefaction calculation, the sample size for each microsatellite was equivalent to the number of strains of *K. brevis* producing an allele for that microsatellite. Diversity results from all cultures were compared to those obtained from two different field samples. One field sample was taken during a bloom off the west coast of Florida (26.555°N, 82.477°W; 2006) and the other sample taken from a bloom in Corpus Christi Bay in Texas (2005). Single cells were isolated from both field samples and genotyped with five microsatellites (*Kbr*5, *Kbr*7, *Kbr*8, *Kbr*9, *Kbr*10) following the method of Henrichs *et al*. (2008). To investigate the possibility of genetic structure between culture isolates from opposite sides of the Gulf, a principal coordinates analysis (PCoA), using the matrix of genetic distances between alleles was run, using GenAlEx v6.41 (Peakall and Smouse 2006) with the following settings: distance calculation set to 'haploid-SSR,' distance output options set to 'output total distance only,' and PCoA method set to 'distance-standardized.'

Diagnostic test

To test the utility of microsatellite markers for diagnostic strain confirmation, cell pellets were obtained from several laboratories studying *K. brevis*. The CCFWC268 strain was obtained from six different laboratories and the CCMP718 strain was obtained once from one laboratory and three times (2000, 2004, 2007) from the National Center for Marine Algae and Microbiota (NCMA). Genotype information was obtained from

CTAB extracted DNA of each strain received as described above and the resulting PCR products were analyzed as detailed above.

RESULTS

New microsatellite markers

Nine new microsatellite markers were identified, bringing the total number of microsatellite markers from *K. brevis* to eighteen. Primer sequences and allelic ranges for the nine new microsatellites can be found in Table 3.2. Of the nine, only one microsatellite produced an allele in all 40 strains of *K. brevis*. The remaining eight each failed to amplify in all 40 strains, in spite of repeated amplification attempts, which suggested the presence of null alleles. Two microsatellites (*Kbr*12, *Kbr*14) each failed to amplify in thirteen strains, though the strains failing to amplify differed between the two microsatellites. Of the six remaining microsatellites, five amplified successfully in thirty nine strains and one amplified successfully in only thirty eight strains. None of the microsatellites produced observable bands after gel electrophoresis (2% agarose) when tested with extracted DNA of other plankton species, which confirmed the specificity of the primers to *K. brevis.*

Table 3.2 Primer sequences for new microsatellite markers.

Microsatellite	Primer sequence	Repeat sequence	T_A	N_A	Size Range
Kbr11[a]	**F**: GGTCACGCTGGTATCATTTGT **R**: GGTGTCATTGAAGGAGTCTGCC	$(GAT)_4$9bp$(GAT)_5$GAC$(GAT)_7$	52	6	169-184
Kbr12[b]	**F**: GCAACAGATGCTGATAGTCCGAAG **R**: GCTGTCTGATTCGTATCCTTC	$(GAG)_9$	52	5	201-213
Kbr13	**F**: TACATATTTGCACGAGAGACACTAC **R**: CTGTGGTCATCGTCATCAAC	$(GAT)_8$	52	12	200-245
Kbr14	**F**: ATTAAAACAACAAAAGGACAAGTG **R**: CGATGAAGATGATGAAGATTGTTAT	$(TAGA)_{10}$	52	14	262-338
Kbr15	**F**: CCTCCTACAAATTGGACCTG **R**: AATTCCCAAGTAGCCCAAGT	$(CT)_{15}$2bp$(CT)_3$	52	9	196-212
Kbr16	**F**: CATGTGTTTTCAACCCAACA **R**: TCCGATTCAGCATCAAATCT	$(AG)_{11}$	52	7	165-179
Kbr17	**F**: CCATGTCCACAAAGCATGTA **R**: TGCCATTCTGGAAAGAAGAG	$(CT)_{17}$	52	13	248-280
Kbr18[b]	**F**: CCTTGAACTGCAAAGAGTGA **R**: TTACAAAAGCAGCAAAGTGG	$(GT)_{17}$	52	16	109-143
Kbr19[b]	**F**: TGTGCATGTAAGAGACTGTGG **R**: GGTTAAAGGGTCTTGGCTTT	$(TG)_{12}$	52	9	116-134

[a] EST from T. McLean.
[b] EST from GenBank.

Table 3.3 Microsatellite allele sizes for 40 cultures of *Karenia brevis* for microsatellites *Kbr*1 - *Kbr*10. Two sizes indicate two alleles detected. No value indicates a failure to amplify an allele.

	Locus								
Culture	*Kbr*1	*Kbr*3	*Kbr*4	*Kbr*5	*Kbr*6	*Kbr*7	*Kbr*8	*Kbr*9	*Kbr*10
CCFWC121	258	274	278	190	231	261	146	161	173
CCFWC122	249	235	266	182	223	261	138	161	177
CCFWC123	261	235	266	184	225	261	134	161	175
CCFWC124	249	253	276	190	225	261	130	158	175
CCFWC125	252	241	262	190	223	261	136	164	189
CCFWC126	261	235	268	182	225	261	126	161	173
CCFWC127	258	241	266	182	223	255	136	158	173
CCFWC128	258	250	268	188	223	261	132	161	171
CCFWC129	258	241	270	188	225	261	134	161	173
CCFWC130	252	238	270	188	223	261	134	164	179
CCFWC250	246	247	270	188	221	261	134	161	169
CCFWC251	255/ 258	235/ 247	264/ 274	190	223	261	128/ 132	161	171/ 175
CCFWC252	258	235	266	188	225	261	132	161	175
CCFWC253	258	235	268	184	225	261	144	161	169
CCFWC254	264	256	268	186	225	261	132	161	175
CCFWC256	258	232/ 247	264	190	224	258/ 261	128/ 138	158/ 164	173
CCFWC257	258	232	264	182	229	261	144	158	175
CCFWC258	261	238	276	182	221	261	108	161	173
CCFWC259	258	235	264	182	219	261	144	161	171
CCFWC260	258	235	264	182	219	261	144	161	171
CCFWC261	255	247	272	188	223	258	146	164	179
CCFWC262	246	232	272	188	225	258	146	164	179
CCFWC263	258	232	270	190	221	261	132	161	173
CCFWC265	264	238	268	190	225	258	138	167	171
CCFWC266	258	235	264	190	221	261	128	161	175
CCFWC267	255	229	264	182	225	261	146	161	177
CCFWC268	267	235	266	186	221	261	377	164	173
CCFWC269	249	229	264	184	221	261	132	164	181

Table 3.3 Continued.

	Locus								
Culture	*Kbr*1	*Kbr*3	*Kbr*4	*Kbr*5	*Kbr*6	*Kbr*7	*Kbr*8	*Kbr*9	*Kbr*10
CCMP2228	258	235	266	190	221	261	134	161	179
CCMP2229	249	244	264	180	221	261	134	161	
CCMP2281	258	235	270	182	219	261	130	161	173
CCMP2820	261	235	268	184	221	261	134	161	171
CCMP718	261	235	252	184	219	261	128	161	181
EPA JR	258	256	268	190	221	264		161	173
NBK	252	232	272	182	221	261	134	161	177
NOAA-1	255	238	270	188	223	261	136	161	173
NSP3[a]	249	235	266	190	221	252	138	161	175
SP1	264	235	272	182	221	261	128	164	177
SP2	258	238	266	182	223	261	132	161	175
SP3[b]	264	235	270	186	223	258	134	161	173

[a] Originally named NTSP3 in Henrichs *et al*. (2008).
[b] Originally named TSP3 in Henrichs *et al.* (2008).

Table 3.4 Microsatellite allele sizes for 40 cultures of *Karenia brevis* for microsatellites *Kbr*11 - *Kbr*19. Two sizes indicate two alleles detected. No value indicates a failure to amplify an allele.

	Locus								
Culture	*Kbr*11	*Kbr*12	*Kbr*13	*Kbr*14	*Kbr*15	*Kbr*16	*Kbr*17	*Kbr*18	*Kbr*19
CCFWC121	184		203		202	175	254	133	116
CCFWC122	182		221	322	206	167		113	116
CCFWC123	184	207	212		204	171	248	113	118
CCFWC124	172	204	215	314	212	169	258	109	128
CCFWC125	178		218		202	165	258	115	134
CCFWC126	172	210	218		202	173	264		122
CCFWC127	181		224	302	204	167	274	117	118
CCFWC128	178	213	245	282	204	171	260	125	118
CCFWC129	184	207	206	318	212	175	256	131	118
CCFWC130	181	207	218	334	196	167	270	115	124
CCFWC250	178		206	330	210	171	262	109	
CCFWC251	181	207	203/ 206		206	173	262/ 276	111/ 117	116
CCFWC252	181			330	202	169	252	133	122
CCFWC253	178		221	338	204	175	270	143	124
CCFWC254	178		218		208		262	119	126
CCFWC256	178	207	206/ 221	314/ 322	204	169	260	125	122
CCFWC257	172	207	206			175	280	109	128
CCFWC258	172	201	227	310	210	179	258	123	118
CCFWC259	178		206	294	202	167	254	123	116
CCFWC260	178		206	294	202	167	254	123	116
CCFWC261	181	207	206		206	169	262	111	116
CCFWC262	181		203	278	206	169	270	127	122
CCFWC263	184	207	206	318	206		260	129	116
CCFWC265	178		221	262	202	169	262	123	122
CCFWC266	178	207	203		204	173	276	117	128
CCFWC267	178	207	221	294	206	173	266	115	118
CCFWC268	178	207	215		204	167	254	111	118
CCFWC269	178	207	200		206	167	262	115	128

Table 3.4 Continued.

	Locus								
Culture	*Kbr*11	*Kbr*12	*Kbr*13	*Kbr*14	*Kbr*15	*Kbr*16	*Kbr*17	*Kbr*18	*Kbr*19
CCMP2228	184	207	200	306	208	165	258	109	122
CCMP2229	178		227		200	165	256	111	122
CCMP2281	184	207	200	322	202	167	256	121	116
CCMP2820	178	204	209	282	198	171	256	137	120
CCMP718	181	204	212	294	204	167	262	137	130
EPA JR	175	204	200	302	208	165	262	135	118
NBK	175	207	236	318	208	173	260	113	118
NOAA-1	181	207	206	314	200	171	260	129	120
NSP3[a]	169	207	224		206	165	262	111	118
SP1	181	207	200	290	204	167	248	115	118
SP2	181	207	215	290	206	169	264	127	120
SP3[b]	184	204	212	294	204	167	262	137	120

[a] Originally named NTSP3 in Henrichs *et al*. (2008).
[b] Originally named TSP3 in Henrichs *et al.* (2008).

K. brevis *haplotypes*

Based on the 18 microsatellite loci, 39 unique haplotypes were identified among the 40 strains of *K. brevis* (~97%). Haplotype information for all 40 strains is summarized in Tables 3.3 and 3.4. Two identical haplotypes were identified from strains established from the same bloom and these strains could have originated from clonal cells. Two strains (CCFWC251, CCFWC256) repeatedly produced two observable bands at each of several loci (Tables 3.3, 3.4). The two bands were from three to fifteen bp apart and were clearly defined when viewed on agarose or polyacrylamide gels. Both cultures produced two bands at three loci (*Kbr*3, *Kbr*8, *Kbr*13). Additionally, CCFWC251 produced two bands at *Kbr*1, *Kbr*4, *Kbr*10, *Kbr*17, *Kbr*18 and CCFWC256 produced two bands at *Kbr*7, *Kbr*9, *Kbr*14. To confirm the observed pattern of two bands, single cells were isolated from CCFWC256 and genotyped at four microsatellites (*Kbr*8, *Kbr*9, *Kbr*13, *Kbr*14). The isolated single cells produced two bands for the four microsatellites tested, though some cells occasionally produced a single band for one microsatellite while the remaining microsatellites produced two bands. The identification of two bands, presumed to be two alleles, even in single cells of a clonal culture, could be the result of a gene duplication event or indicate the presence of diploid cells in these strains. To reduce the chance of bias in the analyses from arbitrarily picking one allele to represent the strain, the microsatellites producing two bands were coded as missing data in the two strains. For the PCoA, both CCFWC251 and CCFWC256 were removed prior to analysis due to the number of loci coded as missing data (eight and six, respectively).

Genetic analysis

The number of different alleles identified for each microsatellite ranged from four to sixteen and unbiased estimates of gene diversity (all 40 strains combined) ranged from 0.323 to 0.945 (0.775 ± 0.170 [mean ± SD]; Table 3.5). The total number of different alleles identified in strains from Florida was higher in 17 of the 18 microsatellite loci. The single remaining microsatellite produced an identical number of different alleles in strains from both Florida and Texas. However, estimates of allelic richness (corrected by rarefaction) showed four microsatellites with greater allelic richness in strains isolated from blooms in Texas. Tests for genotypic disequilibrium resulted in 11 pairs of loci exhibiting significant disequilibrium ($P < 0.05$); however, none remained significant after Bonferroni correction (data not shown).

The PCoA scatterplot showed one main cluster of cultures (Fig. 3.1). The two *a priori* defined geographic groups (Florida, Texas) were not clearly resolved by the PCoA. The first two axes explained more than half (39.8% and 26.4%, respectively) of the variation present in the data (Fig. 3.1).

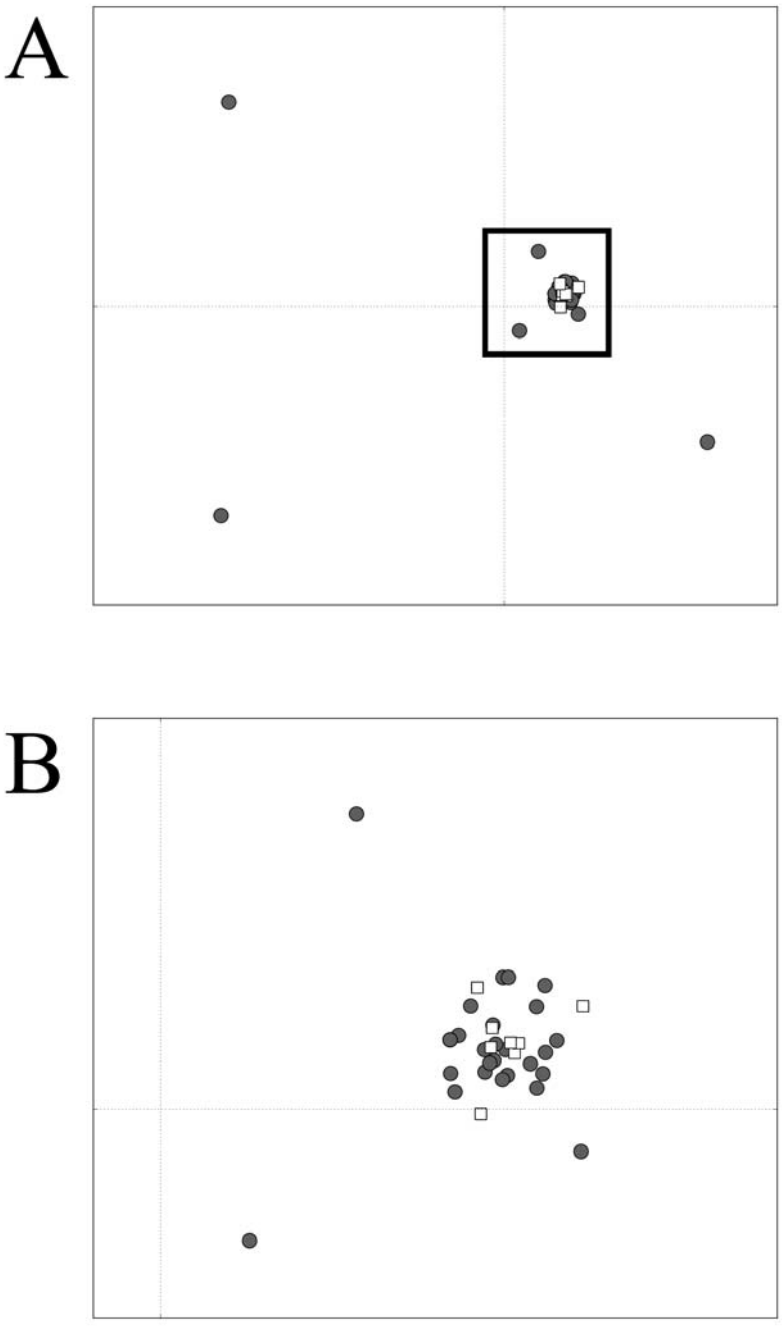

Figure 3.1 Principal coordinates analysis showing the distribution of cultures based on multilocus genotypes from 16 microsatellite markers and 38 cultures. Each dark circle represents a culture from Florida. Each white square represents a culture from Texas. A) View of all 38 cultures. The percentage of variation explained by the x and y axes is 39.8% and 26.4%, respectively. The area within the black box has been enlarged in B to show the tight grouping of cultures from Florida and Texas. B) Enlarged view of A showing no separation between the two geographic groups of cultures, Florida and Texas.

Field samples

A total of 288 single cells (192 FL; 96 TX) of *K. brevis* from field samples collected during blooms were genotyped to compare with the cultures. From these, 192 cells (107 FL; 85 TX) produced an allele at three or more microsatellites (out of the five that were amplified) including 41 cells producing an allele at all five microsatellites. The number of different alleles identified from field populations ranged from five to thirteen and unbiased estimates of gene diversity ranged from 0.280 to 0.843 (0.650 ± 0.216 [mean ± SD]; Table 3.6), comparable to the values identified from cultured strains for those same five loci (N_A: 4-11; $\hat{H}$: 0.323-0.891; 0.657 ± 0.225[mean ± SD]; Table 3.5). The total number of different alleles identified from field samples was higher than the total identified from all cultured strains of *K. brevis* for four of the five microsatellites (*Kbr*5, *Kbr*8, *Kbr*9, *Kbr*10). Estimates of allelic richness (calculated by rarefaction) from field samples were higher than those obtained from cultured strains for two microsatellites (*Kbr*5, *Kbr*10), lower in one microsatellite (*Kbr*7), and approximately the same at the remaining two microsatellites (*Kbr*8, *Kbr*9; Tables 3.5, 3.6).

Table 3.5 Summary of microsatellite information for cultures of *Karenia brevis*.

	Combined				Florida				Texas			
Locus	**n**	**N_A**	***H*[a]**	**Allelic range**	**n**	**N_A[b]**	***H*[a]**	**Allelic range**	**n**	**N_A**	***H*[a]**	**Allelic range**
*Kbr*1	39	8	0.794	246-267	31	8 (4.49)	0.768	246-267	8	5	0.893	249-264
*Kbr*3	37	10	0.787	229-256	29	9 (4.81)	0.803	232-256	8	4	0.750	229-238
*Kbr*4	39	9	0.858	252-278	31	9 (5.12)	0.865	252-278	8	4	0.821	264-272
*Kbr*5	40	6	0.792	180-190	32	6 (4.21)	0.800	180-190	8	4	0.750	182-190
*Kbr*6	40	6	0.764	219-231	32	6 (4.04)	0.786	219-231	8	3	0.607	221-225
*Kbr*7	39	5	0.323	252-264	31	4 (2.12)	0.295	255-264	8	3	0.464	252-261
*Kbr*8	37	11	0.891	108-377	29	11 (5.76)	0.900	108-377	8	5	0.893	128-146
*Kbr*9	39	4	0.457	158-167	31	4 (2.67)	0.475	158-167	8	2	0.429	161-164
*Kbr*10	37	7	0.823	169-181	29	7 (4.43)	0.798	169-181	8	4	0.786	173-181
*Kbr*11	39	6	0.760	169-184	31	5 (3.77)	0.753	172-184	8	5	0.857	169-184
*Kbr*12	27	5	0.484	201-213	19	5 (3.18)	0.579	201-213	8	2	0.250	204-207
*Kbr*13	37	12	0.908	200-245	29	11 (5.73)	0.892	200-245	8	7	0.964	200-236
*Kbr*14	26	14	0.942	262-338	21	13 (4.58)[c]	0.957	262-338	5	3	0.800	290-318
*Kbr*15	39	9	0.841	196-212	31	9 (5.14)	0.860	196-212	8	3	0.679	204-208
*Kbr*16	38	7	0.841	165-179	30	7 (4.84)	0.848	165-179	8	4	0.786	165-173
*Kbr*17	38	13	0.902	248-280	30	10 (5.67)	0.899	248-280	8	6	0.893	248-276
*Kbr*18	38	16	0.945	109-143	30	16 (6.77)	0.952	109-143	8	6	0.893	111-137
*Kbr*19	39	9	0.842	116-134	31	9 (4.871)	0.839	116-134	8	3	0.714	118-128

[a]Unbiased estimates of gene diversity.
[b]Numbers in parentheses are allelic richness estimates calculated by rarefaction to a sample size of eight.
[c]Allelic richness estimate calculated by rarefaction to a sample size of five.

Table 3.6 Summary of microsatellite information for field cells of *Karenia brevis*.

	Combined				Florida				Texas			
Locus	**n**	**N_A**	***H*[a]**	**Allelic range**	**n**	**N_A[b]**	***H*[a]**	**Allelic range**	**n**	**N_A**	***H*[a]**	**Allelic range**
*Kbr*5	162	10	0.772	178-198	90	10 (9.39)	0.821	178-198	72	6	0.685	182-192
*Kbr*7	176	4	0.280	258-267	97	3 (2.97)	0.325	258-264	79	4	0.212	258-267
*Kbr*8	116	13	0.843	108-148	78	12 (10.20)	0.861	108-148	38	6	0.790	124-136
*Kbr*9	173	5	0.531	110-167	91	5 (4.90)	0.561	110-167	82	3	0.497	161-167
*Kbr*10	119	12	0.823	167-213	40	9	0.745	167-185	79	10 (8.01)[c]	0.840	167-213

[a]Unbiased estimates of gene diversity.
[b]Numbers in parentheses are allelic richness estimates calculated by rarefaction to a sample size equivalent to the sample size from Texas at each respective locus.
[c]Allelic richness estimate calculated by rarefaction to a sample size of 40.

Diagnostic test

For the two cultured strains (CCFWC268, CCMP718) tested, different alleles were observed at twelve of the eighteen loci (Tables 3.3, 3.4). The genotype results produced for CCFWC268 (also known as the "Wilson" clone) obtained from all five laboratories were identical. The same result was observed for the strain CCMP718. No new alleles were identified at any microsatellite locus for either strain.

DISCUSSION

The nine new microsatellite markers identified in this study have doubled the number of available microsatellite markers for *K. brevis.* For studies investigating population-genetic structure, increasing the number of loci (and/or alleles) can provide increased power to detect genetic divergence among populations (Kalinowski 2002). The number of different alleles identified and the estimates of gene diversity are comparable to previous genetic work on dinoflagellates. Nagai *et al*. (2004) described 13 microsatellite markers from *A. tamarense* with gene diversity estimates between 0.632 and 0.974. Nagai *et al.* (2007) identified an increased number of alleles (between seven and 42) at nine microsatellite markers when tested on 500 clonal cultures and compared to the 20 originally tested by Nagai *et al*. (2004). A similar result was observed in the present study of *K. brevis*. Four of the five microsatellites amplified from both cultures and field samples had a greater number of alleles in field samples but this result is likely due to the higher number of genotyped individuals (Tables 3.5, 3.6). It is unlikely that 40 clonal cultures would contain all the alleles present in field populations of *K. brevis*. Similar estimates of gene diversity between field samples (unexposed to culturing

biases) and cultures, along with the high number of unique genotypes among the cultures, suggest the clonal cultures currently being grown and studied incorporate much of the diversity present in the field. However, for two (*Kbr*5, *Kbr*8) of the five microsatellites (*Kbr*5, *Kbr*7, *Kbr*8, *Kbr*9, *Kbr*10), the most frequent allele identified from the field samples differed from the most frequent allele identified from cultures (*Kbr*5 shown in Fig. 3.2). The potential biases introduced by the culturing process may have resulted in cultures that, while diverse, are not an accurate representation of the populations present in the field.

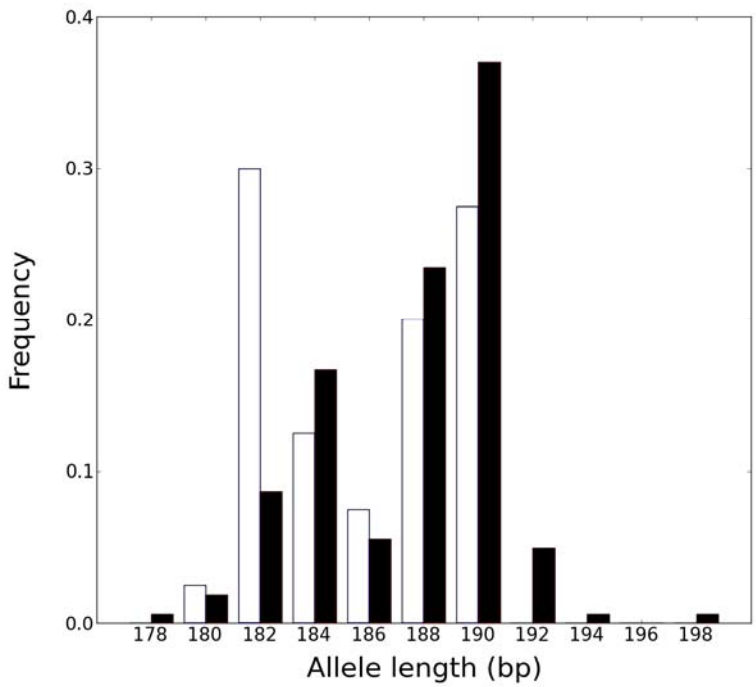

Figure 3.2 Allele frequency distribution for microsatellite marker *Kbr*5 from cultures (white) and field cells (black).

Blooms of *K. brevis* have been identified from many parts of the Gulf (Steidinger *et al*. 1998). With such a widespread distribution, the question arises whether distinct populations of *K. brevis* exist in the Gulf. Steidinger *et al*. (1998) noted the frequent occurrence of harmful algal blooms of *K. brevis* off the west coast of Florida and asked whether this region of the Gulf served as a source for blooms in other parts of the Gulf. If genetic divergence was observed between isolates of *K. brevis* from Florida and isolates from Texas, it would suggest the presence of distinct populations. The inability of the PCoA to distinguish more than one cluster suggests a lack of genetic divergence between isolates from Florida and isolates from Texas (Fig. 3.1). If small parcels of water containing cells from a large algal bloom are being transported westward across the Gulf, genetic drift could reduce the genetic variation of *K. brevis* in the small parcel of water and resulting blooms near Texas would exhibit reduced genetic variation. In such situations, allelic richness is likely to be impacted more than estimates of gene diversity (Leberg 2002). Rarefaction corrected estimates of allelic richness at ten microsatellites were higher for the isolates from Florida, lower at four microsatellites, and approximately the same in the remaining four microsatellites when compared to the isolates from Texas. Estimates of allelic richness for the field sample from Florida were higher in four of the five microsatellites and lower in the remaining microsatellite (*Kbr*7) when compared to the field sample from Texas. The reduction in allelic richness, the overlap of allelic ranges, and the lack of distinct clusters in the PCoA, support the hypothesis of one genetically homogeneous population of *K. brevis* in the Gulf. However, this result is based on the data from a small number of isolates and the two

field samples used for comparison may not represent levels of genetic variation present in field populations. Further work investigating the population-genetic structure of *K. brevis* in the Gulf, incorporating more samples from different geographic areas, has been completed (Ch. 4).

As a diagnostic tool, microsatellite genotyping will also help to eliminate confusion in identifying different isolates. Difficulties in interpreting experimental results may arise if experiments are conducted with different strains. Two strains of *K. brevis* isolated from Florida in 1953 (CCFWC268, CCMP718) are often confused with each other in the literature. Although both are from Wilson's laboratory, they are in fact genetically distinct isolates based on the observed differences in allele size at 12 of the 18 microsatellites (Tables 3.3, 3.4). This result is consistent with observed physiological differences (e.g. growth rates differ between these two strains; Brown *et al*. 2006; Errera *et al*. 2010). Future studies utilizing one or both of these cultures should positively identify the cultures before making comparisons with previous work, especially those studies investigating physiological differences. The confirmation of each strain from several laboratories supports the use of microsatellite markers as a diagnostic tool for researchers who wish to confirm the identity of their strains.

CHAPTER IV

POPULATION-GENETIC STRUCTURE OF *KARENIA BREVIS* IN THE GULF OF MEXICO

INTRODUCTION

Harmful algal blooms in the Gulf of Mexico (hereafter Gulf) can be traced back several hundred years (Steidinger 1998, Magaña *et al.*2003). The major harmful bloom-forming dinoflagellate in the Gulf of Mexico is *Karenia brevis*. While blooms of *K. brevis* occur throughout the northern Gulf, high concentrations of cells of *K. brevis* and their negative impacts (e.g., fish kills, respiratory irritation) are observed commonly off the west coast of Florida (Brand and Compton 2007). This area experiences a bloom almost annually and it is not clear why this region is more prone to experience a harmful algal bloom (Steidinger *et al*. 1998). Harmful algal blooms caused by *K. brevis* are an infrequent occurrence off the coast of Texas (Magaña *et al.* 2003; Walsh *et al.*2006). Steidinger *et al*. (1998) suggested that the west and southwest coast of Florida might serve as a point of origin for blooms of *K. brevis* in other parts of the Gulf, based in part on the frequency of bloom occurrence in this region versus other parts of the Gulf. Accordingly, the conditions necessary for a bloom to move from the west coast of Florida across the Gulf to the coast of Texas may occur only sporadically. An alternative hypothesis is the presence of multiple seed populations where there are two or more populations capable of blooming independently and impacting opposite coasts of the Gulf.

The hypothesis that the west coast of Florida is the origin for blooms throughout the Gulf can be tested by determining whether (i) bloom samples from different regions in the Gulf are genetically distinct, and (ii) bloom samples taken over time from the same location are genetically similar. Genetically distinct samples from different regions would suggest different origins for those samples. If temporal samples taken from the same location are more similar genetically to each other than they are to samples from other regions, this would suggest the presence of localized populations and again indicate multiple origins for blooms of *K. brevis*.

Studies of population-genetic structure of dinoflagellates have been conducted for a number of bloom-forming species, including: *Alexandrium* spp.(Alpermann *et al.* 2009; Masseret *et al.* 2009; Erdner *et al*. 2011; Casabianca *et al.* 2012), *Cochlodinium polykrikoides* (Nagai *et al.* 2009), *Gymnodinium catenatum* (Bolch 1999), *Oxyrrhis marina* (Lowe *et al*. 2010), and *Prorocentrum micans* (Shankle *et al.* 2004). Cultured isolates were examined in all of these studies and both broad- (Nagai *et al.* 2009; Casabianca *et al.* 2012) and fine-scale (Lowe *et al.* 2010; Erdner *et al.* 2011) genetic structure was detected based on using microsatellites. A study of cultured isolates of *K. brevis* did not detect genetic divergence between isolates from Florida (FL) and isolates from Texas (TX), although only 40 isolates were available for study (Ch. 3). For many dinoflagellate species (e.g. *Alexandrium* spp.), new cultures are easy to establish from vegetative cells or resting cysts. New cultures of *K. brevis* are difficult to start from vegetative cells and resting cysts have not been conclusively identified. In this study, population-genetic structure among samples of *K. brevis* across the Gulf was assessed

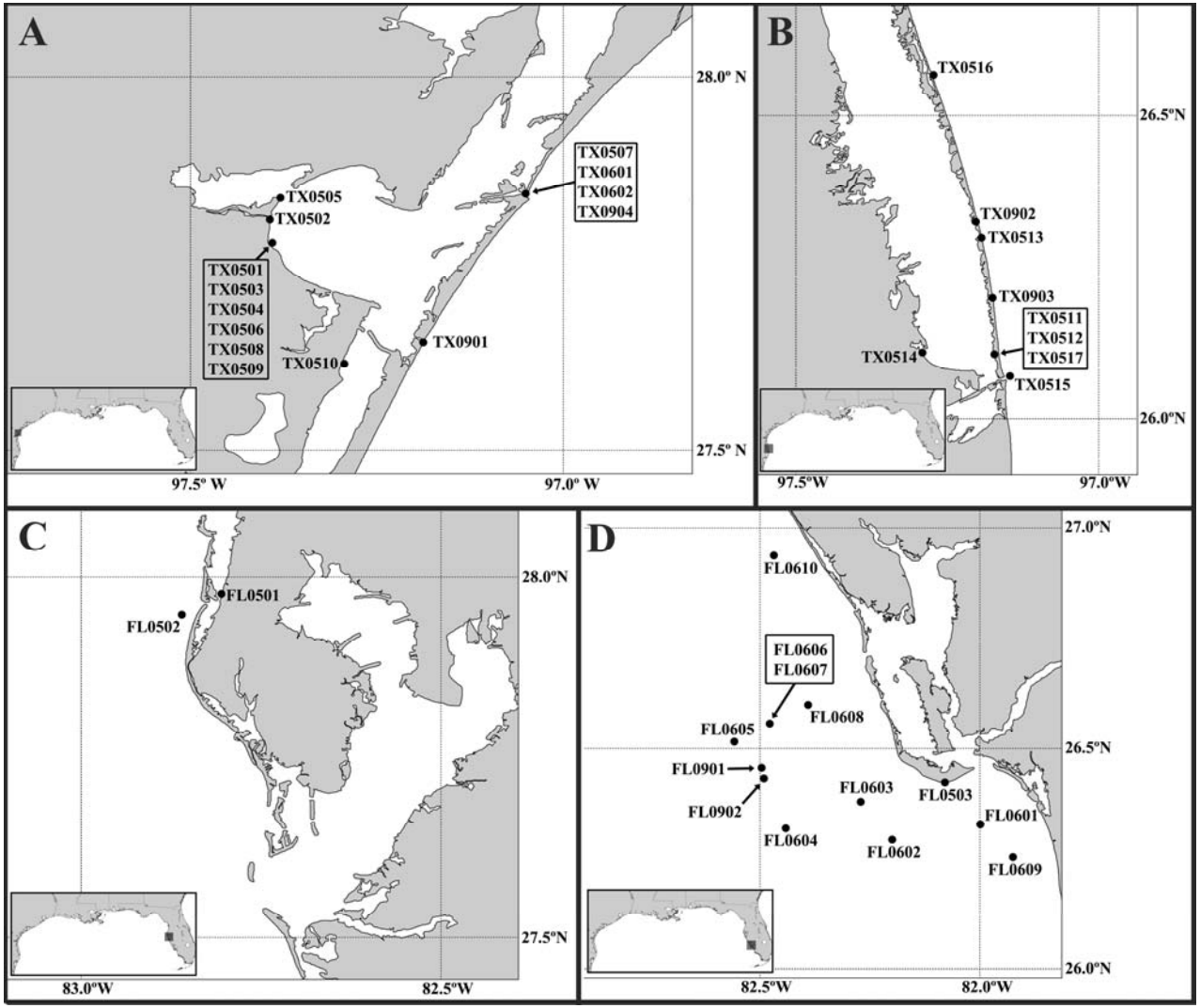

Figure 4.1 Collection locations for samples from the Gulf of Mexico. Black boxes indicate the location of each group of samples. A) Corpus Christi Bay, Texas, USA. B) South Padre Island near Brownsville, Texas, USA. C) Tampa Bay, Florida, USA. D) Charlotte Harbor, Florida, USA.

Table 4.1 Collection location, date, and number of cells included in genetic analyses for surface samples used in this study.

Sample Name	Latitude (°N)	Longitude (°W)	Date Collected	n[a]
FL0501	27.975	82.804	7/04/2005	28
FL0502	27.947	82.860	8/18/2005	20
FL0503	26.422	82.080	9/06/2005	28
FL0601	26.331	82.000	10/05/2006	53
FL0602	26.291	82.201	10/05/2006	56
FL0603	26.380	82.271	10/05/2006	60
FL0604	26.317	82.444	10/05/2006	18
FL0605	26.516	82.557	10/05/2006	16
FL0606[b]	26.555	82.477	10/05/2006	19
FL0607	26.555	82.477	10/05/2006	62
FL0608	26.596	82.389	10/05/2006	34
FL0609	26.251	81.926	10/03/2006	57
FL0610	26.938	82.468	10/02/2006	17
FL0901	26.454	82.493	10/10/2009	116
FL0902	26.432	82.493	10/10/2009	66
TX0501	27.776	97.391	10/13/2005	67
TX0502	27.808	97.392	10/13/2005	44
TX0503	27.778	97.392	10/13/2005	33
TX0504	27.776	97.391	10/17/2005	39
TX0505	27.837	97.381	10/17/2005	17
TX0506	27.776	97.391	10/19/2005	44
TX0507	27.838	97.050	10/20/2005	76
TX0508	27.776	97.391	10/21/2005	81
TX0509	27.776	97.391	11/03/2005	66
TX0510	27.617	97.297	11/04/2005	65
TX0511	26.104	97.170	10/10/2005	46
TX0512	26.103	97.170	10/15/2005	36
TX0513	26.290	97.282	9/29/2005	67
TX0514	26.106	97.293	10/10/2005	24
TX0515	26.068	97.147	10/17/2005	41
TX0516	26.566	97.271	9/28/2005	91
TX0517	26.069	97.164	10/12/2005	36
TX0601	27.837	97.051	10/04/2006	32
TX0602	27.838	97.053	9/29/2006	20
TX0901	27.643	97.187	10/15/2009	56
TX0902	26.325	97.202	10/15/2009	64
TX0903	26.199	97.177	11/24/2009	78
TX0904	27.838	97.050	10/15/2009	31

[a]Total number of cells included in the analyses.
[b]Sample taken at depth of ~10m.

using single-cell haplotypes obtained from spatially and temporally varying bloom samples taken from six bloom events occurring across three different years.

MATERIALS AND METHODS

Field sample collection

Field samples were obtained from six bloom events (3 FL; 3 TX) that occurred over a five year period (2005, 2006, 2009; Fig. 4.1; Table 4.1). Whole water samples were collected from surface blooms, preserved with acidified Lugol's iodine (LI) solution, and stored at 4ºC to preserve morphological characteristics and for DNA extraction. A total of 45 samples were picked for single cell genotyping.

Single cell DNA amplification

Detailed isolation and amplification methods, including primer concentrations and PCR protocols, can be found in Henrichs *et al.*(2008). Briefly, samples were first destained with sodium thiosulfate to remove LI. Single cells were picked by Pasteur pipet into 0.2mL PCR tubes with ~2µL of 0.5X Tris-EDTA. After three rounds of freezing (-85°C) and thawing, lysed cells then underwent two rounds of PCR amplification. The first round reaction (20µL reaction volume) amplified five microsatellite markers (unlabeled primers) multiplexed together; the second round reaction (10µL reaction volume) consisted of five separate simplex reactions each containing a fluorescently-labeled forward primer. The five microsatellites amplified consisted of three dinucleotide repeats (*Kbr*5, *Kbr*8, *Kbr*10) and two trinucleotide repeats (*Kbr*7, *Kbr*9; Renshaw *et al*. 2006).

Allele scoring

Resulting PCR products were visualized with a 5% polyacrylamide gel (Long Ranger Singel Pack, Cambrex Bio Science Rockland, Rockland, ME, USA) and ABI Prism 377 Genetic Analyzer (Applied Biosystems Inc., Foster City, CA, USA). Allele sizes were determined using Genescan 3.1.2 (Applied Biosystems Inc.) and Genotyper version 2.5 (Applied Biosystems Inc.). For each sample, if the number of cells with allele information from at least four microsatellites was less than ten, the sample was removed from the study.

Genetic analyses

Number of alleles, allele frequencies, and unbiased estimates of gene diversity (Nei & Chesser 1983) for each sample were calculated in Arlequin 3.5 (Excoffier & Lischer 2010). To account for differences in number of cells per sample, estimates of allelic richness were determined by rarefaction (El Mousadik and Petit 1996). Tests for genotypic disequilibrium among loci were conducted in Genepop v4.0 (Rousset 2008); sequential Bonferroni correction (Rice 1989) was applied for all multiple tests performed simultaneously.

Pairwise estimates of genetic divergence between samples employed Jost's *D* (Jost 2008) and G''_{ST} (Meirmans and Hedrick 2011). Recent work by Jost (2008) and Meirmans and Hedrick (2011) has shown G_{ST}, an F_{ST}-analogue, to be biased when calculated from genetic markers with high levels of gene diversity within populations (e.g. microsatellites). To account for this, pairwise values of genetic divergence were calculated using custom Python scripts. Multilocus estimates of both *D* and G''_{ST} were

obtained by averaging H_S and H_T over loci as suggested by Meirmans and Hedrick (2011). Pairwise values of F_{ST} (number of different alleles) were calculated using Arlequin 3.5; significance of $F_{ST} = 0$ was tested using 50000 permutations of cells between all pairs of samples and sequential Bonferroni correction was applied for all multiple tests performed simultaneously. Pairwise values of genetic divergence also were calculated for samples grouped according to year and to collection location. An analysis of molecular variance (AMOVA) was performed in Arlequin to determine the amount of genetic variation partitioned into each of three hierarchical levels. Three different grouping strategies were tested in the AMOVA analysis: (i) samples grouped according to year and collection location (six groups, see above), (ii) samples grouped according to location only (two groups: FL, TX) and (iii) samples grouped according to year of collection (three groups: 2005, 2006, 2009). Population structure among samples also was tested using Structure 2.3.3 (Pritchard *et al.* 2000). Each simulation was run for 30000 steps after a 15000 step burn-in. Two models (admixture; no admixture) with correlated allele frequencies were tested with and without the 'locprior' option. Ten simulation runs were performed for *K* values from one to 11. The ΔK statistic of the likely number of true populations was estimated following Evanno *et al*. (2005). Spatial autocorrelation analysis employed GenAlEx v6.41 (Peakall & Smouse 2006) to determine if there was a correlation between haplotype distribution and geographic distance. Geographic distances were measured as straight lines between sample locations. Due to the large distances between many sample locations, all geographic distances were log transformed.

RESULTS

Single cell genotyping

Genotypes with allele information for at least four microsatellites were obtained from 1949 single cells representing thirty-eight samples. For a small number of cells, (n=284; ~14.5%) two bands were observed at one or more microsatellites. It is not known whether the two bands are the result of a gene duplication event or are PCR artifacts. For these cases, the microsatellite(s) having two alleles were coded as missing data, leaving 1804 cells for further analysis. Individual sample sizes ranged from 16 to 116 cells (47.5 ± 23.7 [mean ± SD]; Table 4.1).

Genetic diversity

All samples were highly diverse genetically (Table 4.2). Number of alleles per microsatellite ranged from 13 to 29 (20.2 ± 6.2 [mean ± SD]) while gene diversity ranged from 0.393 to 0.825 (0.682 ± 0.193 [mean ± SD]; Table 4.2). Estimates of allelic richness and gene diversity among groups were similar (Table 4.3).

Prior to Bonferroni correction, there were 24 significant ($P < 0.05$) tests for genotypic disequilibrium; none of the tests were significant after Bonferroni correction.

Table 4.2 Sample diversity statistics for five microsatellite loci.

	*Kbr*5				*Kbr*7				*Kbr*8				*Kbr*9				*Kbr*10			
Sample	**n**	**N_A**	***A***	$\hat{h}$	**n**	**N_A**	***A***	$\hat{h}$	**n**	**N_A**	***A***	$\hat{h}$	**n**	**N_A**	***A***	$\hat{h}$	**n**	**N_A**	***A***	$\hat{h}$
FL0501	21	6	4.4	0.752	26	3	2.5	0.335	28	5	4.5	0.762	27	5	3.6	0.598	16	6	5.2	0.808
FL0502	19	10	6.9	0.906	13	3	3.0	0.500	19	6	5.3	0.819	15	4	3.9	0.667	17	8	6.4	0.875
FL0503	27	7	5.1	0.798	26	4	3.0	0.582	26	9	6.8	0.871	26	5	3.9	0.631	21	8	5.9	0.862
FL0601	51	7	5.0	0.828	48	6	3.5	0.523	47	10	6.0	0.837	52	5	3.3	0.577	37	8	5.0	0.763
FL0602	52	7	4.8	0.816	51	4	2.9	0.445	41	9	6.0	0.824	53	4	2.5	0.494	48	8	5.2	0.794
FL0603	55	8	5.3	0.834	54	6	3.5	0.464	55	13	6.2	0.834	54	4	3.1	0.495	47	7	4.8	0.765
FL0604	16	6	5.1	0.817	15	2	1.9	0.133	16	6	5.4	0.683	16	3	3.0	0.633	11	6	6.0	0.855
FL0605	13	4	3.7	0.679	14	2	2.0	0.363	15	9	8.3	0.914	14	3	3.0	0.385	11	5	5.0	0.818
FL0606	18	6	5.1	0.843	16	4	3.6	0.575	17	6	5.6	0.824	18	3	3.0	0.464	15	7	5.7	0.810
FL0607	58	9	5.2	0.831	60	2	1.9	0.282	53	12	6.9	0.866	59	5	3.2	0.550	31	9	5.3	0.761
FL0608	27	6	4.6	0.783	32	4	3.3	0.567	32	11	7.3	0.871	32	4	3.3	0.577	26	7	5.4	0.843
FL0609	53	8	5.0	0.820	53	4	2.3	0.242	53	12	6.2	0.825	54	4	3.1	0.493	47	8	4.7	0.705
FL0610	15	6	5.4	0.848	13	2	2.0	0.282	13	6	6.0	0.782	17	3	2.8	0.588	12	5	4.8	0.788
FL0901	105	11	5.1	0.818	108	6	2.8	0.387	108	18	7.5	0.888	112	9	3.7	0.525	92	14	5.9	0.842
FL0902	57	8	5.5	0.848	64	5	2.8	0.393	57	13	6.3	0.805	64	6	3.6	0.579	61	11	5.2	0.773
TX0501	65	6	3.8	0.682	66	4	2.0	0.173	34	6	5.0	0.790	67	3	2.2	0.511	64	10	5.6	0.841
TX0502	36	8	5.3	0.843	38	4	3.1	0.471	42	11	7.2	0.878	40	4	3.2	0.627	38	8	5.5	0.835
TX0503	31	5	4.5	0.802	28	3	2.4	0.315	33	7	4.9	0.767	33	4	3.2	0.453	31	7	5.5	0.832
TX0504	35	7	4.4	0.734	26	3	2.0	0.151	39	9	5.0	0.656	39	4	2.7	0.563	34	10	6.2	0.865
TX0505	10	5	5.0	0.844	14	2	2.0	0.264	17	5	4.3	0.507	14	4	4.0	0.648	17	10	7.6	0.919
TX0506	40	7	4.5	0.788	38	2	1.3	0.053	41	8	5.3	0.741	44	5	3.2	0.559	41	7	4.6	0.748
TX0507	59	9	5.4	0.842	67	3	2.6	0.444	75	8	4.0	0.626	72	6	3.4	0.592	65	8	4.9	0.757
TX0508	79	7	4.5	0.758	73	4	2.3	0.248	76	7	3.2	0.419	81	4	2.8	0.564	68	8	5.6	0.838
TX0509	60	8	4.9	0.794	62	8	3.6	0.496	66	9	5.6	0.796	65	6	3.5	0.476	50	11	5.8	0.821
TX0510	63	7	4.3	0.730	62	6	2.9	0.384	63	8	4.0	0.483	63	6	3.6	0.573	48	10	5.2	0.754
TX0511	44	6	4.6	0.797	41	4	3.1	0.446	36	9	6.5	0.867	40	6	3.7	0.464	34	8	4.6	0.642
TX0512	33	6	4.6	0.805	32	3	2.7	0.377	34	10	6.4	0.848	32	4	3.2	0.464	23	9	6.4	0.877
TX0513	67	7	5.0	0.815	67	3	2.7	0.398	62	12	6.4	0.831	65	6	3.6	0.513	58	11	5.6	0.833
TX0514	24	7	5.1	0.837	24	4	3.1	0.471	21	8	6.3	0.819	23	7	5.6	0.711	15	9	7.5	0.914
TX0515	39	5	4.1	0.768	40	4	2.9	0.387	36	11	7.1	0.876	40	5	3.5	0.573	27	7	5.6	0.852
TX0516	86	10	4.5	0.711	84	4	2.7	0.408	91	5	2.9	0.497	87	6	3.2	0.597	84	10	5.4	0.810
TX0517	32	7	4.7	0.808	36	2	2.0	0.356	31	12	7.7	0.890	35	4	3.0	0.487	32	9	5.4	0.804

Table 4.2 Continued.

	*Kbr*5				*Kbr*7				*Kbr*8				*Kbr*9				*Kbr*10			
Sample	**n**	**N_A**	***A***	$\hat{h}$	**n**	**N_A**	***A***	$\hat{h}$	**n**	**N_A**	***A***	$\hat{h}$	**n**	**N_A**	***A***	$\hat{h}$	**n**	**N_A**	***A***	$\hat{h}$
TX0601	31	7	5.3	0.839	26	6	4.6	0.702	30	9	6.4	0.846	29	4	3.4	0.599	23	7	5.2	0.806
TX0602	18	5	4.3	0.791	20	3	2.3	0.195	16	7	6.2	0.817	20	3	2.7	0.511	12	2	2.0	0.409
TX0901	50	7	5.0	0.829	53	5	2.8	0.359	45	14	7.1	0.868	56	6	3.7	0.622	39	9	5.4	0.787
TX0902	60	8	5.0	0.828	64	4	3.0	0.528	46	12	6.3	0.827	64	5	3.6	0.648	55	8	5.3	0.822
TX0903	74	9	5.1	0.826	77	6	3.1	0.414	70	11	6.4	0.856	76	6	4.0	0.657	58	9	5.3	0.806
TX0904	27	7	5.1	0.838	27	4	2.9	0.430	28	9	6.1	0.825	27	3	2.5	0.510	27	6	5.2	0.823
mean	43.4	7.1	4.9	0.803	43.6	3.9	2.7	0.383	42.4	9.3	5.9	0.783	45.4	4.7	3.3	0.557	37.8	8.2	5.4	0.802

Table 4.3 Diversity statistics for grouped samples.

	*Kbr*5				*Kbr*7				*Kbr*8				*Kbr*9				*Kbr*10			
Sample	**n**	**N_A**	***A***	***$\hat{h}$***	**n**	**N_A**	***A***	***$\hat{h}$***	**n**	**N_A**	***A***	***$\hat{h}$***	**n**	**N_A**	***A***	***$\hat{h}$***	**n**	**N_A**	***A***	***$\hat{h}$***
FL2005	67	11	10.1	0.809	65	4	3.8	0.479	73	11	9.3	0.837	68	6	5.6	0.611	54	9	8.2	0.836
FL2006	358	11	7.6	0.819	356	7	4.3	0.399	342	22	11.0	0.837	369	9	4.4	0.532	285	12	8.0	0.773
FL2009	162	11	8.6	0.827	172	7	4.5	0.387	165	20	12.8	0.871	176	10	6.0	0.544	153	14	9.4	0.816
TX2005	803	13	7.3	0.785	798	11	4.1	0.360	797	22	9.2	0.792	840	17	5.2	0.576	729	22	8.6	0.824
TX2006	49	7	7.0	0.820	46	6	6.0	0.526	46	10	10.0	0.863	49	4	4.0	0.558	35	7	7.0	0.716
TX2009	211	9	7.4	0.819	221	7	4.6	0.437	189	19	11.2	0.849	223	7	5.0	0.625	179	10	7.6	0.806
mean	275	10.3	8.5	0.813	276	7.0	4.2	0.431	269	17.3	10.1	0.841	288	8.8	5.4	0.574	239	12.3	8.1	0.795

Genetic divergence

Jost's *D* among samples ranged from -0.055 to 0.316; $G"_{ST}$ values ranged from -0.077 to 0.427 and were generally higher than the corresponding Jost's *D* values (Table 4.4). A total of 125 pairwise comparisons involving Jost's *D* and a total of 122 pairwise comparisons involving $G"_{ST}$ remained significant after Bonferroni correction (120 pairwise comparisons were significant for both *D* and $G"_{ST}$). Pairwise F_{ST} values ranged from -0.075 to 0.162; a total of 103 comparisons remained significant after Bonferroni correction (Table 4.4). The majority of significant comparisons (103 of 125 Jost's *D*; 92 of 103 F_{ST}) included at least one of four samples from the TX2005 group (TX0506, TX0508, TX0510, TX0516). Upon closer inspection, TX0508, TX0510, and TX0516 had lower allelic richness estimates and lower estimates of gene diversity than the remaining thirty-five samples at locus *Kbr8*. Sample TX0506 was dominated by one allele at the *Kbr*7 locus, with the second allele present in only one individual. For each of the four samples, the remaining four microsatellites were not deficient in the number of alleles or gene diversity when compared to the other thirty-four samples. Increased sampling of individuals from the four divergent samples may show increased numbers of alleles and estimates of gene diversity on par with the rest of the samples. For both measures of genetic divergence (Jost's *D*, F_{ST}), all comparisons between samples from the year 2006 or 2009 were nonsignificant. Analysis of genetic divergence among groups showed the FL2005 and TX2005 groups to be significantly genetically diverged from the other groups, consistent with the pairwise comparisons of samples (Table 4.5).

Table 4.4 Pairwise values of genetic divergence between samples. Jost's *D* (above diagonal) and F_{ST} (below diagonal). Shaded boxes indicate comparisons that remained significant after Bonferroni correction.

Sample	FL0501	FL0502	FL0503	FL0601	FL0602	FL0603	FL0604	FL0605	FL0606	FL0607	FL0608	FL0609	FL0610	FL0901
FL0501		-0.021	0.057	0.049	0.031	0.032	0.086	0.047	0.046	0.032	0.056	0.036	0.035	0.041
FL0502	-0.040		-0.021	0.054	0.035	0.034	0.061	0.047	0.028	0.056	0.009	0.074	0.036	0.047
FL0503	0.017	-0.019		0.044	0.071	0.075	0.131	0.128	0.080	0.099	-0.023	0.146	0.057	0.081
FL0601	0.014	0.003	0.016		-0.022	0.002	0.060	0.025	-0.014	-0.006	-0.011	0.027	-0.043	0.008
FL0602	-0.013	0.008	0.026	-0.016		-0.015	0.064	0.029	-0.036	-0.012	0.012	0.009	-0.038	-0.004
FL0603	0.004	0.009	0.031	0.001	-0.009		0.087	0.021	-0.036	0.000	0.026	-0.009	0.001	-0.004
FL0604	0.047	0.021	0.048	0.023	0.024	0.037		0.043	0.078	0.047	0.081	0.067	-0.041	0.077
FL0605	0.029	0.011	0.044	-0.002	-0.004	0.002	0.031		-0.027	0.027	0.040	0.014	0.030	0.012
FL0606	0.012	0.005	0.030	-0.004	-0.013	-0.012	0.038	-0.023		0.012	-0.005	-0.002	-0.024	-0.020
FL0607	0.011	0.001	0.040	-0.004	-0.028	-0.010	0.028	0.003	0.000		0.042	0.005	-0.015	0.005
FL0608	0.020	-0.015	-0.010	-0.012	-0.005	0.007	0.023	0.009	-0.005	0.001		0.071	-0.016	0.014
FL0609	0.007	0.037	0.069	0.016	0.003	-0.003	0.034	-0.002	0.005	-0.009	0.031		0.008	0.021
FL0610	0.006	0.003	0.008	-0.024	-0.020	-0.003	-0.015	0.010	-0.006	-0.013	-0.021	0.005		-0.001
FL0901	0.006	0.013	0.035	0.002	-0.007	-0.002	0.026	-0.008	-0.005	-0.017	0.004	0.010	-0.007	
FL0902	-0.023	0.006	0.049	0.003	-0.005	-0.003	0.027	0.000	-0.002	-0.028	0.026	0.001	-0.006	0.004
TX0501	-0.075	0.000	0.017	0.008	0.025	0.006	0.013	0.014	0.040	-0.042	0.003	0.013	-0.007	-0.004
TX0502	-0.006	-0.014	-0.003	-0.008	-0.008	0.008	0.008	0.018	0.011	-0.020	-0.004	0.029	-0.022	0.012
TX0503	-0.006	0.010	0.032	0.005	-0.008	-0.008	0.030	-0.012	-0.014	-0.028	0.004	0.003	0.002	0.002
TX0504	0.037	0.039	0.009	0.018	0.028	0.039	-0.009	0.054	0.043	0.011	0.015	0.060	-0.017	0.038
TX0505	0.059	0.021	0.026	-0.004	0.017	0.018	-0.041	0.005	-0.007	-0.022	0.023	0.035	-0.028	0.016
TX0506	0.045	0.057	0.068	0.065	0.076	0.087	-0.001	0.091	0.102	0.040	0.062	0.097	0.021	0.079
TX0507	-0.007	0.011	0.069	0.036	0.015	0.032	0.074	0.049	0.034	0.012	0.063	0.036	0.027	0.042
TX0508	0.077	0.068	0.059	0.061	0.070	0.074	0.011	0.085	0.080	0.057	0.055	0.093	0.007	0.074
TX0509	0.009	-0.002	0.020	0.008	-0.008	-0.004	0.038	0.005	-0.007	0.003	0.009	0.013	-0.006	0.008
TX0510	0.081	0.057	0.066	0.049	0.045	0.051	0.023	0.047	0.041	0.052	0.048	0.062	0.005	0.053
TX0511	0.011	0.022	0.043	0.004	0.006	0.007	0.075	-0.001	0.004	0.007	0.013	0.011	0.026	0.007
TX0512	0.011	0.006	0.028	-0.001	-0.022	-0.013	0.049	-0.001	-0.011	-0.009	0.002	-0.002	0.003	-0.008
TX0513	-0.015	0.003	0.029	0.002	-0.006	-0.005	0.021	-0.018	-0.011	-0.014	-0.002	0.004	-0.012	-0.001
TX0514	0.015	-0.018	0.030	0.000	-0.009	-0.005	0.000	-0.020	-0.021	-0.004	-0.007	0.006	-0.014	-0.009
TX0515	0.000	0.003	0.036	0.005	-0.007	-0.002	0.025	-0.014	0.002	-0.005	0.005	0.001	-0.006	-0.005
TX0516	-0.002	0.047	0.089	0.089	0.071	0.085	0.103	0.114	0.104	0.062	0.095	0.092	0.054	0.093
TX0517	0.006	0.022	0.036	-0.012	-0.009	-0.007	0.008	-0.029	-0.015	-0.035	-0.001	-0.002	-0.019	-0.012
TX0601	0.023	-0.003	0.006	-0.002	0.004	0.007	0.038	0.009	-0.009	0.014	-0.009	0.039	0.004	0.012
TX0602	0.049	0.064	0.084	0.015	-0.005	0.008	0.086	0.050	0.039	0.007	0.054	-0.002	0.017	0.009
TX0901	0.001	-0.002	0.034	-0.005	-0.013	-0.001	0.015	0.002	0.010	-0.016	0.010	0.004	-0.014	0.000
TX0902	-0.013	-0.018	0.023	-0.008	-0.006	0.001	0.001	-0.006	-0.010	-0.019	-0.005	0.018	-0.021	0.004
TX0903	-0.004	-0.006	0.029	-0.004	-0.011	0.003	0.012	0.003	0.003	-0.013	0.003	0.013	-0.021	0.004
TX0904	0.005	0.016	0.020	-0.025	-0.017	-0.002	0.014	0.006	-0.007	-0.035	-0.015	0.013	-0.025	-0.003

Table 4.4 Continued.

Sample	FL0902	TX0501	TX0502	TX0503	TX0504	TX0505	TX0506	TX0507	TX0508	TX0509	TX0510	TX0511	TX0512	TX0513
FL0501	0.008	-0.006	0.022	0.040	0.094	0.144	0.123	0.011	0.149	0.037	0.128	0.049	0.030	0.032
FL0502	0.036	0.026	-0.019	0.023	0.084	0.094	0.109	0.046	0.120	0.020	0.119	0.071	0.022	0.022
FL0503	0.127	0.085	-0.002	0.085	0.075	0.137	0.155	0.151	0.107	0.042	0.117	0.111	0.070	0.068
FL0601	0.021	0.070	0.005	0.030	0.081	0.075	0.153	0.079	0.110	0.018	0.074	0.011	0.008	0.014
FL0602	-0.009	0.037	-0.003	-0.004	0.083	0.073	0.140	0.034	0.112	-0.001	0.074	0.017	-0.021	-0.009
FL0603	0.000	0.055	0.028	-0.004	0.098	0.078	0.171	0.061	0.125	-0.006	0.078	0.016	-0.022	-0.007
FL0604	0.092	0.084	0.052	0.078	0.012	0.009	0.036	0.146	0.031	0.099	0.040	0.130	0.096	0.091
FL0605	0.039	0.105	0.064	0.017	0.128	0.031	0.188	0.096	0.160	0.038	0.082	0.015	0.007	0.014
FL0606	-0.009	0.087	0.030	-0.019	0.109	0.025	0.196	0.061	0.126	-0.013	0.058	0.003	-0.025	-0.029
FL0607	0.000	0.037	0.014	0.008	0.079	0.078	0.135	0.048	0.119	0.028	0.080	0.016	0.001	0.025
FL0608	0.073	0.075	-0.006	0.025	0.077	0.084	0.140	0.124	0.106	0.021	0.088	0.058	0.020	0.007
FL0609	0.009	0.064	0.060	0.009	0.104	0.089	0.166	0.058	0.142	0.024	0.087	0.019	0.001	0.016
FL0610	0.011	0.015	-0.015	0.017	-0.012	-0.003	0.048	0.063	0.004	0.006	-0.002	0.042	0.016	0.004
FL0901	0.013	0.048	0.028	0.007	0.090	0.069	0.157	0.081	0.128	0.016	0.091	0.033	-0.004	-0.001
FL0902		0.037	0.029	0.021	0.141	0.123	0.171	0.005	0.182	0.030	0.142	0.023	0.013	0.011
TX0501	0.014		0.021	0.040	0.046	0.128	0.066	0.039	0.104	0.048	0.116	0.094	0.037	0.061
TX0502	0.012	-0.016		0.024	0.059	0.084	0.070	0.054	0.103	0.022	0.100	0.078	0.007	0.017
TX0503	0.009	0.005	0.013		0.088	0.058	0.157	0.048	0.114	-0.009	0.061	0.008	-0.017	-0.004
TX0504	0.062	-0.012	0.022	0.052		0.042	0.029	0.179	-0.007	0.084	0.027	0.156	0.083	0.109
TX0505	0.049	0.037	0.031	0.019	0.001		0.110	0.205	0.042	0.076	0.012	0.127	0.069	0.081
TX0506	0.090	0.022	0.037	0.099	0.023	0.058		0.197	0.070	0.158	0.135	0.251	0.137	0.152
TX0507	0.004	0.008	0.031	0.028	0.100	0.110	0.113		0.221	0.064	0.178	0.058	0.060	0.062
TX0508	0.100	0.044	0.057	0.075	-0.016	-0.009	0.049	0.133		0.104	0.021	0.185	0.123	0.135
TX0509	0.011	-0.006	0.008	-0.014	0.029	0.009	0.080	0.035	0.060		0.058	0.017	-0.013	-0.003
TX0510	0.077	0.042	0.053	0.033	-0.004	-0.043	0.079	0.113	0.011	0.036		0.101	0.079	0.104
TX0511	0.003	0.042	0.021	-0.012	0.068	0.036	0.126	0.026	0.108	0.000	0.066		0.019	0.034
TX0512	-0.006	-0.023	-0.005	-0.024	0.033	0.009	0.067	0.030	0.070	-0.011	0.050	0.004		-0.013
TX0513	0.004	0.013	0.004	-0.006	0.042	0.011	0.075	0.028	0.077	-0.004	0.056	0.005	-0.018	
TX0514	-0.001	0.012	-0.011	-0.024	0.007	-0.059	0.043	0.030	0.052	-0.001	0.030	0.006	-0.012	-0.006
TX0515	-0.004	-0.019	-0.009	0.001	0.030	0.011	0.045	0.029	0.083	0.012	0.071	0.026	-0.009	-0.007
TX0516	0.057	0.018	0.062	0.088	0.120	0.162	0.116	0.019	0.158	0.080	0.160	0.094	0.079	0.080
TX0517	0.006	0.015	0.010	-0.005	0.028	0.010	0.077	0.047	0.067	0.004	0.036	-0.002	-0.017	-0.006
TX0601	0.015	0.030	0.000	0.012	0.034	0.003	0.088	0.049	0.078	0.021	0.064	0.021	0.016	0.018
TX0602	-0.013	-0.004	0.023	0.011	0.062	0.049	0.096	0.037	0.102	0.020	0.088	0.022	0.016	0.012
TX0901	-0.009	-0.005	-0.011	0.002	0.023	0.008	0.048	0.017	0.065	0.010	0.057	0.012	-0.008	0.000
TX0902	0.002	0.037	-0.016	0.004	0.021	0.007	0.055	0.009	0.069	0.012	0.054	0.018	-0.013	0.007
TX0903	-0.004	-0.004	-0.009	0.001	0.024	0.005	0.049	0.019	0.065	0.010	0.054	0.004	-0.005	0.002
TX0904	0.007	0.002	-0.006	0.001	0.021	0.026	0.058	0.044	0.059	0.005	0.045	0.005	-0.013	0.000

Table 4.4 Continued.

Sample	TX0514	TX0515	TX0516	TX0517	TX0601	TX0602	TX0901	TX0902	TX0903	TX0904
FL0501	0.065	0.013	0.037	0.041	0.064	0.102	0.020	0.044	0.027	0.046
FL0502	0.015	0.027	0.098	0.057	0.025	0.190	0.031	0.022	0.024	0.040
FL0503	0.074	0.072	0.177	0.068	0.009	0.214	0.082	0.080	0.067	0.053
FL0601	-0.007	0.006	0.180	-0.025	-0.015	0.038	-0.017	0.004	-0.014	-0.038
FL0602	-0.010	-0.005	0.127	-0.025	0.015	0.029	-0.024	-0.022	-0.016	-0.032
FL0603	-0.002	-0.004	0.156	-0.017	0.013	0.038	-0.002	0.021	0.008	-0.005
FL0604	0.024	0.061	0.207	0.032	0.084	0.138	0.035	0.059	0.055	0.046
FL0605	0.007	-0.007	0.233	-0.024	0.035	0.108	0.024	0.053	0.043	0.029
FL0606	-0.051	-0.002	0.198	-0.043	-0.027	0.081	0.012	-0.021	-0.004	-0.024
FL0607	0.000	0.001	0.142	-0.022	0.023	0.020	-0.023	0.017	-0.005	-0.019
FL0608	0.002	0.015	0.188	-0.009	-0.013	0.154	0.039	0.023	0.019	-0.026
FL0609	0.022	0.009	0.151	-0.002	0.073	0.024	0.008	0.053	0.027	0.020
FL0610	-0.025	-0.003	0.121	-0.034	0.018	0.024	-0.033	-0.016	-0.030	-0.035
FL0901	-0.011	-0.007	0.171	-0.027	0.024	0.070	0.008	0.027	0.011	-0.006
FL0902	0.022	0.006	0.095	0.008	0.042	0.033	-0.007	0.005	0.000	0.017
TX0501	0.069	0.025	0.040	0.058	0.104	0.091	0.029	0.066	0.042	0.040
TX0502	0.005	-0.010	0.113	0.009	0.011	0.113	-0.006	-0.009	-0.005	-0.015
TX0503	-0.018	0.020	0.145	-0.006	0.041	0.089	0.020	0.029	0.014	0.001
TX0504	0.068	0.087	0.202	0.064	0.119	0.159	0.073	0.110	0.085	0.057
TX0505	-0.001	0.080	0.324	0.038	0.083	0.169	0.079	0.096	0.087	0.065
TX0506	0.116	0.101	0.189	0.129	0.199	0.229	0.110	0.136	0.119	0.104
TX0507	0.083	0.059	0.036	0.076	0.104	0.093	0.034	0.037	0.044	0.077
TX0508	0.094	0.144	0.253	0.100	0.141	0.194	0.114	0.137	0.122	0.095
TX0509	0.001	0.024	0.141	0.007	0.049	0.072	0.022	0.046	0.017	0.014
TX0510	0.047	0.115	0.264	0.052	0.105	0.139	0.086	0.112	0.093	0.074
TX0511	0.025	0.060	0.182	0.010	0.049	0.032	0.022	0.066	0.021	0.021
TX0512	-0.012	-0.018	0.157	-0.017	0.039	0.059	-0.009	0.013	0.002	-0.016
TX0513	0.000	-0.004	0.146	-0.013	0.044	0.091	0.017	0.019	0.007	0.005
TX0514		0.002	0.202	-0.033	-0.009	0.110	-0.003	-0.007	-0.012	-0.025
TX0515	0.003		0.130	-0.024	0.021	0.067	-0.017	0.012	0.000	-0.007
TX0516	0.092	0.069		0.188	0.222	0.176	0.120	0.135	0.131	0.176
TX0517	-0.019	-0.014	0.109		-0.007	0.052	-0.018	0.001	-0.013	-0.037
TX0601	-0.002	0.016	0.107	0.003		0.131	0.020	0.001	0.016	-0.015
TX0602	0.041	0.018	0.080	0.000	0.058		0.016	0.107	0.053	0.056
TX0901	0.000	-0.005	0.062	-0.009	0.013	0.001		-0.012	-0.023	-0.019
TX0902	-0.009	-0.001	0.064	0.003	-0.007	0.024	-0.009		-0.012	-0.010
TX0903	-0.005	0.001	0.067	-0.005	0.011	0.009	-0.011	-0.009		-0.019
TX0904	-0.022	-0.008	0.096	-0.013	-0.007	0.004	-0.016	-0.015	-0.014	

Table 4.5 Pairwise values of genetic divergence between groups. Jost's *D* (above diagonal) and F_{ST} (below diagonal). Italicized values indicate comparison remained significant after Bonferroni correction.

Sample	FL2005	FL2006	FL2009	TX2005	TX2006	TX2009
FL2005		*0.052*	*0.051*	*0.024*	*0.056*	*0.048*
FL2006	*0.022*		0.003	*0.015*	-0.005	0.005
FL2009	*0.019*	-0.001		*0.021*	0.001	0.011
TX2005	0.008	*0.005*	*0.010*		*0.029*	*0.017*
TX2006	0.020	-0.002	-0.005	0.007		0.001
TX2009	*0.015*	0.001	0.004	*0.006*	-0.002	

However, the overall magnitude of genetic divergence among groups was low, with maximum divergence values of 0.022 and 0.056 for F_{ST} and Jost's *D*, respectively.

In the AMOVA analysis, all three grouping strategies showed the majority of the genetic variation was found within samples (>97.3%; $P < 0.0001$ for all three groupings). The majority of the remaining genetic variation was attributable to differences among samples within groups (>2.4%; $P < 0.0001$ for all three groupings). There was no significant genetic variation found among groups regardless of the grouping strategy (<0.3%; $P > 0.2000$ for all three groupings).

No significant spatial structure among samples was detected by the spatial autocorrelation analysis (Fig. 4.2). In the analysis with Structure, the peak for ΔK occurred at $K = 3$ for both models (admixture and no admixture) when location information was included. When location information was excluded, the peak for ΔK occurred at $K = 4$ and $K = 5$ for the admixture and no admixture models, respectively. Inspection of individual runs for both models revealed that the ancestry for the majority of individual cells was allocated approximately equally among *K* populations. The calculated values for ΔK for three models (admixture with and without location

information and the no admixture model without location information) were very small, ranging from 0.01 to 4.27, compared to the no admixture model with location information (0.34 to 20.07; Fig. 4.3). For increasing values of *K*, the Ln P(D) values continued to increase slightly while the variance greatly increased, indicating the true value of *K* may have been passed.

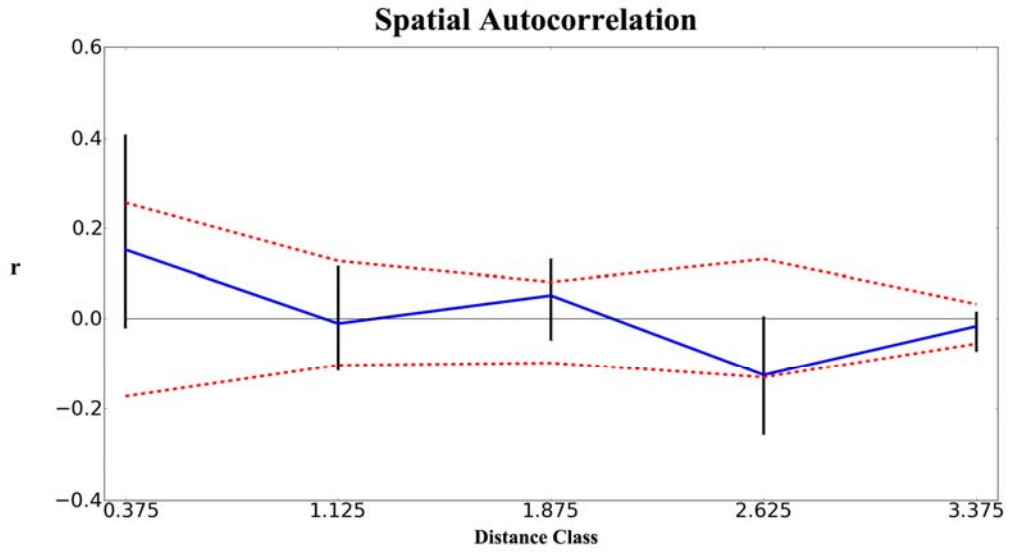

Figure 4.2 Spatial autocorrelation correlogram of correlation (r) at four distance classes. Geographic distance was log transformed in GenAlEx v6.41. Dashed lines indicate the 95% confidence interval of the null hypothesis of no correlation. Geographic distances (km) are ~2.4, 13.3, 75, 421.7, and 2371.4 for the five distance classes, respectively.

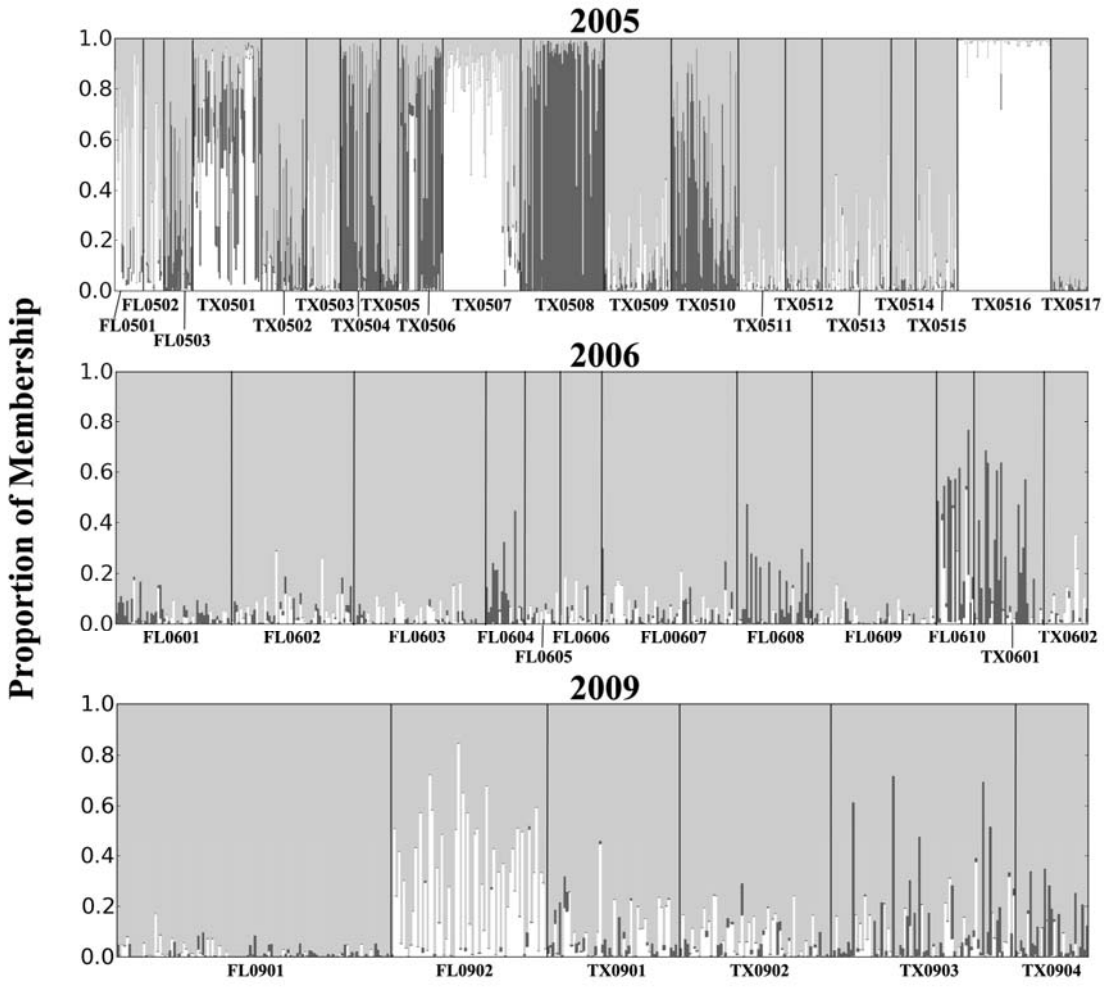

Figure 4.3 Results of the STRUCTURE analysis separated by year for the no admixture model with 'locprior' option invoked ($K = 3$). The three populations are distinguished by color: light grey, dark grey, and white. Vertical black lines separate individual samples.

DISCUSSION

Cells of *K. brevis* are diverse genetically, similar to other dinoflagellates that have been studied (Nagai *et al.* 2007, 2009; Erdner *et al*. 2011). Many studies of dinoflagellate population structure have made use of clonal cultures in order to increase the number of genetic markers due to larger quantities of useable DNA. The difficulty of

establishing clonal cultures of *K. brevis* necessitated use of preserved single cells in this study.

It is not known whether the detection of two alleles at a microsatellite was the result of a previous gene duplication event, an artifact of the amplification process or from two cells inadvertently isolated into the same tube. In other cases, incomplete haplotypes were obtained from single cells when missing alleles were noted. Two possibilities exist to explain missing data: (i) a failed PCR or (ii) the presence of a null allele. In three previous studies of genetic diversity and population structure in dinoflagellates, the number of microsatellite markers used was less than the number tested because of inconsistent amplification from cultured strains of *Alexandrium* spp. (Erdner *et al*. 2011; Casabianca *et al*. 2012) and *Cochlodinium polykrikoides* (Nagai *et al.* 2009). Erdner *et al*. (2011) were testing microsatellite markers originally developed for a closely related species (*A. tamarense*; Nagai *et al.* 2004) so it is not surprising that six of the eleven markers tested amplified in less than 90% of strains tested. Casabianca *et al*. (2012) tested 12 microsatellite markers for *A. minutum* and selected seven based upon the ability to consistently amplify the markers in test strains. It is important to note that in all of these cases, genomic DNA was extracted from pellets of cultured cells. Unfortunately, the single cell amplification protocol used in this study precludes the possibility of attempting the PCR again using the same cell.

No significant genetic divergence was detected among the eight samples from the FL2006 group, suggesting that this geographically large bloom which covered an area greater than 500km^2 may be a single population. This is in contrast to the genetic

divergence detected among several samples in the TX2005 group. Samples from the Corpus Christi Bay area were collected over a three week period of time and indicate there may have been more than one genetically distinct population of *K. brevis* off the coast of Texas in 2005. A similar result was observed by Erdner *et al*. (2011) as they tracked a bloom in the northeastern U.S. for a thirty-seven day period. The environmental data from the sample sites shows varying temperatures and salinities for the locations, likely contributing to the observed genetic divergence and leading Erdner *et al*. (2011) to consider the five samples as having come from the same population.

On a larger geographic scale, Steidinger *et al*. (1998) suggested that the west coast of Florida was a source for blooms of *K. brevis* throughout the Gulf. If this hypothesis is true, no significant genetic divergence should be detected between blooms from Florida and blooms from Texas. Significant genetic divergence was observed between the FL2005 and TX2005 groups but the estimate of genetic divergence was low (*D*: 0.024; F_{ST}: 0.008) and may not be biologically significant. No significant genetic divergence was detected between FL2006 and TX2006 or FL2009 and TX2009, and supports the hypothesis that blooms of *K. brevis* have a common origin. The absence of significant genetic divergence among groups in AMOVA also supports occurrence of a single population of *K. brevis* in the Gulf.

Multiple populations were detected from the results of the Structure analysis. However, the detection of multiple populations appeared to be driven largely by the same four samples responsible for the majority of significant tests of genetic divergence. There was no distinct clustering pattern for the majority of cells from samples in years

2006 and 2009 from both Florida and Texas, supporting the hypothesis of a single origin for blooms. The results of the spatial autocorrelation analysis indicate no significant spatial structure exists among the samples. The asexual nature of *K. brevis* gives rise to clonal cells in close proximity that should provide a strong autocorrelation signal at small distances. Identical haplotypes from cells with no missing data were detected within some samples but, there also were a large number of shared haplotypes among samples, including one haplotype found thirteen times in eight different samples from different years (data not shown).

While significant genetic divergence was observed between some samples, a consistent pattern of genetic divergence between samples from Florida and samples from Texas was not identified. There are at least two possible explanations for this result: (i) blooms have a common geographic origin and are transported to distant regions in the Gulf or (ii) the Gulf contains one large population of *K. brevis* and blooms can occur independently in different regions of the Gulf. Movement of a surface bloom from the eastern side of the Gulf to the western side is possible as evidenced by previous drifter studies (Lugo-Fernandez *et al.* 2001; Morey *et al*. 2003; Ohlmann and Niiler 2005). Whether whole blooms, or seed populations, are actually transported across the Gulf needs additional study. Another, more likely, scenario involves the movement of water by surface currents or eddies in the Gulf of Mexico which continuously mixes cells of *K. brevis* from distant locations, preventing the development of population-genetic structure and resulting in the presence of a single, large population of *K. brevis* in the Gulf (Merrell and Morrison 1981; Elliott 1982). Future work describing bloom initiation

should focus on physical mechanisms of bloom (or cell) transport over long distances in the Gulf of Mexico in addition to coastal areas.

CHAPTER V

CONCLUSIONS:

GENETIC DIVERSITY AND POPULATION STRUCTURE IN *KARENIA BREVIS*

WHAT HAVE WE LEARNED?

INTRODUCTION

Genetic diversity is a key component for the success of any species. By providing a genetic basis for the phenotypic differences upon which natural selection can act, genetic diversity is directly responsible for a species' ability to adapt to a wide range of environments. In an oceanic environment, conditions can change quickly and requires planktonic organisms, including dinoflagellates, to tolerate a broad range of environmental parameters. The global expansion and increase of first reports of dinoflagellate blooms has placed an emphasis on accurate identification of species and their origin, whether endemic or introduced, in addition to toxin determination and physiological variability (Azanza and Taylor 2001; Heil *et al*. 2001; Whyte *et al.* 2001; Smayda 2002; Campbell *et al.*2010). The accurate identification of dinoflagellate species and their putative geographic origin relies on knowing inter- and intra- specific variation present among similar species. The use of highly variable genetic markers (e.g. microsatellites) has greatly increased our knowledge of genetic diversity and population structure in several dinoflagellate species including *Alexandrium* spp. (Alpermann *et al.* 2009; Masseret *et al.* 2009; Erdner *et al*. 2011; Casabianca *et al.* 2012), *Cochlodinium polykrikoides* (Nagai *et al.* 2009), *Oxyrrhis marina* (Lowe *et al*. 2010), and *Karenia*

brevis (Ch. 3,4). *Karenia brevis* is the predominant HAB species in the Gulf of Mexico and much research has been conducted on *K. brevis* investigating physiological variability and toxin production while little investigation into the genetic diversity present in *K. brevis* had been conducted (Steidinger *et al.*1998; Loret *et al*. 2002; Brand *et al.* 2012). Identifying and quantifying the genetic diversity present in *K. brevis* has become an important task as research into toxin production and growth parameters revealed intraspecific variation existed among cultured isolates (Baden and Tomas 1988; Loret *et al.* 2002; Errera *et al*. 2010). This chapter highlights the importance of identifying genetic diversity and population structure in dinoflagellates, with an emphasis on *K. brevis*, and provides a perspective on the anticipated contributions of next-generation sequencing (NGS) technology to both physiological and genetic research on dinoflagellates.

Importance of identifying genetic diversity and population structure

Algal toxins are increasingly responsible for human intoxications and marine mortality events and as global expansion of HAB species continues, these events will likely also continue to increase (Van Dolah 2000). Detection of a HAB event is typically followed by investigation to identify what species are present and their putative origin. It is important to note while not all HABs are caused by dinoflagellates, dinoflagellates are a major contributor to the global expansion of HABs (Smayda 2002). Increased knowledge of the genetic diversity (within and among species) and population structure in dinoflagellates can provide valuable information about the identity and possible geographic origins of the unknown species. Species specific ribosomal markers for

dinoflagellates have been available for almost twenty years and have resolved most of the main branches in the dinoflagellate phylogenetic tree (Scholin *et al.* 1994; Murray *et al.* 2005). In some instances the diversity in ribosomal regions has been too low to distinguish closely related species and mitochondrial markers were needed to provide phylogenetic resolution (Raho *et al.* 2008). The global expansion of dinoflagellate species capable of forming HABs is likely to result in more cases where commonly used markers (e.g. ribosomal, mitochondrial) are unable to resolve the correct identity of the unknown species and places an emphasis on the need to sequence multiple genes from as many dinoflagellate species as possible. There is also a need to identify and genotype highly polymorphic markers within a species. Masseret *et al*. (2009) compared isolates of *Alexandrium catenella* from the Mediterranean with isolates from the coastal waters of Japan and found a close relationship among isolates from both locations when using ribosomal markers however, microsatellite markers were able to distinguish several populations, one from the Mediterranean and two different lineages from Japan that ribosomal markers had not previously identified. More data on the local and global population structure of *A. catenella* could provide greater resolution about the likely origin of the population in the Thau Lagoon (Masseret *et al*. 2009). Identification of the geographic origin of an introduced species is of great importance when that species is capable of forming a HAB. The environmental and economic impacts of HABs from endemic species has spurred research into mitigation techniques that can be applied to blooms in the field (Archambault *et al.* 2003; Lee *et al*. 2008; Kim 2006). Accurate identification of the geographic origin of an introduced dinoflagellate species could

reveal potential mitigation techniques (e.g. top-down control with grazers, Xu *et al.* 2010; clay flocculation, Archambault *et al.*2003; Pierce *et al.* 2004; Lee *et al.* 2008) depending on which species is identified and where it originated. While it is unlikely that the introduced species will be eliminated permanently, mitigation techniques can reduce the impact of HABs on the local environment and economy.

A common result of genetic studies investigating dinoflagellates has been the observation that populations and blooms are very diverse genetically (Nagai *et al.* 2007; 2009; Lowe *et al.* 2010; ch. 3, 4). Vegetative, asexual growth is typically responsible for the increase in cell numbers during the growth stage of a bloom (Steidinger and Garcés 2006). Cells that are better acclimated to environmental conditions should proliferate to the point that few genetically distinct lineages remain and genetic diversity decreases substantially. However, this has not been found to be true. Lowe *et al*. (2010) identified 183 unique haplotypes from 200 isolates (~91%) of *Oxyrrhis marina* and Alpermann *et al*. (2009) identified seventy-seven unique haplotypes from seventy-seven isolates (100%) of *A. tamarense*. In another study of *Alexandrium*, Erdner *et al*. (2011) identified 119 unique haplotypes from 171 isolates (~70%) of *A. fundyense* but noted the proportion of unique haplotypes per sample was higher (83 to 92%; no haplotype occurred more than twice within a single sample). Identical haplotypes (five microsatellite markers) were identified among single cells of *K. brevis* from eighteen bloom samples but the average proportion of unique haplotypes per sample was over 90% (94.0 ± 6.5% [mean ± SD] ; sample size: 36.8 ± 13.8 [mean ± SD]). It is possible increased sampling of cells from these samples would reduce the proportion of unique

haplotypes but the presence of high numbers of unique haplotypes does not support the hypothesis that a few, well adapted lineages are responsible for bloom growth. This pattern of high diversity among isolates and populations has been identified in several phytoplankton species, including a coccolithophore (*Emiliania huxleyi*, Iglesias-Rodriguez *et al*. 2006), diatoms (*Ditylum brightwellii*, Rynearson and Armbrust 2004; *Pseudo-nitzschia* spp., Evans *et al.* 2004, 2005; Adams *et al*. 2009; Casteleyn *et al.* 2009), and raphidophytes (*Chattonella* spp. Demura *et al*. 2007; Nishitani *et al*. 2007). The diploid nature and potential for sexual recombination mean high genetic diversity is likely to be found among the three groups (coccolithophores, diatoms, and raphidophytes) even though asexual reproduction may occur. Finding high levels of genetic diversity among populations of dinoflagellates, though haploid, is therefore not surprising and means they maintain the ability to adapt to a wide range of environments, consistent with the observed global expansion of dinoflagellates.

Genetic diversity and population structure in Karenia brevis

Blooms of *K. brevis* occur almost exclusively in the Gulf of Mexico but can occasionally be carried up the east coast of the United States (Tester *et al*. 1991; Tester and Steidinger 1997). *Karenia brevis* is an oceanic species whose blooms initiate offshore before sometimes being moved onshore, where environmental conditions can vary dramatically (Steidinger *et al*. 1998). Toxin measurements (toxin profile and total toxin) obtained by Errera *et al*. (2010) showed some strains to be very different physiologically, consistent with earlier work on growth and swimming speed (Loret *et al*. 2002; McKay *et al*. 2006). The observed physiological variation is likely a product of

underlying genetic variation among the strains. Renshaw *et al*. (2006) developed microsatellite markers for *K. brevis* and were able to show strains of *K. brevis* to be genetically different from each other. An increased number of microsatellite markers for *K. brevis* and the availability of additional strains of *K. brevis* showed *K. brevis* is very diverse genetically, with only one multi-locus haplotype appearing more than once (Ch. 3; see section, above).

The high amount of genetic diversity present in *K. brevis* raises the question of whether sexual recombination is occurring. Sexual recombination could continually produce new allelic combinations. While sexual reproduction has been documented in dinoflagellates (Walker and Steidinger 1979; Blackburn *et al*. 1989; Parrow and Burkholder 2004; Figueroa *et al.* 2006), the complete sexual cycle has not been documented though presumably diploid cells were identified from cultured cells (Walker 1982). Walker (1982) noted the development of cyst-like structures but was unable positively identify hypnozygotes and it remains unknown how frequently the sexual cycle may occur or if it even occurs in *K. brevis*. Another possible explanation for the high genetic diversity observed in *K. brevis* is mutation during replication of DNA preceding cell division. The extremely large genome size of *K. brevis* (~100pg/nucleus; 1×10^{-11} bp; Rizzo *et al.* 1982) coupled with large population sizes means that genetically distinct cells could be produced from every cell division event (Lakeman 2009 and ref. within). Mutation is an unavoidable consequence of replication but its contribution, and that of the sexual cycle, to the amount of observed genetic diversity in *K. brevis* is not known.

The field study conducted revealed no population structure among blooms of *K. brevis* in the Gulf of Mexico (Ch. 4). The lack of population structure for a planktonic species in the semi-enclosed basin of the Gulf is not surprising given the results of drifter studies revealing the possibility of westward surface transport in the Gulf (Lugo-Fernandez *et al*. 2001; Williams *et al*. 1977, cited in Tester and Steidinger 1997). What remains unknown at this point is where blooms originate in the Gulf. Knowing the population structure of *K. brevis* in the Gulf of Mexico can aid in the identification of physical mechanisms for bloom formation and transport and, if successfully identified, these mechanisms could be utilized in modeling bloom formation. The early warning of a developing bloom provided by such models would be an invaluable tool for reducing human intoxications from brevetoxins.

High-throughput sequencing and dinoflagellates

The widespread availability and decreasing cost of high-throughput sequencing (HTS) will lead to a deeper understanding of the genetic structure and function of dinoflagellates (Lin 2011). The increased number of genetic markers made available by HTS will provide high resolution markers (e.g. single nucleotide polymorphisms) for the study of fine-scale population structure (Holsinger 2010). Single nucleotide polymorphisms are generally limited to a maximum of four alleles but the potential to identify hundreds or thousands of markers will increase the power to detect genetic divergence and undoubtedly result in the identification of previously undetected links between geographically separated dinoflagellate populations (Kalinowski 2002). The higher number of genetic markers means fewer individuals may be required for

genotyping. Though the cost of sequencing is coming down, a population study using HTS and involving numerous individuals can still be cost prohibitive without pooling individuals (Futschik and Schlötterer 2010). The ability to pool individuals during a single HTS run is especially ideal for dinoflagellates capable of forming HABs; a single water sample could provide millions of individuals.

For physiological studies, HTS of RNA (transcriptome) from dinoflagellates is likely to provide a wealth of new information about gene regulation and structure. Differential expression analyses hold the potential to reveal pathways responsible for toxin synthesis or even the genes responsible for toxin production. To date, microarray analysis has been a standard tool for investigation of expression differences and Lidie *et al.* (2005) developed a microarray for *K. brevis* to investigate gene expression changes in *K. brevis* under different environmental conditions. One impediment to microarray analysis is the requirement to know the gene or RNA sequences beforehand and rare transcripts, those that are only expressed under certain conditions, may be missed altogether. HTS datasets are not limited by the requirement to know gene/RNA sequences beforehand because the RNA itself is sequenced. In addition to the identification of rare transcripts, HTS can also identify isoforms of known and unknown transcripts, sequences that may be missed entirely if the hybridization sequence on a microarray differs from that of the isoform. With large sequencing projects currently underway (e.g. Gordon and Betty Moore Foundation's Marine Microbial Eukaryotic Transcriptome Sequencing Project), the next few years will see the availability of large volumes of transcriptome data for dinoflagellates (and other marine microbes). To date,

the large genome sizes found in dinoflagellates has made the possibility of a whole genome sequence unlikely. As technology continues to improve (and costs continue to decrease), the prospect of sequencing entire genomes from multiple dinoflagellate species becomes more probable, making the next few years very exciting in dinoflagellate genomics.

REFERENCES

Adams, N. G., Trainer, V. L., Rocap, G., Herwig, R. P. & Hauser, L. 2009. Genetic population structure of *Pseudo-nitzschia pungens* (Bacillariophyceae) from the Pacific Northwest and the North Sea. *J Phycol* 45:1037-45.

Alpermann, T. J., Beszteri, B., John, U., Tillmann, U. & Cembella, A. D. 2009. Implications of life-history transitions on the population genetic structure of the toxigenic marine dinoflagellate *Alexandrium tamarense*. *Mol Ecol* 18:2122-33.

Archambault, M. C., Grant, J. & Bricelj, V. M. 2003. Removal efficiency of the dinoflagellate *Heterocapsa triquetra* by phosphatic clay, and implications for the mitigation of harmful algal blooms. *Mar Ecol-Prog Ser* 253:97-109.

Azanza, R. V. & Taylor, F. J. R. 2001. Are *Pyrodinium* blooms in the Southeast Asian region recurring and spreading? A view at the end of the millennium. *Ambio* 30:356-64.

Baden, D. G. & Tomas, C. R. 1988. Variations in major toxin composition for 6 clones of *Ptychodiscus brevis*. *Toxicon* 26:961-63.

Blackburn, S.I., Hallegraeff, G.M. & Bolch, C.J. 1989. Vegetative reproduction and sexual life cycle of the toxic dinoflagellate *Gymnodinium catenatum* from Tasmania, Australia. *J Phycol* 25:577-90.

Bolch, C.J.S., Blackburn, S.I., Hallegraeff, G.M. & Vaillancourt, R.E. 1999. Genetic variation among strains of the toxic dinoflagellate *Gymnodinium catenatum* (Dinophyceae). *J Phycol* 35:356-67.

Brand, L.E. & Compton, A. 2007. Long-term increase in *Karenia brevis* abundance along the southwest Florida coast. *Harmful Algae*. 6:232-52.

Brand, L. E., Campbell, L. & Bresnan, E. 2012. *Karenia*: The biology and ecology of a toxic genus. *Harmful Algae* 14:156-78.

Brown, A.F.M., Dortch, Q., Van Dolah, F.M., Leighfield, T.A., Morrison, W., Thessen, A.E., Steidinger, K.A., Richardson, B., Moncreiff, C.A. & Pennock, J.R. 2006. Effect of salinity on the distribution, growth, and toxicity of *Karenia* spp. *Harmful Algae*. 5:199-212.

Campbell, L., Olson, R. J., Sosik, H. M., Abraham, A., Henrichs, D. W., Hyatt, C. J. & Buskey, E. J. 2010. First harmful *Dinophysis* (Dinophyceae, Dinophysiales) bloom in the US is revealed by automated imaging flow cytometry. *J Phycol* 46:66-75.

Casabianca, S., Penna, A., Pecchioli, E., Jordi, A., Basterretxea, G. & Vernesi, C. 2012. Population genetic structure and connectivity of the harmful dinoflagellate *Alexandrium minutum* in the Mediterranean Sea. *Proc R Soc Biol Sci Ser B* 279:129-38.

Casper, E. T., Patterson, S. S., Bhanushali, P., Farmer, A., Smith, M., Fries, D. P. & Paul, J. H. 2007. A handheld NASBA analyzer for the field detection and quantification of *Karenia brevis*. *Harmful Algae*. 6:112-8.

Casteleyn, G., Evans, K. M., Backeljau, T., D'hondt, S., Chepurnov, V., Sabbe, K. & Vyverman, W. 2009. Lack of population genetic structuring in the marine planktonic diatom *Pseudo-nitzschia pungens* (Bacillariophyceae) in a

heterogeneous area in the Southern Bight of the North Sea. *Mar Biol* 156:1149-58.

Cheng, Y. S., Villareal, T. A., Zhou, Y., Gao, J., Pierce, R. H., Wetzel, D., Naar, J. & Baden, D. G. 2005. Characterization of red tide aerosol on the Texas coast. *Harmful Algae*. 4:87-94.

Demura, M., Kawachi, M., Kunugi, M., Nishizawa, T., Kasai, F. & Watanabe, M. M. 2007. Development of microsatellite markers for the red tide-forming harmful species *Chattonella antiqua*, *C. marina*, and *C. ovata* (Raphidophyceae). *Mol Ecol Notes* 7:315-17.

Doyle, J. J. & Doyle, J. L. 1990. A rapid total DNA isolation procedure for fresh plant tissue. *Focus* 12:13-5

El Mousadik, A. & Petit, R. J. 1996. High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theor Appl Genet* 92:832-9.

Elliott, B. A. 1982. Anticyclonic rings in the Gulf of Mexico. *J Phys Oceanogr* 12:1292-309.

Erdner, D. L., Richlen, M., McCauley, L. A. R. & Anderson, D. M. 2011. Diversity and Dynamics of a widespread bloom of the toxic dinoflagellate *Alexandrium fundyense*. *PLoS One* 6:e22965.

Errera, R. M., Bourdelais, A., Drennan, M. A., Dodd, E. B., Henrichs, D. W. & Campbell, L. 2010. Variation in brevetoxin and brevenal content among clonal cultures of *Karenia brevis* may influence bloom toxicity. *Toxicon* 55:195-203.

Evanno, G., Regnaut, S. & Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611-20.

Evans, K. M., Bates, S. S., Medlin, L. K. & Hayes, P. K. 2004. Microsatellite marker development and genetic variation in the toxic marine diatom *Pseudo-nitzschia multiseries* (Bacillariophyceae). *J Phycol* 40:911-20.

Evans, K. M., Kuhn, S. F. & Hayes, P. K. 2005. High levels of genetic diversity and low levels of genetic differentiation in North Sea *Pseudo-nitzschia pungens* (Bacillariophyceae) populations. *J Phycol* 41:506-14.

Excoffier, L. & Lischer, H. E. L. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res* 10:564-67.

Figueroa, R. I., Bravo, I. & Garcés, E. 2006. Multiple routes of sexuality in *Alexandrium taylori* (Dinophyceae) in culture. *J Phycol* 42:1028-39.

Futschik, A. & Schlötterer, C. 2010. The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics* 186:207-18.

Gray, M., Wawrik, B., Paul, J. & Casper, E. 2003. Molecular detection and quantitation of the red tide dinoflagellate *Karenia brevis* in the marine environment. *Appl Environ Microbiol* 69:5726-30.

Heil, C. A., Glibert, P. M., Al-Sarawi, M. A., Faraj, M., Behbehani, M. & Husain, M. 2001. First record of a fish-killing *Gymnodinium* sp bloom in Kuwait Bay, Arabian Sea: chronology and potential causes. *Mar Ecol-Prog Ser* 214:15-23.

Henrichs, D. W., Renshaw, M. A., Santamaria, C. A., Richardson, B., Gold, J. R. & Campbell, L. 2008. PCR amplification of microsatellites from single cells of *Karenia brevis* preserved in Lugol's iodine solution. *Mar Biotechnol* 10:122-7.

Hetland, R. D. & Campbell, L. 2007. Convergent blooms of *Karenia brevis* along the Texas coast. *Geophys Res Letters*. 34:L19604, doi:10.1029/2007GL030474.

Holsinger, K. E. 2010. Next generation population genetics and phylogeography. *Mol Ecol* 19:2361-63.

Hutchison, C. A., Hamilton, O. S., Pfannkoch, C. & Venter, J. C. 2005. Cell-free cloning using {phi}29 DNA polymerase. *Proc Natl Acad Sci USA* 102:17332-6.

Iglesias-Rodriguez, M. D., Schofield, O. M., Batley, J., Medlin, L. K. & Hayes, P. K. 2006. Intraspecific genetic diversity in the marine coccolithophore *Emiliania huxleyi* (Prymnesiophyceae): The use of microsatellite analysis in marine phytoplankton population studies. *J Phycol* 42:526-36.

Jost, L. 2008. G_{ST} and its relatives do not measure differentiation. *Mol Ecol* 17:4015-26.

Kai, A. K. L., Cheung, Y. K., Yeung, P. K. K. & Wong, J. T. Y. 2006. Development of single-cell PCR methods for the Raphidophyceae. *Harmful Algae* 5:649-57.

Kalinowski, S. T. 2002. How many alleles per locus should be used to estimate genetic distances? *Heredity* 88:62-65.

Ki, J., Jang, G. Y. & Han, M. 2005. Integrated method for single-cell DNA extraction, PCR amplification, and sequencing of ribosomal DNA from harmful dinoflagellates *Cochlodinium polykrikoides* and *Alexandrium catenella. Mar Biotechnol* 6:587-93.

Kim, H. G. 2006. Mitigation and controls of HABs. *In* Graneli, E., Turner, J. T. [Eds.] *Ecology of Harmful Algae*, Ecological Studies, Springer-Verlag, Berlin. pp. 327-38.

Lakeman, M. B., von Dassow, P. & Cattolico, R. A. 2009. The strain concept in phytoplankton ecology. *Harmful Algae* 8:746-58.

Leberg, P. L. 2002. Estimating allelic richness: Effects of sample size and bottlenecks. *Mol Ecol* 11:2445-49.

Leblond, J. D. & Chapman, P. J. 2002. A survey of the sterol composition of the marine dinoflagellates *Karenia brevis*, *Karenia mikimotoi*, and *Karlodinium micrum*: distribution of sterols within other members of the class Dinophyceae. *J Phycol* 38:670-82.

Lee, Y. J., Choi, J. K., Kim, E. K., Youn, S. H. & Yang, E. J. 2008. Field experiments on mitigation of harmful algal blooms using a Sophorolipid-Yellow clay mixture and effects on marine plankton. *Harmful Algae* 7:154-62.

Licea, S., Zamudio, M. E., Luna, R. & Soto, J. 2004. Free-living dinoflagellates in the southern Gulf of Mexico: Report of data (1979-2002). *Phycol Res* 52:419-28.

Lidie, K. B., Ryan, J. C., Barbier, M. & Van Dolah, F. M. 2005. Gene expression in Florida red tide dinoflagellate *Karenia brevis*: analysis of an expressed sequence tag library and development of DNA microarray. *Mar Biotechnol* 7:481-93.

Lin, S. J. 2011. Genomic understanding of dinoflagellates. *Res Microbiol* 162:551-69.

Loret, P., Tengs, T., Villareal, T. A., Singler, H., Richardson, B., McGuire, P., Morton, S., Busman, M. & Campbell, L. 2002. No difference found in ribosomal DNA

sequences from physiologically diverse clones of *Karenia brevis* (Dinophyceae) from the Gulf of Mexico. *J Plankton Res* 24:735-9.

Lowe, C. D., Montagnes, D. J. S., Martin, L. E. & Watts, P. C. 2010. High genetic diversity and fine-scale spatial structure in the marine flagellate *Oxyrrhis marina* (Dinophyceae) uncovered by microsatellite loci. *PLoS One* 5:e15557.

Lugo-Fernandez, A., Morin, M. V., Ebesmeyer, C. C. & Marshall, C. F. 2001. Gulf of Mexico historic (1955-1987) surface drifter data analysis. *J Coast Res* 17:1-16.

Magaña, H. A., Contreras, C. & Villareal, T. A. 2003. A historical assessment of *Karenia brevis* in the western Gulf of Mexico. *Harmful Algae*. 2:163-71.

Masseret, E., Grzebyk, D., Nagai, S., Genovesi, B., Lasserre, B., Laabir, M., Collos, Y., Vaquer, A. & Berrebi, P. 2009. Unexpected genetic diversity among and within populations of the toxic dinoflagellate *Alexandrium catenella* as revealed by nuclear microsatellite markers. *Appl Environ Microbiol* 75:2037-45.

McKay, L., Kamykowski, D., Milligan, E., Schaeffer, B. & Sinclair, G. 2006. Comparison of swimming speed and photophysiological responses to different external conditions among three *Karenia brevis* strains. *Harmful Algae*. 5:623-36.

Meirmans, P. G. & Hedrick, P. W. 2011. Assessing population structure: F_{ST} and related measures. *Mol Ecol Res* 11:5-18.

Merrell , W. J. & Morrison, J. M. 1981. On the circulation of the western Gulf of Mexico with observations from April 1978. *J Geophys Res* 86:4181-5.

Mikulski, C. M., Morton, S. L. & Doucette, G. J. 2005. Development and application of LSU rRNA probes for *Karenia brevis* in the Gulf of Mexico, USA. *Harmful Algae*. 4:49-60.

Morey, S. L., Martin, P. J., O'Brien, J. J., Wallcraft, A. A., Zavala-Hidalgo, J. 2003. Export pathways for river discharged fresh water in the northern Gulf of Mexico. *J Geophys Res*. 108:C10, 3303.

Murray, S., Jorgensen, M. F., Ho, S. Y. W., Patterson, D. J. & Jermiin, L. S. 2005. Improving the analysis of dinoflagellate phylogeny based on rDNA. *Protist* 156:269-86.

Nagai, S., Lian, C., Hamaguchi, M., Matsuyama, Y., Itakura, S. & Hogetsu, T. 2004. Development of microsatellite markers in the toxic dinoflagellate *Alexandrium tamarense* (Dinophyceae). *Mol Ecol Notes* 4:83-85.

Nagai, S., Sekino, M., Matsuyama,Y. & Itakura, S. 2006. Development of microsatellite markers in the toxic dinoflagellate *Alexandrium catenella* (Dinophyceae). *Mol Ecol Notes* 6:120-2.

Nagai, S., Lian, C., Yamaguchi, S., Hamaguchi, M., Matsuyama, Y., Itakura, S., Shimada, H., Kaga, S., Yamaguchi, H., Sonda, Y., Nishikawa, T., Kim, C. & Hogetsu, T. 2007. Microsatellite markers reveal population genetic structure of the toxic dinoflagellate *Alexandrium tamarense* (Dinophyceae) in Japanese coastal waters. *J Phycol* 43:43-54.

Nagai, S., Nishitani, G., Sakamoto, S., Sugaya, T., Lee, C. K., Kim, C. H., Itakura, S. & Yamaguchi, M. 2009. Genetic structuring and transfer of marine dinoflagellate

Cochlodinium polykrikoides in Japanese and Korean coastal waters revealed by microsatellites. *Mol Ecol* 18:2337-52.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* 70:3321-3.

Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York, 512 pp.

Nei, M. & Chesser, R. K. 1983. Estimation of fixation indices and gene diversities. *Ann Hum Genet* 47:253-9.

Nishitani, G., Nagai, S., Lian, C. L., Yamaguchi, H., Sakamoto, S., Yoshimatsu, S., Oyama, K., Itakura, S. & Yamaguchi, M. 2007. Development of compound microsatellite markers in the harmful red tide species *Chattonella ovata* (Raphidophyceae). *Mol Ecol Notes* 7:1251-53.

Ohlmann, J. C. & Niiler, P. P. 2005. Circulation over the continental shelf in the northern Gulf of Mexico. *Prog Oceanogr*. 64:45-81.

Parrow, M. W. & Burkholder, J. M. 2004. The sexual life cycles of *Pfiesteria piscicida* and cryptoperidiniopsoids (Dinophyceae). *J Phycol* 40:664-73.

Peakall, R. & Smouse, P. E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288-95.

Pierce, R. H., Henry, M. S., Higham, C. J., Blum, P., Sengco, M. R. & Anderson, D. M. 2004. Removal of harmful algal cells (*Karenia brevis*) and toxins from seawater culture by clay flocculation. *Harmful Algae* 3:141-48.

Pierce, R. H., Henry, M. S., Blum, P. C., Hamel, S. L., Kirkpatrick, B., Cheng, Y. S., Zhou, Y., Irvin, C. M., Naar, J., Weidner, A., Fleming, L. E., Backer, L. C. & Baden, D. G. 2005. Brevetoxin composition in water and marine aerosol along a Florida beach: assessing potential human exposure to marine biotoxins. *Harmful Algae*. 4:965-972.

Pritchard, J. K., Stephens, M. & Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-59.

Raghunathan, A, Ferguson, Jr., H. R., Bornarth, C. J., Song, W., Driscoll, M. & Lasken, R. S. 2005. Genomic DNA amplification from a single bacterium. *Environ Microbiol* 71:3342-7.

Raho, N., Pizarro, G., Escalera, L., Reguera, B. & Marin, I. 2008. Morphology, toxin composition and molecular analysis of *Dinophysis ovum* Schutt, a dinoflagellate of the "*Dinophysis acuminata* complex". *Harmful Algae* 7:839-48.

Raymond, M. & Rousset, F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J Hered* 86:248-9.

Renshaw, M. A., Soltysiak, K., Arreola, D., Loret, P., Patton, J. C., Gold, J. R. & Campbell, L. 2006. Microsatellite DNA markers for population genetic studies in the dinoflagellate *Karenia brevis*. *Mol Ecol Notes* 6:1157-9.

Rice, W. R. 1989. Analyzing tables of statistical tests. *Evolution* 43:223-5.

Richlen, M. L. & Barber, P. H. 2005. A technique for the rapid extraction of microalgal DNA from single live and preserved cells. *Mol Ecol Notes* 5:688-91.

Rizzo, P. J., Jones, M. & Ray, S. M. 1982. Isolation and properties of isolated-nuclei from the Florida red tide dinoflagellate *Gymnodinium breve* (Davis). *J Protozool* 29:217-22.

Robbins, I. C., Kirkpatrick, G. J., Blackwell, S. M., Hillier, J., Knight, C. A. & Moline, M. A. 2006. Improved monitoring of HABs using autonomous underwater vehicles (AUV). *Harmful Algae*. 5:749-761.

Rousset, F. 2008. GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol Ecol Res* 8:103-6.

Rozen, S., Skaletsky, H. 2000. Primer3 on the WWW for general users and for biologist programmers. *In* Krawetz, S., Misener, S. [eds.] *Bioinformatics Methods and Protocols*, Methods in Molecular Biology, Humana Press, Totowa, NJ, pp. 365-386.

Rynearson, T. A. & Armbrust, E. V. 2004. Genetic differentiation among populations of the planktonic marine diatom *Ditylum brightwellii* (Bacillariophyceae). *J Phycol* 40:34-43.

Schaeffer, B. A., Kamykowski, D., McKay, L., Sinclair, G. & Milligan, E. J. 2007. A comparison of photoresponse among ten different *Karenia brevis* (Dinophyceae) isolates. *J Phycol* 43:702-14.

Scholin, C. A., Herzog, M., Sogin, M. & Anderson, D. M. 1994. Identification of group-specific and strain-specific genetic-markers for globally distributed *Alexandrium* (Dinophyceae). II. Sequence-analysis of a fragment of the LSU rRNA gene. *J Phycol* 30:999-1011.

Sebastián, C.R. & O'Ryan, C. 2001. Single-cell sequencing of dinoflagellate (Dinophyceae) nuclear ribosomal genes. *Mol Ecol Notes* 1:329-31.

Selkoe, K. A. & Toonen, R. J. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol Lett* 9:615-29.

Shankle, A. M., Mayali, X. & Franks, P. J. S. 2004. Temporal patterns in population genetic diversity of *Prorocentrum micans* (Dinophyceae). *J Phycol* 40:239-47.

Smayda, T. J. 2002. Adaptive ecology, growth strategies and the global bloom expansion of dinoflagellates. *J Oceanogr* 58:281-94.

Steidinger, K. A., Vargo, G. A., Tester, P. A. & Tomas, C. R. 1998. Bloom dynamics and physiology of *Gymnodinium breve* with emphasis on the Gulf of Mexico. *In* Anderson, D. M., Cembella, A. D., Hallegraeff, G. M. [Eds.] NATO ASI Series: *Physiological Ecology of Harmful Algal Blooms*, Springer-Verlag, Berlin. 41:133-53.

Steidinger, K. A., Garcés, E. 2006. Importance of life cycles in the ecology of harmful microalgae. *In* Graneli, E., Turner, J. T. [Eds.] *Ecology of Harmful Algae*, Ecological Studies, Springer-Verlag, Berlin. pp. 37-49.

Stumpf, R. P., Litaker, R.W., Lanerolle, L. & Tester, P.A. 2008. Hydrodynamic accumulation of *Karenia* off the west coast of Florida. *Cont Shelf Res* 28:189-213.

Tester, P. A., Stumpf, R. P., Vukovich, F. M., Fowler, P. K. & Turner, J. T. 1991. An expatriate red tide bloom - transport, distribution, and persistence. *Limnol Oceanogr* 36:1053-61.

Tester, P. A. & Steidinger, K. A. 1997. *Gymnodinium breve* red tide blooms: Initiation, transport, and consequences of surface circulation. *Limnol Oceanogr* 42:1039-51.

Tittel, J., Bissinger, V., Zippel, B., Gaedke, U., Bell, E. & Lorke, A. 2003. Mixotrophs combine resource use to outcompete specialists: Implications for aquatic food webs. *Proc Natl Acad Sci USA* 100:12776-81.

Van Dolah, F. M. 2000. Marine algal toxins: Origins, health effects, and their increased occurrence. *Environ Health Persp* 108:133-41.

Walker, L. M. 1982. Evidence for a sexual cycle in the Florida red tide dinoflagellate, *Ptychodiscus brevis* (=*Gymnodinium breve*). *Trans Am Microsc Soc* 101:287-93.

Walker, L. M. & Steidinger, K. A. 1979. Sexual reproduction in the toxic dinoflagellate *Gonyaulax monilata*. *J Phycol* 15:312-15.

Walsh, J. J., Jolliff, J. K., Darrow, B. P., Lenes, J. M., Milroy, S. P., Remsem, A., Dieterle, D. A., Carder, K. L., Chen, F. R., Vargo, G. A., Weisberg, R. H., Fanning, K. A., Muller-Karger, F. E., Shinn, E., Steidinger, K. A., Heil, C. A., Tomas, C. R., Prospero, J. S., Lee, T. N., Kirkpatrick, G. J., Whitledge, T. E., Stockwell, D. A., Villareal, T. A., Jochens, A. E. & Bontempi, P. S. 2006. Red tides in the Gulf of Mexico: Where, when, and why? *J Geophys Res* 111:c11003, doi:10.1029/2004JC002813.

Whyte, J. N. C. I., Haigh, N., Ginther, N. G. & Keddy, L. J. 2001. First record of blooms of *Cochlodinium* sp (Gymnodiniales, Dinophyceae) causing mortality to aquacultured salmon on the west coast of Canada. *Phycologia* 40:298-304.

Wynne, T. T., Stumpf, R. P., Tomlinson, M. C., Ransibrahmanakul, V. & Villareal, T. A. 2005. Detecting *Karenia brevis* blooms and algal resuspension in the western Gulf of Mexico with satellite ocean color imagery. *Harmful Algae*. 4:992-1003.

Xu, H. L., Min, G. S., Choi, J. K., Zhu, M. Z., Jiang, Y. & Al-Rasheid, K. A. S. 2010. Temporal population dynamics of dinoflagellate *Prorocentrum minimum* in a semi-enclosed mariculture pond and its relationship to environmental factors and protozoan grazers. *Chin J Oceanol Limn* 28:75-81.

Yeh, F. C. & Boyle, T. J. B. 1997. Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belg J Bot* 129:157.

APPENDIX A

PYTHON PROGRAM CODE FOR ANALYZING DATA

Program for calculating Jost's D and G"ST values and testing their significance from Popgene input file. Language:Python

```
#script to read in the frequency information from a PopGen input(.txt) file
#will pass back a tuple with the frequency information for each allele
#for each locus, in each population
import tkFileDialog
import field_sample_class
import numpy
import random
import time
import math
import scipy.stats
filepathtoread = 'C:/PopGen/Kbr_FS_SC_Popgen_38smpl_reorg.txt'
pairwise_comparisons = False
global_comparison = True
num_iterations = 10000
samples = []
counter = -1
counter_alleles = 0
```

```
current_sample = []

numberofsamples = 0

locus1alleles = []

locus2alleles = []

locus3alleles = []

locus4alleles = []

locus5alleles = []

locus1alleles_count = []

locus2alleles_count = []

locus3alleles_count = []

locus4alleles_count = []

locus5alleles_count = []

locus1 = ""

locus2 = ""

locus3 = ""

locus4 = ""

locus5 = ""

datatopass = []

starttoread = 0

totalcells_sample = 0

sample_cell_counts_locus = []

pairwise_matrix = []
```

```
def jostD_pairwise_estimator(sample1, sample2, index1, index2):
    total_alleles_both = []
    total_allele_freqs_both = []
    total_ht = []
    total_hsest = []
    total_htest = []
    total_dest = []
    total_gststd = []
    total_N = []
    sample1_h = []
    sample2_h = []
    for locus in range(1,6):  #this cycles through each locus in the samples and calculates
        the hs for each
        alleles_both = []
        allele_freqs_both = []
        index = 0
        #sample1
        tempj1 = 0
        for freq in sample1[locus][0]:
            tempj1 += freq**2
            if sample1[locus][1][index] in alleles_both:
                tempindex = alleles_both.index(sample1[locus][1][index])
```

```
            allele_freqs_both[tempindex] += sample1[locus][0][index]
        else:
            alleles_both.append(sample1[locus][1][index])
            allele_freqs_both.append(sample1[locus][0][index])
        index += 1
    hj1 = 1 - tempj1


    #sample2
    index = 0
    tempj2 = 0
    for freq in sample2[locus][0]:
        tempj2 += freq**2
        if sample2[locus][1][index] in alleles_both:
            tempindex = alleles_both.index(sample2[locus][1][index])
            allele_freqs_both[tempindex] += sample2[locus][0][index]
        else:
            alleles_both.append(sample2[locus][1][index])
            allele_freqs_both.append(sample2[locus][0][index])
        index += 1
    hj2 = 1 - tempj2
    hs = (hj1 + hj2)/2.
    total_alleles_both.append(alleles_both)
```

```
total_allele_freqs_both.append(allele_freqs_both)
htfreq = 0
for freq in allele_freqs_both:
    htfreq += (freq / 2.)**2
ht = 1 - htfreq
total_ht.append(ht)
##calculating the hs_est and ht_est
#harmonic mean of the two pops
harm_mean = 2. / ((1./(sample1[locus][2]))+ (1./(sample2[locus][2])))
#estimators
hs_est = (harm_mean / (harm_mean-1)) * hs
ht_est = ht + ((hs_est)/ (harm_mean * 2))
d_est = ((ht_est - hs_est)/(1-hs_est))*2
total_dest.append(d_est)
total_hsest.append(hs_est)
total_htest.append(ht_est)
#printing the results
####putting the gst_std here
#print ht_est, hs_est
if ht_est == 0.0 or hs_est == 0.0:
    gst_std = 0.0
else:
```

```
            gst_std = ((2*(ht_est-hs_est))/((2*ht_est-hs_est)*(1-hs_est)))
        total_gststd.append(gst_std)
    ######
    #trying the average of hs_est and ht_est, then calculating the d_est to see what the
            difference is
    temp_hs_est = numpy.mean(total_hsest)
    temp_ht_est = numpy.mean(total_htest)
    temp_d_est = ((temp_ht_est-temp_hs_est)/(1-temp_hs_est))*2
    temp_gst_std = 2*(temp_ht_est-temp_hs_est)/((2*temp_ht_est-temp_hs_est)*(1-
            temp_hs_est))
    #return total_dest, total_gststd  #<---use this to return the values of d and gst averaged
            over loci
    return temp_d_est, temp_gst_std  #<---use this to return a single value of d and gst
            calculated from the averaged hs and ht from all loci
def jostD_global_estimator(samples, print_info):
    total_alleles = []
    total_allele_freqs = []
    total_ht = []
    total_hsest = []
    total_htest = []
    total_dest = []
    total_gststd = []
```

```
total_N = []
for locus in range(1,6):  #this cycles through each locus in the samples and calculates
        the hs for each
    sample_hs = []
    alleles_samples = []
    allele_freqs_samples = []
    index = 0
    for indiv_sample in samples:
        tempj1 = 0
        for freq in range(len(indiv_sample.outdata[locus][0])):
            tempj1 += indiv_sample.outdata[locus][0][freq]**2
            if indiv_sample.outdata[locus][1][freq] in alleles_samples:
                tempindex = alleles_samples.index(indiv_sample.outdata[locus][1][freq])
                allele_freqs_samples[tempindex] += indiv_sample.outdata[locus][0][freq]
            else:
                alleles_samples.append(indiv_sample.outdata[locus][1][freq])
                allele_freqs_samples.append(indiv_sample.outdata[locus][0][freq])
        hj1 = 1 - tempj1
        sample_hs.append(hj1)
    hs = sum(sample_hs)/float(len(sample_hs))
    total_alleles.append(alleles_samples)
    total_allele_freqs.append(allele_freqs_samples)
```

```
htfreq = 0
for freq in allele_freqs_samples:
    htfreq += (freq / float(len(sample_hs)))**2
ht = 1 - htfreq
total_ht.append(ht)
##calculating the hs_est and ht_est
#harmonic mean of the samples
samplesizes = []
for xval in range(len(samples)):
    samplesizes.append(samples[xval].outdata[locus][2])
harm_mean = scipy.stats.hmean(samplesizes)
#estimators
hs_est = (harm_mean / (harm_mean-1)) * hs
ht_est = ht + ((hs_est)/ (harm_mean * len(samplesizes)))
d_est = ((ht_est - hs_est)/(1-hs_est))*(len(samplesizes)/(len(samplesizes)-1))
total_dest.append(d_est)
total_hsest.append(hs_est)
total_htest.append(ht_est)
####putting the gst_std here
if ht_est == 0.0 or hs_est == 0.0:
    gst_std = 0.0
else:
```

```
            gst_std = ((len(samplesizes)*(ht_est-hs_est))/((len(samplesizes)*ht_est-
            hs_est)*(1-hs_est)))
        total_gststd.append(gst_std)
    ######
    #trying the average of hs_est and ht_est, then calculating the d_est to see what the
            difference is
    temp_hs_est = numpy.mean(total_hsest)
    temp_ht_est = numpy.mean(total_htest)
    temp_d_est = ((temp_ht_est-temp_hs_est)/(1-
            temp_hs_est))*(len(samplesizes)/(len(samplesizes)-1))
    temp_gst_std = len(samplesizes)*(temp_ht_est-
            temp_hs_est)/((len(samplesizes)*temp_ht_est-temp_hs_est)*(1-temp_hs_est))
    if print_info == 0:
        print temp_hs_est, total_hsest
        print temp_ht_est, total_htest
    #return total_dest, total_gststd  #<---use this to return the values of d and gst averaged
            over loci
    return temp_d_est, temp_gst_std  #<---use this to return a single value of d and gst
            calculated from the averaged hs and ht from all loci
def jostd_calculator(samples):
    values = []
    vals_gst = []
```

```
    for sample1 in range(len(samples)):
        for sample2 in range(sample1, len(samples)):
            jost, gst = jostD_pairwise_estimator(samples[sample1], samples[sample2],
            sample1, sample2)
            values.append(jost)
            vals_gst.append(gst)
    return values, vals_gst
def jostd_calculator_permutation(samples, pvalues, original_values, gst_pvalues,
            gst_original_values, num_iterations):
    timetemp = time.time()
    values = 0
    for sample1 in range(len(samples)):
        print "Sample:", sample1+1
        for sample2 in range(sample1, len(samples)):
            for iteration in range(num_iterations):
                temp_samples = create_random_permutation(samples[sample1],
            samples[sample2])
                jost, gst = jostD_pairwise_estimator(temp_samples[0], temp_samples[1],
            sample1, sample2)
                if jost < 0.0:
                    jost = 0.0
                if gst < 0.0:
```

```
                gst = 0.0
            if jost >= original_values[values]:
                pvalues[values] += 1
            if gst >= gst_original_values[values]:
                gst_pvalues[values] += 1
        values += 1
    print "It took", round(time.time()-timetemp, 2), "seconds to run this sample"
    timetemp = time.time()
def create_random_permutation(sample1, sample2):
    #temp_samples = randomize_cells_in_samples_within_sample([sample1.cells,
        sample2.cells]) #use this to sample with replacement from the same sample
    temp_samples = randomize_cells_in_samples_pairwise([sample1.cells,
        sample2.cells]) #use this to shuffle cells between the two samples
    temp_sample1 = field_sample_class.field_sample()
    temp_sample1.initialize(temp_samples[0])
    temp_sample2 = field_sample_class.field_sample()
    temp_sample2.initialize(temp_samples[1])
    return [temp_sample1.outdata, temp_sample2.outdata]
def create_random_permutation_global(input_samples):
    temp_samples = input_samples[:]
    #temp_samples = randomize_cells_in_samples_within_sample([sample1.cells,
        sample2.cells]) #use this to sample with replacement from the same sample
```

```
    temp_samples = randomize_cells_in_samples_global(temp_samples) #use this to
        shuffle cells between the two samples
    outsamples = []
    for smpl in temp_samples:
        temp_sample1 = field_sample_class.field_sample()
        temp_sample1.initialize(smpl)
        outsamples.append(temp_sample1)
    return outsamples
def create_random_permutation_global_first(input_samples):
    temp_samples = input_samples[:]
    #temp_samples = randomize_cells_in_samples_within_sample([sample1.cells,
        sample2.cells]) #use this to sample with replacement from the same sample
    temp_samples = randomize_cells_in_samples_global_first(temp_samples) #use this to
        shuffle cells between the two samples
    outsamples = []
    for smpl in temp_samples:
        temp_sample1 = field_sample_class.field_sample()
        temp_sample1.initialize(smpl)
        outsamples.append(temp_sample1)
    return outsamples
def frequency_printer(samples):
    datafilefreqs = open('c:/cjunk/popgen_freq_reader_freqs.txt', 'w')
```

```
    for sample1 in range(len(samples)):
        datafilefreqs.write('Sample ')
        datafilefreqs.write(str(sample1+1))
        datafilefreqs.write('\n')
        for locus in range(1,5):
            datafilefreqs.write('Locus\n')
            for allele in samples[sample1][locus]:
                datafilefreqs.write(str(allele))
                datafilefreqs.write('\n')
    datafilefreqs.close()
def frequencygenerator(allele_count):
    locus_total = 0.0
    tempfreqs = []
    for count in allele_count:
        locus_total += count
    for count in allele_count:
        tempfreqs.append(round(count/locus_total, 4))
    return tempfreqs
def locus_counter(allele_count):
    locus_total = 0
    for count in allele_count:
        locus_total += count
```

```
    return locus_total
def write_inputdata(datatopass):
    global samples
    [totalcells_sample, locus1alleles_count, locus2alleles_count, locus3alleles_count,
                    locus4alleles_count, locus5alleles_count] = datatopass
    if numberofsamples > 1:
        locus1alleles_freqs = frequencygenerator(locus1alleles_count)
        locus2alleles_freqs = frequencygenerator(locus2alleles_count)
        locus3alleles_freqs = frequencygenerator(locus3alleles_count)
        locus4alleles_freqs = frequencygenerator(locus4alleles_count)
        locus5alleles_freqs = frequencygenerator(locus5alleles_count)
        sample_cell_counts_locus.append([sum(locus1alleles_count),
            sum(locus2alleles_count), sum(locus3alleles_count), sum(locus4alleles_count),
            sum(locus5alleles_count)])
        current_sample = []
        current_sample.append(totalcells_sample)
        locus = []
        locus.append(locus1alleles_freqs)
        locus.append(locus1alleles)
        locus.append(locus_counter(locus1alleles_count))
        current_sample.append(locus)
        locus = []
```

```
    locus.append(locus2alleles_freqs)
    locus.append(locus2alleles)
    locus.append(locus_counter(locus2alleles_count))
    current_sample.append(locus)
    locus = []
    locus.append(locus3alleles_freqs)
    locus.append(locus3alleles)
    locus.append(locus_counter(locus3alleles_count))
    current_sample.append(locus)
    locus = []
    locus.append(locus4alleles_freqs)
    locus.append(locus4alleles)
    locus.append(locus_counter(locus4alleles_count))
    current_sample.append(locus)
    locus = []
    locus.append(locus5alleles_freqs)
    locus.append(locus5alleles)
    locus.append(locus_counter(locus5alleles_count))
    current_sample.append(locus)
    samples.append(current_sample)
datatopass[0]= 0
for data in range(5):
```

```
        datatopass[data+1] = []
    return datatopass
def read_the_input_file():
    global starttoread
    global numberofsamples
    global totalcells_sample
    global locus1alleles_count
    global locus2alleles_count
    global locus3alleles_count
    global locus4alleles_count
    global locus5alleles_count
    global locus1alleles
    global locus2alleles
    global locus3alleles
    global locus4alleles
    global locus5alleles
    global samples

    ####begin reading below here
    filetoread = open(filepathtoread)
    for line in filetoread:
        if line[:-1] == "":
```

```
    starttoread = 0
if "ID =" in line:
    numberofsamples += 1
    continue
    #pass
if 'Name =' in line:
    nameline = 1
elif 'Name =' not in line:
    nameline = 0
if starttoread == 1:
    countcell = 0
    locus1 = line[0:1]
    if locus1 == "." or locus1 == " " or locus1 == "0" or locus1 == " or line[1:2]=='a':
        pass
    elif locus1 in locus1alleles:
        tempindex = locus1alleles.index(locus1)
        locus1alleles_count[tempindex] += 1
        countcell += 1
    elif locus1 not in locus1alleles and nameline == 0:
        locus1alleles.append(locus1)
        locus1alleles_count.append(1)
        countcell += 1
```

```
locus2 = line[8:9]
if locus2 == "." or locus2 == " " or locus2 == "0" or locus2 == '':
    pass
elif locus2 in locus2alleles:
    tempindex = locus2alleles.index(locus2)
    locus2alleles_count[tempindex] += 1
    countcell += 1
elif locus2 not in locus2alleles and nameline == 0:
    locus2alleles.append(locus2)
    locus2alleles_count.append(1)
    countcell += 1
locus3 = line[16:17]
if locus3 == "." or locus3 == " " or locus3 == "0" or locus3 == '':
    pass
elif locus3 in locus3alleles:
    tempindex = locus3alleles.index(locus3)
    locus3alleles_count[tempindex] += 1
    countcell += 1
elif locus3 not in locus3alleles and nameline == 0:
    locus3alleles.append(locus3)
    locus3alleles_count.append(1)
    countcell += 1
```

```
locus4 = line[24:25]
if locus4 == "." or locus4 == " " or locus4 == "0" or locus4 == '':
    pass
elif locus4 in locus4alleles:
    tempindex = locus4alleles.index(locus4)
    locus4alleles_count[tempindex] += 1
    countcell += 1
elif locus4 not in locus4alleles and nameline == 0:
    locus4alleles.append(locus4)
    locus4alleles_count.append(1)
    countcell += 1
locus5 = line[32:33]
if locus5 == "." or locus5 == " " or locus5 == "0" or locus5 == '':
    pass
elif locus5 in locus5alleles:
    tempindex = locus5alleles.index(locus5)
    locus5alleles_count[tempindex] += 1
    countcell += 1
elif locus5 not in locus5alleles and nameline == 0:
    locus5alleles.append(locus5)
    locus5alleles_count.append(1)
    countcell += 1
```

```
            if countcell > 0:
                totalcells_sample += 1
        if "Name =" in line:
            starttoread = 1
            datatopass = [totalcells_sample,
                          locus1alleles_count, locus2alleles_count, locus3alleles_count,
                          locus4alleles_count, locus5alleles_count]
            [totalcells_sample,
                          locus1alleles_count, locus2alleles_count, locus3alleles_count,
                          locus4alleles_count, locus5alleles_count]=write_inputdata(datatopass)
            locus1alleles = []
            locus2alleles = []
            locus3alleles = []
            locus4alleles = []
            locus5alleles = []
    outputdata = write_inputdata(datatopass)
    filetoread.close()
    return outputdata
def randomize_cells_in_samples_within_sample(samples): #use this one to resample
            cells from their original sample
        length_of_samples = []
        total_pool_of_cells = []
```

```
    samples_to_return = []
    for x in range(len(samples)):
        total_pool_of_cells = []
        length_of_samples.append(len(samples[x]))
        for cell in range(len(samples[x])):
            total_pool_of_cells.append(copy.deepcopy(samples[x][cell]))
        random.shuffle(total_pool_of_cells)
        temp_sample = []
        for y in range(len(samples[x])):
            temp_sample.append(total_pool_of_cells[random.randint(0,
      len(samples[x])-1)])
        samples_to_return.append(temp_sample)
    return samples_to_return
def randomize_cells_in_samples_pairwise(samples):  #use this one to shuffle cells
      among samples
    length_of_samples = []
    total_pool_of_cells = []
    samples_to_return = []
    samplesf = samples[:]
    for x in range(len(samplesf)):
        length_of_samples.append(len(samplesf[x]))
        for cell in range(len(samplesf[x])):
```

```
                total_pool_of_cells.append(samplesf[x][cell])
        random.shuffle(total_pool_of_cells)
        for x in range(len(length_of_samples)):
            temp_sample = []
            for y in range(length_of_samples[x]):
                temp_sample.append(total_pool_of_cells.pop())
            samples_to_return.append(temp_sample)
        return samples_to_return
def randomize_cells_in_samples_global(samples):  #use this one to shuffle cells among
    samples
        length_of_samples = []
        total_pool_of_cells = []
        samples_to_return = []
        samplesf = samples[:]
        for x in range(len(samplesf)):
            length_of_samples.append(len(samplesf[x].cells))
            for cell in range(len(samplesf[x].cells)):
                total_pool_of_cells.append(samplesf[x].cells[cell])
        random.shuffle(total_pool_of_cells)
        for x in range(len(length_of_samples)):
            temp_sample = []
            for y in range(length_of_samples[x]):
```

```
                temp_sample.append(total_pool_of_cells.pop())
            samples_to_return.append(temp_sample)
        return samples_to_return
def randomize_cells_in_samples_global_first(samples):  #use this one to shuffle cells
    among samples
        length_of_samples = []
        total_pool_of_cells = []
        samples_to_return = []
        samplesf = samples[:]
        for x in range(len(samplesf)):
            length_of_samples.append(len(samplesf[x].cells))
            for cell in range(len(samplesf[x].cells)):
                total_pool_of_cells.append(samplesf[x].cells[cell])
        total_pool_of_cells.reverse()
        for x in range(len(length_of_samples)):
            temp_sample = []
            for y in range(length_of_samples[x]):
                temp_sample.append(total_pool_of_cells.pop())
            samples_to_return.append(temp_sample)
        return samples_to_return
def get_input_file_data():
    temp_file = []
```

```
    f = open(filepathtoread)
    for line in f:
        temp_file.append(line)
    f.close()
    temp_samples = []
    sample_number = -1
    for line in temp_file[5:]:
        if line[:2] == 'ID':# or line[:2] == 'Na':
            temp_samples.append([])
            sample_number += 1
        elif line == '\n' or line[:2] == 'Na' or line == ' \n':
            pass
        else:
            temp_samples[sample_number].append(line)
    return temp_samples
def main():
    #added the jost's d calculator here to calculate this statistic for the samples
    read_the_input_file()
    print 'samples:', len(samples)
    num_samples = len(samples)
    original_values, gst_original_values = jostd_calculator(samples)
```

```
num_pvalues = num_samples + (math.factorial(num_samples) / (math.factorial(1) *
      math.factorial(num_samples-1)))
pvalues =     [0]* (num_pvalues + 1)
gst_pvalues = [0]* (num_pvalues + 1)
sample_class = []
orig_sample_data = get_input_file_data()
time1 = time.time()
for indiv_sample in orig_sample_data:
   temp_class = field_sample_class.field_sample()
   temp_class.initialize(indiv_sample)
   sample_class.append(temp_class)
#do the global comparison if marked true above
if global_comparison == True:
   orig_d_val = 0
   orig_gst_val = 0
   glob_smpls = sample_class[:]
   firstsample = create_random_permutation_global_first(glob_smpls)
   tempreturn_vals = jostD_global_estimator(firstsample,0)
   orig_d_val = tempreturn_vals[0]
   orig_gst_val = tempreturn_vals[1]
   dest_vals_permutation = []
   gst_vals_permutation = []
```

```
for next_file in range(num_iterations):
    permuted_samples = create_random_permutation_global(glob_smpls[:])
    tempreturn_vals = jostD_global_estimator(permuted_samples,1)
    dest_vals_permutation.append(tempreturn_vals[0])
    gst_vals_permutation.append(tempreturn_vals[1])
print orig_d_val
print orig_gst_val
global_dest_pvalue = 0
global_gst_pvalue = 0
print max(dest_vals_permutation)
print max(gst_vals_permutation)
for destval in dest_vals_permutation:
    if destval >= orig_d_val and orig_d_val >= 0:
        global_dest_pvalue += 1
    elif orig_d_val < 0:
        global_dest_pvalue += 1
print global_dest_pvalue
global_dest_pvalue /= float(num_iterations)
for gstval in gst_vals_permutation:
    if gstval >= orig_gst_val and orig_gst_val >=0:
        global_gst_pvalue += 1
    elif orig_gst_val < 0:
```

```
            global_gst_pvalue += 1
    print global_gst_pvalue
    global_gst_pvalue /= float(num_iterations)
    print "Globals:"
    print "Jost's D: ", round(orig_d_val, 4), "    pvalue:", global_dest_pvalue
    print '    G"st: ', round(orig_gst_val, 4), "    pvalue:", global_gst_pvalue
#do the pairwise comparisons if marked true above
if pairwise_comparisons == True:
    for next_file in range(1):
        if next_file % 100 == 0:
            print 'Permutation', next_file
        jostd_calculator_permutation(sample_class, pvalues, original_values,
        gst_pvalues, gst_original_values, num_iterations)
corr_pvalues = []
for val in range(len(pvalues)):
    corr_pvalues.append((pvalues[val])/float(num_iterations))
gst_corr_pvalues = []
for val in range(len(gst_pvalues)):
    gst_corr_pvalues.append((gst_pvalues[val])/float(num_iterations))
print time.time()-time1, 'seconds'
######################
#print corr_pvalues
```

```
smp1 = 1
smp2 = 1
for val in range(len(original_values)):
    if smp2>num_samples:
        smp1 +=1
        smp2 = smp1
    smp2 += 1
####################
#create CSV files with the data aligned in a matrix
f = open('C:/CJunk/junk_files/output_for_excel.csv', 'w')
smp1 = 1
smp2 = 1
temp_values = ['']
temp_pvalues = ['']
for val in range(len(original_values)):
    temp_values.append('')
    temp_pvalues.append('')
    comma_counter = 0
    while comma_counter < val:
        temp_values[-1] += ','
        temp_pvalues[-1] += ','
        comma_counter += 1
```

```
    if smp2>num_samples:
        smp1 +=1
        smp2 = smp1
    temp_values[smp1] += str(original_values[val]) + ','
    temp_pvalues[smp1] += str(corr_pvalues[val]) + ','
    smp2 += 1
######
#check for lines containing only commas
for lines in range(len(temp_values)-1, -1, -1):
    commas = 0
    for letter in temp_values[lines]:
        if letter != ',':
            commas = 1
    if commas == 0:
        del temp_values[lines]
for lines in range(len(temp_pvalues)-1, -1, -1):
    commas = 0
    for letter in temp_pvalues[lines]:
        if letter != ',':
            commas = 1
    if commas == 0:
        del temp_pvalues[lines]
```

```
for lines in range(len(temp_values)):
    f.write(temp_values[lines])
    f.write('\n')
f.write('\n\n')
for lines in range(len(temp_pvalues)):
    f.write(temp_pvalues[lines])
    f.write('\n')
f.write('\n')
for lines in range(len(temp_pvalues)):
    f.write(temp_pvalues[lines])
f.write('\n')
f.write('\n')
#now the gst values
smp1 = 1
smp2 = 1
temp_values = ['']
temp_pvalues = ['']
for val in range(len(gst_original_values)):
    temp_values.append('')
    temp_pvalues.append('')
    comma_counter = 0
    while comma_counter < val:
```

```
            temp_values[-1] += ','
            temp_pvalues[-1] += ','
            comma_counter += 1
        if smp2>num_samples:
            smp1 +=1
            smp2 = smp1
        temp_values[smp1] += str(gst_original_values[val]) + ','
        temp_pvalues[smp1] += str(gst_corr_pvalues[val]) + ','
        smp2 += 1
######
#check for lines containing only commas
for lines in range(len(temp_values)-1, -1, -1):
    commas = 0
    for letter in temp_values[lines]:
        if letter != ',':
            commas = 1
    if commas == 0:
        del temp_values[lines]
for lines in range(len(temp_pvalues)-1, -1, -1):
    commas = 0
    for letter in temp_pvalues[lines]:
        if letter != ',':
```

```
                commas = 1
        if commas == 0:
            del temp_pvalues[lines]
    for lines in range(len(temp_values)):
        f.write(temp_values[lines])
        f.write('\n')
    f.write('\n\n')
    for lines in range(len(temp_pvalues)):
        f.write(temp_pvalues[lines])
        f.write('\n')
    for lines in range(len(temp_pvalues)):
        f.write(temp_pvalues[lines])
    f.write('\n')
    f.close()
main()
```

The field sample class library for use with the above script. Language: Python

```
import random
import allelic_richness
class field_sample:
    def __init__(self):
```

```
        self.cells = []
        self.numberofcells = len(self.cells)
        self.locus1alleles = []
        self.locus2alleles = []
        self.locus3alleles = []
        self.locus4alleles = []
        self.locus5alleles = []
        self.locus1alleles_count = []
        self.locus2alleles_count = []
        self.locus3alleles_count = []
        self.locus4alleles_count = []
        self.locus5alleles_count = []
    def initialize(self, sample_cells):
        for cell in sample_cells:
            self.cells.append(cell)
        self.create_values()
        self.calc_gene_diversity()
    def calc_gene_diversity(self):
        self.locus1_diversity, self.locus1_eff_num_alleles =
            self.genediversity(self.locus1alleles_freqs)
        self.locus2_diversity, self.locus2_eff_num_alleles =
            self.genediversity(self.locus2alleles_freqs)
```

```
        self.locus3_diversity, self.locus3_eff_num_alleles =
            self.genediversity(self.locus3alleles_freqs)
        self.locus4_diversity, self.locus4_eff_num_alleles =
            self.genediversity(self.locus4alleles_freqs)
        self.locus5_diversity, self.locus5_eff_num_alleles =
            self.genediversity(self.locus5alleles_freqs)
    def frequencygenerator(self, allele_count):
        locus_total = 0.0
        tempfreqs = []
        for count in allele_count:
            locus_total += count
        #locus_total = sum(allele_count)
        for count in allele_count:
            tempfreqs.append(round(count/locus_total, 4))
        return tempfreqs
    def locus_counter(self, allele_count):
        locus_total = 0
        for count in allele_count:
            locus_total += count
        return locus_total
    def write_inputdata(self, datatopass):
        [totalcells_sample, locus1alleles_count, locus2alleles_count, locus3alleles_count,
```

```
                locus4alleles_count, locus5alleles_count] = datatopass
        self.locus1alleles_freqs = self.frequencygenerator(locus1alleles_count)
        self.locus2alleles_freqs = self.frequencygenerator(locus2alleles_count)
        self.locus3alleles_freqs = self.frequencygenerator(locus3alleles_count)
        self.locus4alleles_freqs = self.frequencygenerator(locus4alleles_count)
        self.locus5alleles_freqs = self.frequencygenerator(locus5alleles_count)
        #sample_cell_counts_locus.append([sum(locus1alleles_count),
           sum(locus1alleles_count), sum(locus1alleles_count), sum(locus1alleles_count),
           sum(locus1alleles_count)])
        current_sample = []
        current_sample.append(totalcells_sample)
        locus = []
        locus.append(self.locus1alleles_freqs)
        locus.append(self.locus1alleles)
        locus.append(self.locus_counter(locus1alleles_count))
        current_sample.append(locus)
        locus = []
        locus.append(self.locus2alleles_freqs)
        locus.append(self.locus2alleles)
        locus.append(self.locus_counter(locus2alleles_count))
        current_sample.append(locus)
        locus = []
```

```
        locus.append(self.locus3alleles_freqs)
        locus.append(self.locus3alleles)
        locus.append(self.locus_counter(locus3alleles_count))
        current_sample.append(locus)
        locus = []
        locus.append(self.locus4alleles_freqs)
        locus.append(self.locus4alleles)
        locus.append(self.locus_counter(locus4alleles_count))
        current_sample.append(locus)
        locus = []
        locus.append(self.locus5alleles_freqs)
        locus.append(self.locus5alleles)
        locus.append(self.locus_counter(locus5alleles_count))
        current_sample.append(locus)
        self.outdata = current_sample[:]
    def create_values(self):
        totalcells_sample = 0
        for line in self.cells:
            countcell = 0
            locus1 = line[0:1]
            if locus1 == "." or locus1 == " " or locus1 == "0" or locus1 == " or line[1:2]=='a':
                pass
```

```
elif locus1 in self.locus1alleles:
    tempindex = self.locus1alleles.index(locus1)
    self.locus1alleles_count[tempindex] += 1
    countcell += 1
elif locus1 not in self.locus1alleles:
    self.locus1alleles.append(locus1)
    self.locus1alleles_count.append(1)
    countcell += 1
locus2 = line[8:9]
if locus2 == "." or locus2 == " " or locus2 == "0" or locus2 == '':
    pass
elif locus2 in self.locus2alleles:
    tempindex = self.locus2alleles.index(locus2)
    self.locus2alleles_count[tempindex] += 1
    countcell += 1
elif locus2 not in self.locus2alleles:
    self.locus2alleles.append(locus2)
    self.locus2alleles_count.append(1)
    countcell += 1
locus3 = line[16:17]
if locus3 == "." or locus3 == " " or locus3 == "0" or locus3 == '':
    pass
```

```
elif locus3 in self.locus3alleles:
    tempindex = self.locus3alleles.index(locus3)
    self.locus3alleles_count[tempindex] += 1
    countcell += 1
elif locus3 not in self.locus3alleles:
    self.locus3alleles.append(locus3)
    self.locus3alleles_count.append(1)
    countcell += 1
locus4 = line[24:25]
if locus4 == "." or locus4 == " " or locus4 == "0" or locus4 == '':
    pass
elif locus4 in self.locus4alleles:
    tempindex = self.locus4alleles.index(locus4)
    self.locus4alleles_count[tempindex] += 1
    countcell += 1
elif locus4 not in self.locus4alleles:
    self.locus4alleles.append(locus4)
    self.locus4alleles_count.append(1)
    countcell += 1
locus5 = line[32:33]
if locus5 == "." or locus5 == " " or locus5 == "0" or locus5 == '':
    pass
```

```
        elif locus5 in self.locus5alleles:
            tempindex = self.locus5alleles.index(locus5)
            self.locus5alleles_count[tempindex] += 1
            countcell += 1
        elif locus5 not in self.locus5alleles:
            self.locus5alleles.append(locus5)
            self.locus5alleles_count.append(1)
            countcell += 1
        if countcell > 0:
            totalcells_sample += 1
    datatopass = [totalcells_sample,
                  self.locus1alleles_count, self.locus2alleles_count,
        self.locus3alleles_count,
                  self.locus4alleles_count, self.locus5alleles_count]
    self.write_inputdata(datatopass)
    self.numberofcells = totalcells_sample
def genediversity(self, allele_freqs):
    Na = len(allele_freqs)
    totalallelefreqs = 0
    for allele in allele_freqs:
        totalallelefreqs += allele**2
    genediversity = 1.-totalallelefreqs
```

```
        if totalallelefreqs > 0:
            eff_num_alleles = 1./totalallelefreqs
        else:
            eff_num_alleles = 0.
        return genediversity, eff_num_alleles
class field_sample_culture:
    def __init__(self):
        self.cells = []
        self.numberofcells = len(self.cells)
        self.locus1alleles = []
        self.locus2alleles = []
        self.locus3alleles = []
        self.locus4alleles = []
        self.locus5alleles = []
        self.locus6alleles = []
        self.locus7alleles = []
        self.locus8alleles = []
        self.locus9alleles = []
        self.locus10alleles = []
        self.locus11alleles = []
        self.locus12alleles = []
        self.locus13alleles = []
```

```
self.locus14alleles = []
self.locus15alleles = []
self.locus16alleles = []
self.locus17alleles = []
self.locus18alleles = []
self.locus1alleles_count = []
self.locus2alleles_count = []
self.locus3alleles_count = []
self.locus4alleles_count = []
self.locus5alleles_count = []
self.locus6alleles_count = []
self.locus7alleles_count = []
self.locus8alleles_count = []
self.locus9alleles_count = []
self.locus10alleles_count = []
self.locus11alleles_count = []
self.locus12alleles_count = []
self.locus13alleles_count = []
self.locus14alleles_count = []
self.locus15alleles_count = []
self.locus16alleles_count = []
self.locus17alleles_count = []
```

```
    self.locus18alleles_count = []
def initialize(self, sample_cells):
    for cell in sample_cells:
        self.cells.append(cell)
    self.create_values()
    self.calc_gene_diversity()
def calc_gene_diversity(self):
    self.locus1_diversity, self.locus1_eff_num_alleles =
        self.genediversity(self.locus1alleles_freqs)
    self.locus2_diversity, self.locus2_eff_num_alleles =
        self.genediversity(self.locus2alleles_freqs)
    self.locus3_diversity, self.locus3_eff_num_alleles =
        self.genediversity(self.locus3alleles_freqs)
    self.locus4_diversity, self.locus4_eff_num_alleles =
        self.genediversity(self.locus4alleles_freqs)
    self.locus5_diversity, self.locus5_eff_num_alleles =
        self.genediversity(self.locus5alleles_freqs)
    self.locus6_diversity, self.locus6_eff_num_alleles =
        self.genediversity(self.locus6alleles_freqs)
    self.locus7_diversity, self.locus7_eff_num_alleles =
        self.genediversity(self.locus7alleles_freqs)
```

```
self.locus8_diversity, self.locus8_eff_num_alleles =
    self.genediversity(self.locus8alleles_freqs)
self.locus9_diversity, self.locus9_eff_num_alleles =
    self.genediversity(self.locus9alleles_freqs)
self.locus10_diversity, self.locus10_eff_num_alleles =
    self.genediversity(self.locus10alleles_freqs)
self.locus11_diversity, self.locus11_eff_num_alleles =
    self.genediversity(self.locus11alleles_freqs)
self.locus12_diversity, self.locus12_eff_num_alleles =
    self.genediversity(self.locus12alleles_freqs)
self.locus13_diversity, self.locus13_eff_num_alleles =
    self.genediversity(self.locus13alleles_freqs)
self.locus14_diversity, self.locus14_eff_num_alleles =
    self.genediversity(self.locus14alleles_freqs)
self.locus15_diversity, self.locus15_eff_num_alleles =
    self.genediversity(self.locus15alleles_freqs)
self.locus16_diversity, self.locus16_eff_num_alleles =
    self.genediversity(self.locus16alleles_freqs)
self.locus17_diversity, self.locus17_eff_num_alleles =
    self.genediversity(self.locus17alleles_freqs)
self.locus18_diversity, self.locus18_eff_num_alleles =
    self.genediversity(self.locus18alleles_freqs)
```

```
def frequencygenerator(self, allele_count):
    locus_total = 0.0
    tempfreqs = []
    for count in allele_count:
        locus_total += count
    #locus_total = sum(allele_count)
    for count in allele_count:
        tempfreqs.append(round(count/locus_total, 4))
    return tempfreqs
def locus_counter(self, allele_count):
    locus_total = 0
    for count in allele_count:
        locus_total += count
    return locus_total
def write_inputdata(self, datatopass):
    [totalcells_sample, locus1alleles_count, locus2alleles_count, locus3alleles_count,
                    locus4alleles_count, locus5alleles_count] = datatopass
    self.locus1alleles_freqs = self.frequencygenerator(locus1alleles_count)
    self.locus2alleles_freqs = self.frequencygenerator(locus2alleles_count)
    self.locus3alleles_freqs = self.frequencygenerator(locus3alleles_count)
    self.locus4alleles_freqs = self.frequencygenerator(locus4alleles_count)
    self.locus5alleles_freqs = self.frequencygenerator(locus5alleles_count)
```

```
#sample_cell_counts_locus.append([sum(locus1alleles_count),
    sum(locus1alleles_count), sum(locus1alleles_count), sum(locus1alleles_count),
    sum(locus1alleles_count)])
current_sample = []
current_sample.append(totalcells_sample)
locus = []
locus.append(self.locus1alleles_freqs)
locus.append(self.locus1alleles)
locus.append(self.locus_counter(locus1alleles_count))
current_sample.append(locus)
locus = []
locus.append(self.locus2alleles_freqs)
locus.append(self.locus2alleles)
locus.append(self.locus_counter(locus2alleles_count))
current_sample.append(locus)
locus = []
locus.append(self.locus3alleles_freqs)
locus.append(self.locus3alleles)
locus.append(self.locus_counter(locus3alleles_count))
current_sample.append(locus)
locus = []
locus.append(self.locus4alleles_freqs)
```

```
        locus.append(self.locus4alleles)
        locus.append(self.locus_counter(locus4alleles_count))
        current_sample.append(locus)
        locus = []
        locus.append(self.locus5alleles_freqs)
        locus.append(self.locus5alleles)
        locus.append(self.locus_counter(locus5alleles_count))
        current_sample.append(locus)
        self.outdata = current_sample[:]
    def create_values(self):
        totalcells_sample = 0
        for line in self.cells:
            countcell = 0
            locus1 = line[0:1]
            if locus1 == "." or locus1 == " " or locus1 == "0" or locus1 == " or line[1:2]=='a':
                pass
            elif locus1 in self.locus1alleles:
                tempindex = self.locus1alleles.index(locus1)
                self.locus1alleles_count[tempindex] += 1
                countcell += 1
            elif locus1 not in self.locus1alleles:
                self.locus1alleles.append(locus1)
```

```
    self.locus1alleles_count.append(1)
    countcell += 1
locus2 = line[8:9]
if locus2 == "." or locus2 == " " or locus2 == "0" or locus2 == '':
    pass
elif locus2 in self.locus2alleles:
    tempindex = self.locus2alleles.index(locus2)
    self.locus2alleles_count[tempindex] += 1
    countcell += 1
elif locus2 not in self.locus2alleles:
    self.locus2alleles.append(locus2)
    self.locus2alleles_count.append(1)
    countcell += 1
locus3 = line[16:17]
if locus3 == "." or locus3 == " " or locus3 == "0" or locus3 == '':
    pass
elif locus3 in self.locus3alleles:
    tempindex = self.locus3alleles.index(locus3)
    self.locus3alleles_count[tempindex] += 1
    countcell += 1
elif locus3 not in self.locus3alleles:
    self.locus3alleles.append(locus3)
```

```
    self.locus3alleles_count.append(1)
    countcell += 1
locus4 = line[24:25]
if locus4 == "." or locus4 == " " or locus4 == "0" or locus4 == '':
    pass
elif locus4 in self.locus4alleles:
    tempindex = self.locus4alleles.index(locus4)
    self.locus4alleles_count[tempindex] += 1
    countcell += 1
elif locus4 not in self.locus4alleles:
    self.locus4alleles.append(locus4)
    self.locus4alleles_count.append(1)
    countcell += 1
locus5 = line[32:33]
if locus5 == "." or locus5 == " " or locus5 == "0" or locus5 == '':
    pass
elif locus5 in self.locus5alleles:
    tempindex = self.locus5alleles.index(locus5)
    self.locus5alleles_count[tempindex] += 1
    countcell += 1
elif locus5 not in self.locus5alleles:
    self.locus5alleles.append(locus5)
```

```
            self.locus5alleles_count.append(1)
            countcell += 1
        if countcell > 0:
            totalcells_sample += 1
    datatopass = [totalcells_sample,
                  self.locus1alleles_count, self.locus2alleles_count,
        self.locus3alleles_count,
                  self.locus4alleles_count, self.locus5alleles_count]
    self.write_inputdata(datatopass)
    self.numberofcells = totalcells_sample
def genediversity(self, allele_freqs):
    Na = len(allele_freqs)
    totalallelefreqs = 0
    for allele in allele_freqs:
        totalallelefreqs += allele**2
    genediversity = 1.-totalallelefreqs
    if totalallelefreqs > 0:
        eff_num_alleles = 1./totalallelefreqs
    else:
        eff_num_alleles = 0.
    return genediversity, eff_num_alleles
```