ABSTRACT

Title of Document: SEX, DISPERSAL, AND DEEP DIVERGENCE: THE POPULATION GENETICS OF THE PROTISTAN PARASITE *PERKINSUS MARINUS*.

Peter C. Thompson, Ph.D, 2010

Directed By: Associate Professor Matthew P. Hare,
Department of Biology

Dr. Benjamin M. Rosenthal, Zoologist,
Agricultural Research Service, USDA

The protistan parasite *Perkinsus marinus* causes Dermo disease in the eastern oyster, *Crassostrea virginica*. This parasite causes reduced growth and fecundity in its ecologically and economically important host, and as such has become a focal point for shellfish research. Though much is known regarding the seasonal dynamics and interactions between host and parasite, little research has focused on the basic biology of this parasite. In the research presented here, I used population genetic approaches to investigate the reproduction, dispersal, and origins of extant populations of *P. marinus*. First, I determined the extent of clonality in *P. marinus* populations. Repeated sampling of the same multilocus genotypes and extensive multilocus linkage disequilibrium indicated that clonal reproduction is prevalent. However, genotypic diversity was great and recombination occurred between genetic loci, supporting sexual reproduction as an important source of new genetic variation in *P. marinus*. An interesting consequence of sexual reproduction is that genotypic correlations may be maintained through inbreeding when sex occurs. Next, I investigated the genetic connectivity among locations. Clustering analyses revealed that local geographic samples are collections of independent clonal lineages rather than freely interbreeding populations. Some lineages were widespread while others were found at high frequencies only in specific locations indicating that *P. marinus* has a high capacity for dispersal, but local conditions may determine the success of certain lineages. Finally, I examined an interesting pattern of di-allelism observed in *P. marinus* DNA sequences. Two allelic classes were discovered at six out of seven nuclear loci where large divergences indicated the alleles had been

independent for possibly millions of years. Balancing selection may be responsible for the retention of ancient diversity in this parasite, but it seems more likely that a recent hybridization event has occurred between two formerly allopatric lineages. These results underscore risks in the anthropogenic movement of protistan parasites as there may be no reproductive barriers between ancient lineages. Resulting hybrids could result in increased parasite virulence with increased disease in host populations.

SEX, DISPERSAL, AND DEEP DIVERGENCE: THE POPULATION GENETICS OF THE PROTISTAN PARASITE *PERKINSUS MARINUS*.

By

Peter C. Thompson

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2010

Advisory Committee:
Associate Professor Matthew P. Hare, Chair
Dr. Benjamin M. Rosenthal, Co-advisor
Associate Professor Michael P. Cummings
Associate Professor Eric S. Haag
Associate Professor David J. Hawthorne, Dean's Representative
Associate Research Professor Kennedy T. Paynter

Preface

This dissertation contains a single introductory chapter, three research chapters, and a concluding chapter. Chapters 2, 3, and 4 are presented in manuscript form with abstract, introduction, methods, results, and discussion, followed by tables, figure legends, and figures. Tables and Figures beginning with an "S" are intended as supplementary information, not to be included in published manuscripts. A single bibliography is provided at the end for literature cited throughout the dissertation.

Dedication

For my loving wife Anne and the Furry Beasts who added their pawprints

Acknowledgements

This work would not have been possible without the help and cooperation of numerous individuals and organizations.

I thank my advisor Dr. Matt Hare for giving me this opportunity to apply my molecular biology skills in a new and interesting arena. His wealth of population genetics knowledge and insightful mind greatly improved my research and writing abilities and pushed me to become more attentive to the details that result in good science.

I am deeply indebted to my co-advisor Dr. Ben Rosenthal. I owe him thanks for giving me a home when I chose to be an orphan and welcoming me as family. His patient mentoring allowed me to grow at my own pace, gently guided in a positive direction at all times. His optimism and enthusiasm inspired me daily. I will continue to try to "let go."

I would also like to thank my committee members, Dr. Michael Cummings, Dr. Eric Haag, Dr. Dave Hawthorne, and Dr. Ken Paynter for insightful conversations along the way and encouragement to keep pushing.

For providing vital clonal isolates, my thanks go out to Dr. Gerardo Vasta, Dr. Marta Gomez-Chiarri, Mr. Chris Dungan, and Dr. Dave Bushek. My sincere thanks to Mark Fisher, Jerry Mambretti, and Rebecca Hensley of the Texas Parks and Wildlife Department, Mark VanHoose, Alabama Departmnt of Conservation and Natural Resources, John Gunter, Florida Department of Agriculture and Consumer Science, Aswani Volety, Florida Gulf Coast University, and Paul Kenny at the Baruch Marine Field Lab for assistance in oyster collection. Special thanks to Dr. Gerardo Vasta and Wolf Pecher for sharing infected oyster DNA from northeastern locations.

I would like to thank the former Hare Lab for all the support and discussions along the way. Safra Altman, Gang Chen, Sarah Kingston, Maria Murray, and Colin Rose contributed greatly to the maturation of this document. Thank you for your solidarity, commiseration, and the occasional laugh along the way. I must also thank Jenna Murfree, Katie Schultziski, and Ninh Vu for help in the lab and keeping my world organized.

My sincere thanks go to members of the Animal and Parasitic Disease Lab at the USDA. I would not have completed this without the enthusiasm, common sense, encouragement, and efficiency of Tiger Dunams. Dr. Kristen Webb shared her office space, her untiring work ethic, her insightful mind, and her love of soccer generously. Thanks to Michael Butler Brown for freely lending his time, having good hands, and living an admirable life. I also owe special thanks to Dr. Eric Hoberg whose thoughtful advice expanded my horizons and pushed me to think bigger. My future is brighter for it.

I am greatly appreciative of the BEES community for its support and collegiality along the way. I am especially indebted to Julie Byrd, Jamie Pettengill, Alana Doty, Sky Lesnick, and Jaime Grace for improving my research. Jason Keagy and JF Savard for keeping my spirits up in the office. Thanks to the "Marine People." Finally, I give my sincere gratitude to Gwen Shlicta, Eric Lind, Chris Che-Castaldo, and Jenna Jadin for helping me remember that there are interesting things outside of research.

Special thanks go out to Lois Reid for shepherding me through this process. Your efficiency and knowledge are well-appreciated, but it is your kind heart that will be missed.

Finally, I must thank my family and friends outside of academia for their love and support. My mother and father for their unfailing confidence in my abilities and allowing me to pursue my own path. My sister, Holly, for lending her ear and offering creative solutions to life's little problems. The Posers and Bump Drink Spike for keeping me active. And most especially, my wife, Anne, whose patience and support have gone far beyond any marriage vows.

This research was financially supported by the USDA Agricultural Research Service, NOAA Sea Grant, and fellowships from the BEES program at the University of Maryland.

Table of Contents

List of Tables

List of Figures

Chapter One – Introduction

The modern concept of evolution

Population genetics is one of the foundations upon which modern ideas of organismal evolution is built. Evolution can be defined as a change in allele frequencies in groups of interbreeding individuals from one generation to the next (Futuyma 2006). Genetic variation is produced through mutations, upon which genetic drift and natural selection can act to change population allele frequencies. In order to understand the evolution of particular species, we must understand the reproductive biology, connectivity, and history of extant populations.

The behavior of genes in populations can be compared to a simplified, idealized population model. Under this model, populations are of constant size, randomly mate, and have no mutation, migration, or selection (Hartl and Clark 2007). This model allows a researcher to look for deviations from expectations in order to infer the important population processes that shape genetic variation in extant populations and predict future changes. Mutation and recombination are the sources of new genetic variation while non-random mating, fluctuation in population sizes, genetic drift, and natural selection remove variation through stochastic or consequential interaction with local environments. Very little migration between distinct populations can homogenize allele frequencies if sexual recombination is the norm (Spieth 1974). Investigating patterns of genetic diversity helps us to understand changes in species distributions, interactions, and subdivisions among populations; in short, the process of evolution which ultimately leads to the remarkable biodiversity on Earth.

The Clonal Theory of Parasitic Protozoa

The idealized population model described above provides explicit null hypotheses for genes in interbreeding groups of organisms, but few taxa conform precisely to these expectations. Among these, single celled parasites seem to deviate as a group, but in a consistent way. In 1990, Michel Tibayrence, Finn Kjellberg, and Francisco Ayala put forth a theory that the discrete taxonomic units of parasitic protists were fundamentally different than other taxa (Tibayrenc et al. 1990). They observed that clonal lineages in certain parasitic protists were maintained in natural populations for extended periods of time across large geographic areas, despite the fact that sexual reproduction had been documented in laboratory settings. This suggested that clonal parasite lineages were genetically independent, serving as the evolutionary unit upon which mutation, genetic drift, and selection act. This hypothesis was termed the "clonal theory of parasitic protozoa" and set forth experimental criteria to test for clonality among parasite populations.

Clonal characteristics in parasitic protist populations are a consequence of mixed mating life histories. Many single celled eukaryotes reproduce through both asexual and sexual modes (Dacks and Kasinsky 1999). This dual reproduction strategy engenders populations of protists with evolutionary properties inherently different from organisms exclusively dependent upon either asexual or sexual reproduction. Sexual recombination may bring beneficial alleles together, increasing fitness, while asexual propagation may enhance the ability of recombinant individuals to rapidly exploit favorable environmental conditions (Crow and Kimura 1965). In mixed-mating species such as parasitic protists, clonal lineages may evolve without sexual reproduction for long periods of time. Independent evolution of multiple clonal strains can be punctuated by occasional

outcrossing that resets the clonal evolution clock (Balloux et al. 2003). Depending on the geographic scale of gene flow and its influence on how much variation co-occurs within populations, outcrossing may introduce new beneficial alleles into the lineage and help in repairing or replacing deleterious mutations that have accumulated through time (Ayala 1998). Asexual reproduction of recombinant strains may then spread through the host population, causing morbidity and mortality. The population-level interplay between sexual and asexual reproduction can give rise to varying degrees and patterns of genetic "clonal population structure" that reveal epidemiologically relevant population dynamics of pathogenic organisms, including parasitic protists (Tibayrenc et al. 1990). Furthermore, spatial variation in host-parasite interactions can create a geographic mosaic of within-population patterns (Thompson 2005).

Clonality in populations is characterized by local deviations from panmictic population genetic patterns, signifying an assemblage of genotypic lineages that evolve independently rather than as a group (Tibayrenc and Ayala 2002). Population genetic patterns expected from clonality include departures from Hardy-Weinberg Equilibrium, extensive linkage disequilibrium across the genome, overrepresented multilocus genotypes (MLGs), and an absence of expected recombinant genotypes (Halkett et al. 2005). Clonal population structures have been observed in medically and agriculturally important eukaryotic parasites (Tibayrenc et al. 1990; Anderson et al. 2000; Mallon et al. 2003; Oura et al. 2005; Rougeron et al. 2009). Clonal genetic patterns can arise within mixed-mating populations when asexual reproduction predominates, after epidemic spread of highly successful parasite strains, as a consequence of frequent inbreeding in a sexual population, or a combination of these reproductive processes.

Perkinsus marinus as a study system

Perkinsus marinus is a protozoan parasite of the eastern oyster, *Crassostrea virginica*. Originally identified as a fungal pathogen, this parasite was initially designated *Dermocystidium marinum* (Mackin et al 1950), from which the disease caused by this parasite gets its common name, "Dermo." Subsequent analysis based on morphological characters discovered using electron microscopy (Perkins and Menzel 1966; Perkins 1976) placed *Perkinsus marinus* within the Apicomplexa (Levine 1978). Recent sytematics analysis based on DNA sequence has identified *Perkinsus marinus* as an early branching dinoflagellate (Siddall et al. 1997; Saldarriaga et al. 2003) and most closely related to the congeneric species *P. olseni* and *P. honshuensis* (Dungan and Reece 2006; Moss et al. 2008). This parasite has been occasionally reported in other bivalves (Ray and Chandler 1955; Kotob et al. 1999; Coss et al. 2001; Cáceres-Martínez et al. 2008; Pecher et al. 2008; Reece et al. 2008; Enriquez-Espinoza et al. 2010). Congeneric protozoans can be found in *C. virginica* (Pecher et al. 2008), but these have not been reported to cause disease.

P. marinus damages its economically and ecologically important host. Dermo disease resulting from systemic *P. marinus* infection is responsible for substantial seasonal mortality in adult host populations along the eastern and Gulf of Mexico coasts of the United States (Mackin 1953; Andrews and Hewatt 1957; Quick and Mackin 1971;Andrews and Ray 1988; Craig et al. 1989; Ford 1996; Powell et al. 1996). As *P. marinus* may kill 100% of infected oysters when temperature and salinity are high (Brown et al. 2005), reducing the impact of this parasite on its host is an important goal for managing oyster populations. *P. marinus* induced mortality is interfering with efforts to restore populations in Chesapeake Bay, where water quality concerns have led to an

extensive and expensive environmental rehabilitation program. Recent evidence of disease resistance in certain Chesapeake oysters exposed to high levels of *P. marinus* (Oliver et al. 2000; Ragone Calvo et al. 2003; Brown et al. 2005) have led to the recommendation that oyster sanctuaries be used to promote reproduction by survivors of Dermo epizootics, thus encouraging spread of disease resistance alleles in the host population (US Environmental Protection Agency 2007). Despite these efforts, current restoration of *C. virginica* populations in Chesapeake Bay have only reached 10% of goals and *P. marinus* remains a hindrance to restoration efforts (Chesapeake Bay Program 2009). Though many aspects of the interactions between *P. marinus* and *C. virginica* have been investigated (Paynter 1996; Bushek and Allen 1996; Chu and Volety 1997; Bushek et al 2002; Chintala et al. 2002; Ford et al. 2002), and much is known concerning the seasonality of infection and environmental modifiers (Andrews 1955; Hewatt and Andrews 1956; Andrews 1965; Chu and La Peyre 1993; Oliver et al. 1998), few studies have considered *P. marinus* life history characteristics and their relation to epidemiology.

Laboratory culture of *P. marinus* is accomplished through clonal propagation and clonality is presumed to be the primary mode of reproduction in nature (Ray 1954; Kleinschuster and Swink 1993; La Peyre et al. 1993; Gauthier and Vasta 1995; Bushek and Allen 1996). A sexual life cycle has not been documented for *P. marinus*, but forms resembling gametes have been observed (Perkins 1996), indicating that sexual reproduction may be possible for this protist. As sexual reproduction has evolutionary consequences for host-parasite interactions and responses to environmental change, it is important to determine whether *P. marinus* is capable of sexual reproduction and to what

extent it influences the epidemiology of disease. Observing recombination between pairs of genetic loci would provide evidence that sexual reproduction occurs in *P. marinus*.

P. marinus dispersal mechanisms are not fully understood. New infections have been observed within three months in disease free oysters separated from naturally infected reefs by 5 km (McCollough et al. 2007). There is no known vector for distributing *P. marinus* to uninfected oyster hosts, though natural vectors may exist (White et al. 1989). Oysters are readily infected by ingesting the parasite directly from the water column (Ray 1954; Chu and Volety 1997). Given the small size and limited swimming ability of the parasite (Perkins 1996), it is difficult to understand how this parasite has undergone a 600 kilometer range expansion in less than 30 years. It is possible that the parasite is capable of living independently of the host for extended periods of time, allowing water currents to transport the parasite large distances before infecting a new host. This may also be compounded and complicated by anthropogenic movement of the parasite through intentional (oyster transport for commercial processing) or unintentional (ballast water) means. Genotypic composition of local populations may provide clues as to the scale of *P. marinus* movement across its range.

Deep divergence between alleles has been observed in preliminary samples of *P. marinus* at three loci - actin, ribosomal intergenic repeat sequences (the non-transcribed spacer [NTS] and the internal transcribed spacer [ITS]), and one anonymous nuclear locus (ATAN) (Reece et al. 1997, Robledo et al. 1999, Brown et al 2004). Only two alleles were observed at actin and NTS; these alleles differed by as much as 2.6%. Though more alleles were discovered at the ITSand ATAN loci, the most divergent haplotypes differed greatly (1.5% and 3.5% , respectively). Bi-modal frequencies of the

most divergent alleles at the ATAN locus suggested di-allelism similar to that observed at NTS and actin. This pattern of deep divergence is similar to that seen in trypanosomes (Machado and Ayala 2001) and plasmodium (Roy et al. 2008) and has been attributed to long-term separation of clonal lineages and balancing selection, respectively. These observations suggest a need for more comprehensive examination of the extent and potential causes of di-allelism in *P. marinus*.

The scope of this work

Investigating the population genetic structure of *P. marinus* can provide insights into the demographic and evolutionary dynamics of *P. marinus* populations, and elucidate the natural reproductive biology of the parasite. Previous investigations of the genetics of *P. marinus* populations have been limited to strains that were genotyped only after isolation and clonal propagation in tissue culture (Reece et al. 1997; Reece et al. 2001). This laborious approach provided unambiguous strain genotypes even from oysters infected with more than one parasite strain, but imposed practical limits on the number of isolates that could be examined. Here, I genotyped *P. marinus* directly from infected oyster tissue in order to increase the depth of sampling within populations. I used highly variable microsatellite markers and directly sequenced DNA fragments to investigate the reproductive biology, dispersal, and the extent and origins of di-allelism in *P. marinus* populations.

Oyster samples were collected from fifteen geographic locations, spanning 4800 km of coastline along the Gulf of Mexico and Atlantic coasts of the United States. Locations were selected for comparisons within and among three regions previously determined to contain distinct parasite multi-locus genotype frequencies (Reece et al.

2001): the Gulf of Mexico (five locations), the southeastern Atlantic coast (six sampling locations), and the northeastern United States (four sampling locations). These geographic samples are believed to have different durations of residency (Ford 1996). South of Chesapeake Bay, *P. marinus* has a documented history as far back as the 1940's (Mackin et al. 1950). The parasite has recently expanded into more northern estuaries in conjunction with warmer than average winter water temperatures, extending its range as far as Maine (Ford and Smolowitz 2007). The samples analyzed here from the northeastern U.S. (Delaware Bay, NJ, Narragansett Bay, RI, and Tisbury and Edgartown, MA) may derive from this recent range expansion (Burreson and Ragone Calvo 1996).

I present analyses addressing three main questions: (1) how do *P. marinus* typically reproduce? (2) What is the extent of *P. marinus* gene flow? (3) Is di-allelism pervasive across the *P. marinus* genome and what are its potential origins?

In chapter 2, I present genetic analyses addressing clonality in *P. marinus* and the extent to which the clonal theory of protistan parasites applies to these estuarine parasites. Multilocus microsatellite genotypes are compared between four geographic locations. Asexual reproduction is investigated by looking for repeated multilocus genotypes and lineages of parasites that differ only through recent mutation at individual microsatellite loci. Linkage disequilibrium across the genome is tested using the standardized index of association, rbarD (Agapow and Burt 2001). Local samples were examined for deviations from Hardy-Weinberg equilibrium and the extent of inbreeding was inferred from F_{IS} estimates. Genotypic diversity and Hudson's four-gamete test (Hudson and Kaplan 1985) were used to infer recent recombination. In the course of this chapter, I also address issues of undetermined ploidy in *P. marinus*, the frequency of multiple

infections, and the advantages and pitfalls of genotyping *P. marinus* directly from infected oyster tissue. The interplay of these results reveals a reproductive biology where clonal evolution at ecological timescales may be interrupted by sexual recombination over greater periods of time.

In chapter 3, I investigate the distribution of genetic diversity among locations in order to understand the connectivity of *P. marinus* populations. Several *a priori* regional hypotheses are tested using AMOVA. Genetic distances between geographic samples were used to examine isolation-by-distance in *P. marinus* populations. Three individual clustering analyses were conducted to evaluate the relationships among individual multilocus genotypes. Clustering analyses were conducted based on inter-individual genetic distances, correlations in allele frequencies, and Bayesian population structure analysis aimed at maximizing Hardy-Weinberg Equilibrium among groups of multilocus genotypes. The geographic distribution of distinct clonal clades of parasites was examined in order to infer dispersal among locations. These analyses showed that local samples are generally not interbreeding populations of parasites, but rather a collection of sympatric clonal lineages that have dispersed widely.

Finally, in Chapter 4, I investigate whether previously observed di-allelism is prevalent across the genome and consider the potential origins of deep divergence between allelic classes. Genomic DNA sequences were generated from seven nuclear loci including five putatively neutral microsatellite flanking regions and two loci putatively involved in host-parasite interactions (Schott et al. 2003; Fernandez-Robledo et al. 2008). I describe the variability in DNA sequence between individuals and between geographic samples and provide an estimate of the time of divergence between observed

haplotypes. Recombination within DNA sequences or between loci across the genome was evaluated. I discuss several potential origins for di-allelism within *P. marinus* populations; in particular, ancient asexuality, genome duplication, recent population subdivision, balancing selection, and secondary contact between two ancient allopatric lineages. I argue that the data disfavor balancing selection, genome duplication, prolonged asexual diploidy, or recent population subdivision as explanations for a consistent pattern of di-allelism. Instead, these data suggest that two *P. marinus* populations were subdivided for millions of years, and have subsequently hybridized. Current mosaic genotypes may be the result of contemporary sexual recombination and inbreeding.

This dissertation improves our understanding of the evolution of *Perkinsus marinus* in natural populations. The reproductive biology, dispersal, and origins of genotypic diversity are relatively understudied in protistan parasites outside of those with human health impacts, particularly within the aquatic realm. This parasite is of particular importance due to its negative impacts on growth and reproduction in its ecologically and commercially important host. The results presented here have direct relevance to the epidemiology of Dermo disease and provide tools to monitor future outbreaks in host populations.

Chapter Two - An Evolutionary Legacy of Sex in the Predominantly Clonal Protistan Oyster Parasite *Perkinsus marinus*.

Abstract

Perkinsus marinus, a protozoan parasite of the eastern oyster *Crassostrea virginica*, limits fecundity and causes high mortality in host populations. The long-term efficacy of management strategies for suppressing this disease in both aquaculture and restoration settings depends on the potential rate of evolutionary response by *P. marinus*. Sexual reproduction has never been demonstrated in vitro or in previous population genetic studies in wild *P. marinus*. We developed high resolution microsatellite markers and amplified alleles directly from infected oyster genomic DNA. Of 336 infected oysters from four populations between Massachusetts and Florida, 129 (48%) had no indication of multiple infection and were subjected to population genetic analyses assuming diploidy. A high diversity of multilocus genotypes was observed. Therefore, strictly clonal reproduction seems unlikely. Substantial heterozygote deficits in three populations suggested that sexual reproduction often involves inbreeding. At the same time, significant multilocus linkage disequilibrium occurred in most sampled populations, along with several genotypes found repeatedly in two populations, indicating that asexual reproduction also occurs in *P. marinus* populations. Interestingly, where this parasite has recently expanded its range, lower strain diversity, significant heterozygote excess, and highly heterozygous multilocus genotypes suggests clonal propagation of recent recombinants. These indications of multiple reproductive modes in wild *P.marinus* populations suggest that over short periods of time, selection acts upon independent parasite lineages, rather than a cohesive, interbreeding population of individuals.

Nevertheless, high genetic diversity is the evolutionary legacy of sex in *P. marinus* and anthropogenically elevated opportunities for outcrossing may result in rapid evolution of this parasite.

Introduction

Perkinsus marinus is a protozoan parasite of the eastern oyster, *Crassostrea virginica*. Dermo disease resulting from systemic *P. marinus* infection is responsible for substantial seasonal mortality in adult host populations along the eastern and Gulf of Mexico coasts of the United States (Andrews and Hewatt 1957; Craig et al. 1989; Ford 1996; Powell et al. 1996). This parasite has been occasionally reported in other bivalves (Kotob et al. 1999; Coss 2001; Pecher et al. 2008; Cáceres-Martínez et al. 2008; Reece et al. 2008; Enriquez-Espinoza et al. 2010) *P. marinus* has negative impacts on its economically and ecologically important host. As *P. marinus* may kill 100% of infected oysters when temperature and salinity are high (Brown et al. 2005), reducing the impact of this parasite on its host is an important goal for management of oyster populations. *P. marinus* induced mortality is interfering with restoration of oyster populations in Chesapeake Bay, where poor water quality concerns have led to an extensive and expensive environmental rehabilitation program. Recent evidence of disease resistance in certain Chesapeake oysters exposed to high levels of *P. marinus* (Ragone Calvo et al. 2003; Brown et al. 2005; Oliver et al. 2000) have led to the recommendation that oyster sanctuaries be used to promote reproduction by survivors of Dermo epizootics, thus encouraging spread of disease resistance alleles in the host population (US Environmental Protection Agency 2008). Despite these efforts, current restoration of *C. virginica* populations in Chesapeake Bay have only reached 10% of their goals and *P.*

marinus remains a hindrance to restoration efforts (Chesapeake Bay Program 2009). Though many aspects of the interactions between *P. marinus* and *C. virginica* have been investigated (Paynter 1996; Bushek and Allen 1996; Chu and Volety 1997; Ford et al. 2002; Chintala et al. 2002; Bushek et al. 2002), and much is known concerning the seasonality of infection and environmental modifiers (Chu and La Peyre 1993; Oliver et al. 1998), few studies have considered *P. marinus* life history characteristics and their relation to epidemiology.

Because an organism's mode of reproduction strongly influences its population dynamics and its ability to respond to other taxa with which it interacts, it is important to know whether sexual or asexual reproduction predominates in natural populations. Laboratory culture of *P. marinus* is accomplished through clonal propagation, and clonality is presumed by many authors to be the primary mode of reproduction in nature (Ray 1954; Kleinschuster and Swink 1993; LaPeyre et al. 1993; Gauthier and Vasta1995; Bushek and Allen 1996). A sexual life cycle has not been documented for *P. marinus*, but forms resembling gametes have been observed (Perkins 1996), indicating that sexual reproduction may be possible for this protist.

Many single celled eukaryotes reproduce through both asexual and sexual modes (Dacks and Kasinsky 1999). This dual reproduction strategy engenders populations of protists with evolutionary properties inherently different from organisms exclusively dependent upon either asexual or sexual reproduction. Sexual recombination may bring beneficial alleles together, increasing fitness, whereas asexual propagation may enhance the ability of recombinant individuals to rapidly exploit favorable environmental conditions (Crow and Kimura 1965). In mixed-mating species such as parasitic protists,

clonal lineages may evolve without sexual reproduction for long periods of time. Independent evolution of distinct clonal strains can be punctuated by occasional outcrossing that resets the clonal evolution clock (Balloux et al. 2003). Depending on the geographic scale of gene flow and its influence on how much variation co-occurs within populations, outcrossing may introduce new beneficial alleles into the lineage and help in repairing or replacing deleterious mutations that have accumulated through time (Ayala 1998). Asexual reproduction of recombinant strains may then spread through the host population, causing morbidity and mortality. The population-level interplay between sexual and asexual reproduction can give rise to varying degrees and patterns of genetic "clonal population structure" that reveal epidemiologically relevant population dynamics of pathogenic organisms including parasitic protists (Tibayrenc et al. 1990). Furthermore, spatial variation in these processes can create a geographic mosaic of within-population patterns (Thompson 2005).

Clonality in populations is characterized by local deviations from panmictic population genetic patterns, signifying an assemblage of genotypic lineages that evolve independently rather than as a group (Tibayrenc and Ayala 2002). Population genetic patterns expected from clonality include departures from Hardy-Weinberg Equilibrium, extensive linkage disequilibrium across the genome, overrepresented multilocus genotypes (MLGs), and an absence of expected recombinant genotypes (Halkett et al. 2005). Clonal population structures have been observed and shown to be epidemiologically relevant in medically and agriculturally important eukaryotic parasites (Tibayrenc et al. 1990; Anderson et al. 2000; Mallon et al. 2003; Oura et al. 2005; Rougeron et al. 2009). Clonal genetic patterns can arise within mixed-mating

populations when asexual reproduction predominates, after epidemic spread of highly successful parasite strains, as a consequence of frequent inbreeding in a sexual population, or a combination of these reproductive processes.

At one extreme of the reproductive mode spectrum, strict asexual reproduction will lead, in the near term, to a population comprised of genetically uniform progeny. Over evolutionary time, new mutations will create novel genotypic strains resulting in nonreticulate genealogical inter-relationships. In diploids, heterozygosity created by new mutations will remain fixed in the progeny of strictly asexual organisms, cumulatively producing large inter-allele divergences (Welch and Meselson 2000). Strictly asexual lineages are generally believed to be evolutionarily short-lived because new mutations are usually deleterious, and the accumulation of many deleterious alleles over time ultimately leads to the extinction of the clonal lineage through selection (Lynch et al. 1993).

Even infrequent sexual recombination can erase evolutionary signatures of clonal reproduction (Balloux et al. 2003). At the same time, contemporary population dynamics can locally strengthen the population genetic patterns expected from clonality (Tibayrenc et al. 1990; Anderson et al. 2000; Razakandrainibe et al. 2005). For example, if sexual recombination produces a highly successful parasite strain, its epidemic spread via asexual propagules can generate population genetic patterns reminiscent of long term strictly asexual reproduction (Oura et al. 2005). In such cases, genotypic diversity of the parasite population will be less than that expected for panmictic populations, because successful genotypes will be overrepresented. Frequent inbreeding can also produce patterns reminiscent of strictly asexual populations, because assortative mating among

related individuals limits the production of new recombinants and maintains multilocus linkage disequilibrium within the population (Nordborg 2000; Balloux et al. 2003; Razakandrainibe et al. 2005).

Investigating the population genetic structure of *P. marinus* can provide insights into the demographic and evolutionary dynamics of *P. marinus* populations, and elucidate the natural reproductive biology of the parasite. Previous investigations of the genetics of *P. marinus* populations have been limited to strains that were genotyped only after isolation and clonal propagation in tissue culture (Reece et al. 1997; Reece et al. 2001). This laborious approach provided unambiguous strain genotypes even from oysters infected with more than one parasite strain, but imposed practical limits on the number of isolates that could be examined.

Here, we genotyped *P. marinus* directly from infected oyster tissue in order to increase the depth of sampling within populations, and used highly variable markers to identify and remove data resulting from multiple infections. We report multilocus *P. marinus* genotypes from infected oysters in four geographic locations and assess the dominant reproductive dynamics within these parasite populations. Geographic sampling included three regions previously determined to contain distinct parasite multi-locus genotype frequencies (Reece et al. 2001): the Gulf of Mexico (Port Charlotte, FL), the southeastern Atlantic coast (New Smyrna Beach, FL), and the northeastern United States (Delaware Bay, NJ and Edgartown, MA). These geographic samples also provide a comparison between southern (27°N) and northern (41°N) latitudes and from parasite populations believed to have different durations of residency. South of Chesapeake Bay, *P. marinus* has a documented history as far back as the 1940s (Mackin et al. 1950). In

southern Chesapeake Bay oyster populations, *P. marinus* infections and Dermo disease became more intense during the 1980s, while the parasite also expanded northward into lower salinity reaches of Chesapeake Bay and into Delaware Bay (Burreson and Ragone Calvo 1996; Ford 1996). Further northward population expansion into New England, from 1990 to 1992, appears to have been associated with relatively warm winters (Ford and Chintala 2006; Ford and Smolowitz 2007). The samples analyzed here from the northeastern U.S. (Delaware Bay, NJ and Edgartown, MA) may derive from this recent range expansion (Burreson and Ragone Calvo 1996).

We present analyses addressing four questions: (1) what is the infection prevalence and rate of infection by multiple distinct strains? (2) Are the genetic data more consistent with haploidy or diploidy of *P. marinus* during oyster infection? (3) Are patterns of genetic diversity more consistent with clonality or sexual reproduction, and are there regional differences in these patterns? (4) How robust are the population genetic conclusions made from sampling *P. marinus* from infected oysters using PCR?

Materials and Methods

Sample Collection

A total of 451 live adult oysters were collected from reefs in Port Charlotte, FL, New Smyrna Beach, FL, Delaware Bay, NJ, and Edgartown, MA, generally targeting peak infection seasons. Details of each collection are provided in Table 2-2. To test for temporal heterogeneity and Wahlund effects (incorrect estimation of allele frequencies due to grouping subpopulations into a single population), infected oysters were collected across different months within a single year in Delaware Bay and Edgartown and from

two different years at New Smyrna Beach and Delaware Bay. Oysters were shucked and 1 g of gill and mantle tissue was preserved in 95% ethanol for long term storage.

For comparison with these natural parasite isolates, six clonal *P. marinus* isolates were obtained. Strains NJ-1 (ATCC #50509) from Bed 455 in Delaware Bay, NJ, FL-6 (ATCC #50763) from Fort Pierce, FL, and LA-25 (ATCC #50775) from Mozambique Point, LA were generously provided by D. Bushek. M. Gomez-Chiarri provided reference sample HCTR from Charleston Pond, RI. G. Vasta provided purified DNA from the genomic reference clone isolated from an infected oyster from Bennet Point, MD (ATCC # PRA-240), and a clonal isolate from Texas (ATCC #50983). Additionally, four clonal isolates of other *Perkinsus* species (*P. chesapeaki* (ATCC #50866, ATCC #50864, and ATCC #50807) and *P. olseni* (ATCC # PRA-31)) were obtained from C. Dungan for validation of PCR amplification specificity.

Determination of infection status.

Infection status was established using PCR targeting the multi-copy ribosomal non-transcribed spacer (NTS) locus with primers designed specifically for *P. marinus* as previously described (Robledo et al. 1998). For this purpose, genomic DNA was extracted from oyster gill and mantle tissue using Chelex-100 (Bio-Rad, Hercules, CA) according to Walsh et al (1991). As the proportion of *P. marinus* DNA within any extracted oyster gDNA is directly proportional to infection intensity, 2 µl of undiluted sample DNA was used as template in each primary amplification. Template concentration ranged from 40 ng/reaction to 250 ng/reaction based on spectrophotometric analysis of purified DNA. All PCR reactions contained 1X PCR buffer (Invitrogen, Carlsbad, CA), 2.5 mM magnesium chloride, 0.25 mM mixed dNTP's, 0.1 mg/ml bovine

serum albumin, 0.5 units Taq DNA polymerase (Invitrogen), and 40 nM of each primer. Reactions were subjected to a standard cycling procedure with an initial denaturation at 94 °C for 2 minutes followed by 40 cycles of 94 °C for 30 seconds, 60 °C for 30 seconds, and extension at 72 °C for 1 minute. The cycling was followed by a final extension step at 72 °C for 5 minutes, and reactions were stored at 4 °C until analysis. PCR products were separated on 1.2% agarose gels in TBE buffer, and infection was indicated by the presence of a 300 base pair fragment (data not shown). This amplification protocol has been reported to detect one parasite particle in 30 mg of oyster tissue (Robledo et al. 1998).

From oysters thus diagnosed as infected with *P. marinus*, DNA was again extracted from 100 mg of gill and mantle tissue using the QIAgen DNEasy kit, following the manufacturer's recommended animal tissue protocol, and eluted in a final volume of 200 µl (QIAgen, Hilden, Germany).

Microsatellite genotyping.

Seven microsatellite loci were developed for this study based on sequences identified in the *Perkinsus marinus* genome project database (genome sequences now available in GenBank: Refseq: NZ_AAXJ00000000). Sequences from this database were mined for repetitive sequences using the tandem repeat finder MSATfinder (Thurston and Field 2005). PCR primers for forty of approximately 4000 candidate loci were designed and tested. Primer pairs for seven loci were subsequently selected for use based on their ability to reproducibly amplify variably sized products in an initial subsample of infected oyster DNA based on ethidium-stained agarose gels. Names of loci correspond to the sequence number in the *P. marinus* genome project database. Microsatellite primers and

motifs are shown in Table 2-1. Specificity of primer pairs for their *P. marinus* target locus was demonstrated by e-values less than 0.22. BLAST of primer sequences against *C. virgincia* DNA sequences did not yield any matches with greater than 65% identity or e-values less than 1.8.

Each of the seven loci was assayed in an independent PCR reaction for every genomic DNA that had been diagnosed as containing *P. marinus* on the basis of the NTS assay. Amplification was conducted in 20 µl reactions using the exact same conditions as in the NTS assay. PCR products were separated individually on an Applied Biosystems (Life Technologies, Carlsbad, CA) 3730 DNA Analyzer along with an appropriate LIZ size standard (Life Technologies). Electropherograms were analyzed using Genemapper v.3.7 (Life Technologies). Peaks were only scored as alleles if they exceeded 500 RFU. In order to maximize the possibility of detecting cases where infection with more than one strain of *P. marinus* might be expected, we initially accepted any such peak regardless of its height relative to others in a given chromatogram. Peaks in adjacent size classes were subject to acceptance by the default stutter criteria applied in Genemapper v.3.7 microsatellite settings. Individuals possessing more than two alleles at any single locus were considered to be infected with more than one strain of *P. marinus* and were excluded from subsequent analyses of linkage disequilibria and population differentiation.

Interpretation of population genetic patterns depends on whether the haploid or diploid phase was sampled, so we sought to infer the ploidy of *P. marinus* during oyster infection based on the distribution of *P. marinus* alleles among sampled oysters. Previous genetic data from wild samples have been consistent with diploidy for *P.*

marinus (Reece et al. 1997; Reece et al. 2001), but ploidy has not been confirmed in the laboratory. We reasoned that among singly-infected diploids (≤ 2 alleles per locus), the frequency of completely homozygous multilocus parasite genotypes would be very low in the absence of extreme inbreeding. Sampling haploids, on the other hand, would result in completely "homozygous" multilocus genotypes are both a criterion for single infection and expected to be relatively more common compared with the diploid model. The expected multiplicity of infection for oysters may be described by a Poisson distribution if the outcome of each host-parasite encounter is independent, unrelated to previous infection status of the oyster, and if the binomial probability of becoming infected in any single encounter is constant. The Poisson distribution accounts for variable numbers of lifetime host-parasite encounters (Fig. 2-1a) and provides a predictive model for the number of parasites expected in any single oyster. By assuming the observed proportion of uninfected oysters is an accurate estimate of the probability of the number of successes equal to zero, the mean of the Poisson distribution can be derived from the Poisson equation as -ln(proportion uninfected oysters in sample). Poisson expectations were generated for the entire dataset and for each sampling location. The shape and variance of the Poisson expectations were derived from the mean. The observed *P. marinus* genotypes were then classified as resulting from single infection or multiple infections of an oyster under both haploid and diploid models. In the haploid model, those genotypes with a single allele at all loci were classified as single infections, and all remaining infected oysters were placed in the multiply infected category. Under the diploid model, single infections were those genotypes with two or fewer alleles at all loci, while remaining infected oysters were categorized as multiply infected. Goodness

of fit between Poisson expectations and observed genotype distributions assuming haploidy or diploidy were tested using Chi-square.

Population genetic analysis.

Tests of population differentiation and deviation from Hardy-Weinberg equilibrium were conducted with Genepop on the Web (Raymond and Rousset 1995). GenAlEx 6.1 (Peakall and Smouse 2006) was used to estimate average inbreeding coefficients (F_{IS}), and identify multilocus genotypes represented more than once in the dataset. Allelic richness was calculated in the program FSTAT (Goudet 1995) based on a sample size of 11 to correct for variation in sample sizes and allow comparisons between samples from different locations. Genotypic diversity, defined as the proportion of pairwise genotype comparisions that differ, pairwise linkage disequilibrium (LD) between loci, and the extent of multilocus linkage disequilibrium using the standardized index of association, $\bar{r}_d$, were estimated as implemented in MultiLocus v.1.2.2 (Agapow and Burt 2001) by comparing observed genotypic distributions to those obtained by permuting alleles among genotypes. Where a specific sample did not amplify for a certain locus, these null amplifications were allowed to permute along with alleles in the Multilocus analyses. Permutation of alleles among genotypes in Multilocus v.1.2.2 also provided an expectation for the number of unique multilocus genotypes based on allele frequencies in each population and assuming free recombination among loci. Potential multilocus lineages were identified in the program Genotype v.1.2 (Meirmans and Van Tienderen 2004) with two mutational steps used as a cutoff to define distinct lineages. Subsequently, in order to rule out the the possibility that two genotypes within a lineage were similar by chance, the probability that two genotypes were the product of

independent sexual recombination events (p_{sex}) was calculated in GenAlEx v.6.1 to confirm multilocus lineages as outlined in Arnaud-Haond et al (2007). Significance was accepted at an alpha level of 0.05 for all analyses. When multiple comparisons were made for the same statistic, a sequential Bonferroni correction was used to minimize the risk of Type I errors. Genetic distances between individual genotypes were estimated with MSA (Dieringer and Schlötterer 2003) using the proportion of shared alleles, Dps (Bowcock et al. 1994). A neighbor-joining tree was constructed using the Clustering Calculator (http://www2.biology.ualberta.ca/jbrzusto/cluster.php) based on the resulting Dps distance matrix. Trees were visualized and labeled in MEGA 4 (Tamura et al. 2007).

Results

Markers and Prevalence

An initial evaluation of our microsatellite loci employed six cultured isolates of *P. marinus* and four isolates of other species in the genus *Perkinsus*. Successful amplification was obtained for all seven loci in each isolate of *P. marinus.* None of the loci amplified in *Perkinsus* congeners except for the Pm2232 locus in *P. olseni* and *P. chesapeaki*, yielding a fragment approximately 100 bp larger than any allele found in *P. marinus*. Among six replicate PCR reactions performed for each locus (42 total) from these culture-derived *P. marinus* genomic DNA samples, no amplification attempt failed, and only one discordant genotype was found in clonal sample ATCC#50983 among three amplifications of locus Pm2903. Thus, this genotyping approach yielded considerable reproducibility (98%) when templates were abundant and of high quality.

Among 451 oysters sampled from our four field sites, 336 (74.5%) were positive for *P. marinus* as judged by amplification of the NTS locus (Table 2-2). Prevalence varied temporally and geographically. In 2002, the prevalence in Delaware Bay ranged

from 46.4% in June to 100% in August, whereas the prevalence was consistently high (>75%) in sequential monthly samples from Edgartown of the same year. In Florida, prevalence ranged from 70% in Port Charlotte to a consistently high value (> 85%) across two years at New Smyrna Beach on the Atlantic coast.

Ploidy

Two distinct multilocus genotype patterns were observed in cultured isolates of *P. marinus*. Only one allele was amplified at every locus in isolates derived from Texas, Louisiana, Florida, and Rhode Island, consistent with haploidy or strong inbreeding. However, isolates from Delaware Bay, NJ (ATCC #50509) and Chesapeake Bay (Genome reference PmCV4CB5 2B3 D4) were heterozygous at seven out of seven and six out of seven loci, respectively (data not shown), consistent with diploidy.

In order to assess the ploidy of *P. marinus* amplified from infected oysters, observed counts of single and multiple infections were tallied under both haploid and diploid models and compared to expectations based on a Poisson model of infection multiplicity (see Methods for details of model). When samples from all locations were included, diploid assumptions led to nearly equal counts of single and multiple infections with no significant deviation from Poisson expectations (Fig 1b, chi-square = 0.138, p = 0.714). Under haploidy assumptions, classifying any case of two or more alleles as multiple infection, only 24 oysters (8.9%) were interpreted as singly infected and this ratio significantly differed from Poisson expectations (chi-square = 146.34, $p < 10^{-32}$). However, when considered separately, each location varied in its agreement with either model. The diploid model was more similar to Poisson expectations than the haploid model in three of the four locations (Fig. 2-1c,e,f), though only Edgartown samples fell

within the Poisson expectations according to goodness-of-fit statistics (chi-square = 0.241, p = 0.624). In New Smyrna Beach, classifying genotypes according to a haploid model most closely matched Poisson predictions (Fig 2-1d) though this result was significantly different from predictions (chi-square = 5.45, p = 0.020). All other local comparisons significantly differed from both haploid and diploid multiplicity expectations ($0.006 < p < 10^{-16}$). Based on the agreement between the complete dataset and the diploid model of infection multiplicity, all subsequent analyses were conducted assuming diploidy.

Multiple Infections

The probability of observing a third allele indicative of multiple infection increased with the number of loci genotyped. Because amplification failed for one or more loci in a subset of the 336 NTS-positive individuals, we estimated multiple infection rates from the 271 samples yielding data from any four or more microsatellite loci. Oysters infected with multiple strains of parasite were identified by the presence of three or more alleles at one or more locus. Multiple infections were most common in the oysters of Delaware Bay and Edgartown (77.1% and 67.9%, respectively) and least common in those from Florida (15% to 30.2% from New Smyrna Beach and Port Charlotte respectively). Overall, oysters from Florida had a significantly lower incidence of multiple infections than those from Delaware Bay and Edgartown combined ($chi^2 = 46.7$, $p < 0.0001$). Multiple infection frequency did not vary among years in either Delaware Bay or New Smyrna Beach (Fisher's Exact test, p = 0.207 and 0.229 respectively). Multiple infection frequency also did not vary significantly from month to

month in Delaware Bay or Edgartown (Fisher's exact test, $p = 0.137$ and 0.071), but the trend was for fewer multiple infections in later months (Table 2-2).

Population Genetic Diversity

After excluding oysters determined to be infected with more than one parasite genotype, a total of 129 individuals with four or more loci genotyped were available for population genetic analyses. The number of alleles per locus ranged from 10 to 14, with an average of 11.9 alleles per locus across the entire dataset (Table 2-3). Average allelic richness, following rarefaction for n=11, ranged from 3.2 in the New Smyrna Beach 2005 samples to 4.2 in Delaware Bay 2002, and did not differ significantly among locations (ANOVA, $F = 0.994$, $p = 0.435$). At each locus, null amplification occurred in 5 to 13 individuals despite repeated PCR attempts (detailed in Table 2-3). Genotypes for all seven loci were obtained for 92 individuals (71.3%) and only 7 (5.4%) had four-locus genotpes.

Among the 129 genotyped parasites, 126 unique multilocus genotypes (parasite strains) were identified. Monthly collections in 2002 from Delaware Bay and Edgartown yielded small samples with no significant genic (Raymond and Rousset 1995) or genotypic (G-based test of Goudet 1996) differentiation ($p > 0.39$ for all pairwise comparisons) and were aggregated for all subsequent analyses. Genotypic diversity, calculated as the probability that two randomly chosen genotypes differ, was uniformly high across samples (average of 0.999; Table 2-4). All pairwise geographic and temporal (year) tests for differentiation were significant ($p < 0.003$) both genically and genotypically. Repeated multilocus genotypes were removed for all subsequent ('clone corrected') analyses unless stated otherwise.

Each locus was tested for departures from Hardy-Weinberg Equilibrium (HWE) in each clone-corrected population sample using exact tests (Table 2-5). Eighteen out of 42 tests indicated significant HWE deviations at individual loci following sequential Bonferroni correction. Each locus had at least one significant HWE deviation, but no locus displayed significant HWE rejection in all populations. Tests across all loci using Fisher's method rejected HWE in each population ($p < 0.006$). A significant deficit of heterozygotes was observed in three populations (Port Charlotte, New Smyrna Beach, and Delaware Bay) at several loci, whereas a significant excess of heterozygotes was observed in Edgartown at three of seven loci (one-sided test, Bonferroni corrected critical $p=0.0015$).

Average inbreeding coefficients were estimated by comparing observed and expected heterozygosity (F_{IS}) in each population, treating repeated genotypes as a single genotype. Positive F_{IS} was observed in all samples from Florida, indicating substantial inbreeding in these populations (Table 2-5, Fig. 2-4). Delaware Bay has significant positive F_{IS} values in 2005, but not in 2002. Interestingly, the highest values of F_{IS} occur where multiple infections occur least frequently. In stark contrast, heterozygosity occurs with excess frequency in the Edgartown sample ($F_{IS} = -0.273$). F_{IS} has been shown to have high variance in clonal diploids where sexual recombination is infrequent (Balloux et al. 2003). Standard deviation in F_{IS} ranges from 0.220 to 0.320 in samples studied here (Fig. 2-4).

Multilocus linkage disequilibrium was tested by comparing observed values of the standardized index of association ($\bar{r}_d$) with 1000 datasets generated through random permutation of alleles under a panmictic model. Linkage disequilibrium (LD) was

significant across the genome when all samples were included ($p < 0.001$, Table 2-4, Fig. 2-2). Each genetically distinct sample (as determined above) was also tested independently, as substructure of populations may produce the appearance of multilocus LD despite panmixia within subpopulations (Halkett et al. 2005). All populations showed significant multilocus LD ($p < 0.05$) with the exception of New Smyrna Beach in 2004 ($p = 0.555$). The highest values of $\bar{r}_d$ were observed in Delaware Bay (0.110 in 2002 and 0.160 in 2005, Fig. 2-2) and Edgartown (0.114, Fig. 2-2). In pairwise tests of LD between particular loci using the entire dataset, only six of twenty-one tests were significant following sequential Bonferroni correction. Each instance of LD was associated with either locus Pm2988 or Pm8517. Following removal of these two loci, multilocus LD remained significant within the entire dataset ($\bar{r}_d = 0.061$, $p < 0.001$) and Delaware Bay ($\bar{r}_d = 0.21$ and 0.18 in 2002 and 2005 respectively, $p < 0.003$ for both), but not in any other individual sample.

Despite the high genotypic diversity, permutation of alleles under a panmictic model revealed that there are significantly fewer MLGs than would be expected with free recombination ($p < 0.001$). One MLG was observed in Port Charlotte three times, and another was observed in Edgartown twice (Fig. 2-3). Furthermore, seven individuals in Edgartown share an average of 82.0% of their alleles in pairwise comparisons and differ by fewer than four microsatellite repeats, suggestive of a multilocus lineage with recent mutational derivatives (Fig. 2-3). The probability that any two genotypes in this Edgartown multilocus lineage were generated through independent sexual events is minimal ($p_{sex} < 0.005$ following sequential Bonferroni correction).

As epidemic spread of a single genotype may obscure the processes generating genotypic diversity (Tibayrenc et al. 1991), tests for multilocus linkage disequilibrium were repeated after removal of genotypes represented more than once in our sample. Significant multilocus LD persisted in the entire clone-corrected dataset ($\bar{r}_d = 0.066$, $p < 0.001$, Table 2-4, Fig. 2-2). Likewise, multilocus LD was significant within the Edgartown sample ($\bar{r}_d = 0.104$, $p = 0.005$, Table 4, Fig. 2). On the other hand, though remaining positive in absolute value ($\bar{r}_d = 0.021$), linkage was no longer statistically significant in Port Charlotte after removal of repeated MLGs ($p=0.12$, Table 4, Fig. 2).

Discussion

We have used seven novel microsatellite loci to characterize populations of *P. marinus* from four geographic locations representing three putatively distinct regional populations (Reece et al. 2001). We have shown that sampling *P. marinus* variation directly from infected oyster genomic DNAs is a viable method to genotype *P. marinus* and examine genetic variation at the population level. Amplification of *P. marinus* DNA in this manner forgoes the laborious process of cloning individual isolates prior to genetic analysis, thus allowing many more individuals to be genotyped. PCR assays were specific to *P. marinus* and do not amplify products from other *Perkinsus* species also found in the oyster host (Pecher et al. 2008). The eight RFLP markers used in Reece (2001) could theoretically identify up to 6561 unique MLGs, but revealed only 12 among 86 isolates sampled at a comparable continental scale. The seven microsatellite markers used here are capable of identifying at least 10^{13} MLGs, given the total allelic diversity in our samples, and found 126 unique strains among 129 genotypes. Thus, abundant variability in microsatellite loci has provided greater resolution with which to study

populations of *P. marinus* than was previously available, revealing remarkable genetic diversity and both asexual and sexual reproductive processes.

Diploidy

Ploidy is a critical factor structuring genetic variation, so its determination is an important precursor to population genetic inferences. Phylogenetic analyses place *P. marinus* basal to dinoflagellate lineages (Siddall et al. 1997; Saldarriaga et al. 2003) in which many parasite taxa are known to be haploid with ephemeral diploid stages (Nuismer and Otto 2004). By identifying two alleles per locus in cultured isolates that had been cloned by limiting dilution, our study and previous studies (Reece et al. 1997; Reece et al. 2001) suggest that diploid forms of *P. marinus* may be readily obtained from infected oysters. Out of the 86 isolates subcloned for genetic analysis by Reece et al. (2001), heterozygosity was observed in at least one single copy nuclear RFLP marker in 16 isolates. Among the reference strains used in this study, the isolate of *P. marinus* used for genome sequencing was subcloned through two rounds of limiting dilution (Fernandez-Robledo et al. 2008) yet two alleles were amplified at six of seven microsatellite loci. Though it is formally possible that two different life stages, one haploid and one diploid, could be supported in tissue culture (or oyster infection), it seems reasonable to assume that all cultured isolates have the same ploidy. Therefore, the available evidence from clonal isolates and natural infections indicates that diploidy is likely in *P. marinus.*

Further support for diploidy is provided here based on comparisons of observed and expected frequencies of single and multiple infections using a Poisson model of host-

parasite encounters. This approach does not provide a definitive ploidy inference but shows that patterns of genetic variation are generally consistent with diploidy. This inference is dependent upon the assumptions of independence between infection with different strains and accurate estimates of uninfected oysters in each sample. Furthermore, other population processes causing deviations from Hardy-Weinberg equilibrium may interfere with the ability to distinguish single infection from multiple infections. Nevertheless, the agreement between the diploid model of multiple infections and the data presented here agrees with the presence of two alleles in tissue culture isolates, indicating that diploid forms are common in *P. marinus* found in oysters. Consequently, we assumed for our analyses that *P. marinus* was diploid when sampled (understanding that evidence to the contrary would require revision in our interpretations, particularly with respect to the amount of inbreeding within populations.)

Mixed Mating

Clonal population genetic patterns were observed in all locations examined here, often consistently over time, in agreement with the clonal theory of parasitic protozoa (Tibayrenc et al. 1990). The standardized index of association, $\bar{r}_d$, significantly differed from free recombination expectations in nearly all samples, indicating that allelic correlations among loci extend across the genome and are not explained simply by physical linkage between particular markers. Deficits of heterozygosity relative to Hardy-Weinberg expectations were large in almost every sample, driven by statistically significant departures at different subsets of three to six genetic loci. One exceptional population, Edgartown Massachusetts, had a significant excess of heterozygotes, but even here (as well as in Port Charlotte) the sample included repeated MLGs implicating

asexual reproduction. Given the observed allelic diversity, the probability of identity between any two genotypes is less than 1 in 10,000 for each of the sampling locations assuming a panmictic sexual population. Despite the repeated genotypes and linkage disequilibrium, genotypic diversity is too high to be explained solely by asexual reproduction. Furthermore, pairwise comparisons between loci conform to Hudson's four-gamete test (Hudson and Kaplan 1985) indicating that recombination between loci has occurred unless reversion in allele size in these microsatellite loci is frequent. Thus, clonal population dynamics predominate for *P. marinus,* but sexual reproduction must be occurring in every population to generate observed genotypic diversity.

An interesting repercussion of sexual reproduction in some populations of *P. marinus* appears to be inbreeding, as indicated by high F_{IS} values. Like asexual reproduction, inbreeding can cause correlations among alleles at different genetic loci and are both manifestations of "clonal population structure" (Tibayrenc et al 1991). Under strict clonality with new variation generated solely by mutation, genotypic diversity would be lower than observed. Cumulative mutations in obligately asexual diploids would have created fixed heterozygous loci. The opposite trend is observed here in mid-Atlantic and southern populations.

There were two samples that did not show heterozygote deficiencies, perhaps indicating variation in the propensity for inbreeding. The 2002 Delaware Bay sample had an F_{IS} estimate near zero (only one locus deviated from HWE), suggesting that conditions leading to inbreeding varied between 2002 and 2005 in Delaware Bay. The significant allelic correlations in the 2002 Delaware Bay sample attest to a legacy of clonality even though there is not a significant deviation from HWE. Edgartown had three loci with

significant heterozygote excesses and a negative Fis significantly different from zero ($p = 0.005$). Other factors may be affecting Edgartown heterozygosity in addition to deviations from random mating (see below).

Because heterozygote deficits contribute much of the weight to our conclusion of clonality, and inbreeding in particular, it is important to consider possible biases from the occurrence of non-amplifiable 'null' alleles. No amplification was obtained in 6.3% of amplification attempts from infected oyster genomic DNA extracts used for population genetic analysis. When these occur in samples otherwise yielding full data it is often assumed that they indicate homozygosity for a null allele. Under that hypothesis, amplification failures are likely to occur more in some loci than others, depending on polymorphism at the priming sites. In *P. marinus*, failed amplifications did not differ in frequency among populations or among loci (Two-way ANOVA, $F = 0.544$, $p = 0.776$ and $F = 2.616$, $p = 0.060$ respectively). Examined another way, the number of failed amplifications that would be expected for null alleles to explain HW deviations in the entire dataset (De Meeus et al. 2007) ranges from 1.6 to 5.8 among individual loci (Table 2-S1). Failed amplifications were observed at significantly higher frequencies than these predictions (Table 2-S1, t-test, $p = 0.0060$) suggesting that a large proportion of amplifications failed for reasons other than null alleles, such as low parasite template abundance. This conclusion is reinforced by the fact that Port Charlotte had strong heterozygosity deficits but relatively few failed amplifications and Delaware Bay 2002 had no Hardy-Weinberg deviation and relatively abundant null amplifications. Similarly, there was no positive relationship between the number of expected and observed null amplifications across populations (Fig 2-S1) suggesting that null amplifications do not

explain deviations from HWE. Finally, there were no null amplifications observed among the tissue culture isolates. Though the isolate sample size is small, it covers a wide geographic range, increasing the likelihood that null alleles would be detected in these isolates as well as in population samples. In summary, it is more likely that repeated failure to amplify one or several loci from certain isolates primarily reflects inherent challenges in amplifying *P. marinus* DNA from the milieu of host DNA. We are not able to reject null alleles as a partial cause of null amplifications and heterozygote deficits, however, and therefore consider inbreeding an alternative working hypothesis.

Regional Patterns

In Port Charlotte, clonality is maintained through asexual propagation of some strains and possibly also by inbreeding. One genotype was encountered three times, which is exceedingly unlikely under panmixia ($P_{sex} < 10^{-14}$), and correlations among physically independent loci were observed. Following removal of the genotype observed three times in Port Charlotte, multilocus LD was no longer significant ($\bar{r}_d = 0.021$, $p = 0.12$), indicating that epidemic spread of a single genotype can contribute substantially to the appearance of multilocus LD. At the same time, in order to explain the significant heterozygote deficits ($F_{IS} = 0.432$), selfing would theoretically need to occur in 60% of sexual encounters. Taken together, these results suggest an epidemic clonal population structure (Tibayrenc 1990) in which inbreeding may reinforce independent evolution of lineages in localities such as Port Charlotte.

Significant multilocus LD and departures from HWE each indicate that sexual recombination does not prevail in New Smyrna Beach and Delaware Bay. Significant heterozygote deficits were observed in the 2004 New Smyrna Beach sample, yet

multilocus LD was not present. This is probably due to the multitude of independent parasite strains resident at this location, as independent strains may not share the same correlations between loci. Similarly, in 2002, Delaware Bay parasites show multilocus LD, but have low levels of F_{IS}. This sample is not consistent with HWE when all loci are considered (Fisher's method, $p = 0.0060$), suggesting that random mating has not occurred in this population, but that HW deviations are not due to extensive inbreeding in Delaware Bay in 2002.

Inbreeding structures the populations of even obligately sexual parasites, such as *P. falciparum,* in areas of infrequent transmission (Anderson et al. 2000) and has even been found among the diploid stages of such parasites where transmission is intense (Razakandrainibe et al. 2005). Here, clonality persists in *P. marinus* populations in spite of the frequent occurrence of oysters infected by more than one genotype in locales such as Delaware Bay. Our data indicate that sexual recombination in *P. marinus* most frequently occurs when closely related strains are in proximity.

It is unknown where in its life cycle *P. marinus* may undergo sexual recombination, but if it occurs within the oyster host, the multiplicity of infection could constrain the likelihood of outcrossing. If a given parasite strain represents 90% of the parasite cells in an oyster, random mating would lead to 81% selfing in that oyster in the absence of barriers to self-fertilization. In the microsatellite chromatograms of multiply infected individuals, peak heights were rarely equal, indicating either that some alleles are preferentially amplified or that multiple infections typically have one numerically dominant parasite strain. Delaware Bay had a much higher frequency of apparent multiple infections than New Smyrna Beach (77% versus 17% respectively), providing

more opportunities for outcrossing in Delaware Bay. There is a significant negative linear relationship between F_{IS} and the proportion of multiple infections when all samples except Edgartown are included ($p < 0.025$, $r^2 = 0.8749$) indicating that inbreeding may be more prevalent where there are fewer multiple infections. In order to achieve the observed F_{IS} levels, the selfing rate in Florida populations would need to be greater than 60% at equilibrium (Table 5). Similar extreme inbreeding levels have recently been found in *Leishmania braziliensis* (Rougeron et al. 2009) and *Trypanosoma congolense* (Morrison et al. 2009). In populations characterized by high homozygosity, where multiply infected oysters may go unrecognized, the inbreeding rate may be even higher than estimated here.

Asexual reproduction seems to play a much larger role in shaping the genetic variation in Edgartown, MA. In this study, ten of twenty-six genotypes were heterozygous for all seven loci and an additional seven individuals were heterozygous at six loci; only four individuals had less than four heterozygous loci. This result is in agreement with previous findings where New England samples were unique in being heterozygous at all eight RFLP loci surveyed (Reece et al. 2001). Neither of these observations is consistent with frequent sexual reproduction unless it involves disassortative mating. Also, repeated genotypes provided a direct indication of clonal reproduction. The same MLG was found in oysters collected two months apart in Edgartown, indicating that certain genotypes can persist locally. Seven other co-occurring genotypes differed by less than four microsatellite repeats and shared 82% of alleles among all of the genotypes. This pattern of genotypic relationships might reflect the accumulation of mutations in an asexually propagating strain, because the allelic

composition of this population render it highly unlikely that these seven individual genotypes could have resulted from independent sexual recombination events ($P_{sex} < 0.001$ for all pairwise comparisons).

Any of several explanations might conceivably account for the significant excess of heterozygotes observed in the highly clonal Edgartown population. Recent hybridization between two genetically differentiated parasite strains could account for such a pattern. The recent ability of *P. marinus* to rapidly extend its northern range limit has led others to test for temperature-related differences in performance across parasite populations (Ford and Chintala 2006), but results were negative. Nonetheless, hybridization is a plausible mechanism by which new adaptive variation could promote range expansion (Ayres et al. 2004; Kolbe et al. 2004; Barrett et al. 2007)). Alternatively, if range expansion was accomplished by stepping-stone colonizations, then they likely entailed severe genetic bottlenecks, resulting in a loss of low frequency homozygous genotypes and ending with heterozygote excess. Analysis of the Edgartown sample using the coalescent-based computer program Bottleneck (Piry et al. 1999) indicates that there are fewer heterozygotes than would be expected in a severe population bottleneck, despite the high number of heterozygous loci observed. Conversely, ancient asexual lineages can experience a 'Meselson effect' if, absent recombination, two allelic lineages have been diverging from each other at each locus (Welch and Meselson 2000) though this is unlikely considering the recent emergence of disease in this geographic region. These hypotheses require further investigation in order to identify potential ancestral strains, test for genetic bottlenecks in other New England populations, and systematically evaluate relative fitness of *P. marinus* strains.

Robustness of Conclusions

Our genotyping approach identified a far greater incidence of multiply-infected oysters than a previous report employing isolation of strains by limiting dilution in tissue culture. Reece et al (2001) found 5 out of 20 primary *P. marinus* cultures to be comprised of at least two genetic strains (25% multiple infections). These derived from widely distributed and genetically distinct populations. Multiple infections ranged in our samples from 17%, in New Smyrna Beach, to 77% in Delaware Bay. In this study, highly variable microsatellite markers provided greater resolution to identify multiple infections and reveal spatial, and to a lesser extent temporal heterogeneity in the proportion of multiple infections.

Our criteria for identifying and removing multiple infections from analyses, i.e. treating all peaks ≥ 500 RFU as microsatellite alleles, may have upwardly biased the estimated multiple infection rate if it led to spurious amplification products being accepted as legitimate alleles. Genotypes presumed to represent single infections possessed 1.4 alleles per locus genotyped, which differed significantly (T-test = 14.67, $p < 0.0001$) from those identified as multiple infections (2.1 alleles per locus), raising confidence in our ability to identify true multiple infections (distribution of alleles per locus is shown in Figure S2). However, in order to test the sensitivity of results to the criterion used, alleles were rescored excluding bands at each locus whose height was ≤25% of the highest peak for the individual. After rescoring, the lower peak in heterozygous chromatograms averaged 57% of the tallest peak, so peaks <25% in height were more likely to be multiple infection artifacts than cases of allele dropout. Exclusion

of relatively low signal peaks increased the number of individuals deemed singly infected and having ≥4 loci genotyped (N = 189), and reduced the estimated multiple infection rates down to a maximum of 33% in Edgartown. Multiple infection rates were still significantly heterogeneous across localities (chi-square = 20.0, P = 0.001). Analysis of the rescored dataset did not change the fundamental conclusion that *P. marinus* experiences mixed modes of reproduction. Genotypic diversity remained very high (0.998) while mean expected heterozygosity across loci was unchanged (p = 0.771). The signature of clonal population structure increased in terms of F_{IS} magnitudes and the presence of additional repeated genotypes. F_{IS} in populations from Florida and Delaware Bay in the rescored dataset ranged from 0.339 to 0.671, with the largest increase in Delaware Bay in 2002 (F_{IS} = 0.339 from 0.066). Heterozygosity excess decreased in Edgartown (F_{IS} = -0.100), yet was still significant. Such scoring also resulted in more repeated genotypes being observed (five genotypes occurring at least two times), reinforcing the conclusion that asexual propagation takes place in *P. marinus*.

Conclusions

Taken altogether, *P. marinus* reproduction is predominantly clonal, though the mechanisms engendering recent clonality appear to differ in geographically disparate populations. Asexual reproduction is evident in repeated MLGs and multilocus lineages. Inbreeding may further limit the extent to which sexual recombination shapes *P. marinus* response to selection in the near term. Still, some sexual reproduction was historically necessary to produce so much MLG diversity. Therefore, epidemiology of Dermo disease is dependent upon the short-term success of multilocus lineages. If these lineages differ in their virulence or transmissibility (Bushek and Allen 1996), disease dynamics

may vary according to the composition of clonal parasite lineages present in any given time and place. In both interannual comparisons made in this study, changes in genic and genotypic composition were apparent, underscoring the evolving nature of *P. marinus* population composition. Our data indicate that local populations of *P. marinus* are comprised of a shifting assemblage of myriad lineages with enough diversity to undermine any but the most generalized host resistance mechanisms or disease suppression treatments.

Despite the departure from panmixia in most populations sampled, we conclude that sexual recombination is an evolutionarily important process in *P. marinus*. Even infrequent sexual recombination may provide fitness benefits (Green and Noakes 1995) and can endow populations with evolutionary properties similar to those of panmictic populations (Balloux et al. 2003). Consequently, although *P. marinus* populations have clonal population structures facilitating rapid response to selection from an assemblage of lineages in the short term, recombination may be sufficiently frequent to ensure that *P. marinus* ultimately evolves as an interbreeding species rather than as a collection of isolated strains. Future studies of the genetics of *P. marinus* populations should seek to identify clonal lineages and determine the spatial and environmental context for clonal expansion. Even rare outcrossing in *P. marinus* may give rise to genotypically and phenotypically novel forms, characterized by unpredictable ecological and epidemiological properties (Boyle et al. 2006; Schwenkenbecher et al. 2006; Volf et al. 2007). Efforts should therefore be made to limit anthropogenic movement of *P. marinus* between geographically and ecologically distinct locations.

Tables

Table 2-1. Primers for amplification of *P. marinus* microsatellite loci

Locus	Primer Name	Primer Sequence (5'-->3')	Tm	Repeat Motif	Amplified Product Size Range
Pm2232	Pm2232F	GCAGCTCGCCATTTGATAAT	60	GAT	243-294
	Pm2232R	TCTGTCTGGCGCCTACTTCT	60		
Pm2903	Pm2903F	GTAATGTAGGCCCCGTTAT	60	TCA	188-233
	Pm2903Rint	AGGCCAGAGGCTGTTAGAC	60		
Pm2988	Pm2988F	TGTTGACGGATCACATCTGG	60	TG	255-307
	Pm2988R	ATCGGTATCAGGAGCGTAGC	60		
Pm4488	Pm4488F	GGAGTCGGTCGAGCAGTAAA	60	TAC	272-299
	Pm4488R	TCCTTGGGGATGACTCTTTG	60		
Pm8517	Pm8517F	ATGGCTCAGCTTGCAGGTAT	60	TAG	269-320
	Pm8517R	GCCAAATGCTCTTGTGGAAT	60		
Pm9464	Pm9464F	CAGCTATGGTGTTGGGGAGT	60	TAG	267-309
	Pm9464R	CGACAGAGCCTCCTTCATTC	60		
Pm12067	Pm12067F	ATTCCCTTCGTGTGCTTCTG	60	TAC	287-308
	Pm12067R	TGAGCCGGCTATACAGAGGT	60		

Table 2-2. Prevalence of infection and frequency of multiple infections in each sample

	Date Collected	Total oysters sampled	# NTS positive	Prevalence	# Individuals with >4 uSat loci genotyped	# multiple infections	% multiple infections
Edgartown, MA	June, 2002	30	29	93.3	28	20	71.4
	July, 2002	30	26	76.7	24	19	79.2
	September, 2002	30	30	90	27	14	51.9
Edgartown, MA	**2002 combined**	**90**	**78**	**86.7**	**78**	**53**	**67.9**
Delaware Bay, NJ	June, 2002	28	13	46.4	13	12	92.3
	July, 2002	30	24	80	17	14	82.3
	August, 2002	27	27	100	18	11	61.1
Delaware Bay, NJ	**2002 combined**	**85**	**64**	**75.3**	**48**	**37**	**77.1**
Delaware Bay, NJ	July, 2005	80	38	47.5	30	19	63.3
New Smyrna Beach, FL	October, 2004	48	45	93.8	20	3	15
	September, 2005	48	41	85.4	41	14	34.1
Port Charlotte, FL	September, 2004	100	70	70	53	16	30.2
All Samples		**451**	**336**	**74.5**	**271**	**142**	**52.4**

Table 2-3. Number of alleles at each microsatellite locus for each population sampled. In parentheses, allelic richness is given assuming a sample size of 11.

		Microsatellite Locus							
Sample	**N**	**Pm2232**	**Pm2903**	**Pm2988**	**Pm4488**	**Pm8517**	**Pm9464**	**Pm12067**	**Average**
Edgartown All	26	5 (3.3)	4 (2.8)	6 (3.5)	6 (3.0)	8 (4.7)	5 (2.9)	8 (3.8)	6 (3.4)
Delaware Bay 2002	11	4 (3.0)	6 (4.5)	4 (4.0)	7 (5.2)	7 (5.7)	4 (3.6)	4 (3.6)	5.1 (4.2)
Delaware Bay 2005	11	5 (3.3)	7 (5.3)	4 (3.7)	4 (3.7)	6 (5.0)	5 (3.5)	6 (4.5)	5.3 (4.1)
New Smyrna Beach 2004	17	5 (3.0)	4 (2.4)	5 (3.4)	4 (3.4)	8 (5.1)	2 (1.9)	6 (4.9)	4.9 (3.5)
New Smyrna Beach 2005	27	4 (2.8)	9 (3.2)	6 (2.3)	6 (3.7)	8 (4.7)	3 (2.3)	5 (3.5)	5.9 (3.2)
Port Charlotte 2004	37	3 (1.3)	7 (4.5)	8 (4.0)	5 (3.9)	10 (5.5)	2 (1.2)	8 (4.9)	6.1 (3.6)
All Samples	**129**	**12**	**13**	**12**	**10**	**14**	**11**	**11**	**11.9**
Unsuccessful Amplifications		7	6	13	11	5	9	6	8.1

Table 2-4. Tests of multilocus linkage disequilibrium conducted in MultiLocus v. 1.2.2 using randomized datasets with free recombination to test the significance of $\bar{r}_d$, a measure of the strength of linkage disequilibrium across the genome standardized by correlations between loci. The left portion of the table includes analysis of all genotypes, and, on the right, repeated MLGs were treated as a single genotype to test for epidemic clonal structure.

		All data included				Repeated MLGs treated as single MLG		
Population	**Date**	**N**	**Genotypic Diversity**	**$\bar{r}_d$**	**P-value**	**N**	**$\bar{r}_d$**	**P-value**
Edgartown, MA	2002	26	0.997	0.114	<0.001	25	0.104	0.005
Delaware Bay, NJ	2002	11	1	0.110	0.021			
	2005	11	1	0.160	0.003			
New Smyrna Beach, FL	2004	17	1	-0.004	0.555			
	2005	27	1	0.065	0.011			
Port Charlotte, FL	2004	37	0.996	0.063	<0.001	35	0.021	0.12
All samples		**129**	**0.999**	**0.067**	**<0.001**	**126**	**0.066**	**<0.001**

Table 2-5. Observed (Ho) and expected (He) heterozygosity and resulting average F_{IS} values across all loci with standard deviations (sd). Repeated MLGs were treated as a single genotype. Shown in parentheses are the numbers of loci that reject Hardy-Weinberg equilibrium in each sample by exact tests with sequential Bonferroni correction. The estimate of proportion of selfing required to produce F_{IS} values was calculated as $s = (2 F_{IS})/(1+F_{IS})$ (Hartl and Clark 2007).

Population	Date	Ho	He	F_{IS} (HWE rejections)	sd (F_{IS} across loci)	Selfing
Edgartown, MA	2002	0.794	0.656	-0.254 (3)	0.220	N/A
Delaware Bay, NJ	2002	0.618	0.678	0.066 (1)	0.228	0.124
	2005	0.409	0.647	0.332 (3)	0.226	0.498
New Smyrna Beach, FL	2004	0.235	0.553	0.559 (3)	0.240	0.717
	2005	0.279	0.532	0.474 (3)	0.317	0.643
Port Charlotte, FL	2004	0.223	0.565	0.432 (5)	0.320	0.603
All Samples		**0.403**	**0.703**	**0.424 (7)**	**0.070**	**0.600**

Table 2-S1. Comparison of the observed number of null amplifications with the predicted frequency of null amplifications if null amplification is solely responsible for heterozygosity deficits at each locus when all samples are combined. The predicted frequency of null alleles was calculated according to Hedrick et al. 2005 ((He-Ho)/(1+He)). The number of expected null amplifications is obtained by squaring the predicted null allele frequency and multiplying by the number of individuals sampled at each locus.

Locus	Fis	Expected Null Amplifications	Observed Null amplification
Pm2232	0.162	1.6	7
Pm2903	0.205	2.2	6
Pm2988	0.341	5.0	13
Pm4488	0.365	4.9	11
Pm8517	0.388	5.8	5
Pm9464	0.263	4.0	9
Pm12067	0.383	5.0	6

Figure Legends

Figure 2-1. Multiplicity of infection as predicted by Poisson expectations. A. Poisson expectations for multiplicity of infection were calculated based on the number of uninfected individuals in each locality and in the overall dataset (Poisson mean = -ln(proportion uninfected). Interannual samples were combined in New Smyrna Beach and Delaware Bay. B. Expected numbers of single and multiple infections were calculated by applying the Poisson predictions to the number of individuals that yielded genotypes at four or more loci. Observed parasite genotypes were classified as singly or multiply infected according to haploid and diploid models as outlined in the Methods. In graphs C-F, each sampling location is considered independently, adjusting Poisson expectations for the proportion of uninfected oysters in each location. The diploid model most closely matches Poisson expectations for multiplicity of infection for all sampling locations except New Smyrna Beach.

Figure 2-2. Standardized multilocus linkage disequilibrium ($\bar{r}_d$) in *P. marinus* populations before and after correcting for repeated MLGs. Asterisks indicate values significantly different from zero. Populations are abbreviated as such: Edgartown, MA (EMA), Delaware Bay, NJ 2002 and 2005 (DBA02 and DBA05, respectively), New Smyrna Beach, FL 2004 and 2005 (NSB04 and NSB05, respectively), and Port Charlotte, FL (PCH). All values are significant except for New Smyrna Beach in 2004 and Port Charlotte following correction for repeated MLGs.

Figure 2-3. Neighbor-joining tree of complete genotypes from four geographic locations based on pairwise proportion of shared alleles (Dps). Relationships between individuals are widely variable with little geographical consistency across the tree, indicated by the intermixing of symbols at the branch tips. Most Edgartown samples are at the top of the figure, indicating some consistency among these genotypes, yet there are Delaware Bay genotypes that group with these samples as well. Bootstrap support was only found for nodes close to the tips of the branches and is therefore not shown. Branches leading to repeated MLGs are in bold. The putative multilocus lineage (MLL) in Edgartown has been highlighted with a bracket.

Figure 2-4. Inbreeding coefficients (F_{IS}) for each population in clone-corrected datasets. Populations are arranged from north to south starting with Edgartown (EMA), followed by Delaware Bay 2002 (DBA02) and 2005 (DBA05), New Smyrna Beach 2004 (NSB04) and 2005 (NSB05), and ending in Port Charlotte (PCH). Error bars indicate the standard deviation of F_{IS} among loci for each population. The 2002 sample from Delaware Bay, NJ was the only population that did not significantly differ from panmictic expectations. Edgartown, MA had significantly negative F_{IS} values while the remaining samples showed significantly positive F_{IS}.

Figure 2-S1. Relationship between observed failed amplifications and the number of failed amplifications expected if a null allele is the sole cause of HW deviations in *P. marinus* populations. Each point represents a specific locus from each sampling location except Edgartown, MA. Edgartown samples were excluded from this graph as excess

heterozygosity in this sample does not support null alleles as explaining HW deviations. If a null allele is the sole cause of deviations from HWE, a positive linear relationship would be expected with a slope of 1.0. The relationship is negative here (solid regression line), and does not explain a significant portion of the variation in the data ($p = 0.703$, $r^2 = 0.0429$). Setting the y-intercept at 0 reveals that the slope of the line is less than one (dashed regression line), indicating that there are fewer null amplifications predicted than were observed in the dataset.

Figure 2-S2. Comparison of the distribution of the number of alleles per locus genotyped for samples that were classified as single or multiple infections. The number of alleles per locus was calculated for each individual to take into account the varying number of positively scored loci between individuals. Individuals were then binned in 0.25 increments to display the distribution within each class of individuals (single or multiple infections). The mean of the distributions is significantly different between the two classes (t-test = 14.67, $p < 0.0001$), though there is overlap.

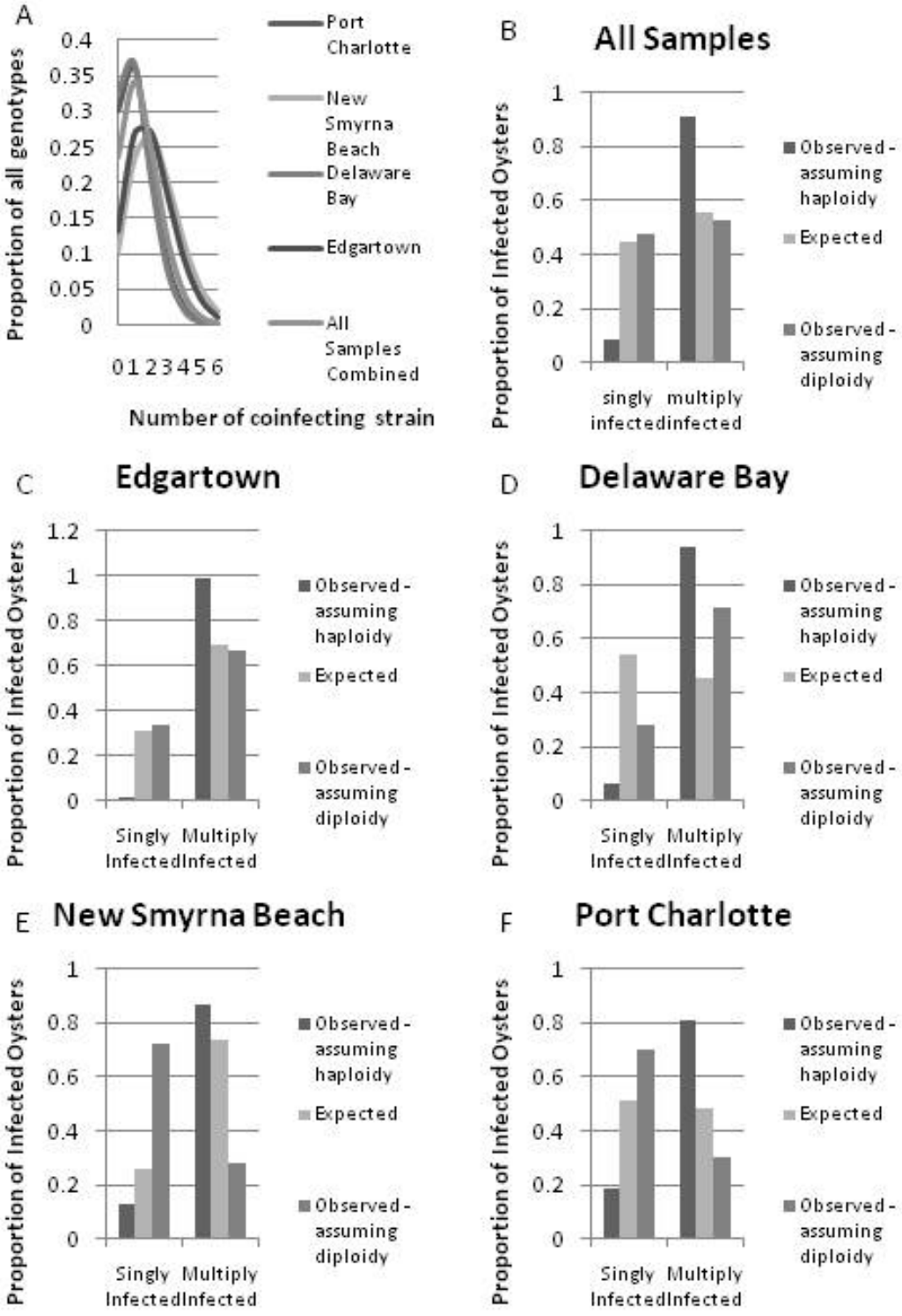
A
Proportion of all genotypes
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0 1 2 3 4 5 6
Number of coinfecting strain
Port Charlotte
New Smyrna Beach
Delaware Bay
Edgartown
All Samples Combined
B
All Samples
Proportion of Infected Oysters
1
0.8
0.6
0.4
0.2
0
singly infected
multiply infected
Observed - assuming haploidy
Expected
Observed - assuming diploidy
C
Edgartown
Proportion of Infected Oysters
1.2
1
0.8
0.6
0.4
0.2
0
Singly Infected
Multiply Infected
Observed - assuming haploidy
Expected
Observed - assuming diploidy
D
Delaware Bay
Proportion of Infected Oysters
1
0.8
0.6
0.4
0.2
0
Singly Infected
Multiply Infected
Observed - assuming haploidy
Expected
Observed - assuming diploidy
E
New Smyrna Beach
Proportion of Infected Oysters
1
0.8
0.6
0.4
0.2
0
Singly Infected
Multiply Infected
Observed - assuming haploidy
Expected
Observed - assuming diploidy
F
Port Charlotte
Proportion of Infected Oysters
1
0.8
0.6
0.4
0.2
0
Singly Infected
Multiply Infected
Observed - assuming haploidy
Expected
Observed - assuming diploidy

Figure 2-1

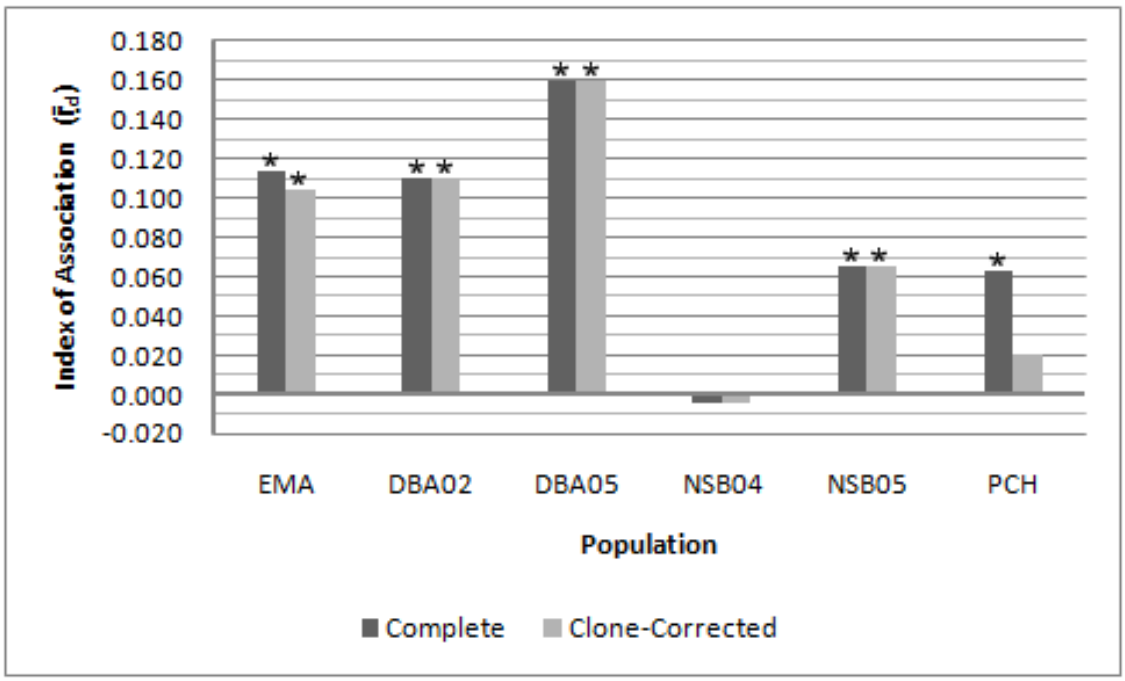
0.180
0.160
0.140
0.120
0.100
0.080
0.060
0.040
0.020
0.000
-0.020
Index of Association ($\bar{r}_d$)
EMA
DBA02
DBA05
NSB04
NSB05
PCH
Population
Complete
Clone-Corrected

Figure 2-2

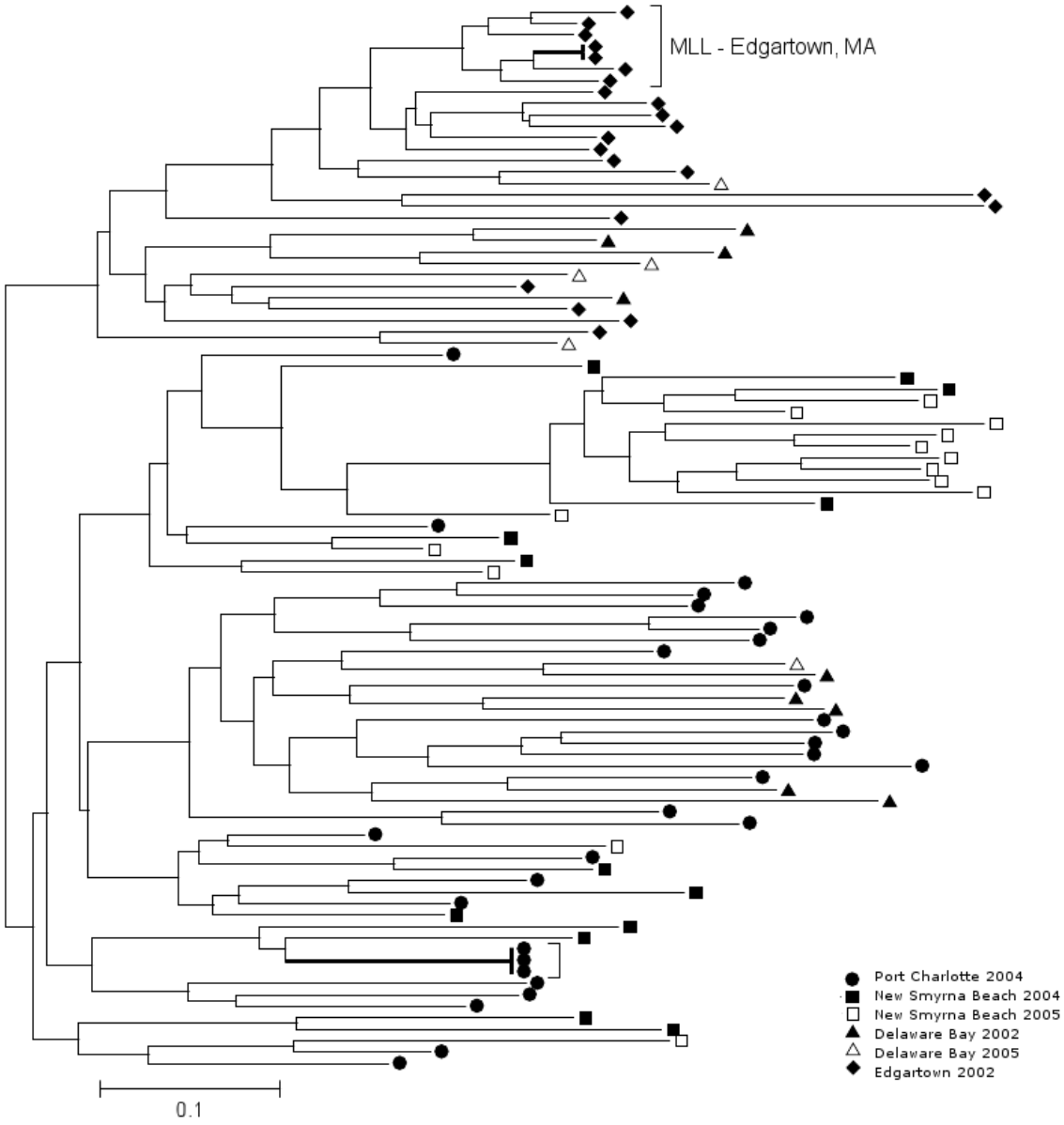
MLL - Edgartown, MA
Port Charlotte 2004
New Smyrna Beach 2004
New Smyrna Beach 2005
Delaware Bay 2002
Delaware Bay 2005
Edgartown 2002
0.1

Figure 2-3

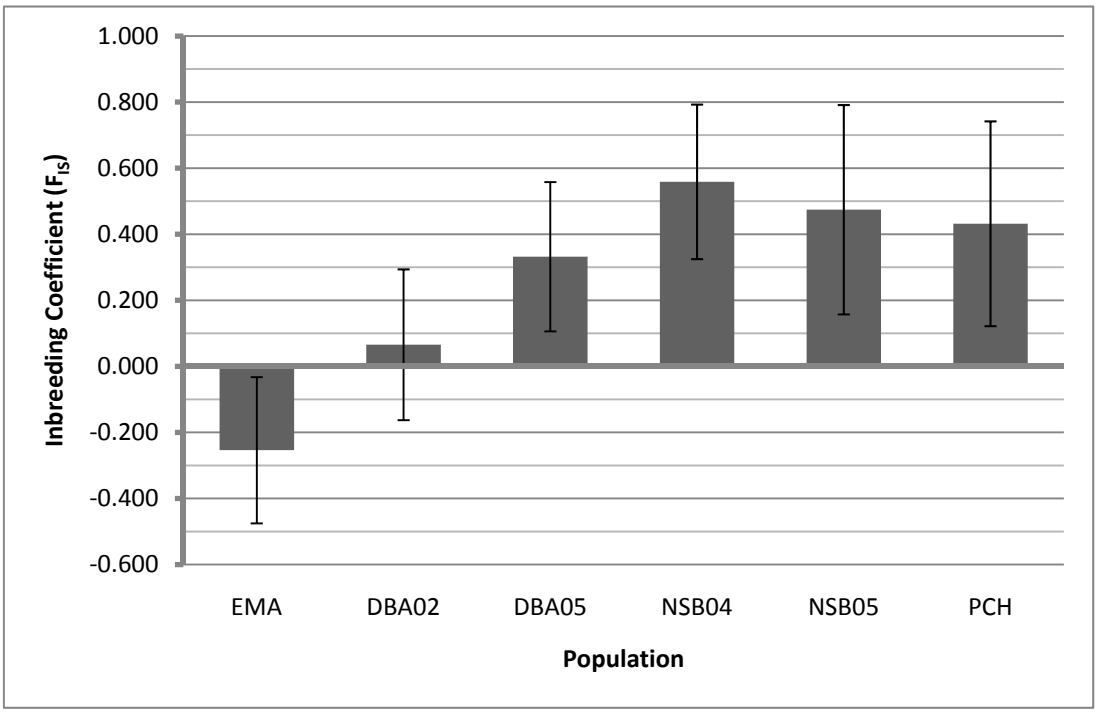
1.000
0.800
0.600
0.400
0.200
0.000
-0.200
-0.400
-0.600
Inbreeding Coefficient (F_{IS})
EMA
DBA02
DBA05
NSB04
NSB05
PCH
Population

Figure 2-4

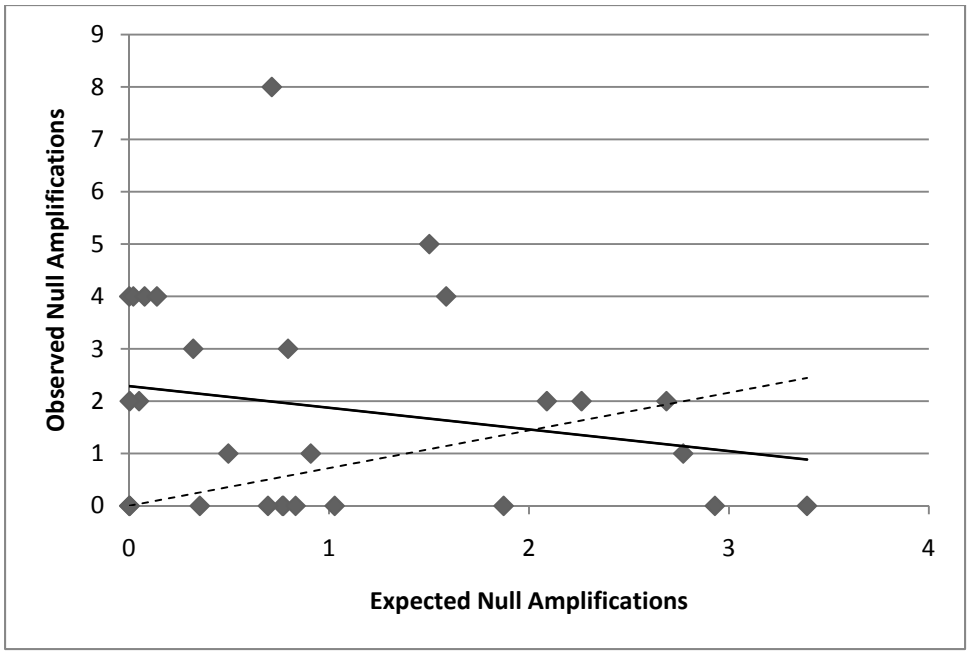
Observed Null Amplifications
Expected Null Amplifications
0
1
2
3
4
5
6
7
8
9
0
1
2
3
4

Figure 2-S1

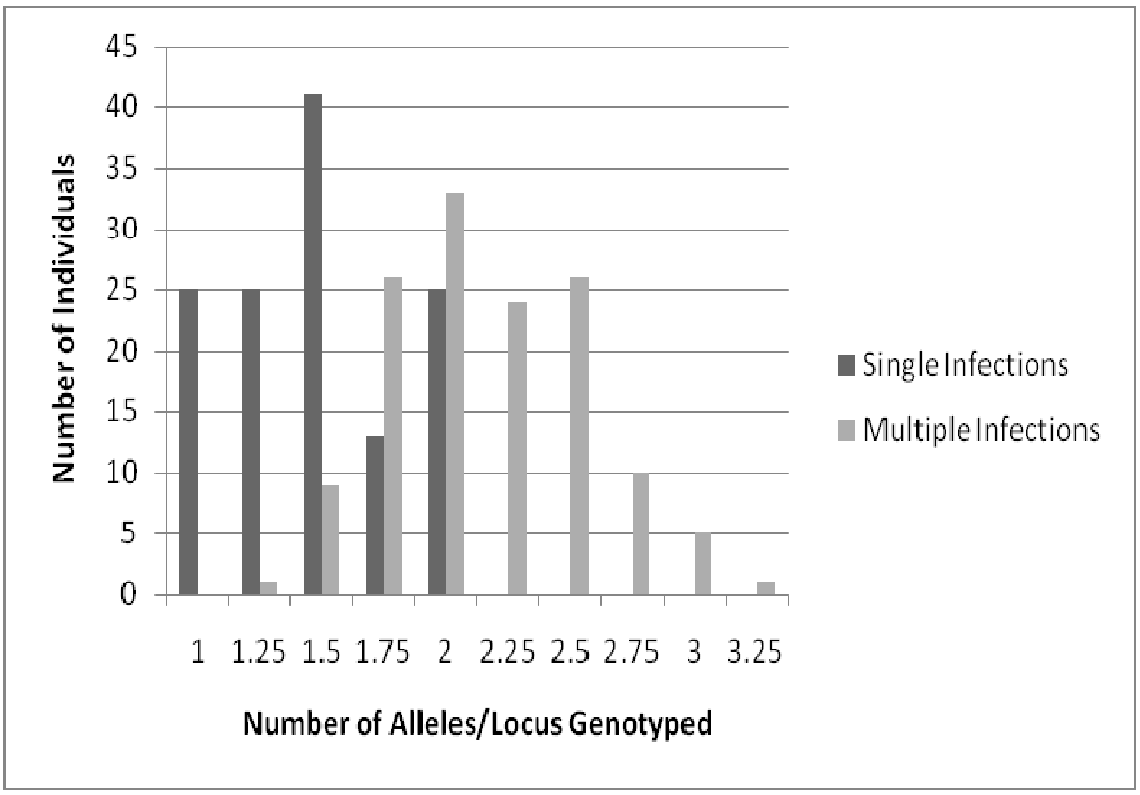
45
40
35
30
25
20
15
10
5
0
Number of Individuals
1
1.25
1.5
1.75
2
2.25
2.5
2.75
3
3.25
Number of Alleles/Locus Genotyped
Single Infections
Multiple Infections

Figure 2-S2

Chapter Three - Microsatellite genotypes reveal that some *Perkinsus marinus* strains disperse widely while others are locally restricted.

Abstract

Perkinsus marinus is a parasite of the eastern oyster *Crassostrea virginica* causing significant mortality and reduced fecundity in host populations. In order to understand the potential for novel strains to invade new territories and cause epizootics, estimates of gene flow among locations are required. Spread of infection between oysters occurs directly, as millions of parasite propagules are released at the death of the host and filtered from the water column by neighboring oysters. This dispersal mechanism is expected to be subject to control by water currents in estuarine systems, limiting connectivity among distant parasite populations. In order to investigate genetic diversity within locations and gene flow among locations, *P. marinus* was genotyped using seven diverse microsatellite loci directly from infected oyster genomic DNA collected from 15 geographic locations encompassing 4800 km of coastline. Among 1082 oysters sampled, 374 were determined to be infected by a single parasite genotype. Most locations were genetically distinct, but genetic distances among populations were not correlated with geographic distances between sampling locations. Genetic diversity was not precisely subdivided into three genetic regions as previously hypothesized. Rather, analyses showed that each location was characterized by multiple genetic lineages in varying proportions. Four distinct clusters of multi-locus genotypes were identified using a

variety of analytic procedures (based on genetic distances between individuals, allele frequencies, and maximization of Hardy-Weinberg equilibrium using a Bayesian algorithm). One of these clusters was widespread among many locations, while the remaining three were geographically limited to a particular region or pair of disjunct locations. There was no evidence for isolation by distance within the widespread cluster, suggesting that dispersal of parasites belonging to this genotypic cluster was recent. Altogether, local assemblages of *P. marinus* are collections of distinct genotypic clusters that coexist within locations with infrequent recombination, such that local parasite assemblages may not be accurately analyzed as a group. Long distance dispersal may play an important role in the spread of *P. marinus,* whereas locally abundant strains may represent focal epizootics.

Introduction

Dispersal of estuarine and marine organisms is a complex process that may depend on many factors. In the particular case of estuarine parasites, passive transport through water movement (Rohde 2005), host transport (Smith 2001; Lafferty 1996), environmental cues (Haas 1992), and larval behavior (Fingerut et al. 2003) have all been implicated as contributing to parasite dispersal. As these parasites may affect community structure through negative effects on their hosts (Minchella and Scott 1991), it is important to estimate the movement of parasites – or evolutionarily effective movement in terms of gene flow - within and among estuarine systems in order to predict outbreaks of disease and understand the coevolutionary potential of host-parasite interactions (Dybdahl and Lively 1996). Recent periods of warmer than average water temperatures

have led to a rise in reports of disease epidemics, often associated with shifts in pathogen distribution (Harvell et al. 1999). *Perkinsus marinus* is one such pathogen, recently expanding its range by over 600 km northward (Ford 1996). This expansion was associated with warmer winter water temperatures in the northeastern United States (Cook et al. 1998; Ford 1996; Ford and Chintala 2006; Ford and Smolowitz 2007). In order to understand recent range expansions and pathogen epidemics, we must first understand local parasite strain assemblagesand degree of connectivity among geographically disparate locations. In the absence of physical experimentation, population genetics provides a tool to investigate the evolutionary history of parasite populations at ecological and evolutionary time scales.

The protozoan parasite *Perkinsus marinus* causes significant disease in populations of the Atlantic oyster *Crassostrea virginica*. The host is a keystone species in estuaries along the east coast of North America, acting as an ecosystem engineer by building reefs and providing substrate on which other organisms depend, clarifying the water column through filtration, and movement of suspended organic material to the benthos (Coen et al. 2007). Despite the negative affects *P. marinus* has on its commercially and ecologically important host, little is known about the genetic diversity within local populations (but see Reece et al. 1997; Robledo et al. 1999; Reece et al. 2001) and the degree to which the parasite can disperse between locations.

Previous genetic studies have shown variation in genotypes among *P. marinus* isolates broadly divided into three geographic regions; the Gulf of Mexico, estuaries of the southeastern Unites States, and a newly expanded range north of Chesapeake Bay extending into Canada (Reece et al. 2001). This study used restriction fragment length

polymorphisms which contain limited variation to genotype relatively few isolates from any single locality. These constraints limited their ability to infer local population processes and connectivity among locations.

P. marinus has a history of epizootic outbreaks associated with warm water temperatures and high salinity; oyster mortality and disease have a positive correlation with these two environmental parameters. In the Gulf of Mexico, prevalence is often high, ranging from 49-100% in 49 sampled locations in one study (Craig et al. 1989). InDelaware Bay, there were sporadic outbreaks of the parasite in the 1950's and 1960's, associated with importation of oysters from southern populations, that lasted for several years and then disappeared without intervention (Ford 1996). Following these outbreaks, movements of oysters were limited to within estuaries in order to prevent further introduction of the parasite. Despite this precaution, *P. marinus* appeared in lower salinity regions of Chesapeake Bay and Delaware Bay in the mid-1980's and has since become established with no signs of abatement (Burreson and Ragone Calvo 1996). The parasite range has continued to expand; *P. marinus* is now found routinely in estuaries as far north as Maine (Ford and Smolowitz 2007).

P. marinus dispersal mechanisms are not fully understood. New infections have been observed within three months in disease free oysters separated from naturally infected reefs by 5 km (McCollough et al. 2007). There is no known vector for distribution of *P. marinus* to uninfected oyster hosts, though natural vectors may exist (White et al. 1989). Oysters are readily infected by ingesting the parasite directly from the water column (Ray 1954; Chu 1996). Given the small size and limited swimming ability of the parasite (Perkins 1996), it is difficult to understand how this parasite has

undergone a 600 kilometer range expansion in less than 30 years. It is possible that the parasite is capable of living independently of the host for extended periods of time, allowing currents to transport the parasite large distances before infecting a new host. This may also be compounded and complicated by anthropogenic movement of the parasite through intentional (oyster transport for commercial processing) or unintentional (ballast water) means. Genotypic composition of local populations may provide clues as to the scale of gene flow among *P. marinus* populations.

Population genetics provides a means to understand the connectivity of populations of organisms. Genotypes of individuals can be used to infer the relatedness among individuals, their geographic origins, and whether populations are freely interbreeding or primarily clonal. If recombination is frequent, a single migrant per generation can limit population differentiation that would otherwise accumulate through genetic drift (Spieth 1974). Therefore, it is expected that populations of organisms in close proximity to each other will share genetic attributes, whereas populations further apart will gradually differentiate over time. Eventually, such processes can lead to a pattern of isolation by distance, in which population differentiation increases with geographic distance (Wright 1943). This pattern has been observed in many organisms, including protistan parasites (Barbujani and Sokal 1991; Anderson et al. 2000; Llewellyn et al. 2009), and may be a reasonable expectation for estuarine species whose movement should be governed by water currents. Prevailing currents may direct the movement of propagules downstream.

Here, I attempt to understand the connectivity of *P. marinus* populations by sampling and genotyping individuals from a wide geographic range. In particular,I

sought to characterize genetic variance within and among local populations to assess evidence for migration among locations. To do so,I used seven highly variable microsatellite loci to characterize each sample and population genetic analyses to investigate the distribution of genotypes geographically. I used a combination of individual clustering analyses and population metrics to understand the relatedness of populations and infer gene flow of the parasite across estuaries of the eastern coast of North America.

Materials and Methods

Sample Collection

Oysters were collected from 15 locations from Texas to Massachusetts along the east coast of the United States of America. Sampling locations are shown in Table 1. Samples from two years were obtained in New Smyrna Beach and Delaware Bay. Samples from at least two months were obtained for one year in Delaware Bay, Narragansett Bay, and the Massachusetts samples. Details of collection sites are found in Table 1 including latitude and longitude, date of collection, and the temperature and salinity at the time of collection. The total numbers of oysters collected are shown in Table 2. Oysters were graciously collected by numerous scientific research institutions and shipped overnight at ambient temperatures or on ice for processing at the University of Maryland College Park.

DNA extraction and amplification

Upon arrival, oysters were shucked and 1 gram of mantle and gill tissue was stored in 95% ethanol for long-term storage. Approximately 50 mg of gill tissue was placed in 1.5 ml tubes for DNA extraction. DNA was extracted from this initial sample

using Bio-Rad Chelex-100 resin (Bio-Rad Laboratories, Hercules, CA). Briefly, 200 µl of a 10% Chelex slurry in TE was added to each tissue sample. Samples were agitated vigorously and heated to 95°C for 20 minutes. After a brief cooling period, tubes were centrifuged at 4000 rpm for 5 minutes. 100 µl of supernatant containing oyster DNA (and potentially *P. marinus* DNA) was transferred to a clean tube making every effort to avoid transferring any chelex resin or oyster tissue.

Individual DNA extracts were tested for infection by amplifying *P. marinus* nuclear ribosomal non-transcribed spacer (NTS) DNA using PCR as described previously (Robledo et al. 1998). Following amplification, DNA was separated on 1.3% agarose gels and stained with ethidium bromide. Infection was indicated by the presence of a 300 base pair band. Each set of PCRs included no-DNA and uninfected oyster DNA controls to ensure that bands were *P. marinus* specific. As Chelex-extracted DNA is reported to have limited stability, success in amplifying the NTS locus was used as a basis to perform, from such oysters, a second DNA extraction procedure. Here, 100 mg of oyster tissue was subjected to extraction using QIAgen DNEasy tissue kits following the animal tissue protocol recommended by the manufacturer (QIAgen, Alameda, CA). DNA was eluted using 200 µl of molecular biology grade water and stored at 4 °C.

Seven microsatellite loci were amplified from DNA as previously described (Chapter Two, Thompson dissertation). Briefly, 35 cycles of PCR was used to amplify each locus independently from 2 µl of template. Amplified DNA was separated using an Applied Biosystems 3730 DNA analyzer along with an appropriate LIZ standard for sizing (Life Technologies, Carlsbad, CA). Chromatograms were analyzed using Genemapper v. 3.7 (Life Technologies). Bins were created to score microsatellite size

automatically with bin size determined by the repeat motif. Samples were considered multiply infected if any single locus had three or more bands. All bands greater than 500 RFU were scored in order to minimize the number of multiple infections scored as single infections. This may have eliminated some samples from further analysis that were singly infected, but there were no adequate criteria to assess whether bands were spurious. Infected samples that did not amplify for at least four loci following repeated attempts were removed from subsequent analyses. Genotype data were organized using GenAlEx 6.1 (Peakall and Smouse 2006) in Microsoft Excel 2007. Identical multilocus genotypes (MLGs) were identified using GenAlEx. Individuals with missing data were not considered for this analysis. The probability that repeated genotypes occurred through independent sexual recombination events (p_{sex}) was calculated based on sampling location allele frequencies. For population genetic analysis identical multilocus genotypes (MLGs) at a single locale were treated as a single genotype as they were most likely the result of asexual reproduction.

Genetic Analysis

Genotypes from individual sampling locations were tested for equilibrium expectations. Hardy-Weinberg Equilibrium was tested for each population in Genepop v.4.0. F_{IS} was estimated in GenAlEx for each locus by sampling location and standard deviations among loci were calculated.

Relationships among multilocus genotypes were examined using three distinct analyses for clustering. Neighbor-joining trees of complete genotypes were constructed from dmu^2 genetic distances using Populations v. 1.3 (Langella, 2001) and bootstrapping across loci. Additionally, Factorial Correspondence Analysis (FCA) was conducted

using Genetix v. 4.03 (Belkhir et al. 1996) to visualize relatedness of individual genotypes in a three-dimensional framework based on allele frequencies. A final clustering analysis was conducted using the Bayesian clustering program Structure v.2.2 (Pritchard et al. 2000) with an admixture model and correlated allele frequencies among populations. Sampling location was ignored in all runs. Ten runs of 1,000,000 generations were conducted for proposed number of populations (K) of one to fifteen. Structure analyses were run using the resources of the Computational Biology Service Unit from Cornell University, which is partially funded by Microsoft Corporation. Ln likelihood values were used to generate delta K in order to determine the most likely number of populations across all Structure runs (Evanno et al. 2005). Consensus among independent Structure runs was evaluated using the full search option in CLUMPP v. 1.1.2 (Jakobsson and Rosenberg 2007). Results are shown for k= 1 to15 in supplementary Figure S1 with error bars corresponding to the standard deviation of likelihood among runs to show consistency of patterns among Structure runs for different values of k. Results of Structure analyses were confirmed by analysis with a different Bayesian population inference program, Structurama (Huelsenbeck and Andolfatto 2007). Five replicates of 1,000,000 generations were used with 50,000 generations used as burn-in, sampling every 10 generations, with gamma shape and scale parameters of 2.5 and 0.5, respectively. A single long run with 10,000,000 generations was run with 100,000 generations of burn-in sampling every 25 generations.

Arlequin v. 3.01 (Excoffier et al. 2005) was used to calculate genetic distances between populations and test *a priori* hypotheses of genetic population structure. Both Fst and Rst genetic distances were estimated. Subsequently, Mantel tests were conducted

to test whether geographic and genetic distances were correlated, using GenAlEx with 999 randomizations. AMOVA was conducted to test three *a priori* hypotheses of genetic population structure. These *a priori* hypotheses included: a single genetic break at Cape Canaveral, coincident with a genetic cline in the host population (Hare and Avise 1996), a genetic break between east and west coast of Florida and one at the mouth of Chesapeake Bay (Reece et al. 2001), and, finally, regions described by Engle and Summers (1999) based on distributions of benthic invertebrates. To test the validity of biogeographic regions hypothesized by Engle and Summers, sampling locations were subdivided into Gulf of Mexico samples, eastern Florida samples, South Carolina samples, Delaware Bay samples, with Rhode Island and Massachusetts forming the final group. AMOVA was conducted using 10000 permutations of the data.

Results

Among 1082 total oysters assayed, 746 were determined by NTS amplification assay to be infected (prevalence = 68.9%) (Table 3-2). Prevalence exceeded 45% in all geographic localities except in Narragansett Bay, RI during June, 2002, where only 3 out of 32 oysters were infected (9.4%). Of the 746 oysters judged infected on the basis of the NTS assay, 574 also sustained amplification at four or more other loci (77%); we do not know whether insufficient template or PCR inhibitors may explain the failure of the remaining 23% to sustain consistent amplifications at microsatellite loci. Doubling the amount of template resulted in amplification of 65% of loci where previous amplification attempts failed. Increasing magnesium concentration to 3.5 mM resulted in amplifying only two of twelve loci from a subset of amplifications that failed initially; this was not pursued further. Three or more alleles at a given locus were observed in 216 oysters

(37.6%) genotyped for 4 or more loci; these were classified as oysters infected with more than one parasite genotype (multiple infections). Multiple infections ranged in frequency from 0% to 77.1% among locations. They were most frequent in Delaware Bay, NJ (77.1%) and Edgartown, MA (67.9%), both locations associated with parasite range expansion. Following removal of multiply infected individuals, 374 multilocus genotypes were used for population genetic analysis.

12 to 18 alleles were observed at each microsatellite locus. Alleles were distributed among populations such that each population had, on average, 3.8 alleles per locus. At each locus, some alleles occurred at very low frequency whereas others occurred at very high frequencies; consequently, the effective number of alleles was considerably lower than the absolute number of alleles in each case (Fig 3-1A). The extent of allelic diversity varied significantly among loci (ANOVA, $F = 14.3$, $p < 0.0001$); the highest diversity occurred at locus Pm8517 ($Ne = 7.2$) and the lowest at locus Pm9464 ($Ne = 1.8$). After correcting for differences in sample size ($N = 11$), using rarefaction in FSTAT v 2.9.3, an ANOVA indicated that the average allelic richness significantly differed among populations ($F = 2.90$, $p = 0.0013$) with Cape Canaveral, Florida having the fewest alleles (2.85 alleles/locus) and Dauphin Island, Alabama having the most (5.02 alleles/locus) (Fig 3-1B). There were no latitudinal trends in allelic richness.

For locations in which collections were made across several months within the same year, there was a trend observed for fewer multiple infections as the year progressed. Logistic regression analysis of this trend treating each infected individual as either being singly or multiply infected indicated that this trend was statistically

significant ($p < 0.0001$, $r^2 = 0.76$). Collection location was not a significant factor for predicting the number of multiple infections ($p = 0.95$). However, when the proportion of multiple infections was plotted by location and month, the sample from June 2002 in Delaware Bay, NJ appeared to be an outlier (lower left, Fig. 3-2). Only three infected oysters were detected in this sample, potentially reducing power to estimate the actual multiple infection rate in Delaware Bay at this time.

Null Alleles

There were consistently loci that did not amplify for certain samples despite multiple PCR attempts. Among 2618 amplifications, 223 PCR reactions yielded no product (8.5%). Thesewere distributed among all loci ranging in frequency from 6% at locus Pm12067 to 11% at locus Pm9464. Every sampling location had at least one such failure. Failed amplification frequency was high in Alabama and South Carolina samples (19 and 21% respectively). In South Carolina, most failed amplifications were observed at locus Pm9464, for which 57% of samples did not amplify. In Alabama, failed amplifications were distributed among all loci except Pm12067, with the highest frequency at locus Pm2232 (34%). The remaining sampling locations varied in failed amplification frequency from 3 to 12%.

Clonality

In order to examine the degree of clonality within sampling locations, repeated genotypes were documented and inbreeding coefficients (F_{IS}) were calculated. Among 374 multi-locus genotypes (MLGs), six were observed more than once. Only one repeated MLG was observed south of Chesapeake Bay, whereas five were observed in locations into which *P. marinus* is believed to have only recently expanded. Three

different MLGs were repeated in Narragansett Bay, RI. Given the allelic diversity within individual sampling locations, it is unlikely that these repeated MLGs were generated through independent sexual recombination events ($p_{sex} < 0.0001$ for all pairwise comparisons). Additionally, most locations had inbreeding coefficients (F_{IS}) significantly greater than 0 (Fig. 3-3), indicating that when sexual recombination occurs, it most frequently involves two closely related genotypes. Positive F_{IS} values ranged from 0.246 in Delaware Bay, NJ to 0.899 in Fort Pierce, FL. In Fort Pierce, self-fertilization may be as high as 94% when recombination occurs. In Massachusetts populations (Tisbury and Edgartown), Fis is significantly negative, with strong linkage disequilibrium across the genome (standardized Index of Association, rbard, $p < 0.001$, Thompson Dissertation Chapter 2).

Cluster Analysis

An unrooted neighbor-joining tree based on inter-individual dmu^2 distance was constructed from those genotypes with no missing data (n= 247). Bootstrap support was not found for internal nodes of the tree, yet four clusters of genotypes were apparent within the tree (Figure 3-4). In 11 of 15 sampling locations, a majority of genotypes belonged to a cluster at the center of the tree. In the remaining 4 locations (Dickinson, TX, New Smyrna Beach, FL, and Edgartown and Tisbury, MA), few isolates were characterized by genotypes represented in that central cluster; instead, most isolates from those populations had genotypes occupying the periphery of the neighbor joining tree. At one end of the tree, genotypes from New Smyrna Beach formed a distinct group, with significant representation of individual genotypes from Port Charlotte, FL and one individual from Dauphin Island, AL. At the opposite end of the tree, a cluster of

genotypes was observed exclusively in locations into which *P. marinus* has recently expanded (Delaware Bay, NJ, Tisbury, MA and Edgartown, MA). Notably, Naragansett Bay, RI was entirely lacking in genotypes corresponding to this northeastern cluster, in spite of its geographic proximity to both Delaware Bay and Massachusetts. Finally, there is an intermediate branch in the tree that is composed primarily of samples from Dickinson, TX (the westernmost sampling locale). Those genotypes falling outside the circled clusters did not occur at high enough frequencies to sustain generalizations regarding their geographic distinctiveness.

Factorial correspondence analysis (FCA), implemented in Genetix, supported the clustering of genotypes observed in the individual distance tree when all genotypes were included (Fig. 3-4). The three primary factors explained 49.32% of the variation in the data with 20.31%, 15.72%, and 13.29% of the variation explained by each component, respectively. Three dimensional graphical representation of FCA showed considerable spread among individual genotypes, yet some clustering of genotypes is evident (Fig 3-4). A large cluster of genotypes was composed of individuals from all populations except Dickinson, TX (located on the left portion of the graph, which corresponds to the central cluster on the individual distance tree). Three other elliptical clusters were composed of fewer individuals, each representing a limited geographic distribution. As in the neighbor-joining tree, one cluster was comprised of samples from New Smyrna Beach and Port Charlotte, locations on opposite sides of the Florida Peninsula. A single sample from Apalachicola, FL clustered with these samples as well. As in the neighbor-joining tree, another cluster was entirely composed of strains from either Delaware Bay or Massachusetts. As in the neighbor-joining tree, a final cluster consisted almost entirely

of samples from Dickinson, Texas (with a single strain from Alabama represented at the edge of the cluster).

Bayesian cluster analysis, using two different algorithms, inferred varying numbers interbreeding populations approximating Hardy-Weinberg Equilibrium frequencies. Structure, using the admixture model with correlated allele frequencies, indicated that the dataset should be subdivided into 6 populations (as judged by the method of Evanno et al 2005). The likelihood of proposed population subdivisions (K-values) rose quickly from one population to six populations, leveled off between six and seven populations, and then continued to rise as more populations were proposed; but the increase in log likelihood was relatively gradual beyond 6 populations (fig. 3-S1a) resulting in a mode for deltaK at 6 populations (Fig. 3-S1b). To confirm this inference, cluster analysis was conducted using Structurama. Among 5 replicates, the average number of inferred populations was 6.6944 indicating higher support for 7 populations than 6, but in general agreement with the Structure results. The Structurama posterior probability distribution shows support for 5 to 9 populations with greater than 90% probability for 6, 7, or 8 populations (Fig. 3-S2).

The Structure results mirrored the results of both phylogenetic trees and FCA analysis of genotypes, with slight deviation regarding the central cluster of genotypes (Fig. 3-4). Considering the general agreement between clustering methods, the data were regrouped into six genotypically distinct groups for further analysis, sorting genotypes by Q in Structure (as this method provides discrete assignment of genotypes, though some individuals are highly admixed). As indicated by the colors in the barplot for K=6 (Fig. 3-3), the Orange, Green, and Blue groups correspond to Dickinson,TX, New Smyrna

Beach, FL, and northeastern genotypes respectively. The Purple, Pink, and Yellow groups correspond to the poorly defined cosmopolitan group of genotypes at the center of the individual genetic distance tree (Fig. 3-4).

After regrouping the genotypes, clusters inferred by Structure showed significant deviations from Hardy-Weinberg Equilibrium (HWE), with high F_{IS} values in 5 groups (F_{IS} range 0.469-0.637) and significant negative F_{IS} (-0.205) in the predominantly northeastern group. No more than one locus was consistent with HWE for any Structure-inferred cluster.

Geographic Distribution of Clusters

Individuals sampled from certain locations possessed genotypes characterized by restricted ranges, but most locations were characterized by several widely distributed genotypic groups with a greater degree of admixture than those endemic genotypes. Barplots assigning each individual genotype to one of 6 populations inferred by Structure are depicted according to geographic sampling location in Fig. 3-5. A distinct population encompassed most isolates from New Smyrna Beach, FL and a few isolates from Port Charlotte, FL (green bars). Dickinson, Texas genotypes were assigned to their own, distinct population (orange bars) with a few representatives in Alabama, Cape Canaveral, FL, and South Carolina. Most genotypes from Delaware Bay, NJ, Tisbury, MA, and Edgartown, MA were assigned to a third distinct population (blue bars) with high probability. Notably, all but one genotype from Massachusetts samples assigned to this population. All genotypes not associated with Dickinson, TX, New Smyrna Beach, FL, or Martha's Vineyard populations clustered into three geographically indistinct populations that collectively correspond to the central cluster in the neighbor- joining

tree. The yellow and pink populations were at highest frequencies in the center of the sampling distribution, while the purple population was observed at high frequencies in the Gulf of Mexico. Interestingly, individuals from Narragansett Bay, RI clustered with the widely distributed populations, and did not resemble individuals from other, geographically proximate northeastern populations. More individuals in the geographically diverse group showed admixture than genotypes from geographically localized clusters, as indicated by the multiple colors in individual bars.

Population Analysis

In order to test for similarities between sampling locations, pairwise comparisons of Fst and Rst using exact tests implemented in Genepop v. 4.0 were conducted (Table 3-3). Fst analysis found that 101 of 105 pairwise population comparisons were significantly different while Rst found differences in 90 of 105 pairwise comparisons. Neither Fst or Rst had a significant correlation with geographic distance based on Mantel tests ($p = 0.334$ or 0.448 for Fst and Rst respectively, data not shown). Dickinson, TX and New Smyrna Beach, FL differed most from all other populations. Alabama samples were least distinctive. Alabama is not genetically distinct from Port Orange, Florida or Delaware Bay, New Jersey despite the observation that Port Orange and Delaware Bay are differentiated from each other.

In order to assess whether there were consistent phylogenetic relationships between particular sampling locations, phylogenetic trees were constructed based on multiple genetic distances (Cavalli-Sforza's Dc, dmu^2, Nei's Ds, and Goldstein's ASD) among locations with bootstraps on loci in order to delineate population relationships not captured by Fst or Rst analysis. In 3 out of 4 distance trees, marginal support was found

for grouping Tisbury and Edgartown, Massachusetts samples (BS = 100, 65, 44, and 56 for each distance respectively). No other pairing was supported consistently across trees. Furthermore, the topology of the phylogenetic tree changed for each genetic distance, suggesting that grouping genotypes by sampling location does not capture the evolutionary history of *P. marinus* populations adequately.

A priori AMOVA models were tested to evaluate explicit geographical divisions among genetic groups. The data do not conform to divisions between the Gulf of Mexico, southeastern United States, and Northeastern United states populations as proposed by Reece, et al (2001) (FCT = 0.021, p = 0.383), nor to a model of coevolution with the host, as no genetic break between parasite populations is observed around Cape Canaveral (FCT=0.091 , p = 0.065). A model based on the zoogeographic province boundaries put forth by Engle and Summers (1999) reveal some differences between regions (7% of total variation), but this differentiation is not a significant portion of the overall variance in the samples (FCT= 0.070, p = 0.18). Taken altogether, regional models do not fit the population genetics of *P. marinus*.

Discussion

Microsatellite genotyping has revealed previously unknown genotypic heterogeneity within local populations of *P. marinus*. The allelic composition of most locations is significantly different from all other locations, as indicated by Fst and Rst genetic distances. This differentiation does not follow an isolation by distance pattern according to a stepping-stone model along the east coast of North America, nor does it correspond precisely to regional patterns previously described (Reece 2001). Rather, each location is characterized by a varying collection of distinct genotypic groups in

differing proportions, as indicated by genetic distance relationships between individual samples, allelic frequency correlations, and Bayesian cluster analysis. Certain locations share high proportions of genotypes belonging to the same group, but frequently those locations are not geographically proximate. In northeastern locations where dermo disease has recently become apparent, there were two genotypic clusters represented in neighboring locations, suggesting at least two distinct parasite introductions into the expanded range. These observations suggest a complex history of dispersal between locations, differing substantially from population genetic equilibrium expectations.

The distribution of genotypic clusters among locations generally matches the regional distributions described by Reece et al 2001, but in all three regions, there is at least one population that stands out as different from neighboring locations in its genotype composition. This does not invalidate previous findings, but rather shows that by sampling more individuals within a given location, a more complex picture emerges of the diversity within regions, confirming previous observations for population samples at a single locus (Robledo et al. 1999). Gulf of Mexico and southeast Atlantic locations had more genotypic clusters in common than did northeastern populations (though proportions of resident clusters varied between most locations), conforming to the previous observations (Reece et al. 2001). However, the predominant genotypic groups in Dickinson, TX, Port Charlotte, FL, New Smyrna Beach, FL, and Narragansett Bay, RI differed from other regional samples, interrupting any regional pattern. Variation in local genotypic composition suggests that local conditions may promote the rise of specific genotypes. However, in the majority of locations, diverse genotypes coexisted. Relative proportions of resident parasite strains may vary depending on prevailing local

environmental conditions and the ecological characteristics of local parasites in a widespread geographic mosaic (Thompson 2005). These were not unusually small sample sizes and in one case (New Smyrna Beach), the unique assemblage was stable for two years.

The use of rapidly mutating microsatellite markers provided greater resolution for distinguishing genetically unique parasites. Great allelic diversity allowed detection of multiple infections in oysters. Frequencies of multiple infections were highest in northeastern populations, but this may be an artifact of detectability, as northeastern genotypes had more heterozygous loci making it more likely that three alleles would be observed in multiply infected individuals.

After removal of multiple infections, 367 unique genotypes were identified among 374 individual genotypes. Six genotypes were found at least twice in the dataset, consistent with asexual spread of these specific genotypes. Among the six repeated genotypes, five were present in northeastern populations where *P. marinus* is believed to have recently arrived, suggesting that asexual spread is more common in these populations. According to three clustering methods, the 367 unique genotypes belonged to four to six genotypic clusters. With the exception of the genotypic group prevalent in New Smyrna Beach, FL, each cluster was widely distributed, though some more widely than others. Thus, despite high microsatellite genotypic diversity, *P. marinus* genotypes can be analyzed as a limited number of parasite lineages with unifying genetic characteristics.

Due to the primarily clonal reproduction in *P. marinus*, multiple genotypic clades can be maintained for long periods of time in particular locations. All locations

contained a collection of disparate genotypic clusters, but not evolving as a group for lack of sex. When any location is considered as a single interbreeding population, significant deviation from HWE is observed (data not shown). HWE deviations may be explained primarily by non-random mating, as indicated by high F_{IS} values in most populations. Mutations in asexual offspring of specific lineages may also limit our ability to remove redundant genotypes from HW analyses, exacerbating the problem. Some individuals show evidence of admixture between lineages in Bayesian clustering results, which may represent new recombinants between genotypic clusters. Among 368 unique multilocus genotypes, 239 individuals (65%) were assigned to a particular cluster with greater than 80% identity. The remaining admixed genotypes were observed primarily in locations typified by the widespread genotypic groups in the Gulf of Mexico and along the southeast Atlantic coast. These correspond to the purple, pink, and yellow clusters in the Structure graphs. Null amplifications may also contribute to deviations from HWE, but these are spread across all loci and evenly distributed among locations suggesting that these may be due to inherent difficulties in amplifying directly from infected host tissue (discussed in Thompson Dissertation Chapter 2).

The widespread geographic distribution of some genotypic clusters indicates that migration is frequent among locations. All groups were found in multiple locations, with two groups represented intwelve locations (encompassing almost 4000 kilometers of coastline). The smallest geographic distribution along the coast for any cluster was 3100 kilometers (yellow cluster from Fig 3-4 at >50% in Port Charlotte on west coast of Florida) indicating that genetically distinct parasites are dispersed widely. Water currents may be primarily responsible for local dispersal of *P. marinus,* as indicated by regional

associations of some groups. Some clusters are shared between geographically distant locations yet are absent in intervening locations. This suggests that the genotypic cluster has gone extinct in the intervening locations, that not enough infected oysters were sampled in intervening locations to detect low frequency clusters, or that long distance dispersal occurred between disjunct locations. Long distance dispersal may occur through anthropogenic movement of parasites between distant locations or through an as yet unidentified natural vector.

The distributions of lineages among populations north of Chesapeake Bay, where the parasite is thought to have invaded only recently, indicates that at least two separate introductions occurred in this region. Genotypes in Narragansett Bay, RI differ markedly from other northeastern genotypes. Parasites in Narragansett Bay were probably introduced directly from a location south of Chesapeake Bay, though the source cannot be pinpointed with the current data. The predominant genotypic cluster observed in Delaware Bay and two locations in Massachusetts were also observed occasionally in samples from the historic range of the parasite (two individuals with >90% assignment to the northeastern clade identified in Alabama). However, no genotypes from the historic range clustered tightly with New Jersey and Massachusetts samples in the genetic distance trees, limiting our ability to identify a source population for introduction into New Jersey and Massachusetts. Interestingly, genotypes in New Jersey and Massachusetts were highly heterozygous, suggesting a recent hybridization event between two divergent parasite lineages. When summarizing independent Structure analyses for K=5 using CLUMPP, the unique genotypes in Delaware Bay and Massachusetts appear to be admixtures between the groups frequent in Dickinson, TX

and New Smyrna Beach, FL (Fig. 3-S3). This hybrid pattern never appeared in an individual Structure run at any proposed value of K. These two populations would be of particular interest in future research regarding source populations for parasites in the newly expanded range. Regardless, genotypes in Rhode Island belong to different genotypic clusters than those observed in New Jersey and Massachusetts, indicating that two separate introductions occurred in the expanded parasite range.

The observation that there was a trend for fewer multiple infections later in the summer may be an artifact of the procedure used to diagnose multiple infections. The rate of multiple infections is likely constant, but the detectability of co-infecting parasite strains may change as intensity of infection and relative proportions of co-infecting strains changes over the course of the year. In early stages of infection, co-infecting parasites may occur in similar numbers, providing opportunities for amplification of alleles from both . If a particular strain becomes numerically dominant due to growth environment, PCR may preferentially amplify the dominant parasite alleles, swamping out the signal of coinfecting strains resulting in a false negative for multiple infections. This should occur at all loci such that the resulting genotype is representative of the dominant parasite strain, rather than a mosaic of coinfecting strain genotypes among loci. As such, this should not affect the population genetic analyses, but will result in under-representation of multiple infections in the population.

Conclusions

All localities were comprised of coexisting mixtures of genetically distinct parasite groups. Over time, these lineages have dispersed widely, resulting in varying local proportions of genetically distinct parasites. Disjunct distributions of some groups

indicate that long distance dispersal may play an important role in the biology of *P. marinus;* this may be due to anthropogenic movement of parasites or to an as yet unidentified natural vector. Predominantly clonal reproduction in *P. marinus* limits gene flow between clades, allowing these genetically distinct entities to evolve in parallel in the same location. These non-equilibrium interactions may result in previously unidentified coevolutionary selection mosaics between parasite and host that may select for increased virulence in some locations through multiple infections. If host resistance is specific to a particular parasite strain, epizootic disease may be caused by increased virulence and spread of a sympatric, genetically divergent group of parasites. This study provides a tool to track the relative proportions of co-occurring parasite strains and could be used to monitor local disease epidemiology. Annual and seasonal assessment of disease causing strains would increase our understanding of the susceptibility of host populations in relation to challenge with unique parasite lineages.

Tables

Table 3-1. Geographic coordinates of sampling locations accompanied by environmental parameters where available.

Location	Abbreviation	Latitude	Longitude	Temp (°C)	Salinity (ppt)
Dickinson, TX	DTX	29° 28' 55.99"	94° 54' 55.00"	28.4	7.8
Sabine Lake, TX	TLB	29° 47' 41.00"	93° 54' 41.00"	28.3	12.2
Dauphin Island, AL	ALA	30° 16' 43.93" N	88° 7' 24.28" W	28	20
Apalachicola, FL	APA	29° 42' 40.80" N	84° 52' 50.40" W	30.3	28.9
Port Charlotte, FL	PCH	26° 57' 7.56" N	82° 5' 58.42" W		
St. Lucie River, FL	SLR	27° 6' 7.49" N	80° 8' 28.93" W	28.5	21
Fort Pierce, FL	FTP	27° 29' 25.36" N	80° 18' 53.12" W	28.5	27
Cape Canaveral, FL	LPA	28° 26' 12.48" N	80° 34' 55.92" W	24	15.9
New Smyrna Beach, FL	NSB	29° 3' 34.16" N	80° 56' 24.05" W	25.8	22
Port Orange, FL	POR	29° 8' 53.52" N	80° 58' 40.08" W	28	19
Georgetown, SC	GSC	33° 20' 57.66" N	79° 11' 19.97" W	27	36.9
Delaware Bay, NJ	DBA	39° 18' 16.67" N	75° 21' 55.69" W		
Narragansett Bay, RI	NRI	41° 26' 58.60" N	71° 26' 51.76" W		
Tisbury, MA	TMA	41° 27' 49.45" N	70° 37' 30.13" W		
Edgartown, MA	EMA	41° 23' 29.44" N	70° 30' 58.95" W		

Table 3-2. Prevalence of infected oysters determined by NTS assay and frequency of multiple infections based on samples that amplified three or more alleles for at least one locus.

Sampling Location	Collection Date	Total oysters sampled	# NTS positive	Prevalence (%)	# Individuals with >4 uSat loci genotyped	# multiple infections	% multiple infections
Dickinson, TX	June, 2007	48	35	72.9	19	3	15.8
Sabine Lake, TX	June, 2007	48	25	52.1	25	4	16.0
Dauphin Island, AL	August, 2007	66	35	53.0	23	0	0.0
Apalachicola, FL	September, 2007	43	37	86.0	37	8	21.6
Port Charlotte, FL	October, 2004	100	70	70.0	53	16	30.2
St. Lucie River, FL	October, 2004	48	36	75.0	25	8	32.0
Fort Pierce, FL	October, 2004	46	21	45.7	21	1	4.8
Cape Canaveral, FL	October, 2004	46	44	95.7	21	4	19.0
New Smyrna Beach, FL	October, 2004	48	45	93.8	20	3	15.0
New Smyrna Beach, FL	September, 2005	48	41	85.4	41	14	34.1
Port Orange, FL	October, 2004	48	36	75.0	32	11	34.4
Georgetown, SC	July, 2008	94	48	51.1	18	4	22.2
Delaware Bay, NJ	July, 2005	80	38	47.5	30	19	63.3
	June, 2002	28	13	46.4	13	12	92.3
	July, 2002	30	24	80.0	17	14	82.4
	August, 2002	27	27	100.0	18	11	61.1
	Total 2002	**85**	**64**	**75.3**	**48**	**37**	**77.1**
Narragansett Bay, RI	June, 2002	32	3	9.4	3	0	0.0
	August, 2002	24	18	75.0	18	10	55.6
	September, 2002	29	18	62.1	18	3	16.7
	Total 2002	**85**	**39**	**45.9**	**39**	**13**	**33.3**
Tisbury, MA	August, 2002	28	19	67.9	16	7	43.8
	September, 2002	31	28	90.3	28	11	39.3
	Total 2002	**59**	**47**	**79.7**	**44**	**18**	**40.9**
Edgartown, MA	June, 2002	30	29	96.7	28	20	71.4
	July, 2002	30	26	86.7	23	19	82.6
	September, 2002	30	30	100.0	27	14	51.9
	Total 2002	**90**	**85**	**94.4**	**78**	**53**	**67.9**
All locations Combined		**1082**	**746**	**68.9**	**574**	**216**	**37.6**

Table 3-3. Fst and Rst values from pairwise tests of genetic distance by sampling location. Fst values are above the diagonal and Rst below. Significance was tested using a sequential Bonferroni correction. Significance was found at $p < 0.006$ for FST and $p < 0.002$ for RST. Larger values of FST and RST are shown in darker shades. Samples from Dickinson, Texas and New Smyrna Beach, Florida had consistently higher values of both Fst and Rst for all pairwise comparisons.

	DTX	TLB	ALA	APA	PCH	SLR	FTP	LPA	NSB	POR	GSC	DBA	NRI	TMA	EMA
Dickinson, TX	0	**0.27208**	**0.11613**	**0.25308**	**0.28825**	**0.26485**	**0.33146**	**0.34423**	**0.30108**	**0.2494**	**0.22247**	**0.16087**	**0.28884**	**0.20961**	**0.2231**
TX-LA Border	**0.5349**	0	**0.05065**	**0.10186**	**0.11996**	**0.11706**	**0.04512**	**0.08374**	**0.2475**	**0.07035**	**0.09328**	**0.07319**	**0.05939**	**0.25231**	**0.20178**
Dauphin Island, AL	**0.34814**	**0.1144**	0	**0.05112**	**0.1**	**0.06798**	**0.0747**	**0.13102**	**0.17715**	0.03864	0.04463	-0.00619	**0.07421**	**0.10267**	**0.08075**
Apalachicola, FL	**0.63365**	**0.12513**	**0.1519**	0	**0.115**	**0.16306**	**0.12794**	**0.19956**	**0.24857**	**0.09399**	**0.10529**	**0.0797**	**0.10701**	**0.19658**	**0.20522**
Port Charlotte, FL	**0.46242**	**0.18931**	**0.19259**	**0.34501**	0	**0.20107**	**0.12941**	**0.18287**	**0.17978**	**0.09953**	**0.0912**	**0.06612**	**0.19273**	**0.1781**	**0.1759**
St. Lucie River, FL	**0.66326**	**0.18409**	**0.08888**	**0.10377**	**0.29464**	0	**0.09524**	**0.13829**	**0.28052**	**0.09406**	**0.1579**	**0.1321**	**0.16485**	**0.19983**	**0.16735**
Fort Pierce, FL	**0.77747**	0.04373	**0.14327**	0.02061	**0.19981**	**0.21054**	0	**0.10876**	**0.28845**	**0.06878**	**0.12163**	**0.11538**	**0.13795**	**0.24114**	**0.1851**
Cape Canaveral, FL	**0.57406**	**0.1828**	0.05826	0.07426	**0.26887**	**0.2488**	**0.2359**	0	**0.32675**	**0.13115**	**0.17287**	**0.16**	**0.18596**	**0.2903**	**0.23831**
New Smyrna Beach, FL	**0.68265**	**0.61088**	**0.59138**	**0.70909**	**0.2859**	**0.67416**	**0.65924**	**0.66552**	0	**0.24472**	**0.19631**	**0.14581**	**0.30034**	**0.18231**	**0.19937**
Port Orange, FL	**0.64324**	**0.10224**	0.05267	0.05064	**0.17718**	**0.08292**	**0.18799**	**0.15893**	**0.62931**	0	0.0321	**0.09269**	**0.11217**	**0.20945**	**0.18354**
Georgetown, SC	**0.63994**	**0.19745**	**0.12032**	**0.15696**	**0.25321**	**0.31597**	**0.34315**	0.02053	**0.64383**	**0.21262**	0	**0.03964**	**0.15639**	**0.19771**	**0.1708**
Delaware Bay, NJ	**0.26037**	**0.21506**	0.03154	**0.28189**	**0.13807**	**0.2278**	**0.25426**	**0.18148**	**0.32852**	**0.20849**	**0.17704**	0	**0.10583**	**0.09292**	**0.07712**
Narragansett Bay, RI	**0.56279**	**0.10956**	0.0389	0.00093	**0.31017**	**0.10435**	**0.11726**	0.03002	**0.6819**	0.0477	**0.12075**	**0.23926**	0	**0.28145**	**0.23867**
Tisbury, MA	**0.34779**	**0.37192**	**0.20792**	**0.43878**	**0.26457**	**0.34532**	**0.36672**	**0.31558**	**0.33243**	**0.31284**	**0.32665**	0.00641	**0.37642**	0	**0.05768**
Edgartown, MA	**0.31491**	**0.32969**	**0.15608**	**0.39354**	**0.22824**	**0.32163**	**0.31271**	**0.25403**	**0.31704**	**0.27563**	**0.26714**	0.00808	**0.33313**	-**0.00206**	0

Figure Legends

Figure 3-1. Allelic distribution by locus and population. A) The absolute number of alleles (Na) ranges from 12 to 18 alleles/locus. Also shown is the effective number alleles (Ne = 1/sum p_i^2). Although many alleles were observed at some loci, biased frequency distributions reduced the information content at all loci. B) Allelic richness for each population was assessed, following correction for a sample size of 11, using rarefaction. Key to sampling location abbreviations are in Table 1. Locations are arranged in a linear fashion along the coast line beginning in Dickinson,TX on the left and ending in Edgartown, MA on the right.

Figure 3-2. Regression analyses of multiple infection frequency by month. For those locations in which multiple timepoints were collected in one year, a trend was observed for decreasing frequency of multiple infections in later months. Regression analysis of all of these samples showed that this trend was not significant (p = 0.3249, dashed line) when all samples were included. However, removal of the June sample from Delaware Bay (red circle on lower left) as an outlier revealed a statistically significant relationship between month of collection and multiple infection frequency (p = 0.0008, solid line).

Figure 3-3. Inbreeding coefficients (F_{IS}) in *P. marinus* sampling locations (error bars represent among-loci standard deviations). Locations are arranged in linear order along the coastline, from the westernmost sampling location in the Gulf of Mexico (DTX) to the most eastern location in Massachusetts (EMA). Inbreeding coefficients were

significantly positive in all populations except for those in Massachusetts, where population F_{IS} values were significantly negative.

Figure 3-4. Comparison of three clustering results based on different genetic criteria. All analyses converge on common clusters of individual genotypes across the sampling range. Genotypes can be broadly categorized as belonging to a northeastern cluster (with representatives from Massachusetts and New Jersey), a cluster characterized by New Smyrna Beach and some Port Charlotte genotypes, a cluster restricted to isolates from Dickinson, TX, and a poorly defined cluster that encompasses many different geographic locations. Arrows highlight the commonalities between the three clustering methods. Colors of individual samples are not consistent across analyses. Patterns of clustering are the intended emphasis of this figure.

Figure 3-5. Geographic distribution of Structure-inferred clusters. Genotypic composition of sampling locations indicates that some *P. marinus* groups are widely distributed but that others occupy a restricted range. Representatives of the pink and purple clusters were found in nearly all populations, whereas the blue cluster was found primarily in the Northeast. The greencluster was found at moderate to high frequencies only in New Smyrna Beach and Port Charlotte, FL with single representatives in Dickinson, TX and Georgetown, SC.

Figure 3-S1. Likelihood distribution of K-values from Structure analysis. A) Likelihood of each K-value based on 10 replicates of Bayesian runs at each value of K using 100,000

generations of burn-in and 1,000,000 generations of exploration. Standard deviations of likelihood between runs are given as error bars. B) Using the method outlined in Evanno et al, 2004, delta K values were calculated showing a mode at k=6as the most likely value of K among all runs.

Figure 3-S2. Distribution of Posterior Probability of differing numbers of populations based on Structurama analysis. A single peak of high probability is centered around seven populations. The probability of the data matching six populations is reasonably high (0.34) with lower support for eight populations (0.18).

Figure 3-S3. Barplots of individual population membership based on a proposed 3 to 6 populations in the Bayesian clustering program Structure. Color of each vertical bar indicates the population assignment of an individual genotype. Sampling locations are provided on the x-axis, organized as a linear array along the coastline from the western extreme in the Gulf of Mexico (Dickinson, TX) to the Atlantic coast, ending in Edgartown, MA at the northeastern end of the sampling distribution. Samples from Dickinson, TX , Tisbury, MA and Edgartown, MA cluster together for K=3 - 5. At K=6, the Dickinson,TX samples are assigned to a population exclusive to them. Isolates of other locations are not consistently assigned to their own populations; instead, those individuals are increasingly assigned to more than one population as additional populations are assumed.

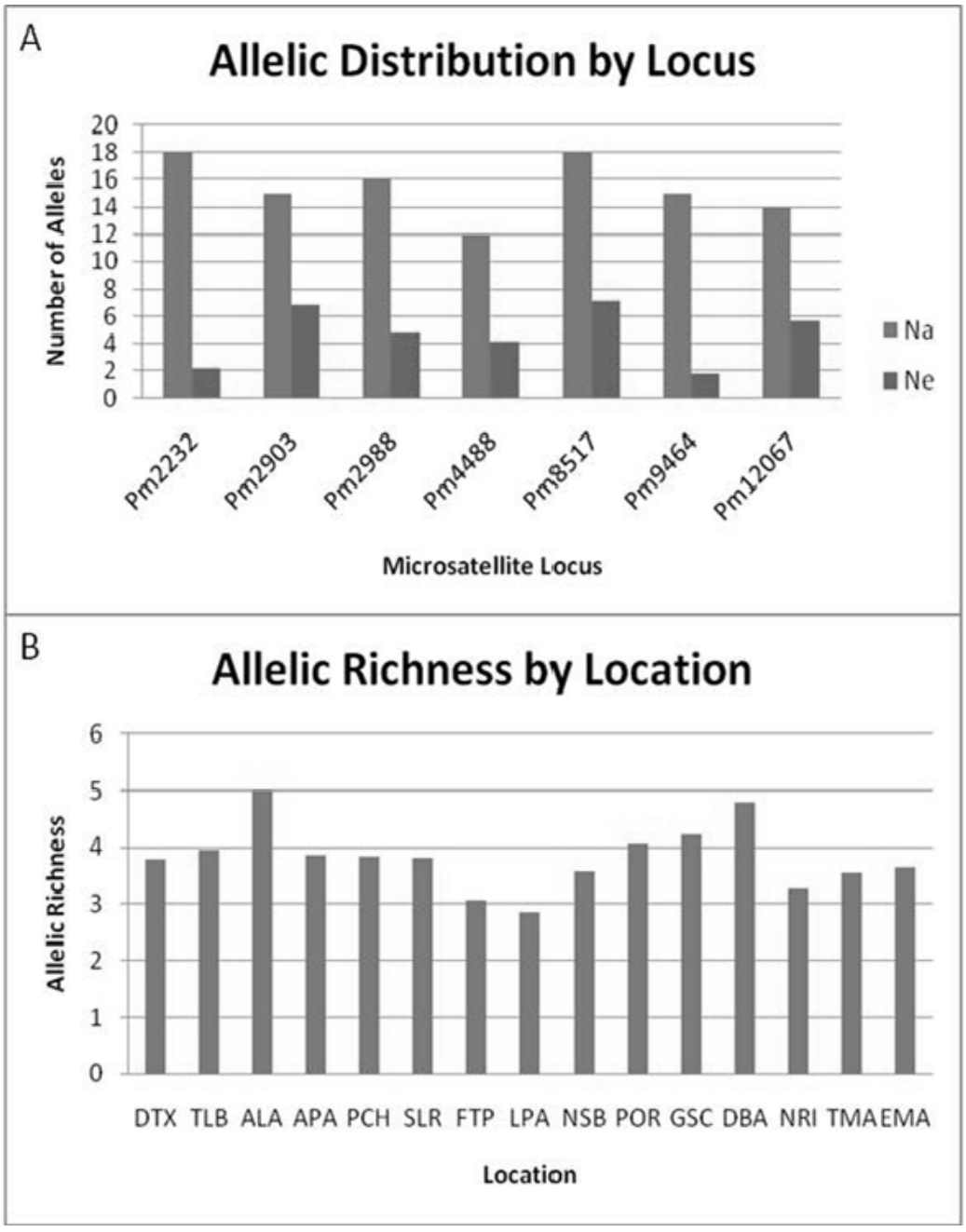
A
Allelic Distribution by Locus
Number of Alleles
20
18
16
14
12
10
8
6
4
2
0
Pm2232
Pm2903
Pm2988
Pm4488
Pm8517
Pm9464
Pm12067
Microsatellite Locus
Na
Ne
B
Allelic Richness by Location
Allelic Richness
6
5
4
3
2
1
0
DTX TLB ALA APA PCH SLR FTP LPA NSB POR GSC DBA NRI TMA EMA
Location

Figure 3-1

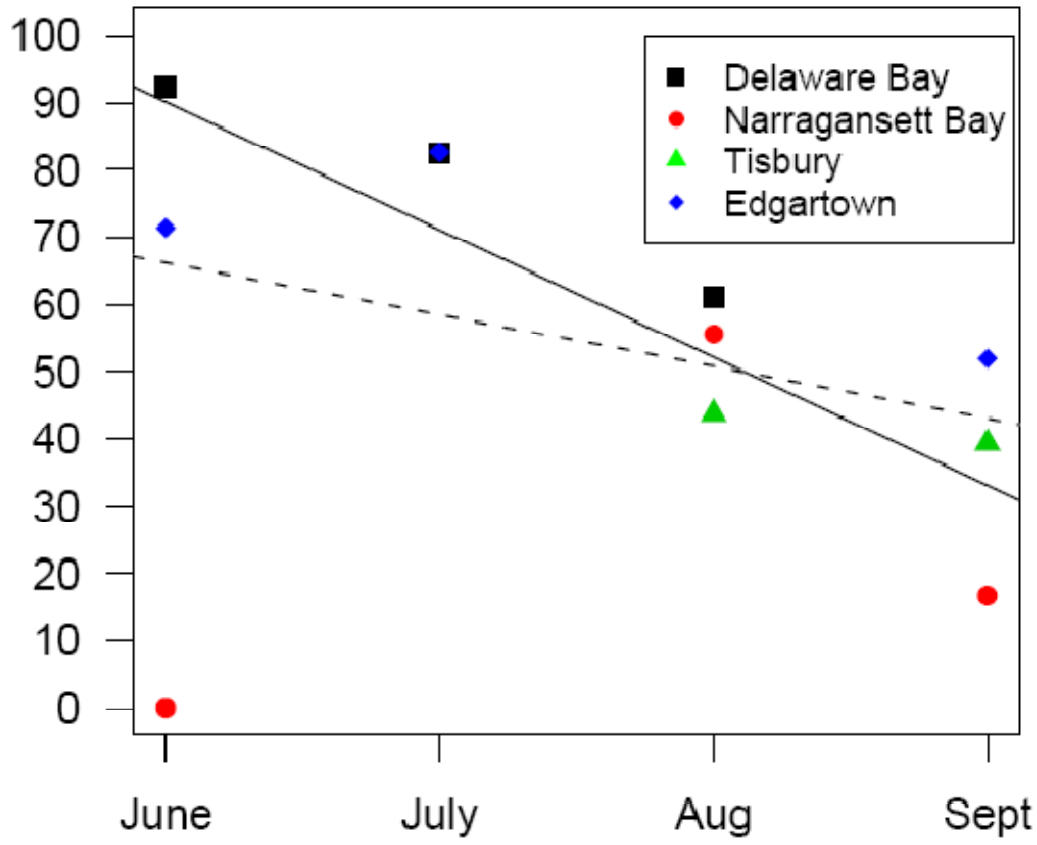
100
90
80
70
60
50
40
30
20
10
0
June
July
Aug
Sept
Delaware Bay
Narragansett Bay
Tisbury
Edgartown

Figure 3-2

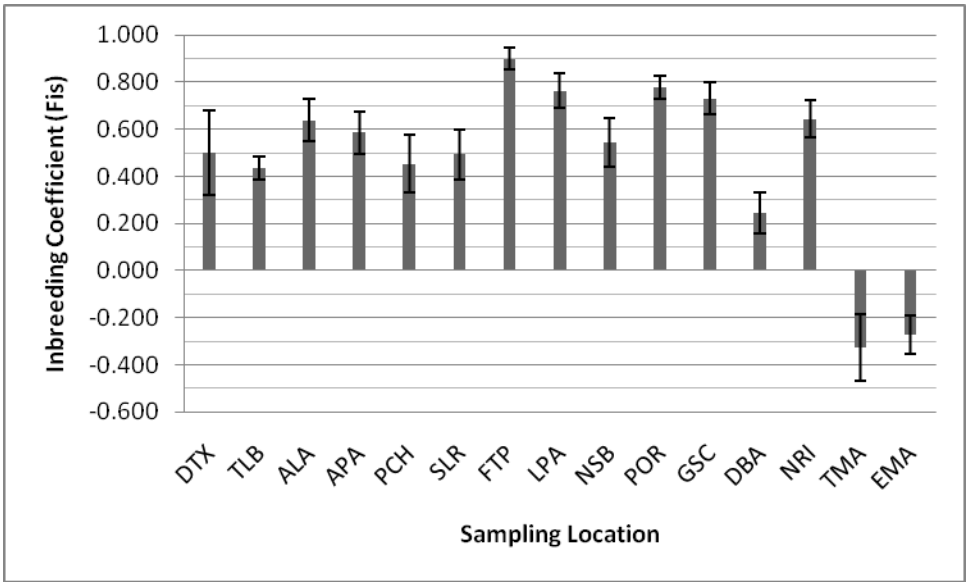
1.000
0.800
0.600
0.400
0.200
0.000
-0.200
-0.400
-0.600
Inbreeding Coefficient (Fis)
DTX
TLB
ALA
APA
PCH
SLR
FTP
LPA
NSB
POR
GSC
DBA
NRI
TMA
EMA
Sampling Location

Figure 3-3

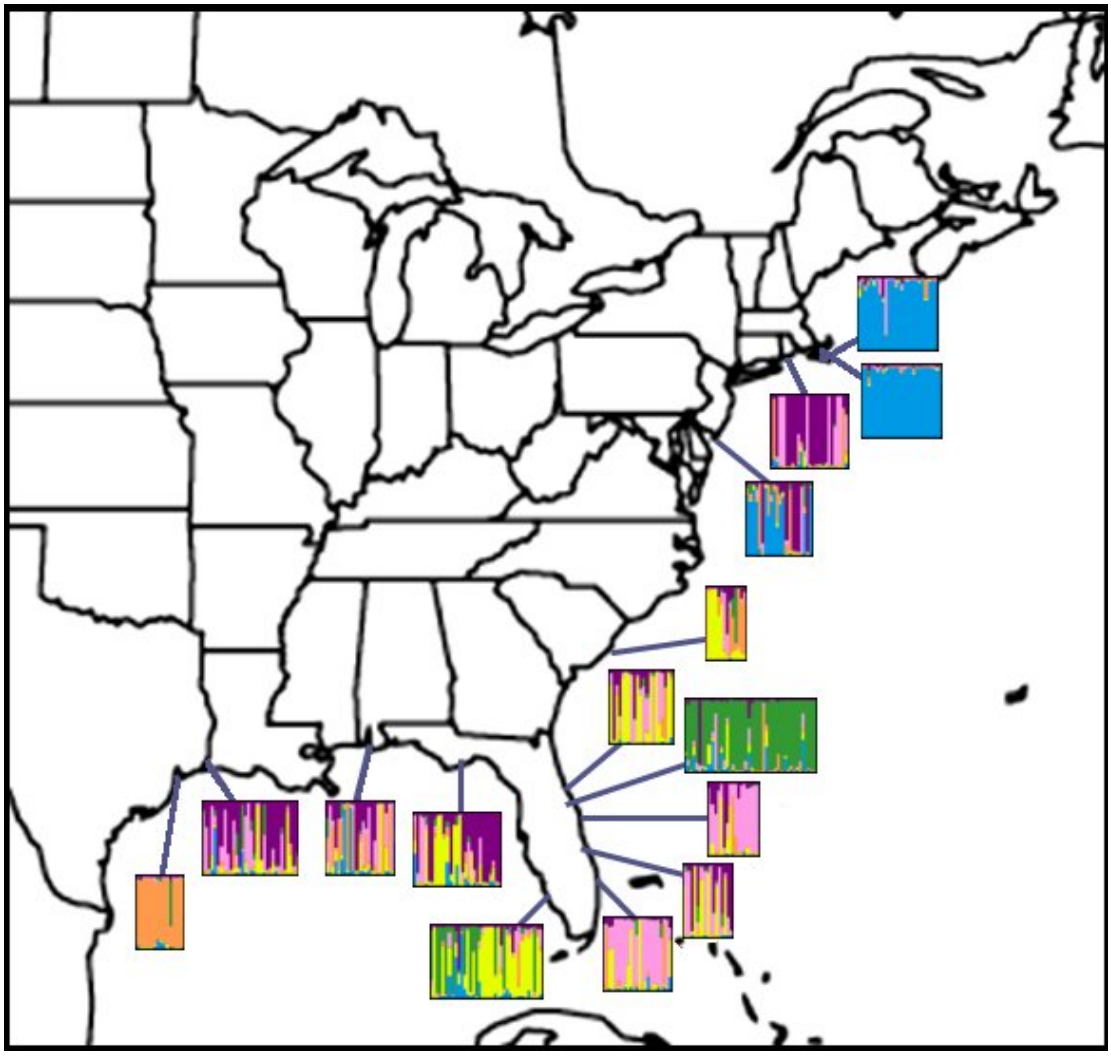

Figure 3-5

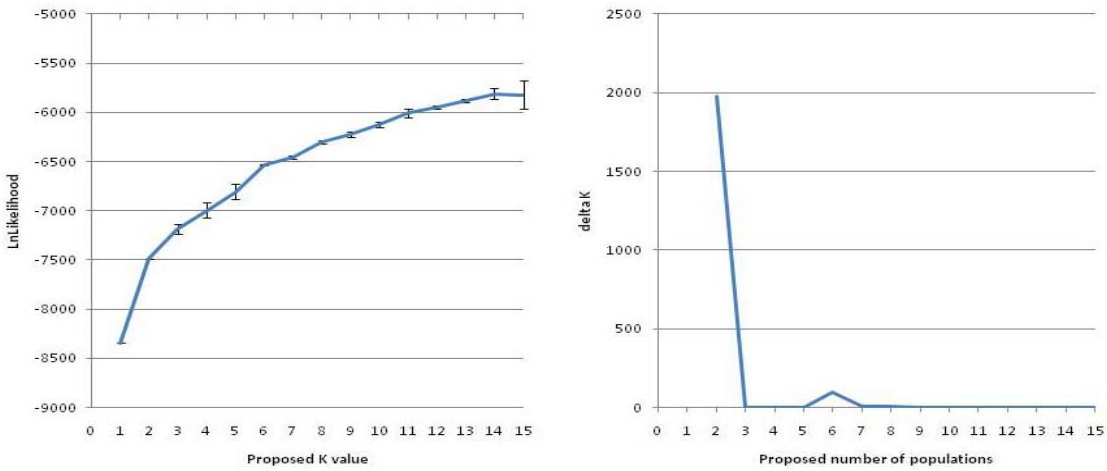
-5000
-5500
-6000
-6500
-7000
-7500
-8000
-8500
-9000
LnLikelihood
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Proposed K value
2500
2000
1500
1000
500
0
deltaK
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Proposed number of populations

Figure 3-S1

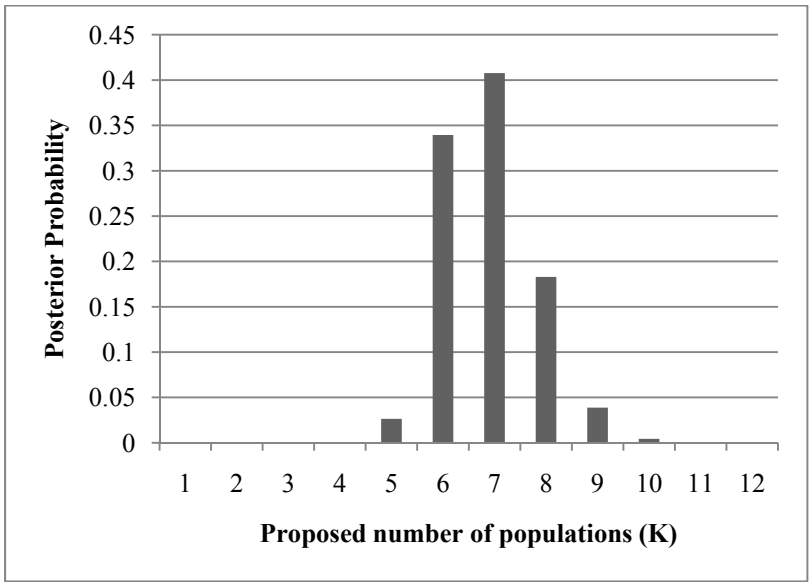
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Posterior Probability
1 2 3 4 5 6 7 8 9 10 11 12
Proposed number of populations (K)

Figure 3-S2

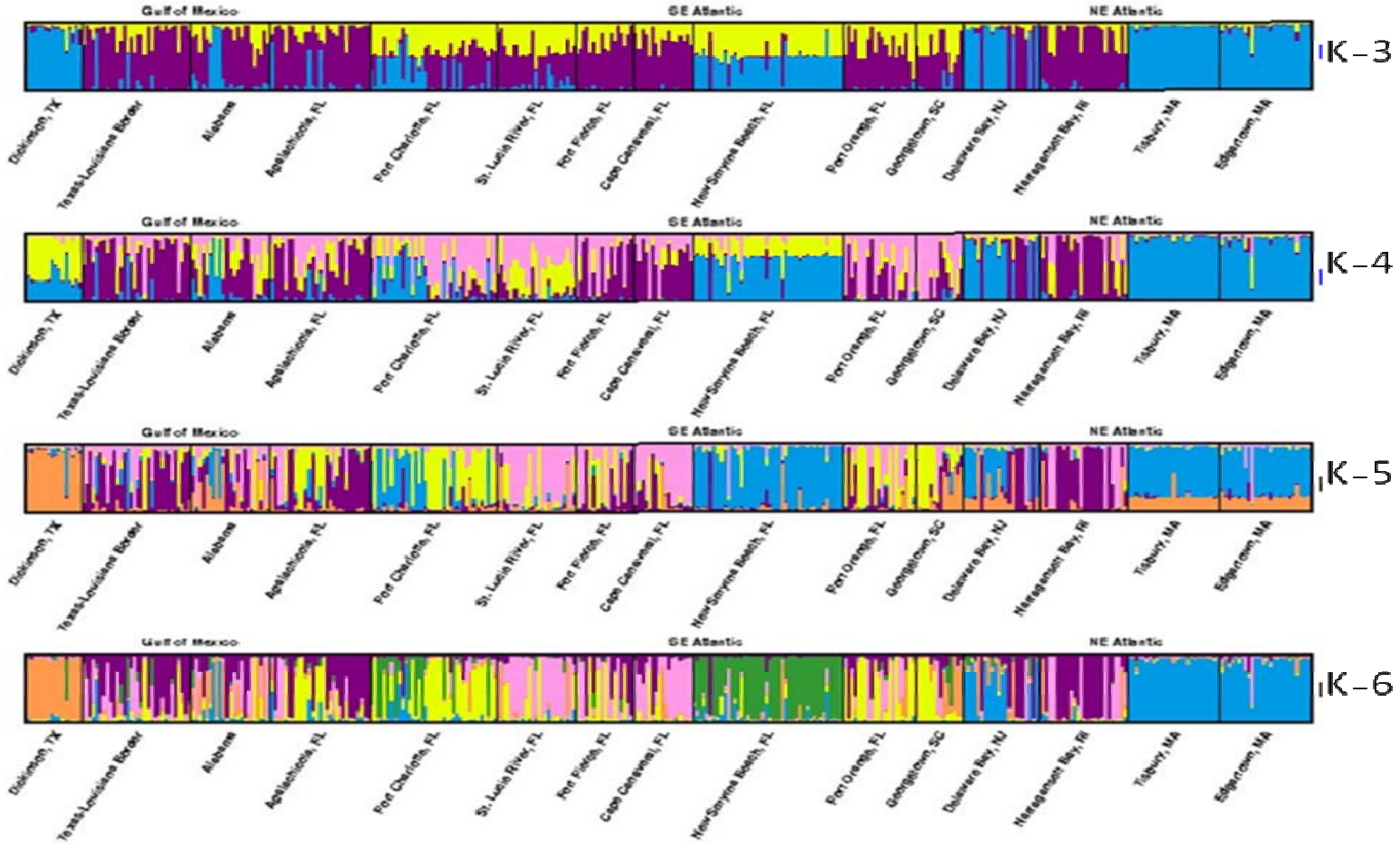
Gulf of Mexico
SE Atlantic
NE Atlantic
K-3
K-4
K-5
K-6
Dickinson, TX
Texas-Louisiana Border
Alabama
Apalachicola, FL
Port Charlotte, FL
St. Lucie River, FL
Fort Pierce, FL
Cape Canaveral, FL
New Smyrna Beach, FL
Port Orange, FL
Georgetown, SC
Delaware Bay, NJ
Narragansett Bay, RI
Tisbury, MA
Edgartown, MA

Figure 3-S3

Chapter Four - Potential causes of di-allelism at six nuclear loci in the protistan parasite *Perkinsus marinus*.

Abstract

Previous studies analyzing DNA sequence data from the parasitic protist *Perkinsus marinus* identified distinct allelic lineages differing by as much as 2.6% but with little variation within either allelic class. Such 'di-allelism' has been reported in other eukaryotic parasites, where it has been attributed either to long term separation of clonal lineages or to balancing selection on loci involved in evading host immune responses. In order to measure the extent of di-allelism across the *P. marinus* genome, and to assess its causes, I collected DNA sequences from eight nuclear loci in samples collected across most of the continental range of this parasite. Five loci are sequences flanking repetitive microsatellite loci, two loci include coding sequences (superoxide dismutase 1 and 2) for genes that may mediate the parasite's interaction with its host, and the last locus is the internal transcribed spacer of rRNA. At six of the seven single-copy loci, haplotypes fell into two discrete allelic classes differing in sequence by 3.8% on average. Little additional variation was observed among the 3198 sequenced nucleotides. At all loci, only two haplotypes were observed more than once. The remaining alleles were singletons differing by no more than two nucleotides from either of the frequently observed haplotypes. The divergence between two allele classes was substantial enough to suggest that they began diverging millions of years ago. Recombination has occurred between loci, generating 13 unique multilocus genotypes. These alleles now circulate within an interbreeding population of parasites in which sexual reproduction occurs to

some degree; in spite of this, no intragenic recombination was observed. Excluding singleton mutations, all alleles were geographically widespread. The genomic di-allelism observed in *P. marinus* was found to be inconsistent with hypotheses of population subdivision, ancient asexuality, or genome duplication potentially generating di-allelism. Balancing selection could explain *P. marinus* di-allelism only if selection were acting on the genome as a whole, or if asexuality directed the entire genome into a pattern driven by strong balancing selection on one or a few loci. However, we found evidence for recombination, which should result in the loss of those alleles not being maintained by balancing selection. I conclude that these distinct allele classes most likely diverged during a long period of allopatry, and that recent hybridization and recombination resulted in the varied multilocus genotypes observed here. Di-allelism is prevalent across the *P. marinus* genome, though the mechanism supporting exactly two lineages remains elusive.

Introduction

Perkinsus marinus is a parasite that causes significant disease in populations of its eastern oyster host, *Crassostrea virginica*. *P. marinus* was identified as an oyster pathogen following outbreaks of disease in Gulf of Mexico oyster populations in the 1940's (Mackin et al. 1950) and is now resident in locations as far north as Maine (Ford 1996). Recent studies using microsatellite loci have revealed that *P. marinus* reproduction is primarily clonal in extant populations, but occasional sexual reproduction occurs, producing recombinant genomes over evolutionary timescales (Thompson Dissertation Chapter 2). *P. marinus* multilocus genotypes bear identifiable signatures of widely dispersed clonal lineages that coexist in varying proportions in different locations

(Thompson Dissertation Chapter 3). Phylogenetic analysis has identified *Perkinsus* as an early branching dinoflagellate (Siddall et al. 1997; Saldarriaga et al. 2003) and most closely related to the congeneric species *P. olseni* and *P. honshuensis* (Dungan and Reece 2006; Moss et al. 2008). In other DNA sequencing studies, an interesting pattern of large divergence between alleles has been observed at multiple nuclear loci (Reece et al. 1997; Robledo et al. 1999; Brown et al. 2004) which was also noted in a bimodal distribution of repeat lengths at microsatellite loci (Thompson Dissertation Chapters 2 and 3). The cause of this deep divergence between alleles deserves further scrutiny and is the subject of this chapter.

In a large neutrally evolving population, multiple haplotypes are expected where most mutations coalesce in the recent past, with fewer coalescent events deeper in the past dependent on the effective size of the population (Hartl and Clark 1997). A large effective population size is suggested by the high genetic diversity observed at microsatellite loci for *P. marinus* (Thompson Dissertation Chapters 2 and 3). If *P. marinus* populations have large effective sizes, selection could be very efficient in changing parasite virulence in different regions where *P. marinus* is resident. This could take the form of purifying selection, overdominance, frequency dependent, or balancing selection. Alternatively, I have characterized *P. marinus* reproduction as predominantly clonal (Thompson Dissertation Chapter 2) which would anticipate small effective population sizes. Selection is less efficient in small populations and should allow genetic drift to be a more powerful evolutionary force. Comparisons between microsatellite data and DNA sequence data may help to define effective population sizes in *P. marinus* and the roles of natural selection and genetic drift in the evolutionary history of this parasite.

An interesting phenomenon in certain protist populations is the division of haplotypes into two discrete classes with a distinct lack of coalescent events at intermediate times in the past. That is, recent mutations are observed, but they are associated with two ancient lineages with no observed intermediate forms or recombinants between the two. This 'di-allelic' pattern may be formally defined by the "¼ rule," where the average difference between haplotypes within a lineage is less than ¼ of the average difference between lineages (Roy et al. 2006). This pattern is a distinct non-neutral pattern as the basal branches of the genealogy are "disproportionately long, with extensive sequence divergence between two allelic classes" (Roy et al. 2006). Di-allelism has been ascribed to numerous causes including both demographic and selective explanations. Processes implicated in establishing di-allelism include balancing selection on immune related loci (Polley and Conway 2001; Roy 2008), gene duplication (Miller et al. 1993; Hartl et al. 2002), ancient asexuality (Birky 1996; Welch and Meselson 2000), population subdivision with genetic drift (Tanabe et al. 1987; Machado and Ayala 2001), and introgression (Polley et al. 2005).

We can discriminate among these alternatives because each would result in certain characteristic patterns distinguished primarily by the genomic extent of di-allelism. For example, balancing selection should maintain distinct classes of alleles only on loci experiencing selection (and closely linked loci), assuming some sexual reproduction and recombination. Gene or genomic duplication and subsequent divergence would be expected to preserve each duplicate in *all* individuals (barring secondary deletion) (Hartl et al. 2002). Obligate asexual diploids should experience continual divergence between homologous chromosomes, ultimately resulting in fixed

heterozygosity across the genome (Welch and Meselson 2000). Population subdivision could encourage the divergence of alleles so long as gene flow was limited or differential selection was strong (Tanabe et al. 1987). Prolonged allopatry could foster divergence of alleles at all loci by genetic drift. Given large enough divergence, subsequent hybridization between a limited number of individuals from formerly allopatric populations would result in population genetic patterns resembling genome-wide di-allelism. Recombination between hybrids or backcrosses to a parental strain may reassort alleles such that genomes become mosaics of parental genomes (Boyle et al. 2006). DNA sequences from multiple loci across the genome are required to distinguish among these alternatives.

Deep divergence between alleles has been observed in preliminary samples of *P. marinus* at three loci, including actin (Reece et al. 1997), ribosomal intergenic repeat sequences (the non-transcribed spacer (NTS) (Robledo et al. 1999)) and the internal transcribed spacer (ITS)), and one anonymous nuclear locus ((ATAN) Brown et al. 2004). Two sequences were discovered at the actin locus; the 2.6% difference between these was attributed to a recent gene duplication event. Only two alleles were discovered at the NTS locus, differing at 6 of 307 nucleotides (1.9%); the relative frequency of each allele differed in two population samples.

Greater numbers of distinct haplotypes have been identified at two additional loci (Brown et al. 2004). From twelve *P. marinus* tissue culture samples, fourteen allelic variants were discovered at the anonymous nuclear ATAN locus. The two most distant sequences among these (termed the A and B alleles) differed by 3.5% and accounted for 86% of all sequences collected. The remaining sequences may have been the result of

recombination during bacterial cloning of PCR products (Brown et al. 2004). Twenty-seven unique ITS haplotypes were observed among the same twelve *P. marinus* isolates. These sequences differed by as much 1.5%, with 10 alleles observed more than once among 84 cloned sequences, and three sequences observed at greater than 10% of the entire sample. If observed haplotypes are the result of cloning errors, the ATAN locus may be classified as di-allelic, whereas the ITS locus has the most haplotypes described thus far.

In order to measure the extent and pattern of di-allelism within *P. marinus* populations, I generated nucleotide sequence data from seven nuclear loci. Putatively neutral microsatellite flanking sequence was collected from five loci and coding and intron sequences were collected from two superoxide dismutase genes (SOD1 and SOD2), loci potentially involved in response to host defense mechanisms (Schott et al. 2003). I describe the variability in DNA sequence among individuals and between geographic samples. I also investigate whether there is recombination within DNA sequences or between loci across the genome. Intragenic recombination was investigated in order to determine whether divergent haplotypes have been in the same breeding population for an extended period of time. Estimates of coalescence time were generated to establish the age of alleles in the samples. In general, to distinguish among the mechanisms described above, I sought to determine whether di-allelism was restricted to a small subset of genes (consistent with locus-specific balancing selection), whether fixed heterozygosity was evident in diploid isolates, or whether distinct alleles occupied discrete geographic locales. My data disfavor balancing selection, genome duplication, prolonged asexual diploidy, or recent population subdivision as explanations for observed

patterns of di-allelism. Instead, these data imply that *P. marinus* populations were subdivided for millions of years, and have subsequently hybridized.

Materials and Methods

Samples

Tissue culture isolates from Mozambique Point, LA (ATCC #50763), Fort Pierce, FL (ATCC #50775), Bennet Point, MD (ATCC #PRA-240), Delaware Bay, NJ (ATCC #50509), and Narragansett Bay, RI (HCTR, graciously provided by Marta Gomez-Chiarri) were used to optimize PCR amplification. Additionally, four clonal isolates of other *Perkinsus* species (*P. chesapeaki* (ATCC #50866, ATCC #50864, and ATCC #50807) and *P. olseni* (ATCC # PRA-31)) were obtained for attempts at amplifying outgroup sequence.

Genomic DNA from six infected oysters from each of 14 geographic locations were selected for PCR amplification and sequencing of microsatellite flanking regions. Only individuals that had been judged by multilocus microsatellite analysis to contain DNA from exactly one strain of *P. marinus* were included for present purposes (Thompson, dissertation Chapter 2). Analyzed individuals were from Dickinson, TX, Sabine Lake, TX, Dauphin Island, AL, Apalachicola, FL, Port Charlotte, FL, St. Lucie River, FL, Fort Pierce, FL, Cape Canveral, FL, New Smyrna Beach, FL, Port Orange, FL, Delaware Bay, NJ, Narragansett Bay, RI, Tisbury, MA, and Edgartown, MA.

SOD2-short sequences were generated from a collection of singly infected oysters from 10 geographic locations. These samples were distinct from those used for sequencing microsatellite flanking regions, though there is some overlap of specific individuals. Locations included Narragansett Bay, RI (22 individuals), Delaware Bay, NJ (1 individual), Skidaway Island, GA (5 individuals), Port Orange, FL (4 individuals),

New Smyrna Beach, FL (13 individuals), Cape Canaveral, FL (11 individuals), Fort Pierce, FL (1 individual), Port Charlotte, FL (26 Individuals), Mozambique Point, LA (1 individual), and Sabine Lake, TX (1 individual).

SOD1-short sequences were generated from a smaller group of individuals from nine locations. Samples included cloned cultures (Mozambique Point, LA (ATCC #50763), Fort Pierce, FL (ATCC #50775), Bennet Point, MD (ATCC #PRA-240), Delaware Bay, NJ (ATCC #50509)), and 17 infected oysters from Florida locations including Port Charlotte (4 individuals), St. Lucie River (2 individuals), Fort Pierce (3 individuals), Cape Canaveral (3 individuals), New Smyrna Beach (2 individuals), and Port Orange (3 individuals).

Loci

Flanking regions from five microsatellite repeat loci were sequenced (details of microsatellite locus discovery can be found in Thompson, Dissertation Chapter 2). These loci are named Pm2232, Pm2903, Pm2988, Pm4488, and Pm8517 based on their contig designation in the *P. marinus* draft genome sequence (GenBank: whole genome shotgun accession AAXJ00000000, which consists of sequences AAXJ01000001-AAXJ01023491). The flanking primers were designed using Primer3 on the web (Rozen and Skaletsky 2000) targeting 300 to 500 bp amplicons. Potential amplicons were limited in size by the amount of flanking sequence available in the draft genome contigs. Primer sequences and amplicon sizes are given in Table 1. The size of each flanking region DNA sequence is reported in Table 2. Microsatellite characteristics have been reported elsewhere (Thompson, Dissertation Chapter 2).

Additionally, portions of two coding loci, superoxide dismutase 1 (*SOD1*) and superoxide dismutase 2 (*SOD2*) were amplified for sequencing. Primers for both loci were designed within exons to amplify across intervening introns using sequences published in Schott et al 2003. Two datasets are analyzed for each of these loci; a short single amplicon sequence obtained from many infected oysters and a longer sequence based on concatenation of multiple, overlapping amplicons and compared across a small number of infected oysters. The short sequence of SOD1 consisted of 273 base pairs, with 225 bp coding sequence and 48 bp intron sequence. The long SOD1 sequence extended from exon 1 through exon 4, consisting of 470 bp of coding sequence and 183 bases of intron sequence. The short sequence from *SOD2* consisted entirely of intron 3 sequence, while the long *SOD2* sequence contained the entire *SOD2* coding sequence plus all intervening introns (Schott et al. 2003). Only a small portion of the 5' and 3' noncoding regions were included in the *SOD2* long sequences.

Sequencing of the entire coding sequence of *SOD2* was accomplished by producing two overlapping (diploid) amplicons that did not require nested PCR as these were only amplified from reference strains. Each amplicon was sequenced using four primers to guarantee sufficient coverage of the entire DNA fragment, and included two allele specific primers designed from results of SOD2-short DNA sequencing efforts.

PCR

Loci were amplified using nested or hemi-nested PCR directly from infected oyster genomic DNA extracts which were described previously (Thompson, Dissertation Chapter 1). Oligonucleotide primers are listed in Table 4-1 along the order in which the primers were used in nested or hemi-nested PCR. Each 20 µl PCR reaction used a

standard amplification protocol with 30 amplification cycles in both the primary and secondary amplification as follows. Each PCR reaction contained 1X PCR buffer (Invitrogen, Carlsbad, CA), 2.5 mM magnesium chloride, 0.25 mM mixed dNTP's, 0.1 mg/ml bovine serum albumin, 0.5 units Taq DNA polymerase (Invitrogen), and 40 nM of each primer. Reactions were subjected to a standard cycling procedure with an initial denaturation at 94 °C for 2 minutes followed by 30 cycles of 94 °C for 30 seconds, 60 °C for 30 seconds, and extension at 72 °C for 1 minute. The cycling was followed by a final extension step at 72 °C for 5 minutes, and reactions were stored at 4 °C until analyzed. For each 96 well amplification plate, a no DNA template and uninfected oyster genomic DNA control was added to control for cross-contamination and non-specific amplifications. Two µl of undiluted sample DNA was used as template in each primary amplification. Genomic DNA template concentration ranged from 40 ng/reaction to 250 ng/reaction. In all cases, 2 µl of the primary amplification were used as template for the secondary reaction. Amplifications of congeneric *Perkinsus* species were attempted using temperature gradient PCR for all microsatellite flanking sequences and the SOD2 locus, varying annealing temperatures from 40 °C to 60 °C and increasing the annealing time to 1 minute per cycle.

Sequencing

PCR products were sequenced directly using Applied Biosystems BigDye v.3.0 chemistry (Life Technologies, Carlsbad, CA). Sequencing primers were those used in the secondary amplification for each locus. Prior to sequencing, PCR products were treated with Exonuclease I and Shrimp Alkaline Phosphatase (Exo/SAP-it, USB Corporation, Cleveland, OH) to remove excess primers and unincorporated dNTPs. Following

Exo/SAP-it treatment, PCR products were diluted 5-fold in TE to produce reliable sequence data based on results of repeated trials using clonal isolate DNA. DNA sequencing reactions followed the manufacturer's recommended protocol, reducing reaction volumes to 10 µl and reaction mix to 0.5X final concentration. Sequencing products were purified using EDTA/ethanol precipitation and resuspended in 10 µl HiDi formamide (Life Technologies) prior to separation. Sequences were generated using an ABI 3730 sequencer in a 96 well format.

Alignment

DNA sequencing chromatograms were edited and aligned in Sequencher v 4.7 (Gene Codes Corporation, Ann Arbor, MI). Sequences were analyzed only where both strands amplified and produced reliable base calls. Individual sequences with overlapping chromatogram peaks of approximately equal height at a particular nucleotide position were scored as heterozygous for that nucleotide. If a particular nucleotide was heterozygous in the chromatogram from one strand, but not from the opposite strand, it was considered heterozygous as well. (Disagreement between strands regarding heterozygosity only affected one sequence for the locus SOD2-short corresponding to Singleton 2). Multiple sequence alignments were then constructed manually using diploid sequences.

Analysis

All DNA sequences were partitioned into two sequences, assuming diploidy in accordance with results from microsatellite amplification (Thompson Dissertation Chapter 2). The number of sequences provided in Table 4-1 reflect the number of

chromosomes sampled. Phase of multiple heterozygous peaks were determined using PHASE as implemented in DNAsp v. 5.0.

A locus was considered di-allelic according the one-fourth rule described in Roy et al 2008. This rule states that if the average divergence between alleles within a class is less than one fourth the average divergence between allele classes, the locus is di-allelic. Hence, haplotypes at each locus were assigned to the two most divergent allele classes according to a mid-point rooted neighbor-joining tree. Average divergence was calculated within each allele class and between the two allele classes as the percent difference.

Divergence time was calculated based on corrected genetic distances. Nucleotide mutation models were compared and tested using FindModel on the web (http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html). The best-fit model was evaluated by AIC and likelihood. A single model was chosen for correcting genetic distances across all loci based on consensus between AIC and likelihood results. Time to the most recent common ancestor was calculated as $t = d/2\mu$ assuming a range of mutation rates from 2×10^{-8} to 1×10^{-9} substitutions per site per year in accordance with mutation rates reported for other single-celled organisms (Machado and Ayala 2001; Kondrashov and Kondrashov 2010).

Summaries of nucleotide sequence diversity were generated in DNAsp v.5 (Librado and Rozas 2009). Tests of neutrality, including Tajima's D, Fu and Li's D* and Fu and Li's F* were conducted in DNAsp. Unrooted neighbor-joining trees were generated in MEGA 4 using the Kimura 2 Parameter model (as determined by FindModel results) to calculate genetic distances (Tamura et al. 2007).

Recombination was tested between loci using linkage disequilibrium based on likelihood as implemented in GenePop on the Web (Raymond and Rousset 1995). Hudson's four gamete test was used to infer whether pairwise comparisons of genotypes were compatible with a history of non-recombining ancestry (Hudson and Kaplan 1985). Intragenic recombination was evaluated based on applying the four gamete test to segregating sites.

Results

474 DNA sequences were collected, covering a total of 3928 bases among seven *P.marinus* loci. In spite of multiple attempts to generate outgroup sequence, no other *Perkinsus* species amplified for any locus. Sequence variation for each locus is described in Table 4-2. In general, there was little diversity within any given locus or allelic class, but marked divergence between the observed alleles. The number of sequences collected at each short locus varied from 22 to 79, but no more than 4 alleles were found at any locus. At each locus, we identified one common allele and one alternative allele that differed by an average of 3.8% (stdev = 2%). The common allele frequency ranged from 72-94% across loci. Seven insertions/deletions (indels) were observed which were limited to two loci. The difference between the two alleles at each locus ranged from 2.2% to 7.4% with the lowest difference observed at the shortest (91 bp) locus Pm2232, and the highest difference observed in 257 bp of Intron 3 sequence of the SOD2-short locus (Table 4-2). Among 474 DNA sequences, only five singleton mutations were observed; these were distributed among four different loci. The allelic divergences and frequency of alleles are diagrammed in Figure 4-1 as unrooted neighbor-joining trees.

Within the coding regions of SOD1 and SOD2, variation was present in both introns and exons. At SOD1, 8 of 23 SNPs were in coding portions of the sequence (35%) and all were synonymous. In contrast, coding regions of SOD2 had equal numbers of synonymous and nonsynonymous mutations out of 30 SNPs, and these coding SNPs represented 37% of all 82 SNPs found at this locus. Six indels were restricted to SOD2 introns and accounted for 14 base pairs total. Transition/transversion ratios in these two coding loci were lower than those observed in microsatellite flanking regions (2.0 avg in coding regions versus 4.0 avg in flanking regions). No locus deviated from neutral expectations by Tajima's D, Fu and Li's D*, or Fu and Li's Fs metrics.

The internal transcribed spacer (ITS) was sequenced from five individuals to provide comparisons to previous work. Only two variants were encountered, differing by seven SNPs (1.0%). The two variants were identical to ITS sequences reported elsewhere (Brown et al 2004). 1.3 transitions were observed for each transversion.

Di-allelism

In order to test for di-allelism, allelic divergence was compared within and between classes of alleles. At locus Pm2232, the common allele was at the midpoint of a neighbor-joining network. The divergence between the common allele and both the alternate and singleton haplotype was identical (one difference out of 91 bp), yielding a within to between allele class divergence ratio of 0.5 regardless of which haplotype was considered "within". For all other loci, haplotypes were clearly divided into two classes by neighbor-joining networks (Figs. 4-1). Excluding Pm2232, the average difference between haplotypes within allele classes ranged from 0% (Pm2988, Pm4488, and SOD1-short) to a maximum of 0.6% (SOD2-short). The difference between allelic classes

ranged from 2.2% to 7.4% (Pm2988 and SOD2-short respectively). The ratio of within allele class divergence to between allele class divergence had a maximum value of 0.1 among all loci except Pm2232, well below the 0.25 threshold for di-allelism.

Time of Divergence

In order to estimate the approximate divergence time between allele classes, the time to most recent common ancestor for the two allele classes at each locus was calculated with several assumptions. FindModel indicated that a Kimura 2-Parameter (K2P) model or a Hasegawa-Kishino-Yano (HKY) model are most appropriate for the data at each subset of loci (K2P for three loci and HKY for four loci). The K2P model of nucleotide substitution was used to calculate genetic distances between common and alternate alleles at each locus in order to limit the risk of over-parameterizing the model. K2P-corrected genetic distances ranged from 0.011 at locus Pm2232 to 0.087 at SOD2-short. The mean genetic distance across all loci was 0.036 with a variance of 0.00068. Using two standard deviations to define the 95% confidence interval, the divergence between alleles ranged from 0.010 to 0.062. Based on the assumed range of mutation rates, this yielded a range of divergence times from 250,000 years to 31 million years of independent evolution between allelic classes. The 120-fold difference in divergence times is a function of the multiplying the extremes of the 95% confidence interval by the fastest and slowest putative mutation rates. That is, the shortest divergence time is calculated by the lower bound of the confidence interval multiplied by the fastest mutation rate and the longest divergence time the result of the upper bound of genetic distance multiplied by the slowest mutation rate.

Recombination

Pairwise tests of linkage disequilibrium were conducted on microsatellite flanking loci to assess whether long-term asexuality could have sustained the divergence among homologous chromosomes. Linkage equilibrium could not be rejected for 4 of 10 pairwise tests ($0.16 < p < 0.68$). Additionally, three of ten pairwise comparisons between loci (Pm2232 and Pm2903, Pm2232 and Pm8517, and Pm2903 and Pm8517) conform to Hudson's four gamete test as being incompatible with a non-recombining ancestral history. However, there was no indication of intragenic recombination at any locus as intermediate haplotypes were not observed in haplotype trees (Fig. 4-1).

Geographic distribution of Genotypes

In order to determine whether particular allelic lineages were restricted to discrete geographic regions, the geographic distribution of alleles was examined. Both common and alternate haplotypes were widely distributed (Table 4-3, Fig. 4-2). The common allele was more prevalent in geographic locations south of Maryland for all loci. Alternate alleles were at higher frequencies in New Jersey and Massachusetts samples, though moderate frequencies were observed in New Smyrna Beach, FL and Port Charlotte, FL (Fig. 4-2).

Among the samples with genotypes for all five microsatellite flanking regions (54% of total), fourteen unique multilocus genotypes were observed (Table 4-3). Genotype 1 was observed in nine locations spanning from Texas to Rhode Island and accounted for 56% of all genotypes. This genotype was homozygous for the common allele at all loci. Two multilocus genotypes differed by only one nucleotide mutation from this common genotype (heterozygosity with a singleton allele), and are represented by Genotypes 12 and 13 respectively. Genotype 2 was found in five samples from New

Jersey or Massachusetts, but was notably absent from Rhode Island. Heterozygous loci were only observed in individuals from Maryland, New Jersey and Massachusetts (both Tisbury and Edgartown). Samples from Massachusetts (both Tisbury and Edgartown) showed the highest diversity, with four unique genotypes among five samples in both locations.

Discussion

Di-allelism characterized six out of seven nuclear loci in the *P. marinus* genome that were examined here, consistent with three other loci previously examined (Reece et al. 1997; Robledo et al. 1999; Brown et al. 2004). Nucleotide differences between allele classes across all loci averaged 3.8%, indicating that alleles had been evolving with no apparent recombination for an extended period of time at all loci, potentially as long as 31 million years but no less than 250,000 years if mutation rates are similar to other single-celled organisms (Machado and Ayala 2001, Kondrashov and Kondrashov 2010).

There was a remarkable lack of haplotype variation in *P. marinus*. 469 of 474 DNA sequences could be categorized as a high frequency "common" allele or an "alternate" allele at low to moderate frequencies, with only a few singleton mutants that differed by a single substitution from common or alternate haplotypes. This lack of haplotype variation suggests a small effective population size and great differences between census and effective population sizes.

Hypothesized causes of genomic di-allelism include genome duplication, ancient asexuality, balancing selection, or introgression between two ancient parasite lineages. Genome duplication seems unlikely because it would be expected that two alleles would be observed at every locus in every individual; homozygosity was encountered for both

the common and alternate alleles at all loci. Likewise, extended asexuality would have been expected to promote heterozygosity throughout the genome, yet heterozygosity was never seen for more than two out of five microsatellite flanking loci among observed multilocus genotypes. Balancing selection seems unlikely because this should only affect specific linked portions of the genome. In our samples, recombination was apparent between loci and thus alleles should have had the opportunity to escape selective pressure, allowing loss of alleles through genetic drift or new mutation to enter the population over time. If balancing selection is driving the diallelic pattern at all loci sampled, the selection must be acting on genomes as a whole, or at multiple loci under balancing selection. As *P. marinus* primarily reproduces in a clonal manner, this hypothesis remains plausible.

We must also consider hybridization between two ancient lineages for explaining this genomic di-allelic pattern. The large divergence between allelic classes at every locus is most consistent with extended isolation between two alleles. In order for this to have occurred by genetic drift, two *P. marinus* lineages most likely would have been physically separated (assuming that the ability for the two lineages to recombine was maintained throughout the period of allelic divergence). There have been two recent reports of *P. marinus* infection in oysters (*C. corteziensis* and *C. gigas*) on the Pacific coast of Mexico (Cáceres-Martínez et al. 2008; Enriquez-Espinoza et al. 2010). The genetic data are limited from these studies, but it is possible that these infections represent endemic infections rather than new introductions. If so, they could represent a potential source of highly diverged alleles in an allopatric population. While balancing selection acting on the entire genome remains a potential explanation for di-allelism in *P.*

marinus, it seems more parsimonious that the observed DNA sequences are a result of recent hybridization between two ancient lineages.

If hybridization between two ancient lineages caused the allelic patterns in *P. marinus*, ecological and epidemiological characteristics may have been altered. Current populations of Spartina hybrids are expanding their geographic range (Ayres et al. 2004). Hybridization apparently played a role in the establishment of anoles in Florida (Kolbe et al. 2004). Hybrid fungal diseases of plants have shown increased virulence (Barrett et al. 2007; Brasier et al. 1999), and changes in life-history, including timing of infection in rust infections (Barrett et al. 2007). Consequently, the putative hybridization of *P. marinus* lineages may have similarly changed the ecological and epidemiological properties of this parasite. This is worth further investigation, comparing and contrasting host-parasite interactions between varieties of distinct genetic lineages. This study provides a tool to genotype isolates prior to infection experiments, targeting distinct lineages for study.

The degree of clonality and the geographic distribution of lineages described here are consistent with what was observed in microsatellite analyses (Thompson Dissertation Chapters 2 and 3). Here, thirteen multilocus genotypes were observed with some recombination evident between loci. Multilocus linkage disequilibrium was apparent among all samples, suggesting that asexual reproduction plays a role in *P. marinus* populations, though nucleotide sequence data do not have the resolution that microsatellite markers provided for these analyses. The repeated genotypes were distributed among locations spanning a wide geographic range, suggesting migration has spread lineages long distances. Yet, there was a lineage restricted to New Smyrna Beach

and Port Charlotte, FL. Similarly, genotypes were more similar between New Jersey and Massachusetts samples than any found in Narragansett, RI indicating two introductions into the newly expanded range of the parasite. The geographic distribution of Genotype 1 matches the poorly defined central clade identified in the microsatellite analysis, though higher resolution for examining contemporary processes was available in the microsatellite data.

Di-allelism may be a common trait in parasitic protists. These findings are similar to those observed in *Plasmodium* and *Trypanosomes* (Rich et al. 2000; Machado and Ayala 2001; Putaporntip et al. 2006). Explanations for di-allelism have been varied and may be specific to each particular taxon. Here, the data seem to be most consistent with hybridization between two ancient lineages, but further research is needed to address the possibility of balancing selection.

Tables

Table 4-1. Primers for amplification and sequencing. Primer names are denoted first by locus name, then by specifics regarding the location within a locus. "F" and "L" refer to forward primers, and "R" refers to reverse primers. Internal primers for nesting were generally given the designation "int." SOD1 primers additionally have numbers indicating the starting position of the primers in the gene sequence provided in Schott et al. 2003.

Locus	Reaction	Primer Name	Sequence	Tm	Amplicon Size
Pm2232	Primary	Pm2232seqF1	tcaacggagcttttctcgat	60	1016
		Pm2232seqR1	tcgagtacactggcagcatc	60	
	Secondary	Pm2232seqFint	ccaggacatatcgtaaagca	56	124
		Pm2232seqR2	tacgctcgcagtgaccatag	60	
Pm2903	Primary	Pm2903seqF2	cagggagccatacgacttgtc	62	1031
		Pm2903Rseq	ctacccaccctatggctgaa	60	
	Secondary	Pm2903seqF	cacgtggtgtcgcatatttc	60	215
		Pm2903seqRint	aacatggcttgggtgatagg	57	
Pm2988	Primary	Pm2988SeqF2	aggcttcaacgcttccaata	60	869
		Pm2988seqR2	gaaatgaatcccggaaaggt	60	
	Secondary	Pm2988SeqF1	tcgatctttgcttgacatcg	60	267
		Pm2988seqRint	gaacgcgtttaacccatagc	60	
Pm4488	Primary	Pm4488seqF1	gacttgatcccttcggtgaa	60	803
		Pm4488seqR1	cctcgctctcattcttggtc	60	
	Secondary	Pm4488seqF1		60	321
		Pm4488seqRint	agctttggagacgtcgttgt	60	
Pm8517	Primary	Pm8517seqF2	gtcaaacggagacgtgtgtg	60	1222
		Pm8517seqR2	cgagcctacgaccaactttc	60	
	Secondary	Pm8517seqFint	ctggaccaggtcaggtttgt	60	405
		Pm8517seqR2		60	
SOD2 short	Primary	SOD2L	ggggagaatgtgttcaatgc	60	442
		SOD2R	gcctttcgcatgaagttctg	60	
		SOD2L		60	317
		SOD2Rint	tgccatacaaagacgaggaa	60	
SOD2-Full Length	5' amplicon	SOD2-5'd	tgatcggagacacgtactgc	60	994
		SOD2R			
	3' amplicon	SOD2L		60	994
		SOD2-3'c	cgtgtcaccacaattcaac	60	
	Sequencing	SOD2-5'd		60	
	5' amplicon	SOD2-x1L	tttcgtcccttcaaggtctg	60	
		SOD2-x2R	ggatgtaaccctggtgatgc	60	
		SOD2R		60	
	3' amplicon	SOD2L		60	
		SOD2-x5R	cgcaagtgaagagtggaat	60	
		SOD2-3'c		60	
		SOD2-x4L	ggggaacgttgatgagatga	60	
		SOD2-AtlL1	gggtcgaccacatatgttgcg	60	
		SOD2-GulfL1	agggtgatctcatatgttggt	60	
SOD1-short	Primary	SOD1-70L	cgttgcttctcgtcagtcac	60	661
		SOD1-731R	actctcccttatcgggttgc	60	

	Secondary	SOD1-290L	actctaccatcgccagcaag	60	441
		SOD1-731R	actctcccttatcgggttgc	60	
SOD1-long	Amplification	SOD1-70L		60	661
		SOD1-731R		60	
	Sequencing	SOD1-70L		60	
		SOD1-501L	aagtttcgagaagttcaagg	60	
		SOD1-542R	acagcagcagcgctgaactcc	60	
		SOD1-731R		60	

Table 4-2. Summary of sequence polymorphism and neutrality test results.

Locus	Pm2232	Pm2903	Pm2988	Pm4488	Pm8517	SOD2	SOD2 short	SOD1	SOD1-short	ITS	Total
# Sequences	142	142	142	146	134	6	158	6	44	10	**948**
Total length	91	181	232	222	298	1521	257	653	273	730	**3928**
# Haplotypes	3	3	2	2	3	2	4	2	2	2	
Singletons	1	1	0	0	1	0	2	0	0	0	
InDels	0	0	1	0	0	6	2	0	0	0	**9**
# bases indel			3			14	11				**28**
Transitions	1	4	4	10	9	55	10	15		4	**112**
Transversions	1	1	1	3	1	27	9	8		3	**54**
Transition/Transversion	1	4	4	3.33	9	2.04	1.11	1.88		1.33	**2.07**
Segregating Sites (S)	2	4	5	11	10	81	21	23	6	7	**143**
Nucleotide Diversity (pi)	0.0036	0.0069	0.0024	0.012	0.011	0.036	0.056	0.019	0.0059	0.0096	**0.038**
Max % difference	2.2	2.7	2.2	5.9	2.7	5.4	7.4	3.5	3.3	1.0	
Non-synonymous						15		0			**15**
Synonymous						15		8			**23**
Tajima's D	-0.327	1.229	-0.822	0.692	1.912		1.324		0.127		
P-value	>0.10	>0.1	>0.10	>0.1	0.05<p<0.1		>0.1		>0.1		
Fu and Li's D*	-0.987	-0.340	1.008	1.404	-0.117		0.811		1.185		
P-value	>0.1	>0.1	>0.1	0.05<p<0.1	>0.1		>0.1		>0.01		
Fu and li's F*	-0.918	0.206	0.471	1.368	0.721		1.217		1.001		
P-value	>0.1	>0.1	>0.1	>0.1	>0.1		>0.1		>0.01		

Table 4-3. Summary of all of the multilocus genotypes observed across all of the sampling locations. "c" is used to denote the common allele, "a" to represent the alternate allele, and "s" for each of the singleton alleles in the dataset. The number of times a genotype was observed is denoted by "N." Locations where each genotype was encountered are listed in the order of western Gulf of Mexico to Northeastern United States in a linear array along the coastline.

	N	Pm2232	Pm2903	Pm2988	Pm4488	Pm8517	Locations Observed
Genotype 1	28	cc	cc	cc	cc	cc	Sabine Lake, TX; Mozambique Point, LA; Apalachicola, FL; Port Charlotte, FL; St. Lucie River, FL; Cape Canaveral, FL; New Smyrna Beach, FL; Port Orange, FL; Delaware Bay, NJ; Narragansett Bay, RI;
Genotype 2	5	aa	aa	cc	cc	aa	Delaware Bay, NJ; Tisbury, MA; Edgartown, MA
Genotype 3	3	cc	aa	cc	cc	cc	Fort Pierce, FL; Dickinson, TX
Genotype 4	2	cc	cc	cc	aa	cc	St. Lucie River, FL; Dauphin Island, AL
Genotype 5	2	cc	aa	aa	cc	aa	Port Charlotte, FL; New Smyrna Beach, FL
Genotype 6	2	aa	ac	cc	cc	ac	Bennet Point, MD; Delaware Bay, NJ
Genotype 7	1	cc	ac	cc	cc	ac	Edgartown, MA
Genotype 8	1	aa	cc	cc	cc	ac	Tisbury, MA
Genotype 9	1	aa	cc	cc	cc	aa	Tisbury, MA
Genotype 10	1	aa	aa	cc	cc	ac	Edgartown, MA
Genotype 11	1	aa	ac	cc	cc	aa	Edgartown, MA
Genotype 12	1	cs	cc	cc	cc	cc	Port Orange, FL
Genotype 13	1	cc	cs	cc	cc	cc	Apalachicola, FL
Genotype 14	1	aa	cc	cc	cc	as	Tisbury, MA

Figure Legends

Figure 4-1. Unrooted neighbor-joining trees for seven loci based on Kimura-2-parameter corrected distances. Haplotypes are labeled as common, alternate, or singleton based on frequency of observances. The number of individuals with identical genotypes is provided in parentheses. Every locus except microsatellite flanking region Pm2232 is characterized by a long internal branch separating two classes of alleles. Few singleton mutations are observed.

Figure 4-2. Haplotype frequencies by location. Sampling locations are provided in a linear array along the coastline from the Gulf of Mexico to Massachusetts. Distribution of haplotypes varied by locus and sampling location, with higher frequencies of common haplotypes in locations south of Maryland.

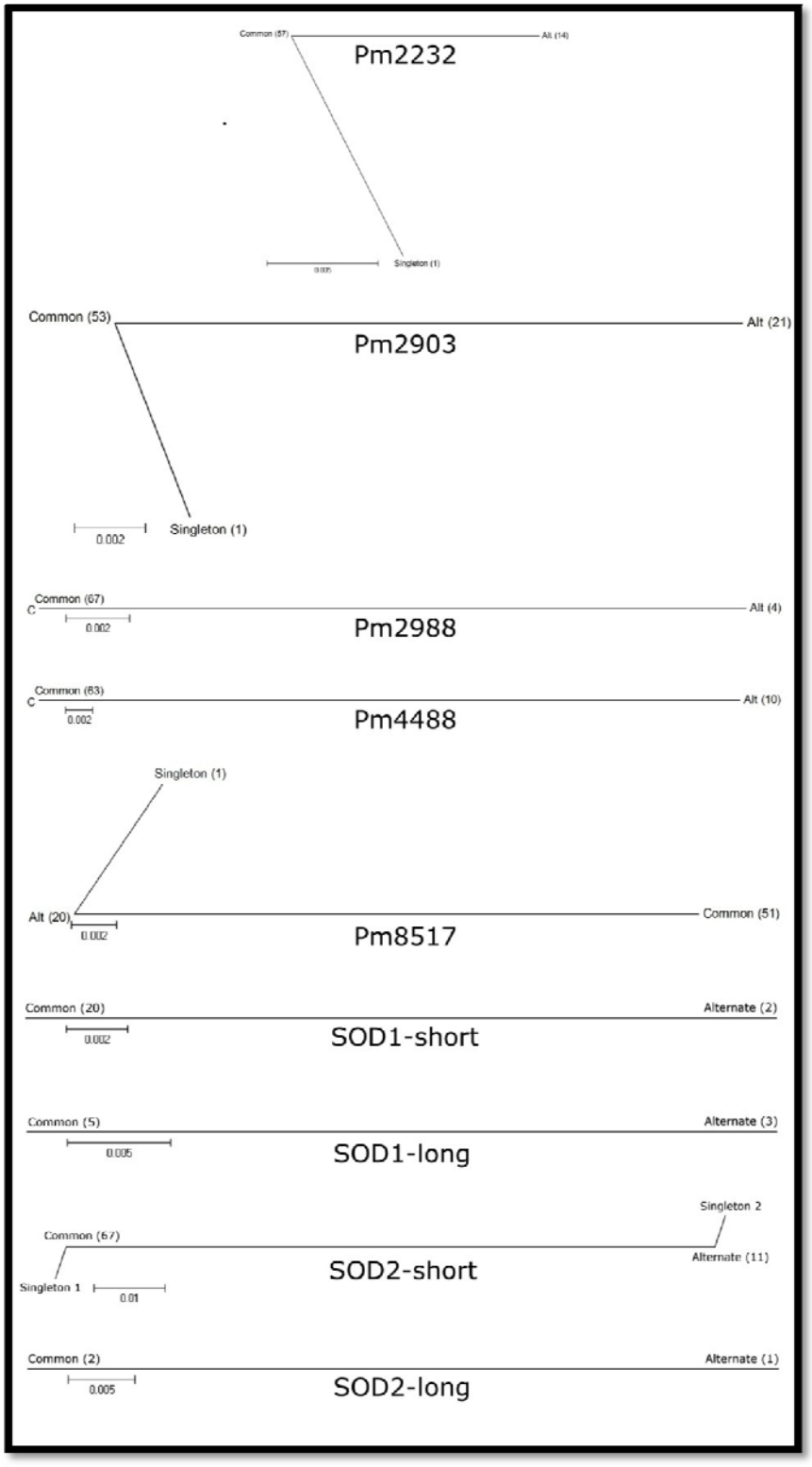
Common (67)
Alt (14)
Pm2232
Singleton (1)
Common (53)
Alt (21)
Pm2903
0.002
Singleton (1)
Common (67)
Alt (4)
0.002
Pm2988
Common (63)
Alt (10)
0.002
Pm4488
Singleton (1)
Alt (20)
Common (51)
0.002
Pm8517
Common (20)
Alternate (2)
0.002
SOD1-short
Common (5)
Alternate (3)
0.005
SOD1-long
Singleton 2
Common (67)
Alternate (11)
Singleton 1
0.01
SOD2-short
Common (2)
Alternate (1)
0.005
SOD2-long

Figure 4-1

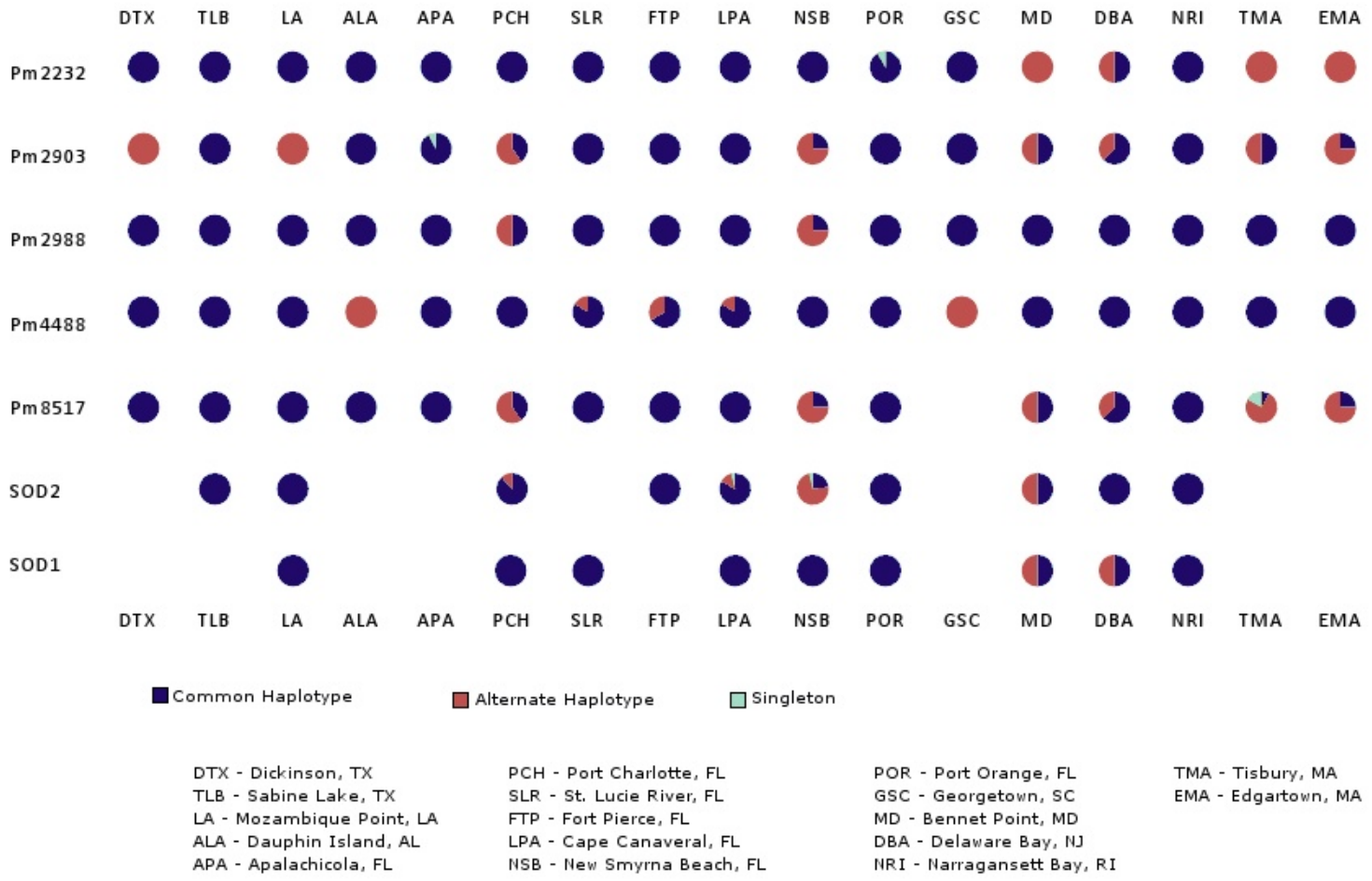
DTX
TLB
LA
ALA
APA
PCH
SLR
FTP
LPA
NSB
POR
GSC
MD
DBA
NRI
TMA
EMA
Pm2232
Pm2903
Pm2988
Pm4488
Pm8517
SOD2
SOD1
Common Haplotype
Alternate Haplotype
Singleton
DTX - Dickinson, TX
TLB - Sabine Lake, TX
LA - Mozambique Point, LA
ALA - Dauphin Island, AL
APA - Apalachicola, FL
PCH - Port Charlotte, FL
SLR - St. Lucie River, FL
FTP - Fort Pierce, FL
LPA - Cape Canaveral, FL
NSB - New Smyrna Beach, FL
POR - Port Orange, FL
GSC - Georgetown, SC
MD - Bennet Point, MD
DBA - Delaware Bay, NJ
NRI - Narragansett Bay, RI
TMA - Tisbury, MA
EMA - Edgartown, MA

Figure 4-2

Chapter Five – General Conclusions and Discussion

In this work, I use population genetics to investigate the evolution of an ecologically and economically important pathogen, *Perkinsus marinus*. Previous work focused on few cloned individuals from many locations and molecular markers with limited resolution. For this study, more individuals were genotyped within locations using highly variable microsatellite markers and DNA sequences in order to examine genetic processes locally as well as across a large geographic area. The goals of this study were to reveal the reproductive biology, dispersal abilities, and origins of extant *P. marinus* populations in order improve our understanding of the epidemiology of disease caused by this single-celled parasite.

The use of highly variable microsatellite markers and DNA sequence has revealed that *P. marinus* has a clonal population genetic structure, though sexual recombination occurs frequently enough that populations may evolve as a panmictic population at evolutionary timescales. This clonal population structure is a product of both asexual reproduction and potentially frequent inbreeding. Asexual reproduction was demonstrated by the repeated sampling of particular genotypes and observed for genotypes consisting of both microsatellite markers and DNA sequences. Multilocus linkage disequilibrium was significant in most sampled locations, also conforming to clonal population expectations. High F_{IS} values were observed in all locations except samples from Massachusetts, suggesting that inbreeding reinforces the maintenance of clonal lineages reproducing asexually. In Massachusetts, F_{IS} was negative suggestive of hybridization with insufficient recombination to break down heterozygous genotypes.

Barring a very high mutation rate, the genotypic diversity in all samples was too great in microsatellite genotypes to be accounted for solely by asexual lineages. Reassortment of alleles was detectable in both microsatellite and DNA sequence genotypes according to Hudson's four-gamete test (Hudson and Kaplan 1985). Given the divergence between common and alternate DNA haplotypes, reversion is unlikely to explain this result. Therefore sexual recombination has occurred among genotypes, but clonal genotypic patterns typify *P. marinus* populations, consistent with the clonal theory of parasitic protozoa.

Analyses of the connectivity among locations indicated that *P. marinus* is capable of rapid dispersal, over long distances. Isolation by distance was not found among locations, and regional patterns of genetic coherence were rejected. Rather, identifiable clades of parasites coexist within locations without forming interbreeding populations of parasites. Each clade was geographically widespread, but three clades had very high frequencies in specific geographic areas. When clades were analyzed independently, isolation-by-distance was still not apparent. This suggests that dispersal is rapid and recent, and does not solely depend on water currents. A parasite vector is more likely. Humans are the most likely vectors as they have moved oysters and associated parasites in recent history for economic reasons. However, efforts are being made to restrict introductions by limiting transport geographically and by quarantining oyster shell used as substrate. Unknown natural vectors may also disperse *P. marinus* long distances.

This study has also revealed an interesting pattern of di-allelism in *P. marinus*. The two allelic classes may have diverged millions of years ago. This may be due to balancing selection or allopatric divergence with recent secondary contact. Ecological

genetics approaches are warranted to ascertain what specific portions of the genome may be undergoing balancing selecton. If this is a recent hybridization event, it serves as a warning for other parasitic protists; nucleotide differences as much as 3.8% have not been accompanied by reproductive barriers. This brings into question what should be used as a species designation for parasitic protists and what are the potential consequences of increasing global connectivity through human transport. New hybrids may arise between divergent lineages that may result in new strains of parasites with unknown ecological and virulence properties. The high frequencies of heterozyotes in New Jersey and Massachusetts in this study may point to just such a case. It is not clear whether the high impact of Dermo disease in the newly expanded range is a result of virulent strains of *P. marinus*, a lack of host defenses, or a combination of both (though past study suggests it may be both (Bushek and Allen 1996).

Non-equilibrium population dynamics in *P. marinus* may have importance in host-parasite interactions. The coexistence of clades within locations provides a situation where one lineage can rise and fall in frequency in response to stochastic events, environmental selection, or host resistance. If host resistance is specific to parasite strain, this dynamic is consistent with a matching-allele model of infection (Howard and Lively 1994). If host resistance is specific to particular parasite clades, matching-allele host-parasite co-evolution may occur (Agrawal and Lively 2002). Recent study has shown that molluscs may be capable of specific immune responses (Bowden et al. 2007), though experimental infection studies suggest it may be more general in *C. virginica* (Bushek and Allen 1996). If there is a specific component to *C. virginica* resistance, selection mosaics may develop among locations where host resistance matches the most prevalent

clade at that particular location. Over time, this may lower prevalence of that clade allowing another to rise in frequency at that location. Over large geographic scales, coevolutionary dynamics will be dependent upon host gene flow for resistance genes or acquisition of new resistance via mutation. Certain oysters have been shown to be more susceptible to disease, differing in responses to challenge with parasites from different geographic locations (Bushek and Allen 1996), rendering this coevolutionary scenario plausible in *P. marinus*-*C. virginica* interactions. This deserves further investigation in natural populations. This suggests that dermo disease is unlikely to disappear in natural populations of oysters and disease severity may rise and fall over time, unless the basis of oyster immunity is very general which may or may not be the case (Bushek and Allen 1996; Tanguy et al. 2004).

The frequency of multiple infections observed in this study may provide insight into the evolution of virulence in *P. marinus*. Where multiple parasites infect a specific host, intraspecific competition among parasites is expected to select for the parasite with the highest reproductive rate; such traits may confer higher virulence (Frank 1996). In the specific case of *P. marinus*, multiple infections were observed more frequently in northeastern populations where highly heterozygous genotypes are common, indicating that selection for virulence should be highest in these locations. However, if parasites are related in multiple infections, it is predicted that selection for virulence should be attenuated by kin selection (Frank 1992). Based on the genotypes of those oysters that were deemed singly infected in the northeast, most parasite gentoypes derive from the same genotypic clade characterized by multiple heterozygous loci in all individuals. These genotypes may belong to the same clonally reproducing multilocus lineage such

that individual genotypes are closely related. Thus, until multiple recombination events occur within this population, selection will be acting on relatives, retarding response to selection for increased virulence in multiply infected hosts. Thus, it is expected that selection for higher virulence will occur more frequently in locations where distinct parasite lineages co-occur, resulting in competition among strains that are unrelated. This may in turn allow for evolution of resistance in host populations.

I have shown that *P. marinus* is consistent with the clonal theory of parasitic protozoa over ecological timescales, though sexual recombination is an important source of novelty for protistan parasite evolution. This parasite can disperse widely over short periods of time, initiating new epidemics when strains reach new locations. In order to limit the impact of this parasite on its host, most likely multiple host resistance phenotypes will be required.

Appendix

Genotype table for all samples determined to be singly infected. Genotypes consist of seven microsatellite loci. Numbers indicate the size of each amplified DNA fragment in basepairs. Three-letter start to each sample designates the sampling location. Abbreviations can be found in Thompson, Dissertation Chapter 3, Table 3-1. "0" indicates an inability to amplify a particular locus for an infected individual.

	Microsatellite Locus													
Sample Name	Pm2232		Pm2988		Pm2903		Pm4488		Pm8517		Pm9464		Pm12067	
EMA-Jl03	249	291	191	191	299	303	287	290	269	308	270	288	293	302
EMA-Jl07	288	288	218	218	281	281	284	287	296	296	279	279	287	299
EMA-Jl13	249	288	191	215	281	303	284	287	293	305	270	279	296	302
EMA-Jl15	246	288	191	218	303	303	284	284	308	308	279	279	296	317
EMA-Jl26	249	285	191	218	281	303	287	287	296	308	270	279	296	302
EMA-S02	249	288	191	218	281	303	284	287	293	302	270	279	296	302
EMA-S05	249	285	191	218	281	305	284	287	305	308	270	288	293	302
EMA-S06	249	288	191	218	303	303	284	287	269	299	270	270	296	302
EMA-S07	249	249	191	215	0	0	0	0	296	305	279	279	293	293
EMA-S13	249	249	0	0	303	303	287	287	269	269	270	270	305	305
EMA-S15	249	288	191	218	281	303	284	287	296	305	270	279	296	302
EMA-S19	249	285	191	218	279	303	284	287	296	308	270	279	296	302
EMA-S21	249	291	191	218	281	303	281	293	296	305	270	279	302	302
EMA-S22	249	285	191	191	281	281	284	287	272	299	270	288	302	302
EMA-S23	249	285	191	215	281	303	281	287	308	308	270	270	293	302
EMA-S26	249	288	191	218	281	303	284	287	296	305	270	279	302	302
EMA-S29	249	288	191	218	279	301	284	287	296	308	270	279	296	302
EMA-S30	249	288	191	218	279	303	284	287	296	308	270	279	302	302
EMA-Jn02	249	288	191	218	281	303	284	287	296	305	270	279	296	302
EMA-Jn05	249	288	191	218	281	303	284	287	305	305	270	279	296	302
EMA-Jn06	249	288	191	218	281	303	284	287	293	305	270	279	296	302
EMA-Jn08	249	288	191	218	281	303	269	287	308	308	0	0	302	302
EMA-Jn12	249	288	191	215	279	279	284	287	0	0	267	279	278	296
EMA-Jn16	249	291	191	218	281	301	284	287	296	308	270	279	296	302
EMA-Jn21	249	288	191	221	279	303	284	287	308	308	270	279	296	302
EMA-Jn26	249	288	191	218	281	303	284	287	296	308	264	270	296	302
TMA-A04	249	285	191	218	281	303	284	287	305	308	270	288	293	302
TMA-A07	249	285	191	218	281	305	284	287	305	308	270	288	296	302
TMA-A10	249	291	191	221	281	301	284	290	296	296	270	282	290	302
TMA-A11	249	288	191	218	281	303	284	287	296	308	0	0	302	302
TMA-A12	249	288	191	218	281	281	284	284	299	314	270	279	296	305
TMA-A13	249	285	191	215	281	301	284	287	296	308	270	279	296	302
TMA-A24	249	288	191	218	281	303	287	287	0	0	270	279	296	296
TMA-A27	249	285	191	218	281	303	284	290	296	302	270	279	296	302
TMA-A28	249	282	191	218	281	305	281	287	296	305	270	279	293	299
TMA-S02	249	282	191	218	281	305	284	287	305	305	270	282	302	302
TMA-S03	249	282	191	218	279	305	284	287	305	308	270	282	299	299

TMA-S04	249	282	191	218	281	305	287	287	305	308	270	282	302	302
TMA-S06	249	282	191	218	279	305	287	287	308	308	270	282	302	302
TMA-S08	291	291	191	218	0	0	0	0	308	308	0	0	305	305
TMA-S11	249	282	191	218	281	305	287	287	305	308	270	282	299	299
TMA-S12	249	285	191	218	281	305	284	287	305	308	270	282	299	299
TMA-S13	249	282	191	218	281	305	284	287	305	311	270	282	302	302
TMA-S14	249	249	191	191	281	305	0	0	0	0	270	282	299	305
TMA-S15	249	282	191	218	281	305	284	287	305	308	270	282	299	299
TMA-S16	249	282	191	218	281	305	284	290	305	305	270	282	302	302
TMA-S17	249	282	191	218	281	305	284	287	305	308	0	0	302	302
TMA-S22	249	282	191	218	281	305	284	287	305	311	270	282	302	302
TMA-S24	249	285	191	218	281	305	284	287	305	308	270	282	302	302
TMA-S28	249	282	191	218	281	305	284	287	305	308	282	282	302	302
TMA-S29	249	279	191	221	281	305	284	287	305	308	270	282	296	299
TMA-S30	249	282	191	218	281	305	287	287	302	308	270	282	293	299
TMA-S31	249	288	191	218	281	305	284	287	296	308	270	279	296	302
NRI-A03	246	249	215	215	305	305	290	290	296	296	270	270	293	293
NRI-A06	249	288	206	218	305	305	284	299	287	287	261	261	287	293
NRI-A07	249	285	215	215	305	305	290	290	296	299	270	270	293	293
NRI-A11	249	249	215	215	303	303	287	287	299	299	270	270	299	299
NRI-A12	249	249	215	233	303	303	287	287	299	302	270	270	299	299
NRI-A13	249	249	215	215	305	305	290	290	296	296	270	270	293	293
NRI-A14	249	249	215	215	305	305	290	290	296	296	270	270	293	293
NRI-A17	249	249	215	215	303	303	0	0	293	293	0	0	293	299
NRI-A21	249	249	215	215	303	305	290	290	296	296	270	270	293	293
NRI-A22	249	288	212	212	303	303	290	290	281	296	270	270	0	0
NRI-S02	249	288	215	215	301	301	290	290	296	296	270	279	293	296
NRI-S04	246	249	206	212	301	303	290	293	293	293	270	270	293	296
NRI-S06	249	249	215	215	303	303	287	287	299	299	270	270	302	302
NRI-S07	249	249	212	212	303	303	290	290	293	293	270	270	293	293
NRI-S13	249	249	209	209	305	305	290	290	296	296	270	270	293	293
NRI-S17	249	249	215	215	303	303	0	0	296	296	270	270	293	293
NRI-S21	246	249	212	215	303	303	287	290	293	296	270	270	293	293
NRI-S23	246	246	209	209	303	303	290	290	293	293	270	270	293	293
NRI-S25	249	249	212	212	303	303	0	0	293	293	270	270	293	293
NRI-S26	249	249	215	215	303	303	287	287	299	299	270	270	299	299
NRI-S27	246	246	209	209	303	303	290	290	293	293	270	270	293	293
NRI-S29	249	249	215	215	303	303	290	290	299	299	270	270	290	299
NRI-Jn11	249	249	0	0	303	303	0	0	299	299	270	270	299	311
NRI-Jn13	249	249	215	215	303	303	287	287	299	299	270	270	302	305
NRI-Jn17	264	291	215	215	0	0	299	299	296	296	0	0	320	320
NRI-Jn28	249	258	215	215	303	303	287	287	293	293	270	270	302	302
DBA-Jn04	249	288	191	218	279	303	284	290	299	314	270	285	293	302
DBA-A01	0	0	191	191	0	0	0	0	302	302	270	285	302	302
DBA-A02	249	285	191	191	281	281	284	287	293	302	270	282	302	302
DBA-A10	249	249	191	191	0	0	293	284	296	296	0	0	296	302
DBA-A11	249	288	233	233	0	0	296	299	308	308	0	0	299	299
DBA-A13	249	249	206	206	303	303	290	290	293	293	270	270	296	296
DBA-A18	249	288	191	218	281	301	287	287	296	308	270	285	293	302
DBA-A23	249	249	191	233	0	0	284	284	296	308	276	276	296	302
DBA-Jl06	249	276	191	191	0	0	0	0	296	308	270	270	0	0
DBA-Jl27	249	288	191	221	279	303	284	287	299	317	270	285	296	302
DBA-Jl28	249	288	197	221	0	0	278	290	299	314	270	282	293	302
BCL002	249	297	191	215	281	305	284	290	308	308	270	285	293	302
NBD004	249	285	191	221	281	303	284	290	293	311	270	282	293	302
NBD011	249	249	212	212	305	305	290	293	290	290	270	300	293	323
NBT001	0	0	194	233	0	0	0	0	308	308	315	315	293	293

NBT002	249	249	215	215	305	305	287	287	290	293	270	270	293	293
NBT003	249	249	215	215	0	0	0	0	0	0	270	270	293	293
NBT006	249	249	209	209	303	303	290	290	293	293	270	270	290	290
NBT013	249	249	212	212	303	303	293	293	290	290	270	270	296	296
NBT015	249	288	191	221	279	279	284	290	299	299	270	285	293	293
NBT019	249	249	215	215	305	305	290	290	293	293	270	270	299	299
SHR007	282	285	191	221	281	303	284	290	311	314	270	285	293	302
GSC001	0	0	0	0	0	0	293	293	287	287	267	267	278	293
GSC024	249	249	209	209	299	299	293	296	0	0	0	0	302	302
GSC025	249	249	209	209	299	299	293	296	293	293	270	270	302	302
GSC029	252	252	212	212	299	299	287	287	293	293	270	270	0	0
GSC032	249	249	209	209	299	299	299	299	281	293	0	0	0	0
GSC033	249	249	233	233	299	299	290	293	293	293	270	270	0	0
GSC038	249	249	209	233	303	303	293	293	290	290	270	270	305	305
GSC048	249	249	0	0	303	303	287	287	290	290	270	270	293	293
Gsc050	249	249	233	233	305	305	290	290	0	0	0	0	296	296
GSC070	246	246	206	233	301	301	290	290	0	0	0	0	293	293
GSC071	246	258	0	0	295	301	290	290	302	302	0	0	290	308
GSC077	255	301	233	233	301	301	293	293	0	0	0	0	278	278
GSC082	0	0	233	233	303	303	290	293	293	293	0	0	278	299
GSC083	246	246	206	206	301	301	290	293	287	296	0	0	278	293
POR-002	249	255	209	215	301	301	293	293	293	293	270	270	296	311
POR-003	249	249	209	209	305	305	299	299	293	293	270	270	293	293
POR-006	249	249	215	215	299	299	287	287	287	287	270	270	293	293
POR-008	249	249	215	215	303	303	290	290	296	296	270	270	302	302
POR-009	249	255	209	209	0	0	278	278	302	302	270	270	290	299
POR-010	249	255	209	209	299	299	287	287	302	302	270	270	308	308
POR-012	255	255	209	209	299	299	287	287	302	305	270	270	299	299
POR-014	249	249	212	212	0	0	293	293	296	296	270	270	299	299
POR-015	249	249	215	215	303	303	0	0	290	296	270	270	0	0
POR-016	249	249	212	212	299	299	290	290	293	293	270	270	302	302
POR-018	249	249	206	206	301	301	287	287	293	293	270	270	293	293
POR-019	249	249	215	215	303	303	287	293	296	296	273	273	299	299
POR-020	246	246	209	209	303	303	293	293	299	299	270	270	296	296
POR-024	255	255	212	212	0	0	290	293	293	293	270	270	305	305
POR-028	249	249	212	215	303	303	287	287	296	296	270	273	296	299
POR-031	255	255	209	209	301	301	293	293	293	293	270	270	308	308
POR-033	249	249	206	206	299	299	287	287	296	299	270	270	299	299
POR-034	249	255	0	0	303	303	0	0	0	0	252	252	296	296
POR-035	249	249	227	227	0	0	287	293	296	296	267	270	293	293
POR-038	255	255	209	230	299	299	287	287	302	302	270	270	308	308
POR-043	255	255	0	0	301	301	296	296	308	308	270	270	308	308
NSB-107	249	285	191	191	281	281	287	287	299	299	270	270	290	308
NSB-114	249	249	191	191	281	281	290	290	296	305	270	273	296	296
NSB-115	246	246	191	191	281	281	287	290	299	302	273	273	290	296
NSB-116	249	249	191	191	281	281	293	293	302	302	270	270	302	302
NSB-121	249	249	191	191	281	281	290	290	299	302	270	270	299	299
NSB-122	249	249	0	0	303	303	290	293	293	293	270	270	293	299
NSB-123	246	246	191	191	281	281	281	281	305	305	270	270	293	293
NSB-124	249	249	191	191	281	281	290	290	308	308	270	270	308	308
NSB-126	255	255	191	191	281	281	287	287	305	305	270	270	293	293
NSB-130	249	249	191	191	279	279	293	293	311	314	270	270	293	296
NSB-131	249	249	212	233	281	303	293	293	302	302	270	270	299	299
NSB-132	246	249	191	203	281	281	290	290	302	305	273	273	290	296
NSB-134	249	249	0	0	281	281	287	287	305	305	270	270	293	293
NSB-136	249	258	0	0	281	281	287	287	302	311	270	270	290	290
NSB-137	249	249	0	0	281	291	287	287	293	302	270	270	290	296

NSB-144	246	246	191	191	281	281	287	287	302	302	273	273	290	290
NSB-148	249	249	191	233	279	279	287	287	299	299	270	270	290	290
NSB2-201	0	0	191	191	281	281	290	290	305	305	270	270	293	293
NSB2-206	246	246	191	191	279	279	290	290	302	308	273	273	290	293
NSB2-207	246	246	191	233	281	281	290	290	302	305	273	273	293	293
NSB2-211	246	246	188	206	281	281	284	284	0	0	273	273	0	0
NSB2-212	249	249	191	209	281	303	293	293	296	308	270	270	296	296
NSB2-216	246	249	191	233	0	0	290	290	293	302	273	273	290	290
NSB2-217	246	246	191	191	281	281	290	290	308	311	273	273	290	290
NSB2-218	249	249	191	203	281	281	290	293	302	305	270	273	290	293
NSB2-219	246	249	191	191	281	281	287	290	302	305	273	273	290	290
NSB2-220	249	249	197	227	0	0	287	314	287	299	300	300	290	296
NSB2-221	249	249	191	191	281	299	290	290	296	302	270	270	302	302
NSB2-222	246	270	191	191	281	281	290	290	302	305	273	273	290	290
NSB2-224	246	246	191	191	281	281	287	293	305	305	273	273	290	290
NSB2-225	246	246	191	191	281	281	293	293	0	0	273	273	290	299
NSB2-226	243	246	194	233	0	0	284	284	305	305	273	273	0	0
NSB2-232	246	246	191	233	281	281	0	0	299	299	0	0	290	290
NSB2-234	0	0	191	191	281	281	287	287	305	305	270	270	293	293
NSB2-236	0	0	191	233	281	287	0	0	308	308	273	273	290	293
NSB2-238	249	249	191	191	281	283	0	0	0	0	273	273	0	0
NSB2-239	246	246	191	191	281	281	290	290	299	302	273	273	293	293
NSB2-240	0	0	191	191	281	281	0	0	302	302	0	0	293	296
NSB2-241	246	246	191	233	281	281	290	296	296	308	273	273	290	290
NSB2-242	252	252	191	233	281	281	287	287	308	308	270	270	296	296
NSB2-243	246	246	191	191	281	281	290	290	305	308	273	273	290	299
NSB2-245	249	249	191	191	281	281	290	290	308	308	270	270	290	293
NSB2-246	246	246	191	203	0	0	290	290	302	305	273	273	290	293
NSB2-247	0	0	191	191	281	281	290	290	308	308	273	273	293	293
LPA105	249	249	224	224	0	0	287	290	290	290	270	270	296	299
LPA106	249	249	212	212	303	303	287	293	293	293	270	270	296	296
LPA107	249	249	212	212	303	303	290	290	290	290	270	270	296	296
LPA108	0	0	212	212	303	303	290	290	287	287	270	270	296	296
LPA112	249	249	212	212	301	301	287	287	302	302	270	270	305	305
LPA114	249	249	215	215	303	303	290	290	293	296	270	270	299	299
LPA118	249	249	215	215	0	0	287	287	293	293	270	270	296	296
LPA123	249	249	212	212	301	301	287	287	302	302	270	270	299	299
LPA127	291	291	0	0	303	303	293	293	287	287	0	0	0	0
LPA128	249	249	0	0	303	303	0	0	287	287	306	306	296	296
LPA132	249	249	0	0	303	303	290	290	287	290	270	270	296	299
LPA134	249	249	212	212	303	303	293	302	287	287	270	270	290	299
LPA136	249	249	212	212	303	303	293	293	287	287	270	279	299	299
LPA137	0	0	224	224	303	303	0	0	293	296	0	0	296	308
LPA138	249	249	212	212	303	303	0	0	290	290	270	270	299	299
LPA145	249	249	212	224	303	303	293	293	290	290	270	270	296	296
LPA146	0	0	212	212	303	303	0	0	287	290	270	270	296	308
FTP 102	249	249	212	212	301	301	287	287	0	0	270	270	296	296
FTP 108	249	249	212	212	303	303	287	287	299	299	270	270	290	290
FTP 109	249	249	212	212	303	303	0	0	296	296	270	270	296	296
FTP 110	249	249	209	209	303	303	284	284	296	296	270	270	293	293
FTP 111	249	249	209	209	301	301	0	0	293	293	270	270	296	296
FTP 115	249	249	215	215	0	0	0	0	293	293	273	273	299	299
FTP 119	249	249	212	212	301	301	0	0	0	0	270	270	296	302
FTP 121	249	249	212	212	301	303	284	284	296	296	270	270	296	296
FTP 122	246	246	209	209	301	301	287	287	293	293	270	270	296	296
FTP 128	249	249	212	212	303	303	287	287	290	290	270	270	302	305
FTP 129	249	252	212	212	303	303	0	0	0	0	270	270	293	293

FTP 130	246	246	209	209	299	299	287	287	290	290	270	270	293	293
FTP 133	249	249	206	206	0	0	281	281	296	296	270	270	296	296
FTP 137	246	246	212	212	0	0	284	284	0	0	270	270	293	293
FTP 139	249	249	209	209	301	301	287	287	290	296	270	270	293	302
FTP 142	249	249	212	212	303	303	284	284	293	293	270	270	293	296
FTP 146	243	243	215	215	303	303	284	284	296	296	270	270	290	290
SLR103	246	252	206	212	303	305	281	284	299	299	270	270	299	299
SLR 107	246	246	206	206	303	305	0	0	275	299	0	0	0	0
SLR 121	249	249	209	218	301	303	287	290	299	299	270	270	293	296
SLR 127	252	252	209	209	303	303	287	287	293	293	270	270	296	296
SLR131	252	252	212	212	303	303	287	287	311	311	270	270	299	299
SLR132	249	249	206	212	303	303	284	284	296	299	267	270	302	308
SLR143	252	252	212	212	303	303	284	284	302	302	270	270	299	299
SLR105	249	249	212	212	0	0	284	284	302	302	270	270	299	302
SLR112	249	249	212	212	299	301	287	287	299	299	270	270	299	299
SLR115	249	249	218	218	301	301	287	287	299	299	270	270	293	296
SLR116	249	249	206	206	303	303	0	0	0	0	0	0	299	299
SLR119	258	258	0	0	0	0	284	284	299	299	0	0	299	299
SLR120	249	249	209	209	301	301	287	287	293	293	270	270	299	299
SLR126	252	252	212	233	303	303	0	0	311	311	270	270	296	302
SLR128	252	255	212	212	303	303	287	287	299	302	270	270	299	302
SLR133	237	249	212	233	303	303	287	287	308	308	270	270	296	296
SLR134	252	252	203	233	303	303	284	284	302	302	0	0	299	299
SLR135	240	249	212	212	303	307	287	287	299	302	270	270	293	302
SLR140	252	252	215	215	303	303	287	287	305	305	270	270	299	299
SLR142	246	246	215	215	303	303	287	287	302	305	270	270	299	299
SLR145	249	249	227	227	301	303	287	287	296	296	0	0	302	302
SLR146	240	249	206	215	303	303	287	287	305	311	270	270	299	305
SLR147	0	0	209	209	0	0	287	287	296	296	270	270	302	302
PCH002	249	249	191	191	281	301	290	293	290	299	270	270	290	290
PCH004	249	249	218	218	301	301	287	290	293	293	270	270	299	302
PCH009	249	249	191	191	281	281	290	290	299	299	0	0	287	293
PCH010	249	249	191	191	281	281	284	284	299	299	270	270	290	290
PCH012	249	249	191	221	301	301	290	290	287	287	270	270	293	293
PCH016	249	249	191	191	281	281	284	284	299	299	270	270	290	290
PCH017	249	249	218	218	281	301	0	0	299	299	270	270	299	299
PCH024	249	249	191	191	281	281	284	284	299	299	270	270	290	290
PCH025	249	249	191	218	281	281	287	287	293	308	270	270	293	293
PCH028	249	249	191	218	305	305	290	293	293	293	270	270	296	296
PCH032	249	249	191	191	281	281	287	287	308	308	270	270	296	299
PCH035	249	249	191	218	281	301	284	293	305	305	270	270	290	290
PCH046	249	249	218	218	281	301	293	293	293	293	270	270	299	299
PCH048	249	249	191	191	281	281	293	293	302	308	270	270	293	296
PCH050	249	249	191	218	281	301	293	293	302	302	270	270	293	299
PCH051	249	249	191	191	281	281	284	290	302	302	270	270	290	293
PCH056	249	249	191	221	281	301	287	293	290	305	270	270	287	296
PCH058	249	249	212	212	299	299	293	293	287	287	270	270	290	290
PCH064	249	249	218	218	301	301	293	293	293	293	0	0	302	305
PCH067	249	252	221	221	301	301	287	293	293	293	0	0	290	293
PCH070	249	249	209	209	299	299	293	293	293	293	270	270	299	299
PCH071	249	249	191	218	281	281	293	293	278	302	270	270	293	293
PCH072	249	249	221	221	299	299	293	296	293	293	270	270	296	296
PCH073	249	249	215	215	301	301	293	293	296	296	270	270	302	302
PCH074	249	249	191	191	279	279	293	293	302	317	270	270	293	293
PCH077	249	249	209	209	299	299	284	284	293	293	270	270	296	296
PCH078	249	249	218	218	281	301	287	296	293	299	270	270	293	293
PCH079	249	249	212	212	299	299	290	290	293	296	270	270	0	0

PCH086	249	249	215	215	299	299	287	287	302	302	270	270	296	296
PCH087	249	249	0	0	277	277	287	287	287	287	0	0	0	0
PCH088	249	249	221	221	279	299	290	290	287	287	270	270	293	296
PCH091	249	249	218	218	301	301	293	293	290	290	270	270	296	296
PCH092	249	249	218	218	301	301	293	293	296	296	270	270	302	302
PCH094	249	249	212	212	299	299	0	0	293	293	270	270	278	293
PCH098	249	249	191	227	281	281	290	290	302	305	270	270	293	293
PCH099	246	249	209	212	299	299	287	293	293	293	267	270	293	293
PCH100	249	249	218	221	303	303	293	293	287	293	270	270	302	302
APA002	249	249	218	218	305	305	278	287	299	299	270	270	302	305
APA003	249	252	212	215	305	309	290	293	290	296	270	270	287	299
APA004	249	249	206	206	307	307	281	281	296	296	270	270	296	296
APA010	249	249	224	224	307	307	287	290	296	296	270	270	296	296
APA017	252	252	212	212	301	301	290	293	305	305	270	270	293	293
APA018	249	249	209	215	305	305	287	290	293	293	270	270	293	296
APA024	246	249	215	215	0	0	290	290	0	0	270	270	0	0
APA025	249	249	212	212	307	307	290	290	293	293	270	273	302	302
APA030	252	252	218	221	307	307	0	0	293	293	0	0	290	293
APA032	252	252	218	221	307	307	287	287	293	293	270	270	290	293
APA033	246	252	218	218	305	307	287	290	293	293	270	270	290	293
APA034	252	252	209	218	281	281	287	290	293	296	282	282	290	293
APA035	249	249	218	218	309	309	290	290	296	296	270	270	293	293
APA036	249	249	218	218	311	311	290	293	293	293	270	270	293	293
APA037	0	0	218	218	281	281	0	0	293	293	270	270	0	0
APA040	252	252	218	218	307	307	287	287	293	293	270	270	293	293
APA041	249	249	206	206	305	305	287	290	293	293	270	270	299	299
APA049	249	252	212	218	305	305	290	290	293	296	270	270	302	302
APA050	249	249	215	215	279	305	290	293	299	299	270	270	299	299
APA052	252	252	215	218	309	309	287	287	293	293	270	270	293	293
APA055	249	249	215	215	305	305	287	287	296	296	0	0	296	296
APA062	243	243	215	215	305	305	287	299	296	296	270	270	293	293
APA065	249	249	215	218	303	303	287	293	293	293	270	270	299	299
APA066	249	249	209	209	309	309	284	290	296	296	270	270	290	290
APA068	249	249	209	212	305	305	0	0	293	293	270	270	293	293
APA069	249	249	212	215	305	305	287	290	293	308	270	270	290	299
APA080	249	249	212	221	305	305	0	0	0	0	270	270	0	0
APA081	249	252	209	212	305	305	290	293	296	296	270	270	293	293
APA095	249	249	215	215	305	305	290	290	290	290	270	270	296	296
ALA034	0	0	0	0	301	305	290	293	293	296	273	273	293	302
ALA035	249	249	203	203	303	303	284	287	296	296	270	270	290	302
ALA036	249	255	215	215	303	303	287	290	293	293	267	267	302	302
ALA039	249	249	218	218	301	303	284	290	296	296	270	270	293	293
ALA040	246	246	209	215	303	307	290	293	0	0	267	267	293	302
ALA041	246	246	188	188	279	299	0	0	296	299	279	279	299	299
ALA042	249	249	191	191	0	0	0	0	302	305	282	282	302	302
ALA043	252	252	194	218	283	305	290	290	302	308	285	285	305	305
ALA044	249	249	191	191	281	301	287	287	302	305	282	282	302	302
ALA046	0	0	0	0	293	305	287	287	281	296	0	0	299	299
ALA049	0	0	0	0	303	303	290	290	0	0	270	270	293	293
ALA050	0	0	221	221	305	305	0	0	293	299	270	270	290	290
ALA053	249	249	212	212	303	303	290	290	0	0	270	270	302	302
ALA055	0	0	0	0	305	305	287	299	281	296	270	270	290	305
ALA056	249	249	0	0	301	303	287	287	296	296	0	0	296	296
ALA060	249	249	0	0	305	305	290	290	290	293	270	270	296	296
ALA067	249	249	215	215	305	307	293	293	296	296	0	0	296	296
ALA068	249	249	212	212	303	303	293	293	296	296	270	270	311	311
ALA070	249	249	218	221	301	301	287	287	302	302	0	0	308	308

ALA074	0	0	0	0	303	303	290	290	0	0	267	267	302	305
ALA078	0	0	0	0	303	303	287	287	293	293	270	270	299	302
ALA082	0	0	0	0	305	305	287	287	281	299	0	0	278	296
ALA090	249	249	0	0	303	303	278	281	299	299	270	270	296	296
TLB002	249	249	212	212	303	303	287	293	299	299	270	270	290	293
TLB003	249	249	206	206	303	303	287	293	296	296	270	270	293	296
TLB005	249	249	209	212	303	303	0	0	293	296	270	270	290	290
TLB006	0	0	209	212	301	307	284	284	296	299	0	0	296	296
TLB007	249	252	191	191	279	303	287	287	296	314	270	270	293	293
TLB008	240	249	209	209	303	303	290	290	284	296	270	270	293	293
TLB009	249	249	0	0	303	307	0	0	290	320	270	270	293	293
TLB010	249	249	203	203	303	303	287	293	296	296	270	270	293	296
TLB011	249	249	212	215	303	303	290	305	290	290	270	270	293	299
TLB012	249	249	212	215	303	303	293	293	293	296	270	270	293	293
TLB013	249	249	212	215	303	303	287	293	293	293	267	267	299	299
TLB014	249	249	212	212	279	305	290	290	302	302	270	270	0	0
TLB015	249	249	212	215	303	303	290	290	296	296	270	270	290	296
TLB016	249	249	191	191	279	279	284	284	0	0	270	270	293	293
TLB018	0	0	212	212	301	303	287	287	296	296	0	0	287	287
TLB020	252	252	212	212	303	303	293	293	296	299	270	270	296	299
TLB021	249	249	212	212	303	303	290	290	305	308	270	270	284	287
TLB022	246	246	212	215	303	303	287	293	293	308	270	270	293	296
TLB028	252	252	0	0	301	301	287	293	0	0	270	270	0	0
TLB029	249	249	215	215	301	301	290	290	293	293	270	270	293	293
TLB030	246	249	191	212	279	303	287	293	302	302	270	270	284	296
TLB031	249	249	215	215	303	303	290	293	293	293	270	270	287	293
TLB035	246	249	212	215	285	303	290	290	296	296	270	270	290	299
TLB036	249	249	212	215	299	303	290	293	296	299	270	270	293	293
TLB039	246	249	212	212	303	303	290	293	293	296	270	279	293	293
TLB040	249	249	209	212	303	303	290	293	290	290	270	270	290	290
TLB041	246	249	206	212	303	303	284	290	296	299	270	270	284	293
TLB043	249	252	215	215	305	305	293	293	290	290	270	270	293	293
TLB044	249	249	209	212	301	301	290	293	302	302	270	270	287	287
TLB045	249	249	206	218	301	301	299	299	296	296	270	270	290	290
TLB046	249	249	206	206	303	303	293	299	293	293	270	270	290	290
TLB048	246	249	197	212	0	0	290	290	0	0	0	0	293	293
DTX003	0	0	194	194	303	303	290	296	305	305	0	0	0	0
DTX005	249	249	194	194	305	305	290	290	293	296	291	291	302	302
DTX009	249	249	194	194	307	307	287	290	296	296	282	282	299	302
DTX010	249	249	194	194	305	305	287	290	296	299	282	282	299	302
DTX011	249	249	194	194	305	305	287	290	299	299	282	282	299	302
DTX013	252	252	194	194	305	307	0	0	0	0	285	285	0	0
DTX014	249	249	194	194	303	303	287	287	296	296	0	0	302	302
DTX016	0	0	194	233	0	0	281	290	305	308	0	0	302	302
DTX021	249	249	191	194	309	309	296	296	299	302	288	288	299	302
DTX022	249	249	197	197	303	303	287	290	305	308	276	276	299	302
DTX023	0	0	194	194	301	307	296	296	296	302	0	0	302	302
DTX029	246	246	191	191	305	305	290	296	0	0	273	273	0	0
DTX030	252	252	194	194	303	303	290	290	296	299	285	285	302	302
DTX040	249	249	194	194	303	303	293	293	0	0	0	0	302	302
DTX042	249	249	194	194	303	305	290	290	0	0	282	282	302	302
DTX043	246	246	194	194	305	305	293	293	287	299	282	282	299	302

Bibliography

Agapow, P. M., and A. Burt. 2001. Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes* 1: 101-102.

Agrawal, A., and C. M. Lively. 2002. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. *Evolutionary Ecology Research* 4: 79-90.

Anderson, T. J. C., B. Haubold, J. T. Williams, J. G. Estrada-Francos, L. Richardson, R. Mollinedo, M. Bockarie, et al. 2000. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Molecular Biology and Evolution* 17: 1467–1482.

Andrews, J. D. 1955. Notes on fungus parasites of bivalve mollusks in Chesapeake Bay. *Proceedings National Shellfisheries Association* 45: 157-163.

Andrews, J. D. 1965. Infection experiments in nature with *Dermocystidium marinum* in Chesapeake Bay. *Chesapeake Science* 6: 60-67.

Andrews, J. D., and W. G. Hewatt. 1957. Oyster mortality studies in Virginia. 2. The fungus disease casued by *Dermocyctidium marinum* in oysters of Chesapeake Bay. *Ecological Monographs* 27: 1-25.

Andrews, J. D., and S. M. Ray. 1988. Management strategies to control the disease cause by *Perkinsus marinus*. *American Fisheries Society Special Publication* 18: 257-264.

Arnaud-Haond, S., C. M. Duarte, F. Alberto, and E. A. Serrao. 2007. Standardizing methods to address clonality in population studies. *Molecular Ecology* 16: 5115-5139.

Ayala, F. J. 1998. Is sex better? Parasites say "no." *Proceedings of the National Academy of Sciences* 95: 3346-3348.

Ayres, D. R., D. L. Smith, K. Zaremba, S. Klohr, and D. R. Strong. 2004. Spread of exotic cordgrasses and hybrids (*Spartina* sp.) in the tidal marshes of San Francisco Bay, California, USA. *Biological Invasions* 6: 221-231.

Balloux, F., L. Lehmann, and T. de Meeûs. 2003. The population genetics of clonal and partially clonal diploids. *Genetics* 164: 1635-1644.

Barbujani, G., and R. R. Sokal. 1991. Genetic population structure of Italy. I. Geographic patterns of gene frequencies. *Human Biology: an International Record of Research* 63: 253-272.

Barrett, L. G., P. H. Thrall, and J. J. Burdon. 2007. Evolutionary diversification through hybridization in a wild host? Pathogen interaction. *Evolution* 61: 1613-1621.

Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste, and F. Bonhomme. 1996. GENETIX 4.05, logiciel sous Windows TM pour la genetique des populations. Montpellier, France: Laboratoire Genome, Populations, Interactions.

Birky, Jr., C. W. 1996. Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. *Genetics* 144: 427-437.

Bowcock, A. M., A. Ruiz-Linares, J. Tomfohrde, E. Minch, J. R. Kidd, and L. L. Cavalli-Sforza. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368: 455-457.

Bowden, L., N. M. Dheilly, D. Raftos, and S. V. Nair. 2007. New immune systems: pathogen-specific host defence, life history strategies and hypervariable immune-response genes of invertebrates. *Invertebrate Survival Journal* 4: 127–136.

Boyle, J. P., B. Rajasekar, J. P. J. Saeij, J. W. Ajiokat, M. Berriman, I. Paulsen, D. S. Roos, L. D. Sibley, M. W. White, and J. C. Boothroyd. 2006. Just one cross appears capable of dramatically altering the population biology of a eukaryotic pathogen like *Toxoplasma gondii*. *Proceedings of the National Academy of Sciences* 103: 10514-10519.

Brasier, C. M., D. E. L. Cooke, and J. M. Duncan. 1999. Origin of a new *Phytophthora* pathogen through interspecific hybridization. *Proceedings of the National Academy of Sciences* 96: 5878-5883.

Brown, B. L., A. J. Butt, D. Meritt, and K. T. Paynter. 2005. Evaluation of resistance to Dermo in eastern oyster strains tested in Chesapeake Bay. *Aquaculture Research* 36: 1544–1554.

Brown, B. L., A. J. Butt, S. W. Shelton, D. Meritt, and K. T. Paynter. 2005. Resistance of Dermo in eastern oysters, *Crassostrea virginica* (Gmelin), of North Carolina but not Chesapeake Bay heritage. *Aquaculture Research* 36: 1391–1399.

Brown, G. D., K. L. Hudson, and K. S. Reece. 2004. Multiple polymorphic sites at the ITS and ATAN loci in cultured isolates of *Perkinsus marinus*. *Journal of Eukaryotic Microbiology* 51: 312-320.

Burreson, E. M., and L. M. Ragone Calvo. 1996. Epizootiology of *Perkinsus marinus* disease of oysters in Chesapeake Bay, with emphasis on data since 1985. *Journal of Shellfish Research* 15: 17–34.

Bushek, D., and S. K. Allen. 1996. Host-parasite interactions among broadly distributed

populations of the eastern oyster *Crassostrea virginica* and the protozoan *Perkinsus marinus*. *Marine Ecology Progress Series* 139: 127-141.

Bushek, D., S. E. Ford, and M. M.Chintala. 2002. Comparison of in vitro-cultured and wild-type *Perkinsus marinus*. III. Fecal elimination and its role in transmission. *Diseases of aquatic organisms* 51: 217–225.

Cáceres-Martínez, J., R. Vásquez-Yeomans, G. Padilla-Lardizábal, and M. A. del Río Portilla. 2008. *Perkinsus marinus* in pleasure oyster *Crassostrea corteziensis* from Nayarit, Pacific coast of Mexico. *Journal of Invertebrate Pathology* 99: 66-73.

Chintala, M. M., D. Bushek, and S. E. Ford. 2002. Comparison of in vitro-cultured and wild-type *Perkinsus marinus*. II. Dosing methods and host response. *Diseases of aquatic organisms* 51: 203–216.

Chu, F. L. E. 1996. Laboratory investigations of susceptibility, infectivity, and transmission of *Perkinsus marinus* in oysters. *Journal of Shellfish Research* 15: 57-66.

Chu, F. L. E., and A. K. Volety. 1997. Disease processes of the parasite *Perkinsus marinus* in eastern oyster *Crassostrea virginica*: minimum dose for infection initiation, and interaction of temperature, salinity and infective cell dose. *Diseases of Aquatic Organisms* 28: 61–68.

Chu, F. L. E., and J. F. La Peyre. 1993. *Perkinsus marinus* susceptibility and defense-related activities in eastern oysters *Crassostrea virginica*: temperature effects. *Diseases of Aquatic Organisms* 16: 223–223.

Coen, L. D., R. D. Brumbaugh, D. Bushek, R. Grizzle, M. W. Luckenbach, M. H. Posey, S. P. Powers, and S. G. Tolley. 2007. Ecosystem services related to oyster restoration. *Marine Ecology Progress Series* 341: 303-307.

Cook, T., M. Folli, J. Klinck, S. Ford, and J. Miller. 1998. The relationship between increasing sea-surface temperature and the northward spread of *Perkinsus marinus* (Dermo) disease epizootics in oysters. *Estuarine Coastal and Shelf Science* 46: 587–598.

Coss, C. A., J. A. F. Robledo, G. R. Vasta. 2001. Fine structure of clonally propagated in vitro life stages of a *Perkinsus* sp. isolated from the Baltic clam *Macoma balthica*. *Journal of Eurkaryotic Microbiology* 48: 38-51.

Craig, A., E. N. Powell, R. R. Fay, and J. M. Brooks. 1989. Distribution of *Perkinsus marinus* in Gulf Coast oyster populations. *Estuaries* 12: 82-91.

Crow, J. F., and M. Kimura. 1965. Evolution in sexual and asexual populations. *American Naturalist* 99: 439-450.

Dacks, J. B., and H. E. Kasinsky. 1999. Nuclear condensation in protozoan gametes and the evolution of anisogamy. *Comparative Biochemistry and Physiology - Part A: Molecular & Integrative Physiology* 124: 287-295.

De Meeus, T., F. Prugnolle, and P. Agnew. 2007. Asexual reproduction: genetics and evolutionary aspects. *Cellular and Molecular Life Sciences* 64: 1355-1372.

Dieringer, D., and C. Schlötterer. 2003. Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*: 167-169.

Dungan, C. F., and K. S. Reece. 2006. In vitro propagation of two *Perkinsus* spp. parasites from Japanese manila clams *Venerupis philippinarum* and description of *Perkinsus honshuensis* n. sp. *Journal of Eukaryotic Microbiology* 53: 316-326.

Dybdahl, M. F., and C. M. Lively. 1996. The geography of coevolution: comparative population structures for a snail and its trematode parasite. *Evolution* 50: 2264-2275.

Engle, V. D., and J. K. Summers. 1999. Latitudinal gradients in benthic community composition in Western Atlantic estuaries. *Journal of Biogeography* 26: 1007-1023.

Enriquez-Espinoza, T. L., J. M. Grijalva-Chon, R. Castro-Longoria, and J. Ramos-Paredes. 2010. *Perkinsus marinus* in *Crassostrea gigas* in the Gulf of California. *Diseases of Aquatic Organisms* 89: 269-273.

Evanno, G., S. Regnaut and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611-2620.

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* 1: 47-50.

Fernandez-Robledo, J. A., E. J. Schott, and G. R. Vasta. 2008. *Perkinsus marinus* superoxide dismutase 2 (PmSOD2) localizes to single-membrane subcellular compartments. *Biochemical and Biophysical Research Communications* 375: 215-219.

Fingerut, J. T., C. A. Zimmer, and R. K. Zimmer. 2003. Larval swimming overpowers turbulent mixing and facilitates transmission of a marine parasite. *Ecology* 84: 2502-2515.

Ford, S. E. 1996. Range extension by the oyster parasite *Perkinsus marinus* into the

northeastern United States: response to climate change? *Journal of Shellfish Research* 15: 45–56.

Ford, S. E., M. M. Chintala, and D. Bushek. 2002. Comparison of in vitro-cultured and wild-type *Perkinsus marinus*. I. Pathogen virulence. *Diseases of aquatic organisms* 51: 187–201.

Ford, S. E., and M. M. Chintala. 2006. Northward expansion of a marine parasite: testing the role of temperature adaptation. *Journal of Experimental Marine Biology and Ecology* 339: 226–235.

Ford, S. and R. Smolowitz. 2007. Infection dynamics of an oyster parasite in its newly expanded range. *Marine Biology* 151: 119-133.

Frank, S. A. 1992. A kin selection model for the evolution of virulence. *Proceedings of the Society for Experimental Biology and Medicine* 250: 195-197.

Frank, S. A. 1996. Models of parasite virulence. *Quarterly Review of Biology* 71: 37-78.

Futuyma, D. J. 2006. *Evolutionary Biology*. Sunderland, MA: Sinauer Associates.

Gauthier, J. D., and G. R. Vasta. 1995. In vitro culture of the eastern oyster parasite *Perkinsus marinus*: optimization of the methodology. *Journal of Invertebrate Pathology* 66: 156-168.

Goudet, J. 1995. FSTAT (Version 1.2): a computer program to calculate F-statistics. *Journal of Heredity* 86: 485-486.

Green, R. F. and D. L. G. Noakes. 1995. Is a little bit of sex as good as a lot? *Journal of Theoretical Biology* 174 (1): 87-96.

Haas, W. 1992. Physiological analysis of cercarial behavior. *Journal of Parasitology* 78: 243-255.

Halkett, F., J. C. Simon, and F. Balloux. 2005. Tackling the population genetics of clonal and partially clonal organisms. *Trends in Ecology & Evolution* 20: 194-201.

Hare, M. P., and J. C. Avise. 1996. Molecular genetic analysis of a stepped multilocus cline in the American oyster (*Crassostrea virginica*). *Evolution* 50: 2305-2315.

Hartl, D. L., S. K. Volkman, K. M. Nielsen, A. E. Barry, K. P. Day, D. F. Wirth, and E. A. Winzeler. 2002. The paradoxical population genetics of *Plasmodium falciparum*. *Trends in Parasitology* 18: 266–272.

Hartl, D. L., and A. G. Clark. 2007. *Principles of Population Genetics*. Sunderland, MA: Sinauer Associates.

Harvell, C. D., K. Kim, J. M. Burkholder, R. R. Colwell, P. R. Epstein, D. J. Grimes, E. E. Hofmann, et al. 1999. Emerging marine diseases - climate links and anthropogenic factors. *Science* 285: 1505-1510.

Hechinger R. F., and K. D. Lafferty. 2005. Host diversity begets parasite diversity: bird final hosts and trematodes in snail intermediate hosts. *Proceedings of the Royal Society B: Biological Sciences* 272: 1059-1066.

Hewatt, W. G., and J. D. Andrews. 1956. Temperature control experiments on the fungus disease *Dermocystidium marinum* of oysters. Proceedings National Shellfisheries Association 46:129-133.

Howard, R. S., and C. M. Lively. 1994. Parasitism, mutation accumulation and the maintenance of sex. *Nature* 367: 554-557.

Hudson, R. R., and N. L. Kaplan. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111: 147-164.

Huelsenbeck, J. P., and P. Andolfatto. 2007. Inference of population structure under a Dirichlet process model. *Genetics* 175: 1787-1802.

Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23: 1801-1806.

Joseph, S., J. Fernandez-Robledo, M. Gardner, N. El-Sayed, C. H. Kuo, E. Schott, H. Wang, J. Kissinger, and G. Vasta. 2010. The alveolate *Perkinsus marinus*: biological insights from EST gene discovery. *BMC Genomics* 11: 228.

Kleinschuster, S. J., and S. L. Swink. 1993. A simple method for the in-vitro culture of *Perkinsus marinus*. *Nautilus* 107: 76-78.

Kolbe, J. J., R. E. Glor, L. R. G. Schettino, A. C. Lara, A. Larson, and J. B. Losos. 2004. Genetic variation increases during biological invasion by a Cuban lizard. *Nature* 431: 177-181.

Kondrashov, F. A. and A. S. Kondrashov. 2010. Measurements of spontaneous rates of mutations in the recent past and near future. *Philosophical Transactions of the Royal Society B-Biological Sciences* 365:1160-1176.

Kotob, S. I., S. M. Mclaughlin, P. van Berkum, and M. Faisal. 1999. Discrimination between two *Perkinsus* spp. isolated from the softshell clam, *Mya arenaria*, by sequence analysis of two internal transcribed spacer regions and the 5.8S ribosomal RNA gene. *Parasitology* 119: 363-368.

Langella O (2001) Populations 1.2.24: population genetic structure (individuals or populations distances, phylogenetic trees). Available: http://bioinformatics.org/~tryphon/populations/.

La Peyre, J. F., M. Faisal, and E. M. Burreson. 1993. In vitro propagation of the protozoan *Perkinsus marinus*, a pathogen of the eastern oyster *Crassostrea virginica*. *Journal of Eukaryotic Microbiology* 40: 304-310.

Levine, N. D. 1978. *Perkinsus* gen. n. and other new taxa in the protozoan phylum Apicomplexa. *Journal of Parasitology* 64:549.

Librado, P., and J. Rozas. 2009. DnaSP (version 5): a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452.

Llewellyn, M. S., M. A. Miles, H. J. Carrasco, M. D. Lewis, M. Yeo, J. Vargas, F. Torrico, et al. 2009. Genome-scale multilocus microsatellite typing of *Trypanosoma cruzi* discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. *PLoS Pathogens* 5: e1000410.

Lynch, M., R. Burger, D. Butcher, and W. Gabriel. 1993. The mutational meltdown in asexual populations. *Journal of Heredity* 84: 339-344.

Machado, C. A., and F. J. Ayala. 2001. Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. *Proceedings of the National Academy of Sciences* 98: 7396-7401.

Mackin, J. G. 1953. Incidence of infection in oysters by *Dermocystidium* in the Barataria Bay area of Louisiana. Proceedings of the National Shellfisheries Association 42: 22-35.

Mackin, J. G., H. M. Owen, and A. Collier. 1950. Preliminary note on the occurrence of a new protistan parasite, *Dermocystidium marinum* n. sp. in *Crassostrea virginica* (Gmelin). *Science* 111: 328-329.

Mallon, M., A. MacLeod, J. Wastling, H. Smith, B. Reilly, and A. Tait. 2003. Population structures and the role of genetic exchange in the zoonotic pathogen *Cryptosporidium parvum*. *Journal of Molecular Evolution* 56: 407-417.

McCollough, C. B., B. W. Albright, G. R. Abbe, L. S. Barker, and C. F. Dungan. 2007. Acquisition and progression of *Perkinsus marinus* infections by specific pathogen-free juvenile oysters (*Crassostrea virginica* Gmelin) in a mesohaline Chesapeake Bay tributary. *Journal of Shellfish Research* 26: 465–477.

Meirmans, P. G., and P. H. Van Tienderen. 2004. Genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology*

Notes 4: 792-794.

Miller, L. H., T. Roberts, M. Shahabuddin, and T. F. McCutchan. 1993. Analysis of sequence diversity in the *Plasmodium falciparum* merozoite surface protein-1 (MSP-1). *Molecular and Biochemical Parasitology* 59: 1-14.

Minchella, D. J., and M. E. Scott. 1991. Parasitism: a cryptic determinant of animal community structure. *Trends in Ecology & Evolution* 6: 250-254.

Morrison, L. J., A. Tweedie, A. Black, G. L. Pinchbeck, R. M. Christley, A. Schoenefeld, C. Hertz-Fowler, A. MacLeod, C. Michael, R. Turner, and A. Tait. 2009. Discovery of mating in the major African livestock pathogen *Trypanosoma congolense*. *PLoS One* 4: e5564.

Moss, J. A., J. Xiao, C. F. Dungan, and K. S. Reece. 2008. Description of *Perkinsus beihaiensis* n. sp., a new *Perkinsus* sp. parasite in oysters of southern China. *The Journal of Eukaryotic Microbiology* 55: 117-130.

Nordborg, M. 2000. Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics* 154: 923–929.

Nuismer, S. L., and S. P. Otto. 2004. Host-parasite interactions and the evolution of ploidy. *Proceedings of the National Academy of Sciences* 101: 11036-11039.

Oliver, J. L., P. M. Gaffney, Jr., S. K. Allen, M. Faisal, and S. L. Kaattari. 2000. Protease inhibitory activity in selectively bred families of eastern oysters. *Journal of Aquatic Animal Health* 12:136-145.

Oliver, L. M., W. S. Fisher, S. E. Ford, L. M. R. Calvo, E. M. Burreson, E. B. Sutton, and J. Gandy. 1998. *Perkinsus marinus* tissue distribution and seasonal variation in oysters *Crassostrea virginica* from Florida, Virginia and New York. *Diseases of Aquatic Organisms* 34: 51-61.

Oura, C. A. L., B. B. Asiimwe, W. Weir, G. W. Lubega, and A. Tait. 2005. Population genetic analysis and sub-structuring of *Theileria parva* in Uganda. *Molecular & Biochemical Parasitology* 140: 229–239.

Paynter, K. T. 1996. The effects of *Perkinsus marinus* infection on physiological processes in the eastern oyster, *Crassostrea virginica*. *Journal of Shellfish Research* 15: 119-125.

Peakall, R., and P. E. Smouse. 2006. Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288-295.

Pecher, W. T., M. R. Alavi, E. J. Schott, J. A. Fernandez-Robledo, L. Roth, S. T. Berg, and G. R. Vasta. 2008. Assessment of the northern distribution range of selected

Perkinsus species in eastern oysters (*Crassostrea virginica*) and hard clams (*Mercenaria mercenaria*) with the use of PCR-based detection assays. *Journal of Parasitology* 94: 410-422.

Perkins, F. O. 1976. Zoospores of the oyster pathogen, *Dermocystidium marinum*. I. Fine structure of the conoid and other sporozoan-like organelles. *Journal of Parasitology* 62: 959-974.

Perkins, F. O. 1996. The structure of *Perkinsus marinus* (Mackin, Owen & Collier, 1950) (Levine, 1978) with comments on taxonomy and phylogeny of *Perkinsus* spp. *Journal of Shellfish Research* 15: 67–88.

Perkins, F. O. and R. W. Menzel 1966. Morphological and cultural studies of a motile stage in the life cycle of *Dermocystidium marinum*. *Proceedings National Shellfisheries Association* 56:23-30.

Piry, S., G. Luikart, and J. M. Cornuet. 1999. BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity* 90: 502-503.

Polley, S., and David J. Conway. 2001. Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 gene. *Genetics* 158: 1505-1512.

Polley, S., G. Weedall, A. Thomas, L. Golightly, and D. Conway. 2005. Orthologous gene sequences of merozoite surface protein 1 (MSP1) from and confirm an ancient divergence of alleles. *Molecular and Biochemical Parasitology* 142: 25-31.

Powell, E. N., J. M. Klinck, and E. E. Hofmann. 1996. Modelling diseased oyster populations .2. Triggering mechanisms for *Perkinsus marinus* epizootics. *Journal of Shellfish Research* 15: 141-165.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

Putaporntip, C., S. Jongwutiwes, T. Iwasaki, H. Kanbara, and A. L. Hughes. 2006. Ancient common ancestry of the merozoite surface protein 1 of *Plasmodium vivax* as inferred from its homologue in *Plasmodium knowlesi*. *Molecular and Biochemical Parasitology* 146: 105-108.

Quick, J. A. and J. G. Mackin. 1971. Oyster parasitism by *Labyrinthomyxa marina* in Florida. Florida Department of Natural Resources, Marine Research Laboratory, Professional Papers Series 13: 1-55.

Ragone Calvo, L. M., G. W. Calvo, and E. M. Burreson. 2003. Dual disease resistance in

a selectively bred eastern oyster, *Crassostrea virginica*, strain tested in Chesapeake Bay. *Aquaculture* 220: 69-87.

Ravel, C., S. Cortes, F. Pratlong, F. Morio, J. P. Dedet, and L. Campino. 2006. First report of genetic hybrids between two very divergent *Leishmania* species: *Leishmania infantum* and *Leishmania major*. *International Journal for Parasitology* 36: 1383-1388.

Ray, S. M. 1954. Experimental studies on the transmission and pathogenicity of *Dermocystidium marinum*, a fungous parasite of oysters. *Journal of Parasitology* 40: 235-235.

Ray S. M., and A. C. Chandler 1955. *Dermocystidium marinum*, a parasite of oysters. Experimental Parasitology 4: 172-200.

Raymond, M., and F. Rousset. 1995. Genepop (version 3.4): population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86: 248–249.

Razakandrainibe, F. G., P. Durand, J. C. Koella, T. De Meeüs, F. Rousset, F. J. Ayala, and F. Renaud. 2005. "Clonal" population structure of the malaria agent *Plasmodium falciparum* in high-infection regions. *Proceedings of the National Academy of Sciences* 102: 17388-17393.

Reece, K. S., C. F. Dungan, and E. M. Burreson. 2008. Molecular epizootiology of *Perkinsus marinus* and *P. chesapeaki* infections among wild oysters and clams in Chesapeake Bay, USA. *Diseases of Aquatic Organisms* 82: 237-248.

Reece, K. S., D. Bushek, and J. E. Graves. 1997. Molecular markers for population genetic analysis of *Perkinsus marinus*. *Molecular Marine Biology and Biotechnology* 6: 197-206.

Reece, K. S., D. Bushek, K. Hudson, and J. Graves. 2001. Geographic distribution of *Perkinsus marinus* genetic strains along the Atlantic and Gulf coasts of the USA. *Marine Biology* 139: 1047-1055.

Reece, K. S., M. E. Siddall, E. M. Burreson, and J. E. Graves. 1997. Phylogenetic analysis of *Perkinsus* based on actin gene sequences. *The Journal of Parasitology*: 417–423.

Reece, K. S., M. E. Siddall, N. A. Stokes, and E. M. Burreson. 2004. Molecular phylogeny of the *Haplosporidia* based on two independent gene sequences. *Journal of Parasitology* 90: 1111–1122.

Rich, S. M., M. U. Ferreira, and F. J. Ayala. 2000. The origin of antigenic diversity in *Plasmodium falciparum*. *Parasitology Today* 16: 390-396.

Robledo, J. A. F., A. C. Wright, A. G. Marsh, and G. R. Vasta. 1999. Nucleotide sequence variability in the nontranscribed spacer of the rRNA locus in the oyster parasite *Perkinsus marinus*. *Journal of Parasitology* 85: 650-656.

Robledo, J. A. F., J. D. Gauthier, C. A. Coss, A. C. Wright, and G. R. Vasta. 1998. Species-specificity and sensitivity of a PCR-based assay for *Perkinsus marinus* in the eastern oyster, *Crassostrea virginica*: a comparison with the fluid thioglycollate assay. *Journal of Parasitology* 84: 1237-1244.

Rohde, K. 2005. *Marine Parasitology*. Victoria, Australia: CSIRO Publishing.

Rougeron, V., T. de Meeûs, M. Hide, E. Waleckx, H. Bermudez, J. Arevalo, A. Llanos-Cuentas, et al. 2009. Extreme inbreeding in *Leishmania braziliensis*. *Proceedings of the National Academy of Sciences* 106: 10224-10229.

Roy, S. W., M. U. Ferreira, and D. L. Hartl. 2006. Evolution of allelic dimorphism in malarial surface antigens. *Heredity* 100: 103-110.

Rozen S, Skaletsky H (2000) primer 3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds Krawetz S, Misener S), pp. 365–386. Humana Press, Totowa, New Jersey.

Saldarriaga, J. F., M. L. McEwan, N. M. Fast, F. J. R. Taylor, and P. J. Keeling. 2003. Multiple protein phylogenies show that *Oxyrrhis marina* and *Perkinsus marinus* are early branches of the dinoflagellate lineage. *International Journal of Systematic and Evolutionary Microbiology* 53: 355-365.

Schott, E. J., J. A. F. Robledo, A. C. Wright, A. M. Silva, and G. R. Vasta. 2003. Gene organization and homology modeling of two iron superoxide dismutases of the early branching protist *Perkinsus marinus*. *Gene* 309: 1-9.

Schwenkenbecher, J. M., T. Wirth, L. F. Schnur, C. L. Jaffe, H. Schallig, A. Al-Jawabreh, O. Hamarsheh, K. Azmi, F. Pratlong, and G. Schonian. 2006. Microsatellite analysis reveals genetic structure of *Leishmania tropica*. *International Journal for Parasitology* 36: 237-246.

Siddall, M. E., K. S. Reece, J. E. Graves, and E. M. Burreson. 1997. 'Total evidence' refutes the inclusion of *Perkinsus* species in the phylum *Apicomplexa*. *Parasitology* 115: 165–176.

Smith, N. F. 2001. Spatial heterogeneity in recruitment of larval trematodes to snail intermediate hosts. *Oecologia* 127:115-122.

Spieth, P. T. 1974. Gene flow and genetic differentiation. *Genetics* 78: 961-965.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24: 1596-1599.

Tanabe, K., M. Mackay, M. Goman, and J. G. Scaife. 1987. Allelic dimorphism in a surface-antigen gene of the malaria parasite *Plasmodium falciparum*. *Journal of Molecular Biology* 195: 273-287.

Tanguy, A., X. Guo, and S. E. Ford. 2004. Discovery of genes expressed in response to *Perkinsus marinus* challenge in eastern (*Crassostrea virginica*) and Pacific (*C. gigas*) oysters. *Gene* 338: 121-131.

Thompson, J. N. 2005. *The geographic mosaic of coevolution*. Chicago: University of Chicago Press.

Thurston, M. I., and D. Field. 2005. MSATFinder: detection and characterisation of microsatellites. Distributed by the authors at http://www.genomics.ceh.ac.uk/msatfinder.

Tibayrenc, M., and F. J. Ayala. 2002. The clonal theory of parasitic protozoa: 12 years on. *Trends in Parasitology* 18: 405-410.

Tibayrenc, M., F. Kjellberg, and F. J. Ayala. 1990. A clonal theory of parasitic protozoa: the population structures of *Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas,* and *Trypanosoma* and their medical and taxonomical consequences. *Proceedings of the National Academy of Sciences* 87: 2414-2418.

Tibayrenc, M., F. Kjellberg, J. Arnaud, B. Oury, S. F. Breniere, M. Darde, and F. J. Ayala. 1991. Are eukaryotic microorganisms clonal or sexual? A population genetics vantage. *Proceedings of the National Academy of Sciences* 88: 5129-5133.

U. S. Environmental Protection Agency. 2008. Chesapeake Bay Health and Restoration Assessment 2007. Annapolis: Chesapeake Bay Program.

Volf, P., I. Benkova, J. Myskova, J. Sadlova, L. Campino, and C. Ravel. 2007. Increased transmission potential of *Leishmania major/Leishmania infantum* hybrids. *Iinternational Journal for Parasitology* 37: 589-593.

Walsh, P. S., D. A. Metzger, and R. Higuchi. 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques* 10: 506-513.

Welch, M., B. David, and M. Meselson. 2000. Evidence for the evolution of *Bdelloid* rotifers without sexual reproduction or genetic exchange. *Science* 288: 1211.

White, M. E., E. N. Powell, E. A. Wilson, and S. M. Ray. 1989. The spatial distribution of *Perkinsus marinus*, a protozoan parasite , in relation to its oyster host (*Crassostrea virginica*) and an ectoparasitic gastropod (*Boonea impressa*). *Journal of the Marine Biological Association of the United Kingdom* 69: 703-717.

Wright, S. 1943. Isolation by Distance. *Genetics* 28: 114-138.