ABSTRACT

Title of Document: POPULATION GENETICS OF THE EASTERN OYSTER IN CHESAPEAKE BAY

Colin Rose, Doctor of Philosophy, 2008

Directed By: Professor Matthew P. Hare
Department of Biology

The eastern oyster, *Crassostrea virginica*, plays an important role in the ecology of Chesapeake Bay. Its large population size, long larval dispersal stage and potential for high variance in reproductive success is representative of many marine invertebrates. Nevertheless, many important aspects of the oyster's biology remain unclear. I investigated how migration, natural selection, and effective population size have shaped the evolution of Chesapeake oysters. First, I examined aspects of genetic connectivity among oysters from rivers throughout the Bay. A correlation between geographic and genetic distance indicated that oyster larval dispersal tends to be local and that migration between Bay tributaries is rare over an ecological time scale. This result contributes to a growing body of literature indicating that larval dispersal is not passive. Next, I showed that a pattern of non-neutral mitochondrial evolution previously observed in different oyster populations also existed in Chesapeake Bay *C. virginica*. Tests of selection indicated that the pattern, in which there is an excess of high frequency and low frequency haplotypes and a deficit of intermediate frequency haplotypes, was the result

of positive selection on the genome. Demographic explanations appear unlikely to account for the mitochondrial haplotype pattern because nuclear loci exhibited neutral patterns of sequence evolution. Estimates of effective population size were several orders of magnitude smaller than census size, indicating that there was variance in reproductive success (sweepstakes reproduction). Nevertheless sweepstakes reproduction was not so severe that individual cohorts of juvenile oysters exhibited reduced levels of variation compared to the adult population. Finally I evaluated the risks associated with a supplementation program in which hatchery-raised oysters bred for disease tolerance were released into wild oyster populations. The results indicated that following supplementation, the wild effective population size remained large despite the danger of severe genetic bottlenecks. Increased hatchery effective population is suggested to prevent future harm to the wild population.

POPULATION GENETICS OF THE EASTERN OYSTER IN CHESAPEAKE BAY

By

Colin G. Rose

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
2008

Advisory Committee:
Professor Matthew P. Hare, Chair and Advisor
Professor Michael P. Cummings
Professor Matthew B. Hamilton
Professor Kennedy T. Paynter, Dean's Representative
Professor Sarah A. Tishkoff

Dedication

For my wife Emily and in loving memory of my mother Kathleen

Acknowledgements

I would like to thank the following people and institutions for help making this dissertation possible.

Dr. Matthew Hare for presenting me with the opportunity to be a part of this project, for his hard work helping me to improve my research and writing, and for giving me the freedom to grow as a scientist.

I also sincerely thank my committee members: Dr. Michael Cummings, Dr. Matthew Hamilton, Dr. Ken Paynter, and Dr. Sarah Tishkoff. Their advice, encouragement and feedback have been invaluable. I thank Dr. Charles Delwiche for providing effort and support that went far beyond the call of duty. I also thank Dr. Mary Christman for serving on an earlier incarnation of my committee, and Dr. Frank Siewerdt for statistical advice.

For help with oyster collections I thank Ken Paynter, Donald "Mutt" Meritt, Stephanie Tobash, James Wesson, the fall survey crew at Maryland DNR, and Heidi Hertler. I am also thankful to Pat Gaffney and Kim Reece for providing samples.

I thank John Wakeley, Robin Waples and Mary Kuhner for extremely helpful correspondence and assistance with data analysis.

I would like to thank Gang Chen, Maria Murray and Peter Thompson for helpful discussions and advice, and to Natasha Sherman, Jenna Murfree, Ninh Vu, and Nathan

Bendik for help in the laboratory. Thanks also to Sky Lesnick, Jibril Hirbo, Chris Che-Castaldo, Inken Kruse, Safra Altman, and the BEES/Biology network of students who have helped me over the years. Special thanks to Lois Reid for making sure that I didn't slip through the cracks of bureaucracy.

Last but not least I thank my wife Emily, my father Gerard, and my sisters Brigid, Anne and Katie for the support that allowed me to complete this work.

Financial support for this research was provided by an NSF Research Training Grant, Maryland Sea Grant, the Maryland Water Resources Research Center, and from the Department of Biology at the University of Maryland.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to Dissertation

Genetics and the Evolution of Populations

A fundamental goal in biology is to understand the evolutionary forces shaping the diversity of life on earth. But to understand the origins of diversity, it is necessary to define the framework in which evolution takes place. Changes in allele frequency that occur over time and that ultimately lead to speciation take place in populations—groups of interbreeding organisms. By studying the genetics of populations, it is possible to better understand how organisms change over time and how life has come to exist as it does today.

Organisms are characterized by diverse forms, but the genetic variation found in populations of every species is determined by the interaction among the same few evolutionary forces. Mutation creates new genetic variation. The stochastic force of random drift and the deterministic force of natural selection change allele frequencies. Migration homogenizes populations and recombination regroups alleles found on chromosomes. The interplay of these forces drives evolutionary changes in populations over time and has created the diversity found in all living things.

Molecular markers have revolutionized the field of population genetics. Electrophoretic enzyme (allozyme) analysis (Hunter and Markert 1957) was the first technique to reveal the extensive molecular variation found in populations. Kimura (1979) recognized that

this high polymorphism is not consistent with equilibrium between strong purifying selection and mutation, and hypothesized that most genetic variation must not have an effect on fitness. This neutral theory of molecular evolution predicts that mutations with little or no fitness effects are governed by random genetic drift rather than natural selection. Today a multitude of genetic markers allow patterns of molecular variation to be analyzed in a number of sophisticated theoretical frameworks. Molecular data can provide insight into population size (Beaumont et al. 2002; Wang 2005; Waples and Do 2008), genetic connectivity (Beerli and Felsenstein 1999, 2001) and population history (Luikart et al. 1998; Kuhner 2006).

Several key concepts are important for understanding the evolution of populations. A theoretically "ideal" population conforms to certain expectations, the most critical of which are a finite and constant size, random mating, and non-overlapping generations (Wright 1931). The effective size of a population is determined by the number of idealized individuals that would have the same amount of genetic drift as the actual population. The effective population size is an important parameter that determines, among other things, the rate at which drift fixes alleles and therefore the level of neutral variation. Since drift moves allele frequencies stochastically, stronger selective force is necessary to change allele frequencies in small populations compared with large populations. The different ways of quantifying genetic drift and measuring effective population size can have subtle but important differences in the interpretation of the parameter. Effective size measured by the change in allele frequencies over time is known as variance effective size (Kimura and Crow 1963). It is known as inbreeding

effective size when it is calculated by the change in the inbreeding coefficient (Wright 1931), and as the eigenvalue effective size when calculated from the rate of loss of heterozygosity (Ewens 1982). These three measures of effective size depend on the assumption that genetic drift due to sampling error in a finite population is the only evolutionary force changing allele frequencies. Since allele frequency, inbreeding, and heterozygosity change at different rates in a dynamic population, however, they can provide different estimates of effective size. Alternatively, effective population size can be calculated from the genealogical relationship of alleles. By calculating the time to the most recent common ancestor, the coalescent effective size (Nordborg and Krone 2002) can be deduced. The coalescent provides a long-term look at effective size since it reflects ancestral patterns of genetic variation that are maintained through the generations. In contrast, estimates based on the magnitude of genetic drift reflect the effective size only in those generations that have been sampled and therefore provide a more recent measure of the parameter.

Gene flow results from the movement of migrants between populations followed by subsequent reproduction. Observing how variation is distributed within and among populations provides insight into their genetic connectivity (Wright 1943, 1951). The scale of gene flow can provide information about the average number of migrants that populations exchange per generation.

Understanding the interplay among effective population size and migration rate can provide insight into the amount of the total genetic variation in a population, which is a

strong indicator of adaptive variation that might be used to respond to stressors that have not yet arisen (Lande 1995).

Population Genetics of Marine Invertebrates

Marine invertebrates provide a unique challenge when interpreting molecular data. Their life history is typically characterized by high fecundity, high early mortality (type III survivorship) and a planktonic larval life stage that potentially allows for long distance dispersal (Caley et al. 1996; Flowers et al. 2002). This life history strategy helps to make them successful, as exemplified by their worldwide distribution, but also differentiates them from many well studied terrestrial organisms. Despite marine invertebrates' ubiquity, major questions about biological processes remain: What is the scale of differentiation in populations? Is the potential for long distance dispersal realized? Does gene flow take place among populations that are distant, or even among those that are geographically close? What effective population size is typical for a species? In species with vast numbers of individuals, is effective population size also very large?

Marine invertebrates are a diverse group of organisms so generalizations are difficult to apply across taxa. Nevertheless, understanding patterns of genetic variation can help to determine how different aspects of life history tend to affect the evolution of species.

The Eastern Oyster in Chesapeake Bay

Oysters are distributed worldwide, and the eastern oyster *Crassostrea virginica* (Gmelin) is found along the North American Atlantic coast (Galtsoff 1964). Like other marine

invertebrates, oysters have the capacity for high fecundity (Hedgecock 1994), type III survivorship (Hunt and Scheibling 1997), and two to three weeks in the plankton before metamorphosis into a sedentary adult stage (Kennedy 1996). They are ecosystem engineers in Chesapeake Bay where their enormous population size helped to regulate water quality and provided three-dimensional structure used as habitat for other estuarine organisms. They were historically an ecologically important species in Chesapeake Bay, which explains the simultaneous decline in the health of the Chesapeake Bay ecosystem with recent reductions in the oyster population (Jackson 2001). The last several centuries have been marked by reductions of the oyster population to perhaps 1% of its former size (Newell 1988b; Rothschild et al. 1994), as a result of the combination of overfishing, habitat degradation and disease (Boesch et al. 2001a). A large, continuously distributed population of oysters once inhabited the Bay; now oysters are found only in patchy distributions and in low densities.

Several studies have been conducted examining the genetics of *C. virginica* in Chesapeake Bay. The first was by Buroker (1983a) who used allozymes to examine the population structure of the oyster population. He found that more than half (23 of 41) of the allozymes demonstrated low but significant differentiation among groups of oysters. Patterns of variation indicated that the 10 oyster bars sampled could be grouped into four latitudinal groups. Buroker argued that an equilibrium state exists between migration and selection for adaptation to local conditions. Next, Rose (1984) sampled oysters from the James and Potomac Rivers and found that two of the four allozyme loci sampled showed significant differentiation between rivers. However, the correlation between water

temperature and allele frequency in one of the significantly differentiated loci calls its selective neutrality into question and undermined the conclusion of population structure. Brown and Paynter (1991) determined the frequency of oyster mitochondrial alleles and failed to detect differentiation among three populations. The use of a single marker and limited sampling, however, raise the possibility that the study could have missed the biological pattern noted by Buroker (1983a).

The other notable study of Chesapeake Bay oyster population genetics was by Hedgecock (1994) who compiled allozyme data from several previous studies (Buroker 1983a; Paynter and DiMichele 1990; Vrijenhoek et al. 1990) to calculate effective population size in two Chesapeake Bay populations. Hedgecock used variance in allozyme allele frequencies over time to estimate that effective population size in upper Chesapeake Bay was 14.9 and in James River was 30.0. These surprisingly low estimates were used to support his sweepstakes hypothesis, which posits that oysters and other marine invertebrates are subject to extreme skew in reproductive success.

Scope of this Study

The goal of this dissertation is to investigate genetic principles of the Chesapeake Bay population of *C. virginica*. To this end, microsatellite genotypes and DNA sequences were collected from oysters collected from the Bay. Using a multilocus approach, the data were analyzed in order to: (i) estimate the effective population size, (ii) test for the effects of natural selection at mitochondrial and nuclear loci, (iii) describe spatially distributed patterns of genetic variation and infer levels of gene flow and (iv) evaluate the

risk of supportive breeding on the Chesapeake Bay population and determine if data confirm the results of predictive models.

Chapter 2 describes the collection of oyster samples from throughout Chesapeake Bay, and the microsatellite analysis of those samples. The distribution of genetic variation within and between sampling sites is characterized, and the relationship between genetic distance and geographic distance is addressed. The differentiation of populations is used to calculate average dispersal distance of planktonic oyster larvae. In addition, two samples of oysters collected more than a decade apart from the same site are genotyped and changes in allele frequencies are used to estimate the variance effective size.

Chapter 3 investigates signals of non-neutral evolution in mitochondrial DNA and determines whether demographic and selective explanations can explain the patterns. DNA sequences are determined for multiple mitochondrial and nuclear loci in *C. virginica* and in an outgroup species. Statistical tests of selection are performed on the DNA sequences and differences in patterns of variation between the mitochondrial and nuclear genomes are interpreted. Silent site DNA sequences are also used in a coalescent analysis to estimate the long-term effective size of the Chesapeake Bay oyster population.

Chapter 4 describes a study of microsatellite variation in wild oysters and hatchery-raised stock used for supportive breeding. The microsatellite data are examined for allelic correlations, which in turn are used to estimate the number of breeding adults. Predictive models determine the impact of supportive breeding on wild effective population size

based on estimates of hatchery contribution to wild reproduction and the number of breeding adults. Data collected after three years of supplementation are then analyzed and interpreted in light of theoretical predictions.

Chapter 5 summarizes and integrates the principle conclusions in chapters 2 through 4. Results are interpreted for oysters in particular and for marine invertebrates in general. Possible directions for future research are discussed in light of the results from this study.

Chapter 2: Population Structure in Chesapeake Bay

Abstract

Intensive efforts are underway to restore depleted stocks of *Crassostrea virginica* in Chesapeake Bay. However, the extent of gene flow among local populations, an important force mediating the success of restoration, is poorly understood. Spatial and temporal population structures were examined in *C. virginica* from Chesapeake Bay using eight microsatellite loci. Deficits in heterozygosity relative to Hardy-Weinberg expectations were seen at all loci and were best explained by null alleles. Permutation tests indicated that heterozygote deficiency reduced power in tests of differentiation. Nonetheless, genotypic exact tests demonstrated significant levels of geographic differentiation overall and a subtle pattern of isolation by distance was observed. Comparisons between age classes failed to show differences in genotype frequencies, allelic richness, gene diversity, or differentiation as measured by F_{ST}, contrary to predictions made by the sweepstakes hypothesis. The isolation by distance pattern could reflect an evolutionary equilibrium established because local gene flow predominates, or be influenced in either direction by recent anthropogenic activities. An evolutionary interpretation appears justified as more parsimonious, implying that local efforts to restore oyster populations will have local demographic payoffs, perhaps at the scale of tributaries or regional subestuaries within Chesapeake Bay.

Introduction

Marine species often have the capability of long distance larval dispersal, and as a consequence show relatively low levels of population structure (Bohonak 1999). Larval duration in the water column prior to settlement explains a substantial fraction of the variation in average effective dispersal distances among species, despite the heterogeneity of methods used to estimate these parameters (Shanks et al. 2003; Siegel et al. 2003). Thus, genetic panmixia over small regional scales is a reasonable null hypothesis for marine species with long larval periods. It is well known that deviations from this trend occur because of rafting of adults (Johannesson 1988) or larval behavior (Hill 1991; Shanks 1995; Baker and Mann 2003), but it is difficult to predict how these mechanisms will interact with hydrography and selection to shape gene flow.

When a marine species is threatened or requires management, it is risky to assume that the potential for long distance dispersal necessarily will be realized as large-scale gene flow (Cowen et al. 2000). The eastern oyster, *Crassostrea virginica* Gmelin, is broadly distributed in the western North Atlantic and was once abundant throughout Chesapeake Bay, a large estuary on the mid-Atlantic coast of the United States (Figure 2-1). The oyster's reef-forming habit and large filter-feeding capacity historically made it a keystone member of the estuarine community (Jackson et al. 2001). Overfishing and disease have reduced the oysters to less than 1% of their historic numbers (Newell 1988a; Jordan and Coakley 2004). Even so, the prolific fecundity of this species might allow for a rapid regeneration of historic numbers if not for the low density of remaining breeders in a severely degraded environment with intense disease pressure (Burreson and Ragone

Calvo 1996; Boesch et al. 2001a; Jackson 2001). This suggests that when local restoration efforts are successful, the geographic scale of their effects will depend on the distribution of improved habitat and the extent of dispersal among those patches.

The eastern oyster has a life history conducive to high gene flow. Oyster larvae spend 2 to 3 weeks in the plankton (Kennedy 1996), conceivably traveling hundreds of kilometers before settlement. Thus, there is the potential that long-range gene flow provides demographic connections between areas with localized restoration activities. In Chesapeake Bay these activities include constructing reef habitat in protected sanctuaries, seeding reefs with broodstock, and developing and releasing disease-tolerant strains of the native oyster (Breitburg et al. 2000; Allen et al. 2003; Mann and Evans 2004). However, large scale gene flow does not assure that restored reefs will be populated by migrants; optimal recruitment depends on matching the size and spacing of sanctuary reefs to the scale and pattern of dispersal (Botsford et al. 2003). Also, the potential genetic impacts from introducing disease-tolerant strains depend on the geographic scale of dispersal and subsequent interbreeding with wild stocks (Hare et al. 2006).

Some Chesapeake Bay tributaries are "trap-like" (Andrews 1979) with respect to oyster recruitment because of low flushing rates, restricted entrances, or retentive local circulation (Southworth and Mann 1998). Swimming behavior of oyster larvae in response to salinity, currents, or other cues (Dekshenieks et al. 1996; Finelli and Wethey 2003) can also promote retention (Tankersley et al. 1995; Southworth and Mann 1998). The best evidence for a trap-like dynamic comes from the Great Wicomico River in

Virginia where surface drifters and concentrations of oyster larvae both maintained their position or moved upstream (Southworth and Mann 1998). Without knowing the relative importance of physical factors and behavior to larval retention, the generality of local recruitment is uncertain. However, only certain tributaries will be retentive of larvae if it depends on hydrography whereas behavioral retention mechanisms should promote local recruitment in most tributaries.

The presettlement movement of larvae only enables gene flow; the processes of natural selection and variance in reproductive success determine which migrants leave offspring and the magnitude of effective gene flow (Palumbi 1994; Hilbish 1996). Hedgecock (1994) suggested that high fecundity and the stochasticity of larval viability can lead to extreme variance in reproductive success (a sweepstakes event) in marine organisms. Sweepstakes events could potentially create genetic heterogeneity among cohorts, or transiently among localities, when there is limited mixing of larvae among breeding populations. Two predictions of Hedgecock's sweepstakes hypothesis are (1) reduced variation within cohorts compared with the rest of the population (Hedgecock 1994), and (2) higher genetic heterogeneity over time in one location than seen spatially among breeding populations (Li and Hedgecock 1998; Flowers et al. 2002). Extreme variance in reproductive success has been hypothesized to explain genetic patterns observed in cod (Ruzzante et al. 1996) and oysters (Hedgecock 1994; Li and Hedgecock 1998; Boudry et al. 2002). In sea urchins sweepstakes events have been proposed in some populations (Addison and Hart 2004) and rejected in others (Flowers et al. 2002).

Selection against migrants can also limit gene flow despite high dispersal (Koehn et al. 1980; Johnson and Black 1984; Bertness and Gaines 1993; Schmidt and Rand 2001). In contrast to the demographic effects of larval retention and sweepstakes reproduction, which shape the distribution of polymorphism across the entire genome, genetic heterogeneity caused by selection is expected only at those loci linked to the genes under selection (Karl and Avise 1992; McGoldrick et al. 2000; Gilg and Hilbish 2003; Johannesson et al. 2004).

Each of the gene flow determinants described above has the potential to generate a complex patchwork of genetic connections that would complicate restoration planning. Using neutral genetic markers to avoid locus-specific patterns, high resolution testing for population structure can be informative about the magnitude and spatial scale of gene flow. It takes very little gene flow over evolutionary time to homogenize populations, however, so population differences might only be expected to accumulate in response to strong and consistent evolutionary barriers to gene flow (Palumbi 2003). Alternatively, anthropogenic effects could have homogenized Chesapeake oyster populations or created a patchwork of genetic differences. Human manipulation of these populations has included transplants within Chesapeake Bay (Carlton and Mann 1996; Mann and Powell 2007), introductions from the Gulf of Mexico (Carlton and Mann 1996; Milbury et al. 2004), and planting of juveniles produced in hatcheries from wild or selected-strain broodstock (Brumbaugh et al. 2000; Sorabella and Luckenbach 2003). Although many of these activities have been extensive in portions of Chesapeake Bay over the last few decades as measured by human efforts and resources expended, their impacts on oyster

population genetics are unknown. It is conceivable that the genetic impacts have been minimal, for all the biological reasons listed above. If most oysters are not contributing offspring in any particular generation (sweepstakes) and intensive fishing pressure quickly culls transplanted oysters, then transplants that have important management benefits could have trivial affects on patterns of gene flow. Nonetheless, with anthropogenic and evolutionary effects confounded, both sources of variation must be considered.

Isolation by distance (IBD), a pattern in which genetic differentiation increases with the geographic scale of comparison, is usually modeled as a stepping-stone pattern of gene flow in which migration only occurs among neighboring demes (Kimura and Weiss 1964). Recent simulations and theory indicate, however, that IBD can also emerge with a low level of long distance migration if most recruitment is local (Palumbi 2003). Thus, if other assumptions are met, an IBD pattern provides a relatively robust indication that local gene flow predominates within the scale of study. IBD has been detected in several high-dispersal marine organisms, including fishes (Gold et al. 2001; Pogson et al. 2001; Riginos and Nachman 2001; Planes and Fauvelot 2002; Castric and Bernatchez 2003; Buonaccorsi et al. 2004), urchins (Palumbi et al. 1997), eels (Wirth and Bernatchez 2001; Maes and Volckaert 2002) and oysters (Launey et al. 2002).

Previous work has examined genetic variation in *C. virginica* from Chesapeake Bay. In a study by Buroker (1983a), samples from ten Chesapeake Bay oyster bars revealed significant genetic differentiation across 32 allozyme loci, with mean $F_{ST} = 0.016$.

Principle component analysis clustered the oysters into four groups whose distribution did not correlate with any obvious environmental variables and isolation by distance was rejected. Other Chesapeake Bay studies found no significant genetic heterogeneity but also had low power (Rose 1984; Brown and Paynter 1991).

Here I tested for population structure in Chesapeake Bay *C. virginica* using eight microsatellite loci. To test for IBD, specimens were collected from various tributaries within Chesapeake Bay at different spatial scales. To test for sweepstakes events, temporal comparisons were made between juveniles and adults. I also tested for anthropogenic effects where possible.

Materials and Methods

Sampling Design

Because temporal and spatial processes of differentiation could act at any geographic scale, spatial samples were collected from sites separated by aquatic distances ranging from one to hundreds of kilometers. A total of 1,228 specimens were collected from 16 locations in or near Chesapeake Bay (Table 2-1, Figure 2-1). All adults were collected by dredge or diver from natural subtidal reefs that to the best of my knowledge have not been manipulated (e.g. transplants, juvenile oyster plantings) for several years prior to my collections (K. Paynter, personal communication). Juvenile oysters (spat) and adults were collected during the same year in the Piankatank, Great Wicomico and Little Choptank Rivers. Spat in the Great Wicomico and Little Choptank Rivers were sampled by serially deploying clean oyster shell "collectors" for two or four week periods, respectively, from

June through September. Collectors were examined by eye for spat. Spat in the Piankatank River were collected by dredge. All oysters were stored on ice until gill and mantle tissue, or whole spat, were preserved in 95% ethanol. Shell height of spat ranged from 2 to 25 mm, consistent with young of the year. Archived samples collected from the James River in 1990 were obtained from P. Gaffney, University of Delaware.

Two *C. virginica* strains artificially selected for disease tolerance have been planted in Chesapeake Bay for restoration purposes since 1999 (Brumbaugh et al. 2000). The two strains, known as CROSBreed and DEBY, were bred for resistance to the protozoan parasites *Perkinsus marinus* and *Haplosporidium nelsoni* (Ragone Calvo et al. 1997). A reference sample of DEBY-strain oysters was obtained in 2002 from the progeny of generation-four broodstock produced at the Center for Environmental Science, University of Maryland. A reference sample of CROSBreed strain, generation 5, was obtained from K. Reece, Virginia Institute of Marine Science (VIMS).

DNA Extraction, Amplification, and Genotyping

Approximately 20 mg of gill or mantle tissue were used for DNA extraction from adults using the DNeasy 96 Tissue kit (Qiagen Inc, Valencia, CA) following the protocol for animal tissues. CROSBreed, GWRa and GWRs samples were extracted with a FastPrep FP120 instrument (BIO 101, Vista, CA) using a FastDNA kit (BIO 101, Vista, CA) (see Reece et al. 2004). Genomic DNA was diluted to 50 ng/µl based on spectrophotometry.

Oysters were genotyped for eight microsatellite loci previously developed by Reece et al. (2004) and Brown et al. (2000). Five loci have perfect repeat motifs (one di-, one tri-, and three tetranucleotide), and three loci have imperfect repeat motifs (1 di-, 1 tri-, 1 tetranucleotide). The primers (reported with optimized annealing temperature and $MgCl_2$ concentration) are *Cvi*9 (52°C, 1.7mM), *Cvi*12 (52°C, 1.7mM), *Cvi*1i24b (52°C, 2.5mM), *Cvi*2g14 (52°C, 2.5mM), *Cvi*2i23 (51.5°C, 1.5mM), *Cvi*2i4 (47°C, 2.8mM), *Cvi*2j24 (touchdown, 1.7mM), and *Cvi*1g3 (touchdown, 1.7mM). Reaction conditions for PCR in a total volume of 7.5 µl included final concentrations of 1× Invitrogen buffer (no $MgCl_2$), 100 µM dNTP, and 200 nM each for forward and reverse primers, one of which was fluorescently labeled. Thermocycling involved one cycle of 95°C denaturing for 1 min; 30 3-step cycles including 95°C for 30 s, annealing temperature for 30 s, and 72°C for 20 s; then a final extension at 72°C for 10 min. Touchdown thermocycling began with ten 3-step cycles in which annealing started at 60°C for 1 min and dropped by 1°C each cycle, followed by 30 cycles of 95°C for 15 s, 50°C for 1 min, and 72°C for 45 s. Following amplification, 8.82 µl HiDi formamide and 0.18 µl Genescan-500 ROX size standard (Applied Biosystems, Foster City, CA) were combined with 1 µl PCR product for fragment analysis.

PCR products were electrophoresed with an ABI-Prism 3100 genetic analyzer (Applied Biosystems) and allele sizes were estimated using Genescan 3.7 and Genotyper 2.5 (Applied Biosystems). Electropherogram peaks were examined before assigning genotypes. If a single peak was detected and it was greater than 500 relative fluorescent units (RFUs), the specimen was labeled a homozygote for that allele. For a specimen to

be labeled a heterozygote, both peaks had to be at least one repeat unit apart and greater than 100 RFUs; if the 2 peaks were of different heights, the shorter peak was scored only if its height was >10% of the taller one. If no peaks were present, a second PCR was performed; if electropherogram peaks were still absent then the locus was considered non-amplifying.

Data Analysis

To minimize missing data, I removed 30 specimens (2.4 %) from the data set, distributed across ten samples, because they had more than two non-amplifying loci. To quantify deviations of genotype frequencies from Hardy-Weinberg expectations among and within samples I calculated the unbiased F_{ST} estimator θ and the F_{IS} estimator f (Weir and Cockerham 1984) using Fstat 2.9.3.2 (Goudet 2001). To test for overall genetic subdivision, θ was calculated for all samples except JR90 (a sample collected in 1990). 95% confidence intervals were calculated using 15,000 bootstrap replicates across loci. Pairwise comparisons were performed between all samples except for JR90, which was only compared with JRD. Number of alleles, allelic richness and gene diversity were also calculated with Fstat. Unbiased estimates of p-values for G-based exact tests of genotypic frequency differentiation (Goudet et al. 1996) were calculated using Genepop 3.4 (Raymond and Rousset 1995b). To determine the independence of the microsatellite loci, I tested for genotypic linkage disequilibrium among each pair of loci using Genepop (10,000 dememorization steps, 1000 batches, 10,000 iterations per batch in the Markov chain).

In order to test for an association between genetic and geographical distances, the natural logarithm of the shortest pairwise aquatic distances (shortest route over water) were correlated with pairwise values of $\theta/(1-\theta)$ between all samples except JR90 (Rousset 1997). Distances between adults and spat in the Great Wicomico River, Little Choptank River, and Piankatank River samples were changed from 0 to 1 km for the log transformation. I used the Mantel test (Mantel 1967) for correlation between the two distance matrices based on 10,000 permutations as implemented in the Isolde program in Genepop (Raymond and Rousset 1995b). Genepop was also used to compute the regression line describing the relationship between $\theta / (1-\theta)$ and the natural logarithm of distance.

I tested for evidence that stuttering or large allele dropout was affecting microsatellite genotypes. Stuttering refers to a tendency by *Taq* polymerase to amplify fragments of multiple sizes in addition to the correct one, especially from di-nucleotide repeats (Shinde et al. 2003). Large allele dropout is the preferential amplification of shorter alleles from heterozygotes (Wattier et al. 1998). Both these artifacts affect the distribution of heterozygosity among allele size classes in predictable ways. To test for a deficiency of heterozygotes carrying alleles differing in size by one repeat unit (stuttering) and for an excess of specimens that are homozygous for small alleles (large allele dropout), I randomized genotypes for each locus within samples using Micro-Checker 2.2.1 (Van Oosterhout et al. 2004).

As a result of several generations of selection, the DEBY and CROSBreed oyster strains have genetic signatures that are distinct from wild oysters (Hare et al. 2006). Because regional plantings of selectively bred oysters could affect an IBD pattern, multilocus genotypes of 49 CROSBreed and 82 DEBY oysters were used as reference samples for assignment tests with the oysters collected in this study. Using the Bayesian method of Rannala and Mountain (1997) in Geneclass2 2.0.b (Piry et al. 2004), each presumed wild oyster was removed from the total collection and treated as unknown for testing against CROSBreed, DEBY, and the remaining N-1 wild samples. This assignment method assumes Hardy-Weinberg equilibrium, but is fairly robust to deviations (Cornuet et al. 1999). Applying a low-stringency assignment criterion to be conservative, specimens that had a lower negative log likelihood assignment score for CROSBreed or DEBY versus the wild reference sample were removed from the data set as possible selected-strain oysters. Similar assignment methods implemented in Immanc 5.0 (Rannala and Mountain 1997) were used to calculate, for each "wild" individual, the probability of being an F1 offspring of a selected strain by wild cross, and individuals with $p > 0.95$ were removed. The IBD analysis was repeated after each culling.

I examined the effect of heterozygote deficiency on tests of genotypic differentiation by randomizing alleles within samples using Genetix 4.05.2 (Belkhir et al. 2001). The randomized data had levels of heterozygosity that were similar to Hardy-Weinberg expectations, but without changing the allele frequencies or homozygous null frequencies. The number of significant pairwise θ and exact tests were compared between the original and permuted data.

A power analysis for F_{ST} was done by randomly subsampling a data set consisting of two identically sized samples (each sample had N = 100). The first sample combined Patuxent River samples (PXD and PXU) and the second combined York River samples (YRD and YRU). Using the Poptools 2.6.2 (Hood 2004) add-in for Microsoft Excel, multilocus genotypes from each sample were randomly subsampled without replacement to create 200 replicate data sets for each of 15 subsample sizes. The subsampled data were analyzed in Fstat and mean F_{ST}, mean upper 95% confidence interval, and mean lower 95% confidence interval were calculated for the replicates at each subsample size.

I used temporally spaced samples from the James River (JRD and JR90) to estimate the effective population size of oysters. N_e was calculated using the moments-based method of Waples (1989) in NeEstimator 1.3 (Peel et al. 2004) and using the pseudo-likelihood method of Wang (2001) in MLNE 1.1 (Wang 2005). Assuming a two-year generation time (Hedgecock 1994), I calculated N_e across 6 generations. Both methods assume that the samples are from a single isolated population; for the pseudo-likelihood method I set the maximum N_e at 10,000 (due to computational constraints).

Average squared dispersal distance between parent and offspring, σ^2, was calculated using the method of Rousset (1997). Under a two-dimensional stepping stone model, the inverse of the IBD regression slope is equal to $4D\pi\sigma^2$, where D is the density of the effective number of individuals (Rousset 2003). The area of Chesapeake tributaries was estimated using Scion Image 4.0.3.2.

Results

Genetic Variation and Hardy-Weinberg Equilibrium

A total of 1,198 individuals were analyzed. All eight microsatellite loci were highly variable in terms of gene diversity (0.618 – 0.947) and number of alleles (10 – 40; Table 2-2). All samples had roughly the same allelic richness and the same proportion of rare alleles to total number of alleles (mean = 30%; Table 2-3). The number of singleton alleles, those observed only once in the entire data set, ranged from 0 to 5 per sample, whereas only two alleles were private, occurring more than once but found only in one sample (Table 2-3).

No significant genotypic disequilibrium was detected between loci ($p > 0.05$ in each case). All samples, including all three spat samples, had significant Hardy-Weinberg deviations in the direction of heterozygote deficiency. Over all samples, the F_{IS} estimator *f* was found to be statistically greater than zero for each locus (all $p < 0.001$) and over all loci ($p < 0.001$; Table 2-2). There were no indications of stuttering or large allele dropout at any loci. Although insertions and deletions are abundant in the flanking sequences of six of the loci (Reece et al. 2004), no correlation was found between the level of polymorphism reported for the flanking regions (2.0 – 5.8%) and F_{IS} (Pearson $r = -0.34$, 5 df, $p = 0.507$) as would be expected if null alleles were caused by polymorphic nucleotides in the PCR priming sites.

Genetic Differentiation

Over all samples (excluding JR90), exact tests of genotypic differentiation detected significant ($p < 0.05$) population structure at one of the eight microsatellite loci, and highly significant ($p = 0.0001$) population structure across all loci (Table 2-2). Genotypic exact tests demonstrated statistically significant differences ($p < 0.05$) at 35 of the 171 pairwise comparisons (21%). After sequential Bonferroni correction for multiple comparisons (Holm 1979), 3 of the pairwise comparisons remained statistically significant ($\alpha = 0.05$). The global estimate of θ was low ($\theta = 0.001$) and not significantly different from zero (Table 2-2). I calculated θ for all pairs of samples, and found that 19 of the 171 comparisons (11%) were significantly different from zero ($p < 0.05$), but none remained statistically significant after sequential Bonferroni correction ($\alpha = 0.05$).

Adult and spat oysters from the same locality showed no significant difference in gene diversity (sign test comparing adults and spat in three locations, eight loci each, $df = 23$, $p = 0.308$), allelic richness ($p = 0.541$), or genotypic frequency as measured by exact tests ($p > 0.05$; Table 2-2). Adult oysters collected in the James River more than a decade apart, JRD and JR90, also did not have significantly different genotypic frequencies (Table 2-2).

Isolation by Distance

A significant association ($p = 0.009$) was found between pairwise estimates of genetic structure ($\theta /(1 - \theta)$) and the natural logarithm of aquatic distance for all samples (excluding JR90) (Table 2-4). A regression of $\theta /(1 - \theta)$ and aquatic distance with all loci

combined is shown in Figure 2-2. Analysis of individual loci revealed statistically significant correlations for 2 of the 8 loci, and 6 of the 8 loci had positive regression slopes (Table 2-4). The IBD pattern remained significant ($p < 0.05$) after setting negative values of $\theta / (1 - \theta)$ to zero (data not shown), after removing any one sampling site from the data set (data not shown), or after removing any one locus (Table 2-4). Furthermore, the pattern of IBD remained significant after combining all downriver and upriver adult samples within tributaries, or combining coincident spat and adult samples in GWR and PTK (data not shown). When all four adult samples from LCR were combined with LCR spat, IBD remained nearly significant ($p = 0.057$). Finally, multilocus assignment tests identified 23 of the study oysters as CROSBreed or DEBY strain individuals, or an overlapping set of 156 as F1 progeny between wild and selected-strain crosses. The IBD slope remained positive and significant ($p = 0.036$) after removing the 23 strain hybrid oysters from the data set, but removing the larger subset of non-wild oysters reduced the slope slightly (0.00066) and made the Mantel test nonsignificant ($p = 0.068$).

Sample Size and Power

Using 100 samples each from the York River and Patuxent River, I detected significant population genetic structure ($\theta = 0.0034$, $p < 0.05$). Randomly drawing 200 replicate samples at each of several subsample sizes, the mean value of θ was unaffected by subsample size (as expected for an unbiased estimator of F_{ST}), but the 95% confidence interval increased as subsample size decreased (data not shown). If I considered θ to be significant when the mean lower confidence interval did not overlap zero, a sample size of 90 or greater was necessary to statistically detect the low observed levels of differentiation.

The data were permuted to investigate the effect of Hardy-Weinberg deviations on power to detect differences by exact tests and θ. A greater number of statistically significant pairwise comparisons were observed in the permuted data than in the original data for both measures of differentiation (Table 2-5). These results suggest that Hardy-Weinberg deficits reduced power to detect population differentiation. Mantel tests detected an IBD pattern in the permuted data set ($p = 0.005$) with slightly greater statistical power than with the original data set ($p = 0.009$). The regression of $\theta / (1 - \theta)$ against the natural logarithm of distance led to nearly identical slope and r^2 compared with the unpermuted data (Table 2-5).

Effective Population Size

The moments-based estimate of oyster N_e in James River was 535 (95% CI: 234 – 6061), whereas the pseudo-likelihood estimate was 1,516 (95% CI: 422 – 10,000). Only the lower limit in the pseudo-likelihood estimate is informative because the upper limit was arbitrarily set to 10,000.

Dispersal Distance

To estimate the variance in dispersal distance (σ^2) from the IBD slope, I first determined the N_e/N ratio for James River from the likelihood estimate of N_e =1,516 and the harmonic mean of James River population estimates over 1998 to 2002, $N = 1.8 \times 10^9$ (VIMS, CBOPE website). This estimate of N_e/N, 8.42×10^{-7}, is consistent with the estimate from Hedgecock et al. (1992). Total N_e for Chesapeake Bay was estimated to be 2611 by assuming the N_e/N ratio is uniform across Chesapeake Bay, and multiplying the James River ratio by the total number of oysters in the Chesapeake, 3.1×10^9 (harmonic

mean of estimates from 1998-2002; VIMS, CBOPE website). The total N_e divided by the area of the tidal waters in Chesapeake Bay, approximately 11,000 km^2 (Boesch et al. 2001b), gives an average density of 0.24 oysters per km^2 assuming all locations within the Bay are equally suitable for oysters. Based on this density estimate and the inverse of the IBD slope (1,429), I estimated average squared dispersal distance to be approximately $\sigma^2 = 479$ km^2.

Discussion

In this study I have examined the magnitude and pattern of genetic differentiation among several eastern oyster populations in Chesapeake Bay. I found evidence for spatial but not temporal genetic heterogeneity. Most significantly, genetic differentiation increased with geographic distance within the Bay. If this pattern represents evolutionary equilibrium, it provides support for the assumption of local recruitment that underlies current strategies for oyster restoration. However, genetic differences contributing to this association were small, so before elaborating on the biological meaning and significance of IBD I discuss the robustness of these findings.

Power

Choice of molecular marker and sampling design both affect the ability to detect differentiation between populations. Microsatellites are markers that permit a high level of statistical power because of their high heterozygosity (Hedrick 1999; Estoup et al. 2002), but homoplasy can downwardly bias F_{ST} estimates for loci with high mutation rates (Balloux et al. 2000; Balloux and Goudet 2002; Adams et al. 2004; O'Reilly et al.

2004). The only individual locus to show significant genotypic differentiation, *Cvi*1g3, also had the lowest number of alleles and gene diversity (Table 2-2), implicating homoplasy caused by high mutation rates as a constraint on differentiation at the other seven loci. However, F_{ST} was not any higher for *Cvi*1g3, so in this case the different statistical results may depend on the relative power of rare and moderate frequency alleles (Waples 1998).

For highly differentiated populations, population structure can be statistically detected even with small sample sizes. However, this is generally not the case for marine populations with high gene flow (Ruzzante 1998). When differentiation is low, exact tests of differentiation have greater power than F_{ST} to reject homogeneity (Raymond and Rousset 1995a; Goudet et al. 1996), and this was true here with testing at the genotypic rather than allelic level. While exact tests provide a powerful statistical test, they do not provide information about the degree of gene flow. In order to analyze geographic patterns of gene flow, I relied on estimates of F_{ST} under the assumptions of an infinite island model of gene flow (Wright 1951).

When the number of migrants per generation is high ($N_em > 10$), which my data suggest was the case for Chesapeake Bay oysters, F_{ST} is estimated with low precision (Neigel 1996). Thus, with typical sample sizes, values of θ may be statistically indistinguishable from zero (Kalinowski 2002). Nonetheless, the power requirements differ for statistically testing F_{ST} between any pair of populations versus testing for an association between F_{ST} and another variable across many pairwise comparisons. Specifically, in this latter case

the statistical significance of any particular pairwise comparison is less important than the absence of a systematic bias related to sample sizes, as with θ (Weir and Cockerham 1984; Cockerham and Weir 1993). In addition, Peterson and Denno (1998) found that the likelihood of detecting IBD increases with the number of populations sampled in a study, and power also depends on adequate sampling at multiple spatial scales (Palumbi 2003). Therefore, rather than sampling the >90 specimens per location that my power analysis showed were necessary to statistically detect differences between pairs of populations, I sampled approximately 50 specimens from a larger number of populations at multiple spatial scales to test overall patterns of gene flow in Chesapeake Bay. Of course, low precision could obscure an IBD pattern, but it should not falsely generate IBD because sampling error is independent of the proximity of the collection sites.

Heterozygote Deficiency

The microsatellite loci used here showed large heterozygote deficits relative to Hardy Weinberg expectations (positive F_{IS}). Heterozygote deficiency can be explained by Wahlund effects, inbreeding, natural selection, or null alleles (and other technical artifacts), but the large F_{IS} values make some of these hypotheses untenable. The Wahlund effect, a reduction in heterozygosity resulting from sampling across subdivided populations, cannot be a major contributor to the heterozygote deficiency because dramatic population structure is lacking. Even the selected strains of oysters that have been released into Chesapeake Bay are only moderately differentiated from wild stocks, so Wahlund-induced F_{IS} could not have been higher than about 0.05. Finally, inbreeding was an unlikely source for such large F_{IS} values because sweepstakes events were negligible and because *C. virginica* is dioecious.

Natural selection cannot be rejected as easily, but it seems unlikely to be the sole cause for heterozygote deficiencies because it would have to be acting in a similar manner across all eight unlinked loci. Zouros et al. (1980) proposed that background selection against deleterious alleles might cause heterozygote deficiency at linked genetic markers in oysters. Strong selection against deleterious alleles (genetic load) has been used to explain segregation distortion in studies of the Pacific oyster *Crassostrea gigas* (McGoldrick and Hedgecock 1997; Bierne et al. 2000; McGoldrick et al. 2000; Launey and Hedgecock 2001; Boudry et al. 2002) and the eastern oyster (Yu and Guo 2003). Background selection reduces within-deme heterozygosity, potentially amplifying between-deme population structure (Charlesworth et al. 1997; Pamilo et al. 1999). However, there is no reason to expect that population structure induced by background selection would have been positively associated with geographic distance between populations.

Null, or non-amplifying alleles are a plausible explanation for heterozygote deficiencies in the data. Polymorphisms at priming sites could have created differences among alleles in their amplification efficiency in PCR, resulting in a global deficiency of microsatellite heterozygosity. Hedgecock et al. (2004) detected null alleles in 49 of 96 microsatellite loci in the Pacific oyster, and calculated that the minimum level of sequence polymorphism in the priming region was 1.2%. Data from Reece et al. (2004) showed that the DNA sequences flanking six of the microsatellite loci used in this study (*Cvi*1i24b, *Cvi*2g14, *Cvi*2i23, *Cvi*2i4, *Cvi*2j24, and *Cvi*1g3) had a mean sequence polymorphism of 3.6% (2.0 - 5.8%). Given this high level of polymorphism surrounding

PCR priming sites for these loci, it is likely that null alleles contributed to the heterozygote deficiency described in this study. Although I failed to detect a correlation between sequence variation near the priming sites and F_{IS} in the six loci, the estimates of sequence variation from Reece et al. (2004) were relatively imprecise, based on an average of 3.7 alleles sequenced per locus.

Primer redesign can correct heterozygote deficiency if null alleles are to blame, as was done for an anonymous nuclear locus by Hare et al. (1996). In highly polymorphic species, however, a large number of sequences must be considered to assure that polymorphisms do not affect priming sites. Thus, the redesign of PCR primers for the *Cvi*2g14 and *Cvi*1g3 loci by Reece et al. (2004) corrected some null alleles discovered by pedigree analysis but may not have prevented additional PCR null alleles in wild populations.

Corrections for null alleles are frequently applied to data sets, but common methods (e.g. Chakraborty et al. 1992; Brookfield 1996) assume a single null allele. These correction methods are inappropriate for the oyster data because multiple PCR null alleles may occur at different frequencies, and Hardy-Weinberg deviations may derive from both technical and biological factors (Foltz 1986; McGoldrick et al. 2000). Instead, I performed a permutation test to address whether heterozygote deficiency introduced bias or reduced statistical power in my estimates of F_{ST} and exact tests of genotypic differentiation. Results indicated that heterozygote deficits led to reduced power in tests of differentiation, but the IBD pattern was not sensitive to these effects.

Sweepstakes

Comparisons of adult and juvenile oysters in three tributaries of Chesapeake Bay demonstrated that sweepstakes events, if they happened, have not been frequent or strong (see Table 2-3). The strongest evidence against sweepstakes reproduction was that juvenile and adult oysters had no statistical difference in allelic richness, the most sensitive indicator of recent bottlenecks (Spencer et al. 2000). I may have missed sweepstakes events either because they are rare, in which case additional breeding seasons need to be examined, or because sweepstakes events were local and ephemeral, which would require sampling at a finer scale (both spatially and temporally) to detect. In either case, given these results and the high polymorphism observed at the microsatellites and their flanking sequences (Reece et al. 2004), I conclude that sweepstakes reproduction did not lower effective population size of Chesapeake Bay oysters as dramatically as previously hypothesized (Hedgecock 1994). This conclusion does not negate the expectation that high fecundity elevates variance in reproductive success and lowers N_e/N (Hedrick 2005), but merely rejects extreme sweepstakes events.

Effective Population Size

Hedgecock (1994) measured temporal genetic variance between eastern oyster adults sampled one generation apart in the James River, Virginia, and used moments-based methods (Pollak 1983; Waples 1989) to estimate $N_e = 30.0$ (95% CI: 13.5 – 60.8). Both my estimates of James River N_e are substantially larger than Hedgecock's. Using the same estimation method as Hedgecock (1994), I rejected $N_e < 234$ in James River. However, when many alleles are at low frequency, as with my data, moment-based

estimates can be biased (Waples 1989; Turner et al. 2001) whereas likelihood-based estimates perform well (Wang 2001). Thus, the likelihood estimate of N_e = 1,516 was probably more accurate. This estimate of effective size is still consistent with a very small N_e/N ratio and high variance in reproductive success, but not compatible with extreme sweepstakes events.

Isolation by Distance

The populations of *C. virginica* in Chesapeake Bay have statistically significant population structure consistent with isolation by distance, but genetic differentiation explained a small fraction of variation in aquatic distance and the regression slope was shallow. Hedrick (1999) raised the question of whether subtle microsatellite divergence has evolutionary meaning. Faint substructure could result from recent nonequilibrium processes or from random noise due to sampling error (Waples 1998). However, Palumbi (2003) suggested that low levels of genetic differentiation can be verified by demonstrating a relationship between genetic relatedness and distance, because sampling error is unlikely to produce a significant IBD pattern. In this study, locus- and site-specific artifacts seem unlikely to have created IBD, because the pattern remained significant after individual populations or loci were removed.

Oyster transplants among Chesapeake tributaries could have genetically homogenized populations, reducing the strength of an evolutionary equilibrium IBD pattern. Alternatively, it is conceivable that the IBD pattern was created by a particular combination of anthropogenic impacts. There are several reasons why anthropogenic effects are likely to be minimal in my data. First, I took great care to collect oyster

samples from locations with no recent history of restoration activities or oyster transplanting. Second, plantings and transplants are often designed to have large impacts on local census numbers for fisheries or restoration, but this doesn't necessarily mean that the planted oysters successfully reproduce at a scale that would leave a genetic trace. This may be especially true when there are targeted harvests of transplanted oysters. Finally, in those cases where the population genetic consequences of oyster manipulations were predictable, I tested for those effects to assess the magnitude of their impacts.

The hatchery mass spawns that produce oysters for planting use a limited number of parents and have a potential for skewed parental contributions that lower allelic diversity (Launey et al. 2001). Thus, plantings of hatchery-produced oysters, done on a large scale, are predicted to lower allelic richness near the planting site. This could increase genetic relatedness locally while accentuating differences regionally (through independent, hatchery-induced bottlenecks), conceivably generating a pattern of IBD. However, allelic richness was uniformly high across sites, and equally high inside and outside the Chesapeake Bay.

Large-scale plantings of genetically distinct disease tolerant *C. virginica* (DEBY and CROSBreed selection lines, Ragone Calvo et al. 1997), could also create an IBD pattern if different practices in Maryland and Virginia contributed to regional differentiation, while individual plantings homogenized local populations. My results were equivocal on this matter because some loss of power is expected when over 10% of the total sample is removed to conservatively eliminate the affects of selected-strain introgression. Thus, it

is possible that a combination of management activities has created an IBD pattern, but an evolutionary explanation for IBD seems more parsimonious. Finding an IBD pattern in Chesapeake Bay with other kinds of genetic markers, such as mitochondrial DNA, or among populations along the U.S. Atlantic coast, would help confirm the appropriateness of applying an evolutionary interpretation here.

Evolutionary IBD develops as equilibrium is reached between gene flow and genetic drift (Wright 1943). The modern distribution of oysters in Chesapeake Bay arose in the last 12,000 to 18,000 years after the most recent glacial advance (Grumet 2000). Assuming a generation time of four years, there may have been as few as 3,000 generations for Chesapeake oysters to reach equilibrium. While this is implausible for species with low levels of migration, it is possible when the proportion of migrants (m) is high because the time to equilibrium is inversely related to migration (Crow and Aoki 1984). More specifically, if mutation rate is much smaller than m and $1/N_e$ is much smaller than 1 (both reasonable assumptions for oysters), then the time required for F_{ST} to go half way to equilibrium is approximated by $(\text{Ln } 2)/(2m + 1/2N_e)$. To illustrate the strong dependence on migration rate, suppose that oysters in Chesapeake Bay have $N_e = 100{,}000$. Then $m = 0.0001$ ($N_e m = 10$) requires 3,381 generations to get half way to equilibrium, while $m = 0.001$ ($N_e m = 100$) requires only 346 generations. My low estimates of F_{ST} for Chesapeake Bay oysters reflected high rates of migration ($N_e m \cong 250$) that could have generated migration-drift equilibrium since the Pleistocene. Furthermore, during the approach to equilibrium, IBD is manifest initially at relatively small spatial scales (Slatkin 1993). Thus, it is feasible for oysters to be at migration-drift equilibrium and

show IBD within Chesapeake Bay. Also, under an equilibrium interpretation, IBD within Chesapeake Bay should not have been sensitive to the degree of local recruitment occurring within the tributaries I sampled. That is, the pattern of IBD would have been dictated by the least retentive tributaries whether they were sampled or not.

Spatial Scale of Dispersal

If an IBD pattern indicated that local gene flow predominates within Chesapeake Bay, how local is local? The average squared dispersal, $\sigma^2 = 472$ km^2, is roughly equivalent to 4% of the entire Chesapeake Bay, or the area within a large tributary (e.g., area of James/Elizabeth Rivers = 747 km^2). Since σ^2 defines a geographic scale encompassing the bulk of dispersal from a central point source, the data suggest that recruitment of oysters in Chesapeake Bay is local within tributaries or regional subestuaries. This is consistent with the conclusions of Southworth and Mann (1998) who found that larvae and surface drifters were entrained within a Chesapeake Bay tributary, and may result from an interaction between trap-like hydrography and larval behavior (Tankersley et al. 1995; Dekshenieks et al. 1996; Finelli and Wethey 2003). The single-generation value of σ^2 calculated here is a long-term evolutionary average that may encompass some inter-annual variation in dispersal distances. A two-dimensional IBD measure of σ^2 depends on population density, but is independent of the shape of the distribution of dispersal distances (Rousset 1997). At higher oyster densities characteristic of Chesapeake Bay before 1900, the same slope would indicate a smaller average squared dispersal. A given σ^2 can result from lots of short range dispersal or a little longer-range dispersal (Rousset 1997), so a measure of average dispersal distance is impossible to calculate from σ^2 without simulations based on particular distributions of dispersal distances. Ongoing

studies are expected to help define average dispersal distance by contributing direct estimates from a point source (Hare et al. 2006) and by estimating dispersal distributions from individual-based models of larval behavior and hydrographic mixing (North et al. 2008).

Conclusion

There are many potential explanations for differences between potential and realized dispersal (Ehrlich and Raven 1969; Slatkin 1987; Hilbish 1996; Cowen et al. 2000; Pogson et al. 2001). Hydrodynamic features within Chesapeake Bay tributaries are often cited as a primary mechanism determining local recruitment (Andrews 1979; Mann 1988). However, retentive characteristics such as low flushing rate or tidal gyres are only strongly expressed in a few tributaries, and are therefore not likely to be the primary factor generating IBD at the scale of Chesapeake Bay. This reasoning implies that larval behavior may be as important as hydrography, making local recruitment the rule, not a tributary-specific phenomenon.

What is the relevance of this evolutionary equilibrium pattern of gene flow to restoration practices? Very few successful migrants are needed on average to homogenize populations over an evolutionary time scale (Wright 1931), so even slight genetic differentiation (such as at larger scales in Chesapeake Bay) indicates that gene flow is trivial over the ecological time scale relevant to restoration (Waples 1998; Palumbi 2003). IBD in Chesapeake Bay oysters therefore suggests that impacts from population enhancement efforts will be concentrated near where resources are invested.

Tables

Table 2-1. Oyster sample information. Spat and adult samples are designated with an ‘s’ or ‘a’ in the sample code. Sample sizes indicate the number of individuals analyzed.

Sample Code	Description	Sample Size	Date Collected	Latitude, Longitude
AIN	Assateague Island	50	14 July 2002	38°14.39′ N, 75°08.74′ W
GWRa	Great Wicomico River adults	90	1 November 2002	37°49′ N, 76°18′ W
GWRs	Great Wicomico River juveniles	102	22 July 2002	37°49′ N, 76°18′ W
HRC	Harris Creek	50	16 September 2002	38°45.05′ N, 76°17.75′ W
JR90	James River (Archived)	48	1990	37°03′ N, 76°41′ W
JRD	James River downriver	50	10 December 2003	36°03′ N, 76°41′ W
JRU	James River upriver	38	10 December 2003	37°04.17′ N, 76°35.12′ W
LCRa1	Little Choptank River adults – site 1	59	9 April 2002	38°32.02′ N, 76°14.64′ W
LCRa2	Little Choptank River adults – site 2	50	9 April 2002	38°32.61′ N, 76°13.62′ W
LCRa3	Little Choptank River adults – site 3	46	9 April 2002	38°32.91′ N, 76°13.07′ W
LCRa4	Little Choptank River adults – site 4	57	9 April 2002	38°34.08′ N, 76°10.57′ W
LCRs	Little Choptank River juveniles	163	June-August 2002	38°34′ N, 76°10′ W
PTKa	Piankatank adults	47	5 November 2002	37°30.58′ N, 76°20.53′ W
PTKs	Piankatank juveniles	48	5 November 2002	37°31.35′ N, 76°21.2′ W
PXD	Patuxent River downriver	50	18 October 2002	38°23.51′ N, 76°33.53′ W
PXU	Patuxent River upriver	50	18 October 2002	38°30.44′ N, 76°40.19′ W
RPD	Rappahannock River downriver	50	10 December 2003	37°36.32′ N, 76°24.75′ W
RPU	Rappahannock River upriver	50	10 December 2003	37°50.67′ N, 76°45.67′ W
YRD	York River downriver	50	10 December 2003	37°15.25′ N, 76°31.43′ W
YRU	York River upriver	50	10 December 2003	37°30.25′ N, 76°47.85′ W

Table 2-2. Per-locus and global allelic richness, gene diversity, Weir and Cockerham estimates of F_{IS} (f) and F_{ST} (θ), and exact tests of genotypic differentiation.

	*Cvi*9	*Cvi*12	*Cvi*1i24b	*Cvi*2g14	*Cvi*2i23	*Cvi*2i4	*Cvi*2j24	*Cvi*1g3	All loci
Number of alleles[§]	24	32	26	37	40	28	21	10	218
Allelic richness[§]	14.2	14.1	14.6	21.8	21.1	17.0	11.1	6.47	15.0
Gene diversity[§]	0.905	0.853	0.888	0.947	0.895	0.923	0.866	0.618	0.862
F_{IS}[§]	0.182*	0.176*	0.404*	0.087*	0.017*	0.164*	0.163*	0.233*	0.175*
F_{ST}[§]	0.001	0.002	0	0	-0.002	0.001	0.002	0.002	0.001
Exact test[§] (p value)	0.460	0.146	0.064	0.798	0.991	0.052	0.056	**<0.0001**	**0.0001**
F_{ST} (GWRa and GWRs)	-0.001	0.003	0.006	0	-0.001	0	-0.003	-0.004	0
Exact test (p value)	0.167	0.155	0.220	0.440	0.649	0.051	0.891	0.868	0.253
F_{ST} (LCRa1-4 and LCRs)	-0.002	-0.001	0.001	0.001	-0.001	0.002	0.008	-0.003	0.001
Exact test (p value)	0.623	0.767	0.058	0.262	0.547	0.340	**0.013**	0.987	0.146
F_{ST} (PTKa and PTKs)	-0.005	-0.007	-0.001	-0.005	-0.006	0.014	-0.005	-0.001	-0.002
Exact test (p value)	0.655	0.754	0.473	0.832	0.977	**0.018**	0.632	0.660	0.662
F_{ST} (JRD and JR90)	-0.005	0	0.02	-0.003	-0.003	-0.003	0.006	0.008	0.002
Exact test (p value)	0.703	0.253	0.118	0.548	0.417	0.400	0.294	**0.020**	0.119

[§]All populations except JR90; * $p < 0.001$; Bold type indicates $p < 0.05$ (Fisher's method)

Table 2-3. Number of alleles and mean allelic richness across all loci in each population. Singleton alleles were those that appeared only once in my analysis. Private alleles occurred more than once but were found only within one sample. Rare alleles were at less than 2% overall frequency.

		Number of Alleles				
Population	N	Singletons	Private	Rare	Total	Mean allelic richness
AIN	50	0	0	36	128	14.8
GWRa	90	3	1	55	148	15.1
GWRs	102	2	0	54	147	15.2
HRC	50	2	0	28	120	14.0
JR90	48	2	0	47	136	15.0
JRD	50	0	1	39	129	14.8
JRU	38	2	0	32	119	15.0
LCRa1	59	2	0	43	134	14.9
LCRa2	50	1	0	37	127	14.4
LCRa3	46	0	0	30	120	14.8
LCRa4	57	1	0	38	130	14.6
LCRs	163	5	0	62	155	14.2
PTKa	47	0	0	31	120	14.9
PTKs	48	2	0	35	127	14.7
PXD	50	1	0	35	126	14.6
PXU	50	0	0	35	125	15.0
RPD	50	3	0	38	129	15.5
RPU	50	0	0	43	134	15.3
YRD	50	3	0	43	133	15.0
YRU	50	2	0	41	130	16.0
All populations	1198	31	2	126	220	15.1

Table 2-4. IBD parameters for each locus, all loci combined, and all loci minus one. IBD slope and r^2 were calculated from regression of $\theta/(1 - \theta)$ against log distance. Mantel's test was performed to determine the significance of the relationship (P) between genetic differentiation and aquatic distance.

	Per Locus			One Locus Excluded		
Locus	IBD slope	r^2	P	IBD slope	r^2	P
*Cvi*9	0.0010	0.030	**0.032**	0.0006	0.076	**0.025**
*Cvi*12	0.0017	0.056	0.059	0.0005	0.057	**0.037**
*Cvi*1i24b	0.0002	0.001	0.415	0.0007	0.114	**0.002**
*Cvi*2g14	0.0004	0.016	0.160	0.0007	0.082	**0.016**
*Cvi*2i23	-0.0002	0	0.683	0.0008	0.096	**0.006**
*Cvi*2i4	-0.0001	0.002	0.706	0.0008	0.119	**0.001**
*Cvi*2j24	0.0009	0.024	0.231	0.0006	0.065	**0.019**
*Cvi*1g3	0.0020	0.020	**0.046**	0.0005	0.075	**0.035**
All loci	0.0007	0.093	**0.009**			

Bold type indicates uncorrected $p < 0.05$

Table 2-5. Effect of permutation on tests of differentiation and IBD. Exact tests and θ were calculated for 171 pairwise comparisons of 19 populations. Sequential Bonferroni correction was calculated with $\alpha = 0.05$.

		Pairwise test of differentiation					
		θ		Exact test			
	F_{IS}	# significant ($p < 0.05$)	# significant (Bonferroni corrected)	# significant ($p < 0.05$)	# significant (Bonferroni corrected)	IBD slope	IBD r^2
Original data	0.175	19	0	35	3	0.0007	0.093
Permuted data	0.003	78	6	117	85	0.0007	0.100

Figures

Figure 2-1. Map of Chesapeake Bay showing sampling sites and location abbreviations used in Table 2-1 and the text.

Figure 2-2. Isolation by distance in Chesapeake Bay oysters. Multilocus estimates of pairwise differentiation are plotted against logarithm of aquatic distances. The regression is $y = 0.0007\,x - 0.0023$ and the distance between subpopulations ranges from 1 to 345 km.

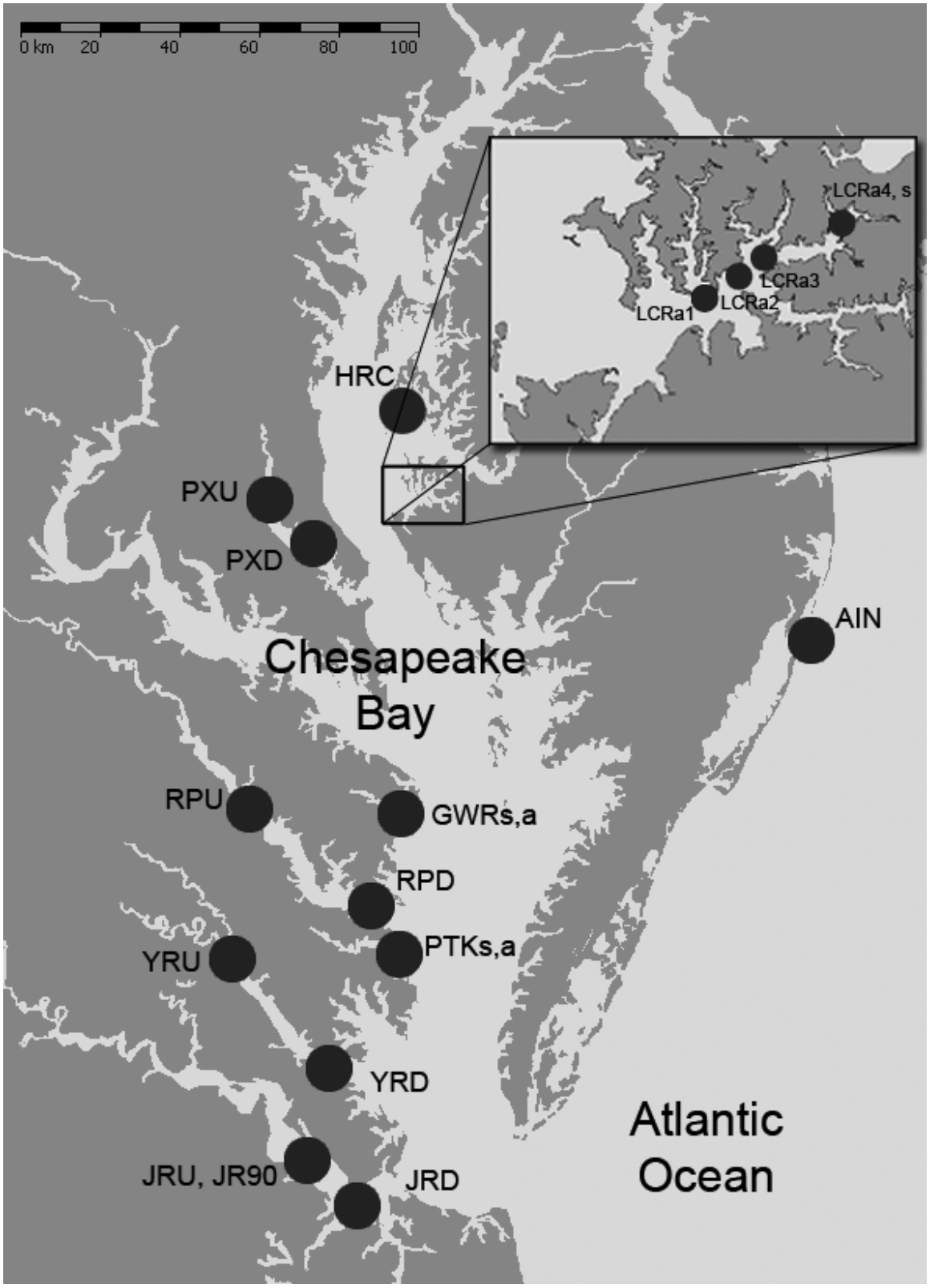
0 km
20
40
60
80
100
LCRa4, s
LCRa3
LCRa2
LCRa1
HRC
PXU
PXD
AIN
Chesapeake
Bay
RPU
GWRs,a
RPD
PTKs,a
YRU
YRD
Atlantic
Ocean
JRU, JR90
JRD

Figure 2-1

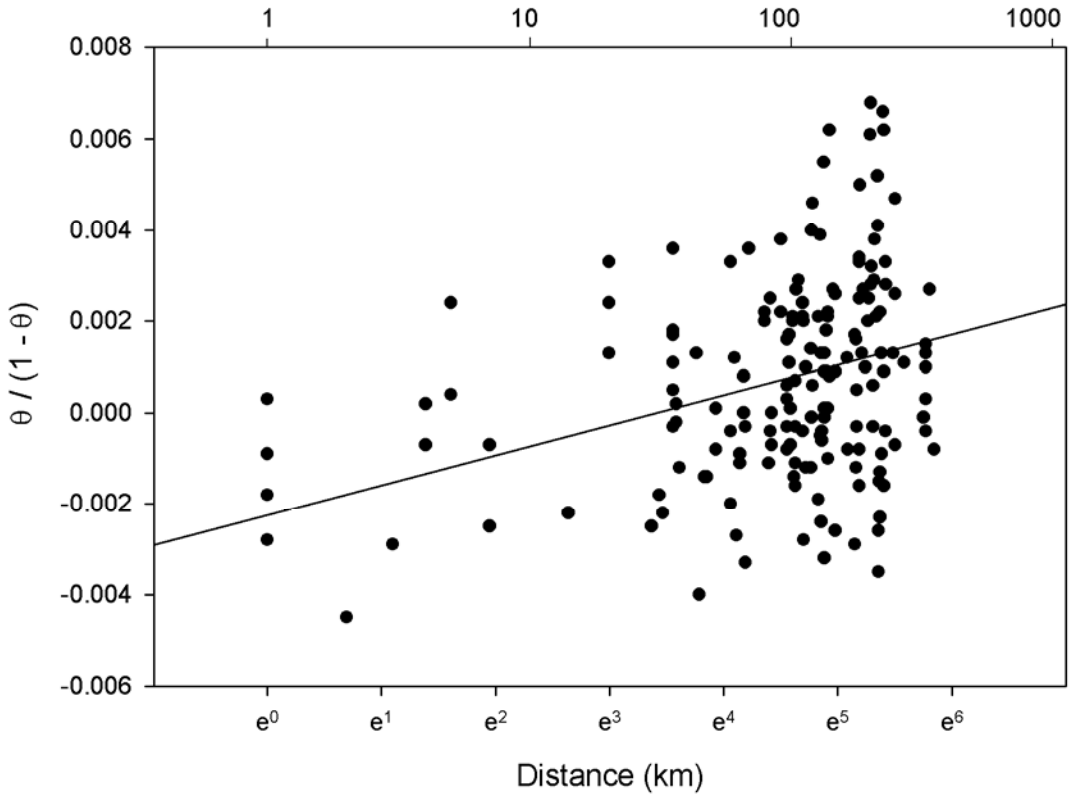
1
10
100
1000
0.008
0.006
0.004
0.002
0.000
-0.002
-0.004
-0.006
θ / (1 - θ)
e0
e1
e2
e3
e4
e5
e6
Distance (km)

Figure 2-2

Chapter 3: Positive Selection on Mitochondrial DNA

Abstract

Previous studies of oysters have revealed non-neutral patterns of allele frequency distributions characterized by an excess of high frequency and rare haplotypes and a deficit of intermediate haplotypes. This non-neutral pattern could be explained by demographic processes like sweepstakes reproductive events or recent bottlenecks, or could result from natural selection. I collected DNA sequence data from a population of oysters in Chesapeake Bay and found fundamentally different patterns of nuclear and mitochondrial DNA evolution. I observed non-neutral patterns of variation in mitochondrial DNA that could be due to demography or selection, but mostly neutral patterns at nuclear loci. Because demographic events are expected to create non-neutral patterns in mitochondrial and nuclear loci alike, it appears likely that mitochondrial DNA have undergone a recent selective sweep. An analysis of the nuclear loci in a standard coalescent framework indicated an effective population size on the order of 10^5 that was nevertheless many orders of magnitude smaller than census size, consistent with sweepstakes reproduction. Thus, large population size and selective sweeps were consistent with genetic draft, despite high variance in reproductive success and a small N_e/N ratio.

Introduction

Mitochondrial DNA (mtDNA) seems ideally suited for population genetic studies. Its high copy number makes it easy to amplify in the laboratory using PCR on even very small tissue samples (Avise et al. 1987). The maternal inheritance of animal mitochondria means that the genome is haploid, eliminating analytical problems associated with recombination and the determination of haplotypic phase. A high mitochondrial mutation rate in animals (relative to nuclear DNA) makes mtDNA markers highly polymorphic in many cases. Until recently it was widely assumed that most mitochondrial variation segregating in populations was selectively neutral (Brown 1983; Ballard and Kreitman 1995). Because of these perceived strengths, mtDNA has been used alone or in combination with nuclear DNA in countless studies of population genetics, phylogeography and systematics (reviewed in Avise et al. 1987; Moritz et al. 1987; Funk and Omland 2003).

The neutral theory of evolution (Kimura 1979) predicts that the amount of genetic variation (polymorphism) in a neutral marker is proportional to the effective size (N_e) of a population. Neutral evolution is a powerful theoretical framework that is utilized for increasingly sophisticated methods for estimating evolutionary N_e (Felsenstein 1992; Hey and Nielsen 2004), inferring historical population growth or decline (Kuhner 2006) and distinguishing recent migration from ancestral polymorphism (Beerli and Felsenstein 1999, 2001). Many studies have relied solely on mtDNA markers for estimating these parameters and processes in wild populations (Moritz et al. 1987; Avise et al. 1988; Neigel et al. 1991; Neigel and Avise 1993) despite the risk of bias if non-neutral forces

like natural selection shape the evolution of mitochondrial variation (Ballard and Kreitman 1995).

It has long been acknowledged that mtDNA could be subject to selective sweeps (Moritz et al. 1987; Rand 2001), but most studies using mitochondrial loci to make demographic inferences have assumed that segregating mtDNA variants are neutral or nearly neutral (Avise et al. 1987). Recent work, though, has shown that positive selection on mtDNA may be common in taxa with large population sizes. Bazin et al. (2006b) performed a meta-analysis of mitochondrial population genetic studies and found that across different classes of animals, population size does not correlate with mitochondrial genetic diversity. Tests of selection indicated that mtDNA frequently undergoes adaptive evolution, and that the strength of positive selection is greater in invertebrates (whose census populations tend to be large and therefore may have a large effective size) compared with vertebrates (with relatively small population sizes and, presumably, small effective sizes). The authors concluded that mtDNA haplotypes conferring higher fitness often sweep to fixation due to positive selection in species with large N_e (Bazin et al. 2006a). This scenario fits the predictions of the "genetic draft" hypothesis (Gillespie 2000, 2004), which posits that in wild populations selection events and their stochastic effects on linked loci are more important than genetic drift for shaping genetic variation. Species with large population sizes have a higher frequency of new favorable mutations and a higher likelihood of selective sweeps, so polymorphism should be relatively insensitive to population size for genomic regions with low recombination rates. The

draft hypothesis predicts that above a certain population size positive selection will maintain similar levels of mitochondrial variation across species.

The conclusion of Bazin et al. (2006b) that mtDNA shows signs of positive selection across animal classes is contentious and has been subject to scrutiny (Mulligan et al. 2006; Wares et al. 2006). For example Mulligan et al. (2006) argued that for mammals, whose effective sizes tend to be smaller than those of invertebrate species, mtDNA diversity corresponds to population size and therefore mtDNA appears to evolve neutrally. Other evidence, however, suggests that even mammalian mitochondria can be subject to strong positive selection (Elson et al. 2004). Stronger studies demonstrating positive selection on mtDNA in individual taxa are needed to confirm the claim by Bazin et al. (2006b). Unfortunately, among non-insect invertebrates there have been very few studies providing robust tests of mtDNA selection by comparing patterns of DNA sequence variation across multiple independent loci.

Here I address whether positive selection is an important factor in the mitochondrial evolution of the Chesapeake Bay (USA) population of the eastern oyster, *Crassostrea virginica*. *C. virginica* is a protandrous, long-lived bivalve that lives in subtidal and intertidal mesohaline waters along the east coast of North America, from New Brunswick in Canada to the Yucatan Peninsula in Mexico (Galtsoff 1964). Non-neutral distributions of mitochondrial haplotypes have been reported for both Atlantic and Gulf of Mexico populations of *C. virginica* (Reeb and Avise 1990) and in the Pacific oyster, *C. gigas* (Boom et al. 1994). Contrary to Ewens's (1972) null expectations for the haplotype

frequency spectrum of neutral alleles, these crassostreid oyster populations were shown to have an excess of both high frequency and very rare mitochondrial haplotypes, and a deficit of intermediate frequency haplotypes. This pattern, typically recognized in terms of a negative Tajima's *D* value, is associated with "star-like" genealogies in which the high frequency common allele(s) are separated by relatively few mutational steps from a large number of rare alleles (Beckenbach 1994). These patterns can be produced by any of a number of processes that deviate from the assumptions of a Wright-Fisher population (i.e., non-neutral processes). When non-neutral evolution takes place, estimates of historical timing or N_e made with an assumption of neutrality can be misleading (Rosenberg and Hirsh 2003) and the proper correction is unknown without reference to patterns at independent loci. It is therefore informative to use nuclear data to address the question of mtDNA neutrality in oysters because effective population size in this and other high-fecundity marine invertebrates has been a matter of considerable debate.

The skewed mitochondrial haplotype spectrum in oysters was the inspiration for the "sweepstakes" hypothesis (Hedgecock 1994; Li and Hedgecock 1998), which describes reproduction in some marine invertebrates as a sweepstakes with a few very fecund winners, high larval mortality (type III survivorship) and the unpredictability of a marine environment. The degree to which sweepstakes can explain patterns of genetic variation in marine invertebrates is contentious (Flowers et al. 2002) but it nevertheless provides an explanation for skewed haplotype frequencies and for the low ratio of effective population size to census size (N_e/N ratio) found in many marine organisms (Hedrick 2005). Beckenbach (1994) hypothesized that sweepstakes reproduction leading to low N_e

in combination with a high mitochondrial mutation rate can account for the observed haplotype frequency skew: a female who is disproportionally successful at reproduction would pass her mitochondrial haplotype to a large proportion of the next generation, while meiotic mutations carried by her gametes would lead to unique variations of the common haplotype. In order to explain the extreme skew found in oyster populations, Beckenbach's simulations demonstrated that female effective population sizes must be very small ($N_e < 300$) and mutation rate must be very high ($\mu \sim 10^{-2}$). Assuming sexual reproduction, even sex ratio and a diploid genome, the total N_e of a population can be no more than 4 times its female N_e (Hartl and Clark 1997), so under the Beckenbach hypothesis a very small effective number of females would necessarily mean that the total N_e is also small. Therefore, under the Beckenbach hypothesis I predict one of two patterns at nuclear loci. First, if the nuclear mutation rate is very high I should see a haplotype spectrum skewed in a similar way to the mitochondria, with an excess of common and rare alleles, and a deficiency of intermediate frequency alleles; nuclear haplotype networks should be star-like and resemble mitochondrial networks. Alternatively, if nuclear mutation rates are low, then nuclear loci should show very low levels of genetic diversity. Evidence of large N_e or of neutral patterns of haplotype diversity in nuclear loci would indicate that the Beckenbach hypothesis cannot explain the mitochondrial haplotype skew in oysters.

The non-neutral mitochondrial pattern in oysters also could be understood through the multiple-mergers coalescent model (Eldon and Wakeley 2006). Unlike Beckenbach (1994), Eldon and Wakeley found that reproductive skew could explain the non-neutral

mtDNA patterns even without elevated mutation rate. The multiple-mergers coalescent is based on a modified Moran (1958) model in which a fraction ψ of the population comes from a single parent in the previous generation, a departure from the standard Kingman coalescent model (Kingman 1982b, a) and its Wright-Fisher population assumptions (Wright 1931). Eldon and Wakeley (2006) used the mitochondrial data of Boom et al. (1994) to show that the mitochondrial haplotype skew in Pacific oysters is consistent with a scenario in which 8% of the population was replaced by the offspring of a single, highly successful female. The multiple-mergers coalescent model has only been rigorously developed for haploid genomes, but the extreme reproductive skew suggested by mtDNA data under the multiple mergers theory should also be reflected in patterns of diversity at nuclear loci.

Small N_e due to recent bottlenecks, rather than oyster biology in general, could also produce a skew in mitochondrial haplotype frequency. A bottleneck can reduce mitochondrial variation to just one or a few haplotypes, with new mutations appearing during subsequent population growth. New allelic variants would remain at low frequency in a population until sufficient time has passed to allow some rare mutations to reach intermediate or high frequencies, as expected under a neutral distribution (Kelly 1977). A population that has undergone a recent bottleneck is expected to exhibit the negative Tajima's *D* values and the star-like haplotype networks found in oyster mtDNA (Tajima 1989). To explain the simultaneous haplotype skew in both *C. virginica* and *C. gigas*, however, a coincidence of bottleneck timing must have occurred. Beckenbach (1994) speculated that this coincidence is unlikely, especially given that the *C. gigas*

population was introduced, but patterns in multilocus data make it possible to test whether bottlenecks are the best explanation for the mitochondrial haplotype skew.

Finally, the skewed haplotype frequency spectrum could be the result of recent selective sweeps on the mitochondrial genome. Because the mitochondrial genome does not recombine, a strongly selected advantageous variant at any mitochondrial locus could "sweep" away linked variation and if driven to fixation, reduce a population's variation to just one genomic haplotype. Neutral or nearly-neutral (Ohta 1992) mutations that have accumulated since a selective sweep can explain the excess of rare alleles, particularly if the sweep was relatively recent and new alleles have not had time to drift to higher frequencies. This pattern might appear in different oyster species and in other marine invertebrates with large effective population size if, as Bazin et al. (2006b) suggest, selective sweeps are common in mitochondrial DNA of species with large N_e. A difficulty with single locus tests of selection is that the signature of selection at any one locus could also be explained by demographic events (Tajima 1989; Braverman et al. 1995; Simonsen et al. 1995; Fu 1997; Galtier et al. 2000; Hahn et al. 2002). For example a negative value of Tajima's *D* can result from selective sweeps or from recent genetic bottlenecks (Tajima 1989). But natural selection is locus-specific, with effects at one part of a recombining genome generally not extending to unlinked loci (Nielsen 2005). Therefore, a consistent non-neutral pattern across mitochondrial and nuclear loci in oysters would imply that forces other than natural selection explain the mitochondrial skew (Stajich and Hahn 2005).

Here I present DNA sequence data for several mitochondrial loci to determine if the Chesapeake Bay population of *C. virginica* exhibits the non-neutral mitochondrial patterns found in other oyster populations. I then make comparisons with sequence variation at multiple nuclear loci to assess whether the demographic explanations outlined above can explain the combined genomic data, or whether the data are more consistent with oyster population sizes large enough to make nonrecombining regions such as mtDNA susceptible to positive selection and Gillespie's genetic draft.

Materials and Methods

Sampling

I collected live adult *C. virginica* from four locations in Chesapeake Bay. Sampling sites were from an unknown site in Virginia, Drum Point in the Little Choptank River (38°38.68' N, 75°57.55' W) in October 1999, Deep Neck in the Choptank River (38°44.38' N, 76°14.70' W) in April 2001, and Bolingbroke Sands in the Choptank River (38°34.68' N, 76°01.68' W) in June 2005. Previous to my collection, Bolingbroke Sands was planted with hatchery oysters produced from wild broodstock (K. Paynter, personal communication), so the sample is likely to include individuals that were hatchery-raised. I also collected outgroup samples of the mangrove oyster, *C. rhizophorae*, the closest known species to *C. virginica* (Buroker et al. 1979). Samples of *C. rhizophorae* were collected by H. Hertler (University of Puerto Rico) in Boquerón Bay, Puerto Rico (18°00.48' N, 67°10.73' W) in August 2006. Oysters were kept on ice until they were brought into the laboratory, where gill and/or mantle tissue was removed and preserved in

ethanol. *C. virginica* samples from Bolingbroke Sands and *C. rhizophorae* samples were also preserved in RNAlater (Applied Biosystems).

DNA Extraction, Amplification and Sequencing

Genomic DNA was extracted with DNeasy 96 Tissue Kits (Qiagen Inc., Valencia, CA) using the protocol for animal tissues. Samples preserved in RNAlater were used for RNA extraction and cDNA synthesis. RNA was extracted using an RNeasy Mini Kit (Qiagen) using the protocol for animal tissues with frozen tissue ground with mortar and pestle, QIAshredder homogenization, and a DNAse step. RNA was used for cDNA synthesis using an oligo$(dT)_{18}$ primer with a First Strand cDNA Synthesis Kit (Fermentas Inc., Glen Burnie, MD). After synthesis, cDNA was purified with a MinElute PCR Purification Kit (Qiagen).

Names and descriptions of the loci used in this study are found in Table 1. I used the polymerase chain reaction (PCR) to amplify DNA. Reaction conditions for PCR in a total volume of 25 µl included 1 µl DNA, 0.12 µl Bovine Serum Albumin, and 0.12 µl Invitrogen *Taq* polymerase with final concentrations of 1 × Invitrogen buffer (without $MgCl_2$), 250 µM deoxynucleotide triphosphate, 4 mM $MgCl_2$ and 200 nM each of forward and reverse primers. Thermocycling involved one cycle of 95°C denaturing for 1 min; 35 three-step cycles including 95°C for 30 s, annealing temperature for 30 s, a 1°C/s ramp to 72°C for 2 min; and a final extension at 72°C for 5.5 min. Annealing temperatures and other information for each primer pair are listed in Table 2. Because large introns inhibited PCR amplification when using genomic DNA, I used cDNA to amplify all nuclear loci except *Vg*. *Cox1* primers developed by Folmer et al. (1994)

amplified DNA from both *C. virginica* and *C. rhizophorae*. Primers for all other *C. virginica* mitochondrial markers were developed from the *C. virginica* mitochondrial genome sequence (GenBank accession #AY905542). Degenerate primers for amplifying *C. rhizophorae Atp6* and DNA encompassing *Cytb*, the *Cytb* 3' untranslated region (UTR) and *Cox2* were developed after aligning the *C. virginica* mitochondrial genome with the *C. gigas* mitochondrial genome sequence (GenBank accession #NC_001276). *Nad2* and *Atp6* 3' UTR were extremely variable between *C. virginica* and *C. gigas* and degenerate primers did not amplify *C. rhizophorae* DNA. Primers amplifying nuclear loci were developed from EST sequences developed by Peatman et al. (unpublished, 2003; GenBank accession numbers reported in Table 2). EST sequences were chosen from "housekeeping" genes and from genes that are potentially involved in disease response.

After amplifying DNA with PCR, samples were prepared for sequence analysis using each of the PCR primers. I cleaned PCR samples by adding 0.33 µl each of shrimp alkaline phosphatase (SAP), exonuclease I, and SAP dilution buffer to 20 µl PCR product, then incubating the mixture at 37°C for 30 min and 80°C for 15 min. Sequencing reactions were carried out using a BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). I prepared 10 µl sequencing reactions with 0.5 µl DNA, 2 µl BigDye and with final concentrations of 1.5 × BigDye buffer and 330 nM primer. Thermocycling involved 25 three-step cycles of 96°C for 10 s, 50°C for 30 s, and a 0.9°C/s ramp to 60°C for 4 min. After thermocycling the DNA was washed with 40 µl 75% isopropanol, vortexed and incubated at room temperature for 15 min, then centrifuged for 30 min at 3770 RPM. Immediately following the spin the lids were

removed, PCR tubes were inverted and the isopropanol was removed by centrifugation at 2230 RPM for 1 min. I performed a second rinse using 75 µl 70% isopropanol, vortexed and centrifuged the samples at 3770 for 10 min, then inverted tubes and removed the isopropanol with a 2230 RPM spin for 1 min. Following the wash, PCR products were resuspended in 10 µl Hi-Di formamide (Applied Biosystems), heated at 95°C for 2 min, placed on ice for 2 min, then analyzed with an ABI-Prism 3100 or 3730xl genetic analyzer (Applied Biosystems). I collected at least 2 overlapping sequences for each sample. Base calls were made using Sequencher 4.6 (Gene Codes, Ann Arbor, MI).

Because the *Cox1* PCR primers generated poor quality sequence, I cloned *Cox1* PCR products with a TOPO TA Cloning Kit (Invitrogen, Inc.) and sequenced the clones with M13 forward (-20) and reverse primers. In order to account for *Taq* error I collected sequences from at least 3 clones per sample and made sequence calls based on the majority consensus. For all other loci I sequenced the PCR products directly. I estimated haplotype phase of nuclear sequence data using Phase 2.1.1. Haplotypes with confidence less than 70% in Phase were confirmed by sequencing clones produced from heterozygous PCR products. After confirming low confidence haplotypes I reestimated haplotypes with Phase.

Analysis of Genetic Variation

DNA sequences were aligned with ClustalX in MEGA4 (Tamura et al. 2007), and adjustments were made by eye to correct alignment errors. *Cytb* 3' UTR amplified for *C. virginica* and *C. rhizophorae* but high sequence divergence precluded unambiguous alignment, so outgroup sequences were not included in the *Cytb* 3' UTR analysis. All

indices and tests of selection were based on alignments that excluded sites with missing data or gaps. MEGA4 was used to calculate nucleotide diversity indices including number and proportion of segregating sites; proportion of variable sites; $\hat{\theta}_{\omega}$, Watterson's (1975) per site heterozygosity based on variable sites; $\hat{\theta}_{\pi}$, Nei's (1987) per site heterozygosity based on average pairwise proportion of differences between sequences; and between-species divergence, the average number of nucleotide substitutions per site between species. For *C. virginica* nucleotide diversity indices were calculated for replacement sites (zerofold degenerate coding) and silent sites (noncoding, twofold and fourfold degenerate) using the flatworm mitochondrial genetic code for mitochondrial loci or the standard genetic code for nuclear loci.

I used Arlequin 2.000 (Schneider et al. 2000) to calculate mitochondrial haplotype frequencies for *C. virginica* mitochondrial loci. *Cox1* was removed from the analysis because of small sample size, and 4 haplotypes were removed because of missing data. I calculated expected haplotype frequencies and standard deviations with 10,000 simulations under the infinite-allele model based on Ewens's (1972) distribution. I created haplotype networks, based on the parsimony method described in Templeton et al. (1992), using TCS 1.21 (Clement et al. 2000). Each nucleotide insertion or deletion was considered to be a mutational step.

Tests of Selection

I performed tests of selection using DnaSP 4.50.3 (Rozas et al. 2003). I calculated Tajima's *D*, which is a scaled measure of the difference in the compound parameter

$\theta = 4N_e\mu$ as calculated by $\hat{\theta}_\pi$ and $\hat{\theta}_\omega$. Significance was determined from confidence limits estimated from a beta distribution (Tajima 1989). I also calculated Fu's (1997) F_S statistic, a test of population growth or positive selection which compares observed and expected haplotypic frequency distributions. Significance was determined for Fu's F_S using 10,000 coalescent simulations based on the calculated value of $\hat{\theta}_\omega$, using DnaSP to estimate recombination rate *R* for nuclear loci or assuming no recombination for mitochondrial loci. A combined mitochondrial data set was created by concatenating sequences from each of the mtDNA loci (except *Cox1*, which could not be concatenated) from each individual. Arlequin was used to calculate Tajima's *D* and Fu's F_S for the combined *C. virginica* mitochondrial data set (excluding 4 individuals with missing sequences).

For those loci with outgroup sequences, I performed additional tests of selection with DnaSP. I determined d_N/d_S, the ratio of replacement to silent polymorphisms. The ratio is expected to equal 1 under neutral conditions, < 1 with negative (purifying) selection, and > 1 with positive selection. However, because some purifying selection is expected and will lower d_N/d_S, this ratio provides a very conservative test of positive selection and has low power for many data sets (Nielsen 2001). The test is most useful when the use of multiple loci allows the identification of outliers (Barrier et al. 2003). I also performed Fu and Li's *D* (*FL-D*) and *F* (*FL-F*) tests (Fu and Li 1993), variations of Tajima's *D* that are powerful tests for detecting background selection (Fu 1997). Statistical significance was determined for *FL-D* and *FL-F* using critical values from Fu and Li (1993). I performed the Hudson, Kreitman and Aguadé (HKA) test (Hudson et al. 1987), which measures

whether the rate of evolution correlates with within-species polymorphism. This test determines if the data fit the expected pattern of equal evolutionary rates in which loci that evolve quickly (those with high levels of divergence between species) exhibit high levels of within-species polymorphism. If forces other than mutation are affecting interspecific divergence or intraspecific polymorphism, then the null hypothesis of equal rates of evolution is rejected. I performed the HKA test in DnaSP using the "direct mode" for each pairwise locus comparison by entering intraspecific data including the number of segregating sites and sample sizes, and interspecific data including average number of differences and total number of sites. For the HKA test mitochondrial loci were characterized as Y chromosome data to compensate for their haploid, uniparental inheritance. Finally, I performed the McDonald-Kreitman (MK) test (McDonald and Kreitman 1991), which tests for selection independently from genealogical patterns. The MK test compares the ratio of nonsynonymous to synonymous polymorphisms within species to the ratio of nonsynonymous to synonymous fixed differences between species. Like the HKA test, the MK test assumes that mutation and genetic drift are the only forces affecting the between-species ratio of fixations to polymorphisms in silent sites. Under neutrality the same ratio is expected in silent and replacement sites. Deviations in the ratio of fixations to polymorphisms between silent and replacement sites suggest the action of natural selection. After performing the MK test I used a G test to determine whether 2x2 contingency tables deviate from expectations. For each locus the type of selection affecting the MK test was assessed with the neutrality index (N.I.), as described by Rand and Kann (1996), where

$$\text{N.I.} = \frac{\left(\frac{\text{\# replacement polymorphisms}}{\text{\# replacement fixed differences}}\right)}{\left(\frac{\text{\# silent polymorphisms}}{\text{\# silent fixed differences}}\right)}.$$

The ratio of silent polymorphisms to silent fixed differences found in the denominator represents neutral expectations, so N.I. = 1 indicates that replacement sites are evolving neutrally. Positive selection is inferred when N.I. < 1 because there are more replacement fixed differences than expected, and purifying selection is inferred when an excess of replacement polymorphisms leads to N.I. > 1. The HKA and MK tests of selection are highly reliant on an appropriate outgroup: if the outgroup is too distant then the infinite alleles assumption may be violated; an outgroup that is too closely related will not have enough mutations to distinguish shared polymorphisms with fixed differences (Wayne and Simonsen 1998). Finding an appropriate outgroup can be especially problematic when comparing across nuclear and mitochondrial loci if they experience different rates of evolution.

Coalescent Analyses

I analyzed mitochondrial loci from *C. virginica* under the multiple-mergers coalescent framework using the computer program Multicoal2 (generously provided by J. Wakeley, Harvard University). Under selective neutrality the program estimates the mutation parameter θ and ψ, the scaled family size measured as the fraction of the total population that the offspring of a single female replaces. The mutation parameter θ in the multiple-mergers coalescent is analogous (though not directly comparable) to θ under Kingman's coalescent, since deviations from the assumptions of a Wright-Fisher population mean that N_e cannot be defined in the traditional sense. I compiled Multicoal2 using GCC, and used data from *Atp6*, *Atp6* 3' UTR, *Nad2*, *Cytb*, *Cytb* 3' UTR, and *Cox2*; I excluded *Cox1* data due to small sample size. The data input to Multicoal2 are 44 sequences, 33 variable sites, and 17 singletons. I performed an investigation of 20,000 gene genealogies for each

point in a grid with 150 values of ψ and 120 values of θ. Results were exported to Mathematica 6.0.1.0 (Wolfram Research, Champaign IL) for analysis and visualization.

Estimations of θ ($4N_e\mu$) and recombination rate were performed using the standard coalescent. I used maximum likelihood and Bayesian frameworks in Lamarc 2.1.2b (Kuhner 2006) to analyze twofold and fourfold degenerate coding sites and noncoding sites from the seven nuclear loci. I performed 3 maximum likelihood and 3 Bayesian runs of 10 initial chains with 2,000 trees, then 2 final chains with 20,000 trees. Maximum likelihood analyses were used to generate point estimates for θ and recombination rate, and Bayesian analyses generated 95% confidence intervals. Loci were analyzed using the F84 model of sequence evolution (Kishino and Hasegawa 1989) and because of computational constraints, simulations were carried out with equal relative effective population sizes and equal relative recombination rates. Also because of computational constraints, loci with more than 20 sequences were randomly subsampled to create sample sizes of 20 (Kuhner 2006).

Results

Numbers of sequences, which varied by collection site for most loci, are described in Table 3. Sequence diversity for the seven mitochondrial and seven nuclear loci are described in Table 4. For all loci, $\hat{\theta}_\omega$, $\hat{\theta}_\pi$, and interspecific divergence were higher at silent than replacement sites. Overall divergence between *C. virginica* and *C. rhizophorae* was significantly higher at mitochondrial loci (mean value 0.17) than at nuclear loci (mean value 0.04, t-test assuming unequal variance, $p < 0.01$). Divergence

between species was significantly higher for mitochondrial loci than nuclear loci at silent sites (mean mitochondrial divergence = 0.54, mean nuclear divergence = 0.13, t-test assuming unequal variance, $p < 0.01$), but not at replacement sites (mean mitochondrial divergence = 0.05, mean nuclear divergence = 0.01, t-test assuming unequal variance, $p > 0.05$). There was also a significant difference in *C. virginica* estimates of $\hat{\theta}_\pi$ between mitochondrial (mean value 0.0018) and nuclear loci (mean value 0.0071, paired two sample t-test, $p < 0.05$). In contrast, $\hat{\theta}_\omega$ and proportion of polymorphic sites showed no statistical differences between nuclear and mitochondrial loci (paired two sample t-tests, $p > 0.05$).

Observed and expected mitochondrial haplotype frequencies for *C. virginica* are presented in Figure 1. In agreement with previous studies (Reeb and Avise 1990; Boom et al. 1994) I detected an excess of high frequency alleles, a deficit of intermediate frequency alleles, and an excess of singleton alleles. Testing the significance of this deviation from neutral expectations in both mitochondrial and nuclear loci (Table 5), I found the non-neutral patterns to be nearly absent from nuclear genes. Three of seven mitochondrial loci showed significantly negative values of Tajima's *D*, in contrast to just one significant test out of seven for nuclear loci. The combined mitochondrial data had Tajima's $D = -2.58$ ($p < 0.001$). Using the more powerful F_S test, six of seven mitochondrial loci showed significant results, whereas one of seven nuclear loci significantly departed from neutral expectations. For the combined mitochondrial data set I found $F_S = -27.2$ ($p < 0.00001$).

Tests of selection with an outgroup were performed on the four mitochondrial loci for which sequences from *C. rhizophorae* could be obtained and aligned to *C. virginica*, and are reported in Table 6 and Table 7. A 2-sample t-test (assuming unequal variances) showed no significant difference between mitochondrial and nuclear loci for d_N/d_S ($p > 0.05$). All values of d_N/d_S (Table 6) were much closer to 0 than to 1, suggesting that purifying selection has been a strong evolutionary force on the *C. virginica* loci studied. *FL-D* and *FL-F* (Table 6) each demonstrated significance at 1 out of 4 mitochondrial loci and 1 of 7 nuclear loci. Results from the HKA test are reported in Table 7. Of the 55 pairwise comparisons, 10 indicated significantly different ratios ($p < 0.05$) of silent to replacement variation between loci: 1 of 6 (17%) mitochondrial-mitochondrial, 2 of 21 (11%) nuclear-nuclear, and 7 of 28 (25%) mitochondrial-nuclear comparisons. Highly significant ($p < 0.01$) results were found only in mitochondrial-nuclear comparisons, but no tests were significant after Bonferroni correction for multiple tests. MK tests were all nonsignificant ($p > 0.05$). Neutrality index values are reported in Table 6. *Atp6* and *Cytb* had N.I. = 0 due to a lack of polymorphic replacement substitutions. N.I. was undefined for the nuclear loci *Acadm* and *Gpi* because there were no fixed replacement substitutions.

Haplotype networks are reported in Figure 2. Mitochondrial haplotype networks were generally star-like and characterized by a high frequency central allele that was closely related to multiple low frequency alleles. All mitochondrial alleles were separated from central haplotypes by at most two mutational steps (including *Atp6* 3' UTR whose apparent long branch was actually caused by a single, 3 bp insertion). In contrast, nuclear

loci had haplotype networks that included long branches and frequent recombination (loops in the networks). Only the nuclear loci *Vg* and to a lesser extent *Gapdh* exhibited the star-like pattern common to mitochondrial loci.

Results from the analysis of mitochondrial loci under the multiple-mergers framework are shown in Figure 3. There were two maximum likelihood points found, approximated by a single "x" in the figure. The first maximum likelihood point was $\theta = 0.056$ and $\psi = 0.10$ and the second is $\theta = 0.067$ and $\psi = 0.10$. Both points were consistent with a sweepstakes event in which a single individual replaced 10% of the population.

Using the standard coalescent on nuclear data, independent comparisons across 3 Bayesian and 3 maximum likelihood runs provided nearly identical estimates of θ and recombination rate. The mean across runs for multilocus nuclear DNA θ was 0.032, with a mean 95% Bayesian confidence interval of 0.028 to 0.044. Values of θ varied moderately from locus to locus, with a range of 0.011 to 0.056. Recombination was estimated to occur at an average frequency of 0.74 with multilocus Bayesian 95% confidence interval estimated to be 0.441 to 1.07, indicating that recombination was 74% as frequent as mutation. Recombination rates were highly variable among loci and ranged between 0.001 and 1.127. I assumed a silent site mutation rate $\mu = 1 \times 10^{-8}$ nucleotide substitutions per generation (Yang et al. 2001) and a generation time (T) of 5 years (Galtsoff 1964). Rearranging $\theta = 4N_e\mu$ and accounting for generation time, I calculated $N_e = \frac{\theta}{4\mu T} = 1.6 \times 10^5$.

Discussion

Multilocus sequence data from a population of *C. virginica* indicate patterns of non-neutral mitochondrial evolution but mostly neutral nuclear patterns of variation. Using these data to understand the evolutionary forces shaping oyster genetic variation is a difficult task because differences in mutation rate, inheritance pattern and cellular copy number complicate comparisons of genetic variation across genomes. I will revisit and evaluate the demographic and selective explanations for patterns in the data, and focus on the relationship between effective population size and natural selection in the context of sweepstakes reproduction and the potential for genetic draft.

Does Oyster Demography Create Mitochondrial Skew?

For most single locus tests of selection, a severe genetic bottleneck and a strong selective sweep are statistically indistinguishable. The key to distinguishing the two hypotheses is to test for consistency across loci: a non-neutral pattern that is evident throughout the genome is likely due to demography, whereas a non-neutral pattern that is restricted to a small number of loci is probably due to selection. Here the nuclear loci were for the most part evolving neutrally while the mitochondrial loci (which are in reality one large, unlinked locus) were not. The difference between patterns of nuclear and mitochondrial diversity was further illustrated by haplotype networks. Since they depict the frequencies and relatedness of DNA sequences, haplotype networks provide a visual summary of genetic variation, even in the presence of recombination and polytomies. Differences between mitochondrial and nuclear haplotype networks (Figure 2) were indicative of very different patterns of nuclear and mitochondrial diversity. The homogeneity of the shape

and distribution of variation among mitochondrial networks was contrasted by the variation found among nuclear networks. The relatively uniform nuclear haplotype frequencies in the nuclear networks were consistent with a large population, not one that has undergone a recent bottleneck. Because most nuclear loci conformed to neutral expectations, bottlenecks do not provide the most parsimonious explanation for non-neutral mitochondrial patterns.

The Beckenbach hypothesis, that mitochondrial haplotype skew is due to an extremely small effective population size coupled with a high mutation rate, is inconsistent with results presented here. The standard coalescent results indicated that effective population size was on the order of 10^5 although this estimate is sensitive to the assumed nuclear mutation rate. Furthermore, empirical evidence rules out a high mitochondrial mutation rate necessary for Beckenbach's hypothesis to explain the mitochondrial distribution. A recent study of *C. virginica* used temperature gradient capillary electrophoresis to compare mothers and offspring and determined that the mitochondrial mutation rate is between 1×10^{-5} and 6×10^{-5} substitutions per site per generation (Milbury 2007), far lower than the mutation rate of 10^{-2} that Beckenbach's model requires.

A recent genetic bottleneck also can explain the mitochondrial haplotype skew that I observed. Because the diploid nuclear genome has a larger N_e than mtDNA, a bottleneck of intermediate severity could leave neutral variation at nuclear loci and remove non-neutral variation at mitochondrial markers. In many organisms the uniparental inheritance and haploid nature of mtDNA lead to a fourfold lower mitochondrial N_e compared with

the nuclear genome (Palumbi et al. 2001). However, protandrous reproduction permits all oysters to reproduce as females, so mitochondrial N_e may be closer to one half nuclear N_e. The relatively small difference in effective population size between the two genomes reduces the likelihood that a bottleneck would have drastically different effects on their genetic variation.

The multiple-mergers coalescent, perhaps the most important theoretical development motivated by the idea of sweepstakes reproduction, has the potential to change the interpretation of effective population size in species with high variance in reproductive success. Sargsyan and Wakeley (2008) suggested that the multiple-mergers might be applicable to marine organisms like the oyster, and indeed the mitochondrial data were consistent with sweepstakes events in which a single parent gave rise to 10% of all offspring in the population. But is sweepstakes reproduction the best explanation for mitochondrial haplotype distributions in *C. virginica*? The multiple-mergers model, like most existing coalescent models, is based on selective neutrality. Positive selection can generate patterns of DNA polymorphism identical to what would be expected with sweepstakes reproduction under the multiple-mergers framework (Eldon and Wakeley 2006). Given the neutral patterns of sequence evolution at nuclear DNA, it is unlikely that a multiple-mergers analysis optimized for nuclear loci would lead to high values of ψ. I conclude that mitochondrial selection, rather than sweepstakes reproduction or any other demographic explanation, is the most likely reason for observed patterns of oyster mitochondrial evolution.

Evidence of Mitochondrial Selection

Tests of natural selection indicated that the nuclear loci chosen for this study were generally evolving according to neutral expectations, but that the mitochondrial genome has undergone a selective sweep. The pattern of highly significant Fu's F_S for most mitochondrial loci was a strong indicator of positive selection after demographic explanations were rejected. In contrast, only one nuclear locus showed the same statistically significant pattern of positive selection. The remaining nuclear loci had negative values of Fu's F_S, but this is expected when recombination takes place (Fu 1997) and the values were not significantly different from neutral expectations. The HKA test also indicated a difference in evolutionary patterns between mitochondrial and nuclear loci. The test produced the greatest number and proportion of significant results in pairwise comparisons between mitochondrial and nuclear loci; there were relatively few cases in which tests between mitochondrial loci or tests between nuclear loci showed significantly different patterns of evolution. The d_N/d_S ratio, on the other hand, indicated that both mitochondrial and nuclear loci experienced purifying selection that led to higher rates of evolution at silent sites compared with replacement sites.

MK tests did not reveal a statistically significant signal of natural selection at any nuclear or mitochondrial loci. This result is surprising given that the other tests rejected the null model, and may reflect the fact that the MK test has difficulty detecting positive selection when purifying selection reduces levels of replacement polymorphisms (Fay et al. 2002). Even under ideal circumstances the test is conservative and can miss a signal of non-neutral evolution (Akashi 1999; Hahn 2008). Nevertheless, the related neutrality index

provided some provocative results. First, N.I. > 1 at *Cox1* suggested that negative selection may have been especially important in the evolution of the gene. This conclusion was reinforced by other indications of negative selection, including the lowest observed value of d_N/d_S and significant results from *FL-D* and *FL-F* tests. The signature of *Cox1* negative selection, superimposed on the mitochondrial pattern of positive selection at other mitochondrial genes, suggest that there has been strong functional constraint preventing the accumulation of deleterious mutations in the *Cox1* gene. Negative selection on *Cox1* may have been recent because its signal necessarily must have arisen since a selective sweep on the mitochondrial genome. These *Cox1* results are in agreement with those of Meiklejohn et al. (2007), whose meta-analysis showed that in invertebrates N.I. tends to be higher for the *Cox* genes than for other mitochondrial genes, and those of Pesole et al. (1999) who found heterogeneous rates of evolution. I am unable to thoroughly test an alternative hypothesis that *Cox1* is an outlier among mitochondrial genes because of sampling error. *Cox1* sequences were determined for samples mostly collected from Bolingbroke Sands. If these samples included oysters that were derived from a hatchery, it is possible that a hatchery-induced bottleneck could have affected levels of genetic variation, but without comparative data this possibility is difficult to test conclusively. Given the intermediate levels of silent site $\hat{\theta}_\omega$ and $\hat{\theta}_\pi$ for *Cox1* relative to the other mitochondrial loci, there were no strong indications of bottleneck effects in Bolingbroke Sands samples. In any case, while bottlenecks can lead to negative values of *FL-D* (Depaulis et al. 2003) and *FL-F* (Simonsen et al. 1995), bottlenecks are expected to affect d_N and d_S equally (Tajima 1989) and therefore should not affect the d_N/d_S ratio or the neutrality index associated with the MK test. However, selection for hatchery

conditions has been documented previously (Ford 2002) and perhaps could have contributed to a pattern of negative selection at the *Cox1* locus in oysters.

Another notable result from the MK tests were that N.I. = 0 for the mitochondrial genes *Atp6* and *Cytb*. While this could be interpreted to be the result of positive selection, there are other explanations as well. Strong functional constraint may have prevented the two loci from accumulating replacement substitutions (Meiklejohn et al. 2007), though non-significant *FL-D* and *FL-F* tests failed to support this explanation. Alternatively, N.I. = 0 at *Atp6* and *Cytb* could reflect a high level of mitochondrial divergence between *C. virginica* and *C. rhizophorae*. High mitochondrial divergence (relative to nuclear divergence) is expected if the mtDNA mutation rate is higher than the nuclear mutation rate or selective sweeps are differentially fixing mitochondrial haplotypes in one or both oyster species. If N.I. = 0 because of high mitochondrial divergence, then the use of a less distant outgroup (such as *C. virginica* from the Gulf of Mexico) could provide more polymorphic sites between groups and improve the test. MK tests were better suited for the nuclear loci because of their lower interspecific divergence (Table 4). In fact the neutrality index was undefined for the nuclear loci *Acadm* and *Gpi* because low divergence has precluded any fixed replacement substitutions. Most nuclear loci with defined values of N.I. were close to 1, indicating neutral evolution. The exception to this pattern was the nuclear locus *Usp* whose low N.I. is typical of loci under positive selection, though non-significant values of Tajima's *D* and F_S failed to support adaptive evolution. Alternatively, the mixed signals may have been due to limitations of individual tests of selection or statistical chance.

Murray and Hare (2006) demonstrated the difficulty in differentiating neutral stochasticity from natural selection when comparing genetic variation across loci. Mixed signs of selection in the nuclear data presented here may have reflected true heterogeneity in evolutionary processes, insufficient power for detecting actual patterns of natural selection, or could have indicated sensitivity of individual tests to violations of demographic assumptions. However, my analysis benefited from the use of a combination of tests that focus on different molecular signatures of selection. Tajima's *D*, Fu's F_S, *FL-D* and *FL-F*, and the HKA test were calculated from the distribution of alleles, while the MK test, N.I. and d_N/d_S ratio were calculated by comparing variability in different classes of mutations. The distinction is an important one because tests based on allelic distributions are powerful, but they can be sensitive to underlying Fisher-Wright assumptions; in contrast tests that make use of mutation classes are not constrained by demographic assumptions because they are independent of genealogical history (Nielsen 2001). I have detected non-neutral evolution of mitochondrial DNA with both types of tests, strengthening the conclusion that selection, not statistical chance or violations of demographic assumptions, shaped the mitochondrial patterns.

Mitochondria play an important role in osmoregulation for oysters and other bivalves (Zurburg and De Zwaan 1981; Powell et al. 1982; Paynter et al. 1985), perhaps contributing to the signal of positive selection at mtDNA loci found here. Oysters are incapable of regulating the salinity in their shells (Shumway 1996) and occasionally are exposed to extremes in salinity, so tolerance to salt stress appears to be a strongly selected trait (Butler 1949). When stressed by high salinity, bivalves accumulate high

levels of alanine in their hemolymph by inhibiting the mitochondrial pyruvate dehydrogenase complex (PDC) and reducing levels of pyruvate (Paynter et al. 1984; Ellis et al. 1985). The same pathway is used to reduce levels of alanine in hemolymph when salinity is low. In addition, the PDC plays an important metabolic role in hypoxic conditions by anaerobically converting carbohydrate-derived pyruvate into acetate (Kluytmans et al. 1978; Ho and Zubkoff 1982, 1983). The PDC may be a target of natural selection, particularly in stressful conditions, and could make a useful candidate for studying mitochondrial selection.

Other authors also have suggested that *C. virginica* mitochondrial DNA may be subject to positive selection. Inconsistent patterns of nuclear differentiation between Atlantic and Gulf of Mexico oyster populations (Buroker 1983b; Karl and Avise 1992; Hare and Avise 1996; McDonald et al. 1996; Hoover and Gaffney 2005; Murray and Hare 2006) are contrasted by dramatic differences in allele frequencies at mtDNA loci (Hare and Avise 1996). Ballard and Rand (2005) suggested that there may be strong selection for thermal adaptation in oysters, and that positive selection on mtDNA in response to temperature may explain the exaggerated mitochondrial divergence between Atlantic and Gulf of Mexico populations.

Effective Population Size

I have made the first multilocus estimate of evolutionary N_e for a *C. virginica* population based on a coalescent analysis of silent site sequence data. My estimate of $N_e = 1.6 \times 10^5$ could represent one of two scenarios. First, variation present in these data could predate

contemporary population structure. In this case the data suggest that an ancestral population with N_e on the order of 10^5 gave rise to the modern Chesapeake Bay oyster population. The presence of high ancestral polymorphism also would imply that neither sweepstakes reproduction nor bottlenecks have led to reductions in Chesapeake Bay N_e to levels observed in the Dabob Bay population of *C. gigas* (Hedgecock 1994). Alternatively, if the variation present in the data reflects current deme structure, then high levels of gene flow among demes may have introduced divergent haplotypes into the samples, artificially increasing the estimate of N_e (M. Kuhner, personal communication). A previous study suggests that the Chesapeake Bay oyster population is composed of demes (rivers or subestuaries) that exchange migrants in a stepping-stone fashion over evolutionary time (Rose et al. 2006). Maruyama and Kimura (1980) showed that when extinction is rare the effective size of a metapopulation is equal to the sum of its demes. If my estimate of N_e reflects current deme structure, then Chesapeake Bay may be made up of demes with N_e on the order of tens of thousands. However since coalescent theory predicts that alleles take on average $4N_e$ generations to coalesce, it appears that there has been insufficient time for deep genealogical splits to have developed between demes. Chesapeake Bay as it exists today was formed from rising sea levels after the Pleistocene (Hobbs 2004), so modern population structure must have developed within the past 10,000 years. With N_e on the order of tens of thousands, 10,000 years (or approximately 2,000 generations) is too short a time period for lineage sorting to have taken place. The data do not seem to represent current patterns of diversity, and therefore are likely to reflect ancestral patterns of variation and therefore a large ancestral effective population size.

Previous studies of *C. virginica* have reported dramatically different estimates of effective population size. Hedgecock (1994) used levels of heterozygosity in allozyme data collected by Buroker (1983b) to estimate N_e for *C. virginica* populations in the Atlantic ($N_e = 2.9 \times 10^5$) and the Gulf of Mexico ($N_e = 2.4 \times 10^5$). For Chesapeake Bay oysters allozymes were used to estimate $N_e = 14.9$ (Hedgecock 1994) and microsatellites were used to estimate $N_e = 1516$ (Rose et al. 2006). Each of these values of N_e apply to different temporal and spatial scales for which the parameters were calculated. For example, Hedgecock's (1994) estimates of Atlantic and Gulf of Mexico N_e were based on levels of heterozygosity, which reflect levels of identity-by-descent due to inbreeding. Hedgecock (1994) and Rose et al. (2006) estimated Chesapeake Bay N_e from the change in allele frequencies in temporally discrete samples; each of these measures of variance N_e reflected the effective size of the population only for the time interval and location sampled. In contrast, N_e calculated from the coalescent reflects much older patterns of genetic variation. It should be noted that changes in the mutation rate and generation time would have a linear effect on my coalescent estimate of N_e. A tenfold increase in mutation rate above what I have assumed would lead to an equivalent decrease in effective size to $N_e = 1.6 \times 10^4$. Similarly, a longer average generation length would directly decrease my estimate of N_e. I note that my estimate of generation time, 5 years, was longer than the 2 years listed for *C. virginica* in a previous publication (Rose et al. 2006). This approximation is justified by the fact that the coalescent framework reflects evolutionary patterns that evolved mostly before human intervention reduced the average age of oysters in Chesapeake Bay (Rothschild et al. 1994), and more closely matches historical estimates of generation length (Galtsoff 1964).

Low N_e/N ratios ($< 10^{-5}$), found in many marine invertebrates and fishes (Hedgecock 1994; Hauser et al. 2002; Árnason 2004; Hedrick 2005; Hoarau et al. 2005), are sometimes explained by sweepstakes reproduction. In Chesapeake Bay the census size of the oyster population was estimated at 3.1×10^9 (harmonic mean of estimates from 1998-2002; VIMS, CBOPE website) and may have been as much as 100 times larger in pre-industrial times (Newell 1988b; Rothschild et al. 1994). Based on my estimate of effective size and a historical census size of 3.1×10^{11} the pre-industrial N_e/N ratio was as small as 1.8×10^{-6}, suggesting that sweepstakes events may have contributed to a reduced effective size in Chesapeake Bay oysters. It does not appear that sweepstakes events are strong enough to have reduced effective population size to the order of tens or hundreds, however, so sweepstakes events are unlikely to be as extreme as Hedgecock (1994) estimated. Hedrick (2005) suggested that sweepstakes reproduction may be explained by increased standardized variance in offspring number due to large population size, a proposal supported by the trend in which N_e/N decreases with increasing census size (Pray et al. 1996). Another contributing factor to the low ratio may be variance in productivity among demes in a subdivided population (Turner et al. 2002), which can dramatically reduce N_e in a large population.

Gillespie (2000; 2001; 2004) described the process by which genetic draft supersedes drift in genomic regions of low recombination as rare alleles increase in frequency. In large populations, there is an increased probability of new favorable mutations and a higher probability that positive selection will lead to a selective sweep. In regions that do not experience recombination (like the mitochondrial genome), low variation is expected

for small populations subject to genetic drift, and for larger populations that undergo selective sweeps. An implication of sweepstakes reproduction—which appears to have taken place in oysters—is that effective population size can be very small, reducing the role of selection. However I have found indications that positive selection has determined patterns of mitochondrial variation. In oysters it appears that genetic draft can be an important evolutionary force shaping mitochondrial variation despite a low N_e/N ratio associated with sweepstakes reproduction.

Conclusion

I described a mitochondrial haplotype skew in Chesapeake Bay oysters similar to the pattern found in other oyster studies (Reeb and Avise 1990; Boom et al. 1994) and has previously been explained by strong reproductive skew (Hedgecock 1994; Eldon and Wakeley 2006). The pattern of excess high frequency and singleton mitochondrial haplotypes was unlikely to be caused by sweepstakes reproduction or a recent genetic bottleneck because nuclear loci as a whole did not show the effects expected from these demographic processes. Instead, non-neutral patterns of oyster mitochondrial nucleotide variation appeared to be due to a recent selective sweep. Because a high mutation rate has been rejected by a recent empirical study, I hypothesize that singletons are relatively new allelic variants of a mitochondrial haplotype that swept to fixation. The data supported the conclusions of Bazin et al. (2006b), who argued that positive selection is common in species with large N_e, and that mitochondrial diversity is governed by genetic draft. Additionally, I found that the historic effective population size in Chesapeake Bay oysters was on the order of 10^5, which was approximately six orders of magnitude smaller than the historic census size, perhaps consistent with sweepstakes events. Patterns

of nuclear variation indicated that sweepstakes events contributed to a low N_e/N ratio. At the same time, the data indicated that the effective population size was large enough for genetic draft to determine levels of mitochondrial variation. Thus, oysters show that genetic draft can be an important evolutionary force despite sweepstakes reproduction. These findings help frame our understanding of mitochondrial selection and sweepstakes reproduction in oysters and other marine invertebrates that have the potential for high variance in reproductive success.

Tables

Table 3-1. DNA markers used in Chapter 3.

Mitochondrial Loci

Locus	Description	Function
Atp6	ATP synthase, subunit 6	Oxidative phosphorylation
Cox1	Cytochrome C oxidase, subunit 1	Oxidative phosphorylation
Cox2	Cytochrome C oxidase, subunit 2	Oxidative phosphorylation
Cytb	Cytochrome B	Oxidative phosphorylation
Nad2	NADH dehydrogenase, subunit 2	Oxidative phosphorylation
Atp6 3' UTR	3' untranslated region of ATP synthase, subunit 6	Noncoding
Cytb 3' UTR	3' untranslated region of Cytochrome B	Noncoding

Nuclear Loci

Locus	Description	Function
Acadm	Acyl-coenzyme A dehydrogenase	Fatty acid metabolism
Ctsy	Cathepsin Y	Cysteine protease activity
Gapdh	Glyceraldehyde 3-phosphate dehydrogenase	Glycolysis
Gpi	Glucose-6-phosphate isomerase	Glycolysis
Hadh	Short chain 3-hydroxyacyl-CoA dehydrogenase	Oxidoreductase activity
Usp	Universal stress protein	Unknown
Vg	Vitellogenin	Glyco-lipo-protein

Table 3-2. Primer information.

Primer	Product(s)	Species	DNA in PCR	Annealing Temperature	Primer Length	Primer Sequence (5' to 3')	Fragment Size (bp)	Notes
mt118	*Atp6*, *Atp6* 3' UTR, *Nad2*	*C. virginica*	genomic	55°C	24	CTAGAGAAGGAACCGGATGAGTGT	624	
mt119					24	TGAAATTAGTAAAGCGCCATAATG		
DegCras9785F	*Atp6*	*C. rhizophorae*	genomic	57°C	22	CCTCRTGAGARRTYGTGGCKGG	~355	
DegCras10156R					23	CCCARAAACARAGARCTTTGGAC		
LCO 1490	*Cox1*	*C. virginica* / *C. rhizophorae*	genomic	55°C	25	GGTCAACAAATCATAAAGATATTGG	~785	Primers described in Folmer et al. 1994
HCO 2198					26	TAAACTTCAGGGTGACCAAAAAATCA		
mt120	*Cytb*, *Cytb* 3' UTR, *Cox2*	*C. virginica*	genomic	55°C	24	TAATGCGGGATGCCAATTATGGAT	567	
mt121int					18	CGTAATCAATGGCGTCTC		
DegCras4445F	*Cytb*, *Cytb* 3' UTR, *Cox2*	*C. rhizophorae*	genomic	60°C	23	CGYGTWATTAAGCCTGAGTGATA	~1085	
DegCras5534R					23	ACTTCWACWCGAATYGGCATRAA		
Cv862L	*Acadm*	*C. virginica* / *C. rhizophorae*	cDNA	60°C	23	ACTCCAGGGATCACAGTGGGCAG	543	Developed from GenBank EST Accession # CD648862
Cv862R					21	TCTCTGAATCTGGGCTGTTCC		
Cv036L	*Ctsy*	*C. virginica* / *C. rhizophorae*	cDNA	60°C	22	TCAATACTCTACCCGCAAGCTG	602	Developed from GenBank EST Accession # CD649036
Cv036R					21	AGTACTCTGTTTTGGTGGCTG		
Cv312L	*Gapdh*	*C. virginica* / *C. rhizophorae*	cDNA	58°C	22	TGTTCTAAGGGCCGCACTGGAC	670	Developed from GenBank EST Accession # CD648312
Cv312R					22	AGAGACGTCTGGAACTGGCACG		
Cv019L	*Gpi*	*C. virginica* / *C. rhizophorae*	cDNA	56°C	21	TCAAGTCTGGAATGTCGCAGG	343	Developed from GenBank EST Accession # CD649019
Cv019R					21	ACAGACTCGATTTGGGGGCGT		
Cv179L	*Hadh*	*C. virginica* / *C. rhizophorae*	cDNA	58°C	21	AGGTATTGCTCAGGTTGCTGC	630	Developed from GenBank EST Accession # CD648179
Cv179R					22	CCTAGCTTCATAGCAGTGTCCA		
Cv124L	*Usp*	*C. virginica* / *C. rhizophorae*	cDNA	59°C	22	ACGCGTCCGCTTACACAGCGAG	485	Developed from GenBank EST Accession # CD649124
Cv124R					23	TCTAATCATGAATCGCGGACCAC		
Cv526L	*Vg*	*C. virginica* / *C. rhizophorae*	genomic	58°C	22	TCGTGGGTGTCTAGCTTGGTGG	519	Developed from GenBank EST Accession # CD647526
Cv526R					21	TGACGTTGCTCTTCGAGGCAC		

Table 3-3. Number of chromosome sequences determined for each locus and from each sampling site. *C. virginica* were collected from sites in Chesapeake Bay, USA and *C. rhizophorae* samples were collected from Boquerón Bay, Puerto Rico.

	Mitochondrial Loci							**Nuclear Loci**						
	Atp6	*Cox1*	*Cox2*	*Cytb*	*Nad2*	*Atp6* 3' UTR	*Cytb* 3' UTR	*Acadm*	*Ctsy*	*Gapdh*	*Gpi*	*Hadh*	*Usp*	*Vg*
C. virginica														
Drum Point	12	4	13	13	14	13	14	-	-	-	-	-	-	-
Deep Neck	13	1	14	13	14	13	14	-	-	-	-	-	-	-
Virginia	19	-	20	20	20	19	20	-	-	-	-	-	-	-
Bolingbroke Sands	-	13	-	-	-	-	-	27	22	26	24	15	28	26
C. rhizophorae	2	2	4	4	-	-	-	4	4	6	4	3	3	2
Total	**46**	**20**	**51**	**50**	**48**	**45**	**48**	**31**	**26**	**32**	**28**	**18**	**31**	**28**

Table 3-4. Indices of nuclear diversity. The first 7 loci are mitochondrial markers, and the last 7 are nuclear markers. *bp*=base pairs, *S*=number of segregating sites, P_S=proportion of variable sites, $\hat{\theta}_\omega$=per site heterozygosity based on variable sites, $\hat{\theta}_\pi$=per site heterozygosity based on average number of nucleotide substitutions per site

		C. virginica					*C. rhizophorae*					**Interspecific**
locus	**sites**	***bp***	***S***	P_S	$\hat{\theta}_\omega$	$\hat{\theta}_\pi$	***bp***	***S***	P_S	$\hat{\theta}_\omega$	$\hat{\theta}_\pi$	**Divergence**
Atp6	all	127	2	0.016	0.004	0.001	79	0	0.000	0.000	0.000	0.177
	silent	47	2	0.043	0.010	0.004						0.564
	replacement	80	0	0.000	0.000	0.000						0.035
Cox1	all	658	3	0.009	0.003	0.001	658	0	0.000	0.000	0.000	0.169
	silent	233	3	0.026	0.008	0.003						0.621
	replacement	425	0	0.000	0.000	0.000						0.021
Cox2	all	216	5	0.029	0.006	0.002	216	0	0.000	0.000	0.000	0.120
	silent	73	4	0.069	0.016	0.004						0.439
	replacement	143	1	0.009	0.002	0.000						0.030
Cytb	all	348	7	0.020	0.005	0.001	307	0	0.000	0.000	0.000	0.202
	silent	121	6	0.050	0.011	0.002						0.541
	replacement	227	1	0.004	0.001	0.000						0.098
Nad2	all	411	10	0.026	0.006	0.001	-	-	-	-	-	-
	silent	145	9	0.065	0.015	0.003						-
	replacement	266	1	0.004	0.001	0.000						-
Atp6 3' UTR	all	108	8	0.077	0.018	0.005	-	-	-	-	-	-
Cytb 3' UTR	all	70	2	0.029	0.006	0.001	-	-	-	-	-	-

Table 3-4, continued.

		C. virginica					*C. rhizophorae*					**Interspecific**
locus	**sites**	***bp***	***S***	P_S	$\widehat{\theta}_\omega$	$\widehat{\theta}_\pi$	***bp***	***S***	P_S	$\widehat{\theta}_\omega$	$\widehat{\theta}_\pi$	**Divergence**
Acadm	all	498	23	0.047	0.012	0.007	498	10	0.020	0.011	0.012	0.025
	silent	163	21	0.134	0.035	0.021						0.103
	replacement	335	2	0.006	0.002	0.000						0.001
Ctsy	all	555	21	0.038	0.010	0.009	555	0	0.000	0.000	0.000	0.040
	silent	182	19	0.106	0.029	0.026						0.150
	replacement	373	2	0.005	0.001	0.001						0.009
Gapdh	all	624	6	0.011	0.003	0.002	624	2	0.004	0.002	0.002	0.033
	silent	205	5	0.028	0.007	0.005						0.119
	replacement	419	1	0.003	0.001	0.000						0.005
Gpi	all	297	18	0.109	0.019	0.016	297	1	0.003	0.002	0.002	0.034
	silent	105	12	0.130	0.035	0.029						0.108
	replacement	192	6	0.036	0.010	0.009						0.010
Hadh	all	585	12	0.021	0.006	0.007	585	2	0.005	0.003	0.003	0.048
	silent	207	11	0.054	0.017	0.019						0.128
	replacement	378	1	0.003	0.001	0.000						0.025
Usp	all	438	12	0.027	0.008	0.006	438	0	0.000	0.000	0.000	0.055
	silent	146	11	0.076	0.019	0.018						0.182
	replacement	292	1	0.003	0.001	0.000						0.018
Vg	all	547	12	0.030	0.008	0.003	547	3	0.006	0.006	0.006	0.034
	silent	262	10	0.069	0.018	0.007						0.102
	replacement	285	2	0.008	0.002	0.001						0.012

Table 3-5. Tests of selection without an outgroup. Combining mitochondrial loci, both Tajima's *D* and Fu's F_S were highly significant ($p < 0.001$). *$0.01 < p < 0.05$, **$0.001 < p < 0.01$, ***$p < 0.001$

		Locus	**Tajima's *D***	**Fu's F_S**
Mitochondrial DNA	coding	*Atp6*	-1.14	**-1.67***
		Cox1	-1.714	**-2.60****
		Cox2	**-1.82***	**-5.34*****
		Cytb	-2.18	**-5.23*****
		Nad2	**-2.30****	**-10.75*****
	noncoding	*Atp6* 3' UTR	**-1.99***	**-8.02*****
		Cytb 3' UTR	-1.47	-2.98
Nuclear DNA		*Acadm*	-1.52	-13.29
		Ctsy	-0.52	-0.32
		Gapdh	-1.18	**-3.20***
		Gpi	-0.54	-4.80
		Hadh	0.29	-2.98
		Usp	-0.65	-6.59
		Vg	**-2.08***	-2.35

Table 3-6. Tests of selection with an outgroup. d_S = rate of synonymous substitutions , d_N = rate of nonsynonymous substitutions, d_N/d_S = ratio of nonsynonymous to synonymous substitution rate, *FL-D* = Fu and Li's D, *FL-F* = Fu and Li's F, N.I. = Neutrality Index. *$0.01 < p < 0.05$

	Locus	**d_S**	**d_N**	**d_N/d_S**	***FL-D***	***FL-F***	**N.I. (McDonald-Kreitman Test)**
Mitochondrial DNA	*Atp6*	0.564	0.035	0.035	-1.83	-1.89	0 (no polymorphic replacement substitutions)
	Cox1	0.621	0.020	0.016	**-2.51***	**-2.69***	5.10
	Cox2	0.439	0.030	0.046	-0.11	-0.69	1.00
	Cytb	0.540	0.098	0.110	-0.10	-0.76	0 (no polymorphic replacement substitutions)
Nuclear DNA	*Acadm*	0.103	0.001	0.013	-2.02	-2.35	undefined (no fixed replacement substitutions)
	Ctsy	0.150	0.009	0.057	-0.61	-0.71	0.78
	Gapdh	0.119	0.005	0.042	-2.04	-2.09	1.20
	Gpi	0.114	0.010	0.083	-0.94	-0.93	undefined (no fixed replacement substitutions)
	Hadh	0.128	0.025	0.182	-0.14	-0.01	0.82
	Usp	0.182	0.018	0.088	-0.64	-0.98	0.21
	Vg	0.106	0.012	0.107	**-2.46***	**-2.80***	1.07

Table 3-7. Hudson, Kreitman and Aguadé (HKA) tests of selection. I report χ^2 values above the diagonal and p values below the diagonal. * $0.01 < p < 0.05$, ** $0.001 < p < 0.01$

	Atp6	*Cox1*	*Cox2*	*Cytb*	*Acadm*	*Ctsy*	*Gapdh*	*Gpi*	*Hadh*	*Usp*	*Vg*
Atp6	-	0.096	**5.048***	0.013	**4.488***	2.005	0.049	**7.767****	0.821	0.340	1.473
Cox1	0.7566	-	2.127	0.194	**5.842***	3.069	0.283	**8.502****	1.158	0.868	2.561
Cox2	**0.0247***	0.1447	-	1.464	1.800	0.266	0.806	**4.026***	0.007	0.292	0.100
Cytb	0.9105	0.6594	0.2263	-	**6.351***	2.899	0.023	**9.858****	1.096	0.402	2.136
Acadm	**0.0341***	**0.0156***	0.1797	**0.0117***	-	0.567	3.737	0.258	1.499	2.954	0.730
Ctsy	0.1568	0.0798	0.6060	0.0886	0.4516	-	1.683	1.603	0.281	0.987	0.022
Gapdh	0.8242	0.5948	0.3693	0.8794	0.0532	0.1945	-	**5.615***	0.594	0.165	1.250
Gpi	**0.0053****	**0.0035****	**0.0448***	**0.0017****	0.6115	0.2054	**0.0178***	-	2.858	**4.873***	1.779
Hadh	0.3648	0.2095	0.9339	0.2952	0.2208	0.5964	0.4409	0.0909	-	0.174	0.137
Usp	0.5598	0.3516	0.5890	0.5262	0.0857	0.3205	0.6849	**0.0273***	0.6766	-	0.653
Vg	0.2250	0.1095	0.7519	0.1439	0.3929	0.8809	0.2635	0.1822	0.7110	0.4190	-

Figures

Figure 3-1. *C. virginica* mitochondrial haplotype frequencies in Chesapeake Bay, rank ordered from highest to lowest frequency. Black bars are observed frequencies and gray bars are expected frequencies based on Ewens's (1972) sampling distribution. Samples were simulated 10,000 times to provide standard deviations for the expected frequencies.

Figure 3-2. Haplotype networks for mitochondrial and nuclear loci. Open circles represent observed haplotypes, with oval area representing haplotypic frequency. Solid dots signify hypothetical haplotypes. Each connection between haplotypes corresponds to a single mutation.

Figure 3-3. The log-likelihood surface of mitochondrial sequence data under the multiple-merger framework. Contour lines are drawn every two log-likelihood units from the maximum, indicated by an "x". The maximum likelihood was shared by one value of ψ (0.10), the proportion of the population that a single individual's offspring replaces, and two values of θ (0.056 and 0.067).

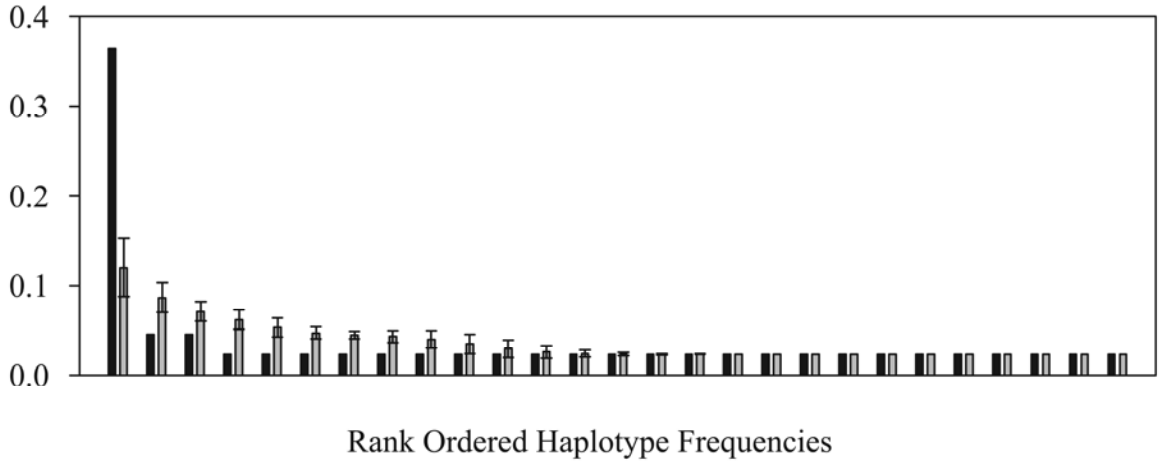

Figure 3-1

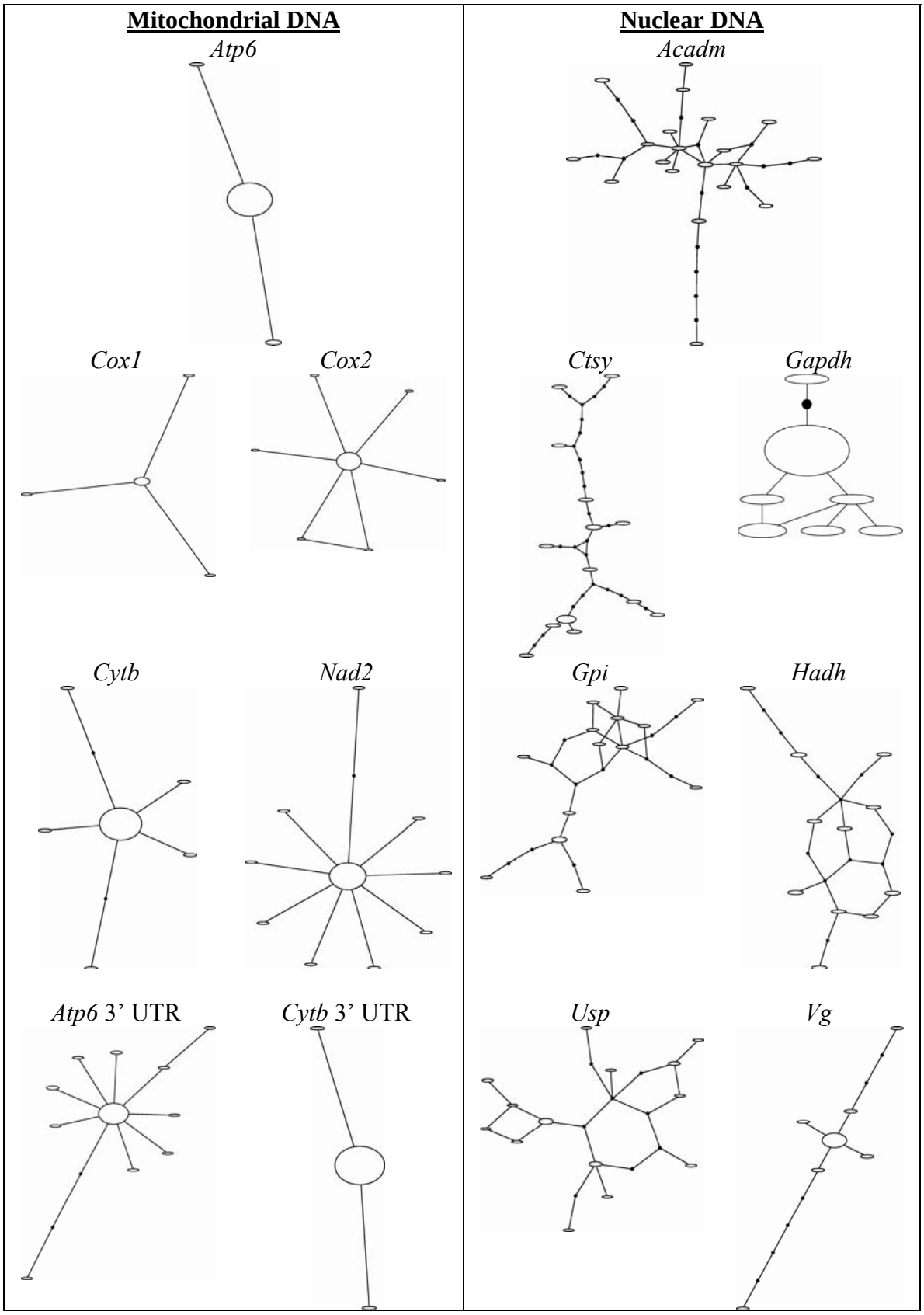
Mitochondrial DNA
Nuclear DNA
Atp6
Acadm
Cox1
Cox2
Ctsy
Gapdh
Cytb
Nad2
Gpi
Hadh
Atp6 3' UTR
Cytb 3' UTR
Usp
Vg

Figure 3-2

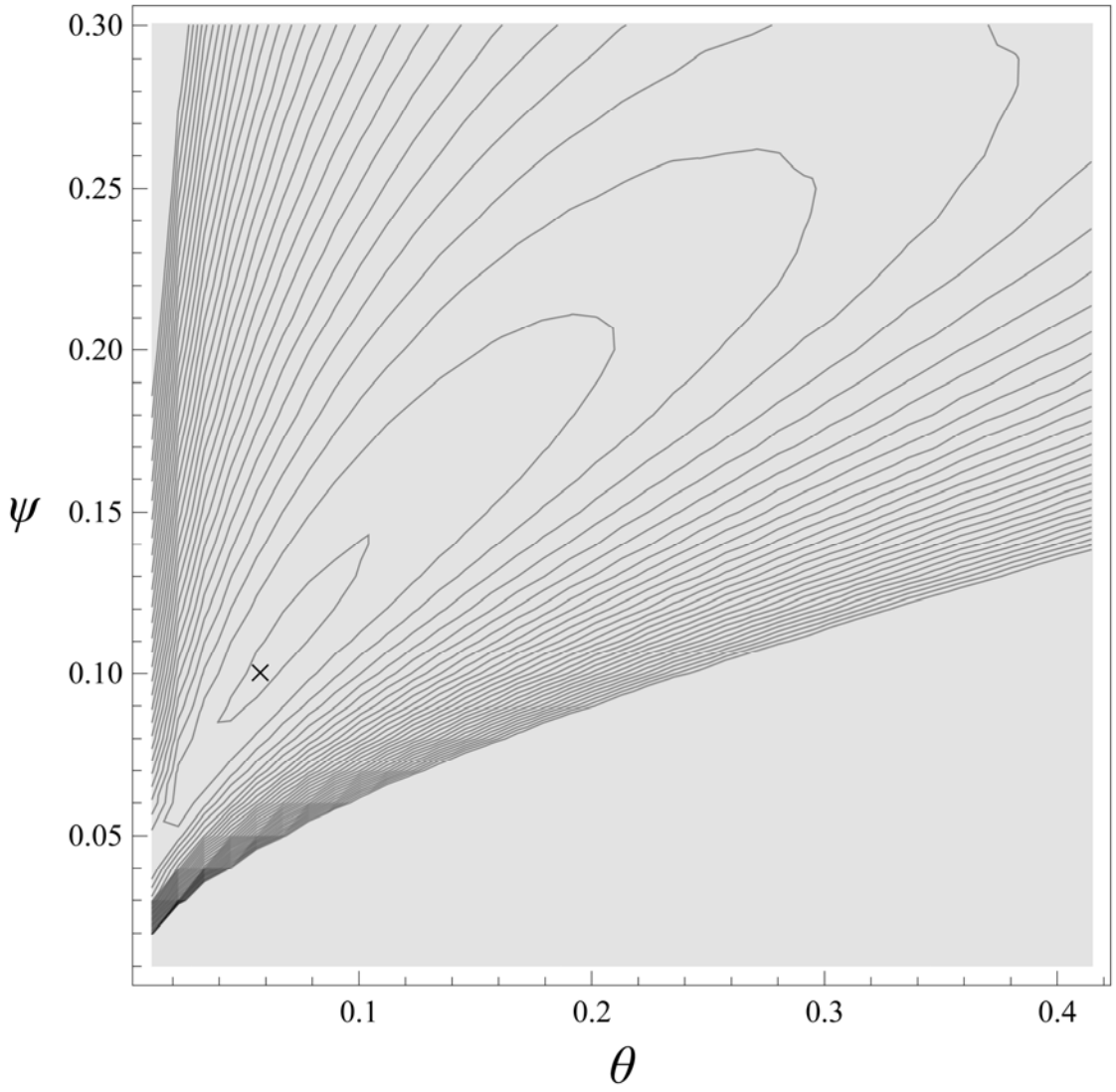

Figure 3-3

Chapter 4: Supportive Breeding and the Risk of Genetic Bottleneck

Introduction

Oysters are intensively managed worldwide because of their economic importance, but nevertheless have a history of fisheries collapse (Kirby 2004). The eastern oyster (*Crassostrea virginica*) has traditionally played an important role in the ecosystem of Chesapeake Bay (Jackson et al. 2001), North America's largest estuary but overfishing, habitat loss and disease have caused the Chesapeake Bay oyster population to shrink to approximately 1% of its historic size (Newell 1988a; Rothschild et al. 1994). In response, hundreds of millions of dollars have been invested by U.S. federal, state and private organizations to support the oyster population and the oyster fishing industry (Mann and Powell 2007). With the goal of population restoration, managers have used supportive breeding to supplement the wild Chesapeake Bay population with large numbers of hatchery-raised, disease tolerant oysters (Hare et al. 2006), but the potential effect of this supplementation remains poorly understood. Here I present genetic data indicating that the hatchery-raised oysters have a small effective population size compared to wild oysters. In the Great Wicomico River, one of the tributaries in Chesapeake Bay where selected hatchery oysters have already been released, I estimated that a single generation of hatchery supplementation (supportive breeding) may lower the wild effective population size from on the order of 5,000 to 2,500 and that the continued use of a

"closed" hatchery line can potentially cause a severe genetic bottleneck. After three years of supplementation the wild effective population size remained high, suggesting that hatchery oysters have not yet contributed substantially to wild reproduction. These data suggest that even expertly managed supplementation programs can have unintended, adverse consequences for wild populations, and indicate the importance of management that explicitly accounts for the genetic properties of populations.

Discussion

The health of a population cannot be judged solely by how large it appears. Population genetics distinguishes between census size—the total number of living organisms—and the genetic effective population size, a measure that takes into account the average level of inbreeding (specifically the rate of change in average homozygosity, which increases as population size decreases) (Crow 1954). The effective size is often much smaller than the census size in wild populations (Frankham 1995a), but a population risks extinction when effective size becomes too small (Hedrick and Kalinowski 2000). Reduced effective size can lead to a genetic bottleneck that lowers average fitness and fecundity, increases inbreeding and homozygosity, and reduces the amount of genetic variation in a population reducing its ability to adapt to new stressors (Ryman and Laikre 1991; Lande and Shannon 1996; Araki et al. 2007). Maintaining or raising the effective size of a population is therefore a top priority of population managers (Templeton and Read 1983).

Oyster fisheries are managed worldwide, but populations are still at risk of collapse due to overfishing and habitat degradation. Despite regulation, fishery declines have taken

place for *Ostreola conchaphila* along North America's Pacific coast (Kirby 2004), *Crassostrea virginica* along North America's Atlantic coast (MacKenzie 1996), *Saccostrea glomerata* along Australia's eastern coast (Kirby 2004), and *Ostrea edulis* in Europe (Lapègue et al. 2007). Supportive breeding programs have been employed for many oyster species, though low effective population size has been found in hatchery populations (Saavedra 1997; Launey et al. 2001). Use of hatchery lines with low effective population sizes could lower the effective size of wild populations, putting them in danger of extinction. Quantifying any changes in effective size of wild populations will be important for evaluating the benefits and risks of supportive breeding programs.

In Chesapeake Bay the eastern oyster (*C. virginica*) has been managed for at least a century. For more than 100 years oyster fishermen have relocated oysters to regions of the Bay that were more favorable for growth (Carlton and Mann 1996; Mann and Powell 2007). Fishermen also return large oysters to the Bay through state run buyback programs and recently managers have begun hatchery supplementation of the wild population (Figure 4-1). In recent years supportive breeding has taken place in greater numbers and with hatchery strains that have been selected for tolerance to disease (Ragone Calvo et al. 2003; Hare et al. 2006), making the deployment of selected hatchery lines an important part of the oyster restoration strategy (Figure 4-1). Since 2003 selected hatchery oysters have been used exclusively for supplementation in Virginia regions of Chesapeake Bay, though Maryland's larger supplementation program has mostly used hatchery lines bred from wild broodstock that have not been selected for disease tolerance. Estimates of oyster effective population size do not exist for hatchery populations of *C. virginica* and

vary considerably for wild Chesapeake populations (Hedgecock 1994; Rose et al. 2006), so it is unknown how supportive breeding will influence the effective population size of wild oysters.

Here I investigate the effective sizes of the wild oyster population in the Great Wicomico River, found on Virginia's western shore of Chesapeake Bay, and the disease tolerant hatchery line that was used for its supplementation. In 2002 the Great Wicomico River was home to nearly 35 million oysters and was supplemented with 790,000 hatchery-raised oysters selectively bred for disease tolerance (Hare et al. 2006). I determined genotypes for 7 unlinked microsatellite loci from 82 hatchery-raised oysters from the 2002 selected hatchery line, from 732 wild Great Wicomico River juvenile oysters collected in 2002, and 571 wild Great Wicomico juveniles collected in 2005. Using these samples, I estimated the number of breeding adults (N_b) using gametic disequilibrium (the nonrandom association of alleles at unlinked loci due to inbreeding, sometimes called linkage disequilibrium). N_b, which is related to inbreeding effective population size (Laurie-Ahlberg and Weir 1979), was estimated from genotypic correlations (r^2). Theoretical predictions of inbreeding effective population size indicate that two generations are necessary to show the effects of identity-by-descent, but estimates of N_b from allelic correlations take only one generation to indicate a reduction in effective size (Waples 2005). Juvenile oysters from 2005 therefore reflected the effective size of the population in 2003, and represented post-supplementation patterns of genetic variation. Assuming Wright-Fisher conditions, N_b can be converted to contemporary per-generation effective size by multiplying generation length (in years) by the harmonic mean of N_b

across generations (Waples 2005). Because interannual variation in N_b is unknown, I have made the simplifying assumption that N_b approximates the per-generation effective population size.

My analysis indicates that in 2002 the hatchery-raised oysters had N_b = 12 (95% confidence interval: 6-21), and the wild population had N_b = 4,925 (95% confidence interval: 1,578-infinity). I also estimate that in 2005 the wild population had N_b = 11,677 (95% confidence interval: 3,882-infinity). An earlier study used assignment tests to determine hatchery contribution to wild reproduction in the Great Wicomico River in 2002 (Hare et al. 2006); after accounting for the false discovery rate I estimated hatchery contribution at 5% (see Supplemental: Hatchery Contribution).

Given the large disparity between hatchery and wild effective population sizes and the relatively large contribution that hatchery oysters contribute to reproduction, what are the potential effects of supplementation? I used the empirical estimates of 2002 N_b and the best population models available to predict the effects of supportive breeding. First I investigated the effect of a single generation of supportive breeding beginning at generation t using the Ryman-Laikre model (Ryman and Laikre 1991). Given a hatchery population with effective size of 12, I predicted that a 5% contribution to reproduction (hatchery oysters produce 1.8×10^6 offspring at generation t+1) would lower wild effective size from 4,925 to 2,587, a 47% drop at generation t+2 (Figure 4-2, solid black line). However, the severity of predicted bottlenecks increases with greater hatchery contribution to reproduction (Figure 4-3). Because inputs of selected hatchery oysters

have increased dramatically since 2002, there is a strong possibility that hatchery contribution to wild reproduction also could increase. If the hatchery oysters were to double their reproductive output and to contribute 10%, I predict that the wild effective size would instead fall to 1,023, a 79% decline in a single generation (Figure 4-2, dashed black line).

I also modeled the effects that sustained use of the hatchery population would have on the wild population with the Wang-Ryman model (Wang and Ryman 2001) under two scenarios. In the first case the hatchery population is "open" and wild oysters are randomly sampled for use in the hatchery line. In the second scenario, hatchery oysters are bred in a "closed" line that is genetically isolated from the wild population. Under both scenarios I predicted a dramatic decrease in wild effective size if newly introduced hatchery oysters continue to contribute 1.8×10^6 offspring each generation (initial contribution of 5%, which decreases proportionally as the population grows) (Figure 4-4). With an open hatchery line I predicted that the effective size would fall 82% to 891 at generation t+18 before slowly recovering. I calculated the harmonic mean effective population size and found that the effective population size would remain below the pre-supplementation value for 1,332 generations. If on the other hand, the hatchery line is closed then the wild effective population size would drop 95% to 245 at generation t+2 (Figure 4-4), and will not reach a pre-supplementation effective size until there have been 325 generations of successful supportive breeding. These recovery predictions were based on the assumptions that the wild oyster census size would remain stable without supplementation, and that supplementation would increase census size each generation.

Any asymptote or reduction in census size would lower the wild effective population size and lead to a longer recovery time.

Although my estimate of N_b was 12, managers bred oysters in the hatchery with a goal of producing a line with an effective population size of 50 (S. Allen, personal communication). I modeled the genetic consequences for supportive breeding using this target effective population size. With a hatchery effective size of 50 and a 5% contribution to reproduction, a single generation of supplementation to the wild Great Wicomico population would be expected to decrease the effective size by 13% from 4,925 to 4,287 at generation t+2 (Figure 4-2, solid gray line). A single generation in which hatchery oysters with effective size of 50 contribute 10% to the total population would decrease the wild effective size by 44% from 4,925 to 2,744 (Figure 4-2, dashed gray line). If the supportive breeding program were to continue at the target hatchery effective population size of 50 with 5% contribution to the wild population, I predict that the wild effective size would drop to a minimum of 3,067 at generation t+13, down 38% from its initial value (Figure 4-4). The harmonic mean of the effective size indicated that the wild population would have a reduced effective population size until generation t+139, or 278 years assuming a two-year generation time (Hedgecock 1994).

Risk of extinction due to accumulation of mildly deleterious alleles, known as mutational meltdown, increases as effective population size shrinks (Frankham 1995b; Lynch et al. 1995). However the magnitude of effective size reduction that constitutes a dangerous genetic bottleneck is debatable (Franklin 1980; Soulé 1980; Lande 1995). Since oyster

populations tend to harbor a large number of deleterious recessive alleles (Bierne et al. 1998; Launey and Hedgecock 2001; Yu and Guo 2003) they may be especially vulnerable to extinction by mutational meltdown following a bottleneck (van Oosterhout et al. 2000). While I cannot precisely predict the degree to which a drop in effective population size increases risk of local extinction, simulations indicate that an effective size of 500 is too small for the maintenance of adaptive variation (Lande 1995). Following successful supplementation with the closed hatchery line (as measured by measurable reproductive contribution), I predict that the oyster population would have a harmonic mean effective population size below 500 until generation t+21.

Models forecast a dramatic and potentially dangerous drop in effective population size after the F1 hybrid oysters (those with wild and hatchery parents) begin randomly mating in the wild. Given a two-year generation length, the models predicted a reduction in wild effective population size by 2005. Contrary to these predictions, samples collected in 2005 indicated that N_b in Great Wicomico River was as large as—or larger than—what it was in 2002. Two scenarios may account for the maintenance of a large wild effective population size in spite of supplementation with a small closed hatchery line. First, Juvenile F1 oysters detected in 2002 may have died before reproduction, or survived but failed to reproduce. If juvenile oysters in 2002 were subject to high levels of predation, then F1 oysters may have been eliminated before they reached reproductive age. Genetics could also play a role in reduced survivorship or fecundity. It is not uncommon for hatchery organisms to have low fitness in the wild because of inadvertent selection for hatchery conditions or accumulation of deleterious alleles (Lynch and O'Hely 2001; Ford

2002; Araki et al. 2007). Also, F1 oysters could conceivably have low fitness due to harmful epistatic interactions between wild and hatchery alleles. Second, the juvenile F1 oysters detected in 2002 may eventually contribute to wild reproduction, but have not yet done so. Under ideal circumstances (density of 250 oysters per m^2, equal sex ratio, complete mixing of gametes, adult mortality of 15% per year), simulations suggest that maximum reproductive potential is not reached until oysters are six years old (K. Paynter, personal communication); thus, three years may have been an insufficient amount of time to have allowed F1 oysters to fulfill their reproductive potential. If so, the sample from 2005 could have been collected too soon after supplementation to demonstrate an impending bottleneck. Simulations have shown that age-dependent fecundity can lead to uneven patterns of inbreeding, particularly at the beginning of a restoration program (Waples and Do 1994), which could have obscured the genetic signature of a bottleneck.

Despite its importance for interpreting the effects of supportive breeding, the estimation of effective population size is not a straightforward exercise. Recent work in population genetic theory has improved gametic disequilibrium estimators of effective size from single generation samples (Hill 1981; Waples and Do 2008). My estimates relied on the expectation that gametic disequilibrium between neutral, unlinked loci should increase as effective size decreases. Nevertheless, assumptions underlying my estimates of effective size may have deviated slightly from the model expectations; I estimated the number of breeding adults producing the sampled cohort and assumed that patterns are representative of typical reproduction. Based on this assumption I approximated the per-generation effective population size in the Great Wicomico River. Another estimation

method, however, makes use of samples collected from multiple generations of a focal population to calculate an estimate of effective size that is integrated across generations (Vucetich and Waite 1998). A previous study used this temporal sampling method to estimate that the effective size of oysters in James River, a Chesapeake Bay tributary near the Great Wicomico River, is 1,516 (Rose et al. 2006). The difference between the temporal estimate of effective size and my estimate could be due to variation in effective size between tributaries, or could indicate that the number of breeding adults is occasionally very low. On the other hand there is reason to believe that the effective size estimated by Rose et al. (2006) could be downwardly biased because the sampling scheme violated several assumptions of the temporal estimation method. First, sample sizes may have been too small to precisely estimate allele frequencies. Second, the estimate did not account for sampling in populations with overlapping generations of adults (Waples 2005). Third, allele frequencies may have been affected by migration from outside populations in the period between the sampling points. However, even accepting this smaller estimate of effective population size, hatchery supplementation was still expected to have an adverse effect on the wild effective population size. Assuming that a hatchery with an effective size of 12 contributes 5% to wild reproduction, a single generation of supplemental breeding would reduce a population with an effective size of 1,516 by 17% to 1,252 (Figure 4-2). If the same hatchery population were instead to contribute 10% the wild effective size would shrink by 49% to 742.

I recognize that my estimates of effective size are subject to a great deal of uncertainty due to the stochastic nature of genetic drift, and due to the underlying assumptions of the estimates. In particular, it has been shown that oyster microsatellites are subject to segregation distortion, perhaps due to natural selection (Launey and Hedgecock 2001; Reece et al. 2004). If natural selection has acted on the unlinked loci that I used here, there is the possibility that it created gametic disequilibrium (Allendorf 1983) and downwardly biased my estimates of effective population size. Even with this level of uncertainty, the discrepancy between the large effective size in the wild Great Wicomico population and the small effective size of the hatchery population could make the current supplementation strategy risky.

Given my best approximations of hatchery population parameters (5% contribution to wild reproduction, effective size 12), I have determined that any wild oyster population with an initial effective size greater than 480 will experience a drop in effective population size after successful supplementation. However, if the hatchery stock has a sufficiently large effective size, it is not expected to reduce the wild effective population size. In the Great Wicomico River managers can maintain the wild effective population size at 4,925 if they use a hatchery stock with effective size of 126 (assuming 5% hatchery contribution) or 259 (assuming 10% contribution). This would be a difficult task to achieve because high variance in parental contribution can dramatically reduce a hatchery's effective population size (Launey et al. 2001; Appleyard and Ward 2006; Petersen et al. 2008). However, risk of extinction is expected to increase as the hatchery effective size drops below these thresholds.

The impact that a supplementation program would have on non-target populations is determined by genetic connectivity. Other studies of oysters have shown that there are barriers to dispersal between tributaries in Chesapeake Bay (Rose et al. 2006; Mann and Powell 2007; North et al. 2008), meaning that a genetic bottleneck in the Great Wicomico River will not immediately affect other populations. A pattern of isolation by distance between bay tributaries (Rose et al. 2006) indicates that relatively few migrants are exchanged between Chesapeake Bay populations. This conclusion is substantiated by studies suggesting that oyster larval transport may be limited to within tributaries of Chesapeake Bay due to retentive hydrodynamic forces (Southworth and Mann 1998) and larval behavior that minimizes dispersal (Tankersley et al. 1995; Dekshenieks et al. 1996; Finelli and Wethey 2003; North et al. 2008). Assuming that larval transport is indeed limited, then few oysters would emigrate from supplemented populations in any given generation. Any negative consequences of supplementation would then be localized to supplemented rivers in the short term, while other populations would be relatively safe from abrupt, severe bottlenecks. However, the effective size in other populations would slowly fall—and bottlenecks in supplemented populations would be ameliorated—as migration-genetic drift is reached within the metapopulation. As migrants are exchanged over hundreds and thousands of generations, their cumulative effect might induce bottlenecks in populations that were never supplemented with hatchery oysters. Thus a supportive breeding program that targets even a few tributaries could reduce the effective size of the entire Chesapeake Bay oyster population over a long time scale.

Supplementation of wild fish and shellfish populations with hatchery-raised stocks is becoming increasingly common, but supportive breeding is not without its risks. My results mirrored those from other studies of fish and shellfish showing that hatchery-raised populations have smaller effective sizes than the wild populations that they support (Bartley et al. 1992; Hedrick et al. 1995; Launey et al. 2001); here I have provided a quantitative prediction of the risk that this scenario presents to a wild population. Use of a closed hatchery line would not be harmful if the hatchery oysters do not reproduce in the wild, but their reproductive success could mean reduced effective size for hundreds of years, potentially leading to protracted, detrimental consequences. Furthermore the numbers of hatchery oysters introduced to the wild population have increased dramatically since 2002, so these simulations may have underestimated the potential severity of bottlenecks. Three years after the implementation of a supportive breeding program, effects of a bottleneck are not yet evident. While it is unclear if wild effective population size remains high because hatchery oysters have failed to survive or failed to reproduce, this could be an opportunity to change the supplementation program before long-term damage is done. This case is an extreme example of how supportive breeding can harm a wild population, and reflects the discrepancy between the extremely large wild effective size and low hatchery effective size, and the potential for large variance in reproductive success in oysters (Hedrick 2005; Petersen et al. 2008). If supportive breeding is to continue in Chesapeake Bay, resources should be invested to dramatically increase hatchery effective sizes and to make use of open hatchery lines. The use of an open hatchery line bred from many pairwise crosses of wild oysters rather than from mass spawns might raise hatchery effective size, thereby mitigating the risk of a severe

bottleneck. A truly successful supportive breeding program will be one that increases the census size of the wild population without reducing its effective population size. These data serve as a reminder to conservation managers that careful planning should be done to prevent supportive breeding from placing wild populations at risk.

Supplementary Material

Sample Collection and DNA Analysis

Oysters were collected from the "DEBY" hatchery line that was selected for disease tolerance and from wild oysters. The "primary" DEBY line used to breed working hatchery lines was sampled in 2005 (adult samples were provided by Stan Allen, Virginia Institute of Marine Science). DEBY oysters were collected from the 2002 Great Wicomico River planting and from Little Choptank River plantings in 2002 and 2004. Wild juvenile oysters were collected from the Great Wicomico River in 2002 and 2005. Oysters were genotyped for the microsatellite loci *Cvi*9, *Cvi*12, *Cvi*1i24b, *Cvi*2g14, *Cvi*2i23, *Cvi*2j24, *Cvi*1g3 as previously described (Rose et al. 2006). Wild 2005 oysters were additionally genotyped for the microsatellite loci *Cvi*5*VIMS* and *Cvi*12*VIMS* as described in Carlsson et al. (2006). In the "wild" population samples any individuals with multilocus genotypes assigned as DEBY or DEBY × wild F1 were removed according to Hare et al. (2006).

Estimation of Effective Population Size

I calculated the number of breeding adults (N_b) according to Waples (2006) using LDNE 1.31 (Waples and Do 2008). Calculating N_b I assumed a random mating model and

excluded rare alleles (those at frequency less than 0.01). Rare alleles provide greater precision in the estimate of N_b, but also lead to an upward bias. Changing the frequency that constitutes a rare allele had only a marginal effect on my estimates of N_b; by excluding alleles at frequency less than 0.02 I estimated N_b to be 4,491 for wild Great Wicomico River juveniles collected in 2002, 6,643 for wild Great Wicomico River juveniles collected in 2005, and 10 for hatchery oysters released into Great Wicomico River in 2002. I calculated 95% confidence intervals by jackknifing on loci. N_b, confidence intervals around N_b, harmonic mean sample size, and number of independent comparisons are reported for each population in Table 4-1.

Large variation in N_b existed between two cohorts of wild juvenile oysters from the same river (Table 4-1). I used N_b from the 2002 wild population in the Ryman-Laikre and Wang-Ryman models but note that using the larger N_b from 2005 (or effective size based on the harmonic mean of the two estimates multiplied by a 2-year generation length) resulted in the prediction of a more dramatic bottleneck following supplementation. All three DEBY lines used for supplementation had smaller N_b than the primary line from which they were derived, consistent with the expectation that hatchery lines are a genetic subset of the primary lines.

Demographic Information

Data on restoration activities in Virginia waters of Chesapeake Bay were provided by T. Leggett, Chesapeake Bay Foundation. Estimates of hatchery plantings in Maryland waters of Chesapeake Bay were provided by Steve Allen, Oyster Recovery Partnership.

Great Wicomico River census size was estimated from fishery independent data available at http://www.vims.edu/mollusc/cbope/VAPDFfiles/VABasin2002.pdf. I calculated wild oyster census size by multiplying the density of oysters found in the Great and Little Wicomico Rivers by the area of the oyster reefs that were surveyed. I estimated that the 2002 census size for the Great Wicomico River was 3.46×10^7 oysters, though model results were insensitive to changes in the estimate of census size.

Hatchery Contribution

Contribution of hatchery oysters to reproduction in 2002 in the Great Wicomico River (x_c) was previously estimated to be 10% using assignment tests (Hare et al. 2006). I used Qvalue (STOREY 2002) to calculate the false discovery rate based on the distribution of assignment test p-values. I input the distribution of p values from 1,579 assignment tests for juvenile oysters collected in 2002, 153 of which had DEBY × wild F1 assignments ($p<0.05$). I applied the smoother method which calculated a false discovery rate of 51%, suggesting that only 75 of the DEBY × wild assignments are actual F1 offspring. Assuming equal survivorship and fertility, and that the sample was representative of the cohort, so the hatchery contribution to recruitment in 2002 was 75 out of 1579, or $x_c = 0.05$.

Predictive Modeling

I predicted the effect of a single generation of hatchery supplementation on wild oyster inbreeding effective size according to the Ryman-Laikre model (Ryman and Laikre 1991). I estimated change in effective size by calculating the harmonic mean of the two populations while accounting for their relative contributions with

$$\frac{1}{N_e} = \frac{x_c^2}{N_c} + \frac{(1-x_c)^2}{N_w} \quad \text{(equation 1)}$$

where N_e was the total effective size after supplementation, N_c was the hatchery effective size and N_w was the wild effective size. I calculated the Ryman-Laikre model for $N_w = 4{,}925.1$, $x_c = 0.05$ and $x_c = 0.1$, and $N_c = 12.3$ and $N_c = 50$.

I also predicted the effects of multiple generations of hatchery supplementation according to two models from Wang and Ryman (2001). Wang-Ryman model 1 simulates the effects of supplementation from an open hatchery line drawn at random from the wild population. It is an extension of the Ryman-Laikre model that calculates change in inbreeding effective size iteratively over generations. Supplementation first occurs at generation t but change in the inbreeding effective size does not occur until generation t+2. I modeled change in effective population size using the equation

$$\frac{1}{N_{w,t+2}} = \frac{x_{c,t+1}^2}{N_{c,t+1}} + \frac{\left(1-x_{c,t+1}\right)^2}{N_{w,t+1}} \quad \text{(equation 2).}$$

For Wang-Ryman model 1 I assumed $N_c = 12.3$ and $N_c = 50$ with $x_c = 0.05$ and $N_w = 4{,}925.1$.

Wang-Ryman model 2 predicts change in effective size of a wild population when it is supplemented with a closed hatchery line whose population at generation t+1 is a random sample from the hatchery population at generation t. With model 2 I calculated the inbreeding effective size by determining ΔIBD, the change in allelic identity-by-descent over time. I used the equation

$$N_{w,t+2} = \frac{1-F_{w,t+1}}{2\left(F_{w,t+2}-F_{w,t+1}\right)} \quad \text{(equation 3)}$$

where

$F_{w,t+2} = x_{c,t+2}F_{c,t+2} + \left(1 - x_{c,t+2}\right)F^*_{w,t+2}$ (equation 4),

$F_{c,t+2} = G_{c,t+1} = 1/2\,\beta_{c,t}\left(1 + F_{c,t}\right) + \left(1 - \beta_{c,t}\right)G^*_{c,t}$ (equation 5),

$F^*_{w,t+2} = G_{w,t+1} = {x_{c,t+1}}^2 G^*_{c,t+1} + 2\left(1 - x_{c,t+1}\right)x_{c,t+1}G_{w,t} + \left(1 - x_{c,t+1}\right)^2 G^*_{w,t+1}$

(equation 6),

$G^*_{w,t+1} = 1/2\,\beta_{w,t}\left(1 + F_{w,t}\right) + \left(1 - \beta_{w,t}\right)G_{w,t}$ (equation 7),

$G^*_{c,t+1} = 1/2\,\beta_{c,t}\left(1 + F_{c,t}\right) + \left(1 - \beta_{c,t}\right)G^*_{c,t}$ (equation 8),

$\beta_{w,t} = \frac{{\sigma_w}^2/N_{w,t} + \mu_w - 1}{\mu_w N_t - 2}$ (equation 9),

$\beta_{c,t} = \frac{{\sigma_c}^2/N_{c,t} + \mu_c - 1}{\mu_c N_t - 2}$ (equation 10),

and where N was the wild census size, and μ and σ^2 were the mean and variance of the number of gametes that an individual successfully contributes to the next generation. I began the model at generation t with

$F_{w,t} = G_{w,t} = \frac{1}{2N_{w,t}}$ (equation 11)

and

$F_{c,t} = G^*_{c,t} = \frac{1}{2N_{c,t}}$ (equation 12).

For the wild population I set μ_w=2 to ensure no population growth without supplementation. For the hatchery population I calculated gametic contribution as

$\mu_c = \frac{2N_{seed}}{N_{c,t}}$ (equation 13)

where N_{seed}, the number of hatchery offspring in generation t+1, was calculated as

$N_{seed} = \frac{x_c N_t}{1 - x_c}$ (equation 14).

I calculated ${\sigma_w}^2$ by rearranging equation 8 from Ryman and Laikre (1991) to

$\sigma_w{}^2 = \frac{N_t \mu_w{}^2 - 2\mu_w}{N_{w,t} + \mu_w(1-\mu_w)}$ (equation 15)

and calculated $\sigma_c{}^2$ according to binomial expectations such that

$\sigma_c{}^2 = N_{c,t}\left(1 - 2/N_{c,t}\right)$ (equation 16).

I calculated Wang-Ryman model 2 with initial parameter estimates of $N_{w,t} = 4{,}925.1$, $\mu_w = 2$, $\sigma_w{}^2 = 28{,}075$, $F_{w,t} = 1.0 \times 10^{-4}$, $N_{c,t} = 12.3$, $\mu_c = 295{,}855$, $\sigma_c{}^2 = 247{,}749$, $F_{c,t} = 4.1 \times 10^{-2}$, and $N_{seed} = 1.82 \times 10^6$.

I derived two equations that restoration managers can use as tools to evaluate the utility of hatchery supplementation. By substituting N_w for N_e in the Ryman-Laikre equation (equation 1 above) and rearranging, I showed that for an open hatchery with a known effective size N_c that contributes the proportion x_c to reproduction, the maximum size of a wild population to which hatchery organisms can be added without lowering the wild effective population size is

$N_{w(max)} = \frac{N_c\left(1-(1-x_c)^2\right)}{x_c{}^2}$ (equation 17).

I also rearranged the equation to determine that the minimum open hatchery size necessary to avoid lowering wild effective population size when it contributes x_c to a population with effective size N_w is

$N_{c(min)} = \frac{N_w x_c{}^2}{1-(1-x_c)^2}$ (equation 18).

Tables

Table 4-1: Estimates of the number of breeding adults (N_b) with 95% confidence intervals, harmonic mean sample size, number of independent genotypic comparisons, and genotypic correlation coefficient (r^2) from hatchery (DEBY) and wild oysters from the Great Wicomico River (GWR) and the Little Choptank River (LCR). Hatchery populations released into rivers were bred from oysters in the "primary" DEBY line.

Population	Year	N_b (95% confidence interval)	Harmonic mean sample size	Independent comparisons	r^2
Primary					
DEBY	2005	23 (12-53)	48	1,506	0.0354
GWR DEBY	2002	12 (6-21)	77	1,321	0.0360
LCR DEBY	2002	18 (9-36)	90	1,127	0.0278
LCR DEBY	2004	20 (15-25)	99	1,946	0.0257
GWR wild	2002	4,925 (1,578-∞)	692	4,587	0.0015
GWR wild	2005	11,677 (3,882-∞)	514	8,523	0.0020

Figures

Figure 4-1: A recent history of hatchery supplementation in Chesapeake Bay. Over the past decade conservation priorities have shifted from the buyback of large oysters to supportive breeding programs that use hatchery-raised oysters from wild stock and, increasingly, stock selected for disease tolerance.

Figure 4-2: Change in wild oyster effective population size (N_e) from a single generation of hatchery supplementation according the Ryman-Laikre model. The horizontal axis represents the initial wild effective population size and the vertical axis is percent change in wild effective size after supplementation. The data presented here suggest that the effective size of hatchery oysters was 12 and that they contributed 5% to the Great Wicomico River wild population (solid black line). I estimated that wild effective size in the Great Wicomico River was 4,925 and predict a 48% reduction to 2,587. A 2006 study (Rose et al. 2006) estimated that the wild effective population size is 1,516; this effective size would shrink by 17% to 1,252 after one generation of supportive breeding in the Great Wicomico River.

Figure 4-3: Effect of hatchery reproductive contribution on the effective size (N_e) of the wild oyster population after a single generation of supplementation according the Ryman-Laikre model. When hatchery contribution is very small, wild effective size was expected to remain relatively constant. When hatchery oysters contribute more to wild reproduction their effective size were expected to dramatically change the wild effective population size.

Figure 4-4: Change in the effective size (N_e) of the wild population after multiple generations of supplementation with hatchery oysters according to the Wang-Ryman model. An open hatchery line is made up of wild oysters collected from the wild, and a closed line uses the same lineage each generation. If census size increases each generation due to supplementation, then given enough time populations are expected to recover and surpass initial values of effective size. The crossover between two lines after generation 19 reflected differences in how effective size is calculated for the open and closed models.

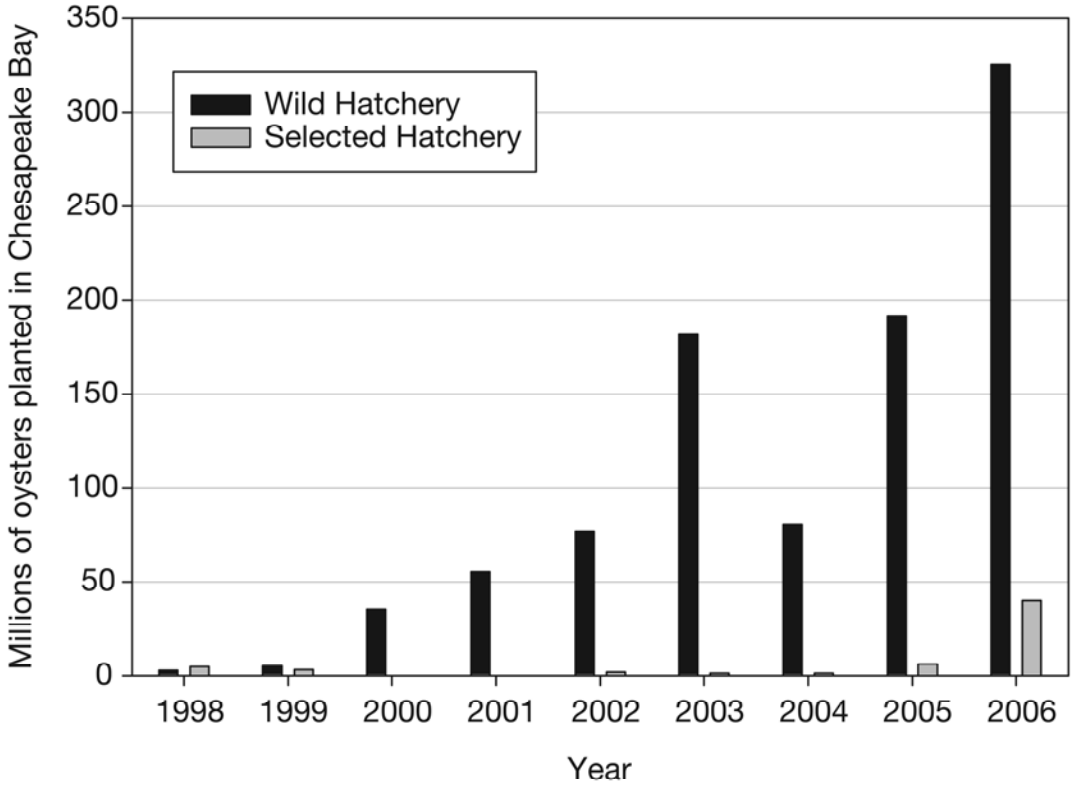
Millions of oysters planted in Chesapeake Bay
350
300
250
200
150
100
50
0
Wild Hatchery
Selected Hatchery
1998
1999
2000
2001
2002
2003
2004
2005
2006
Year

Figure 4-1

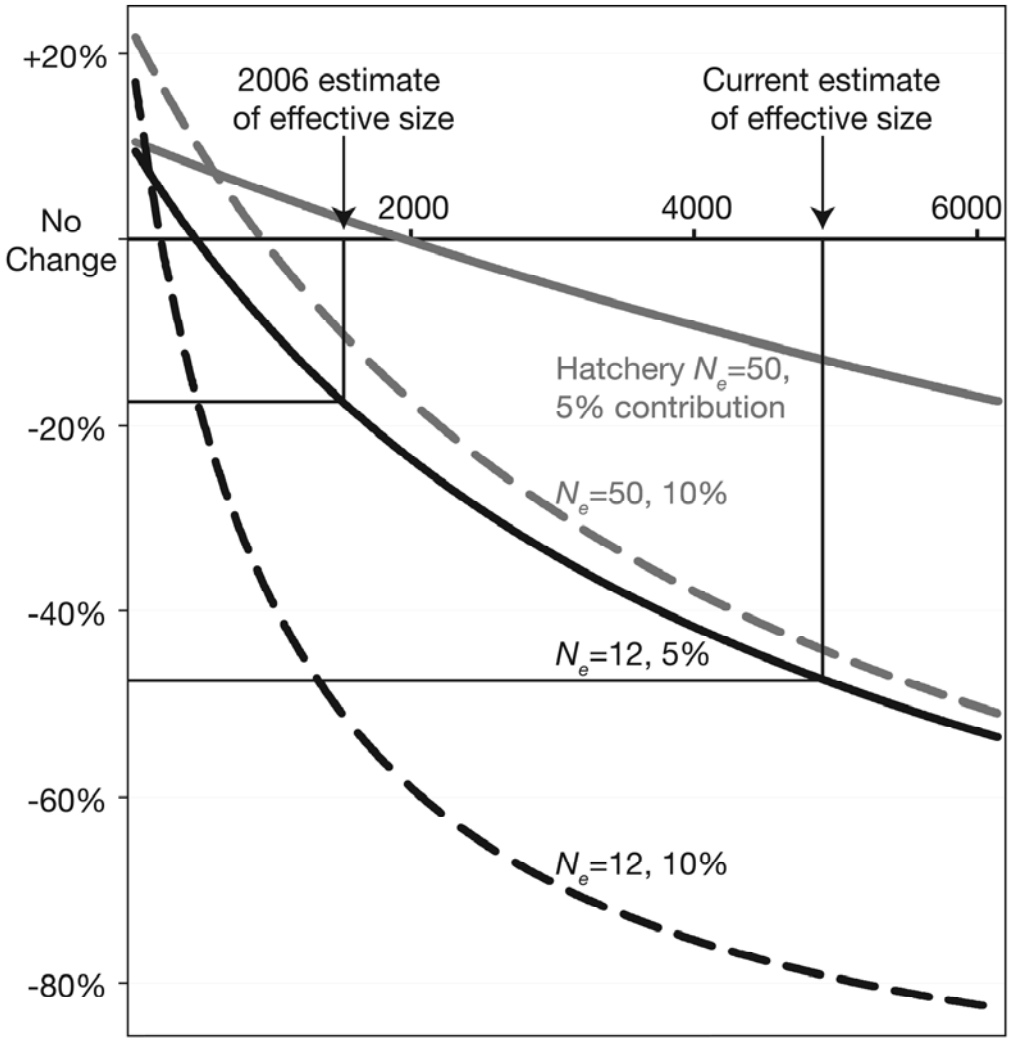
+20%
No Change
-20%
-40%
-60%
-80%
2006 estimate of effective size
Current estimate of effective size
2000
4000
6000
Hatchery N_e=50, 5% contribution
N_e=50, 10%
N_e=12, 5%
N_e=12, 10%

Figure 4-2

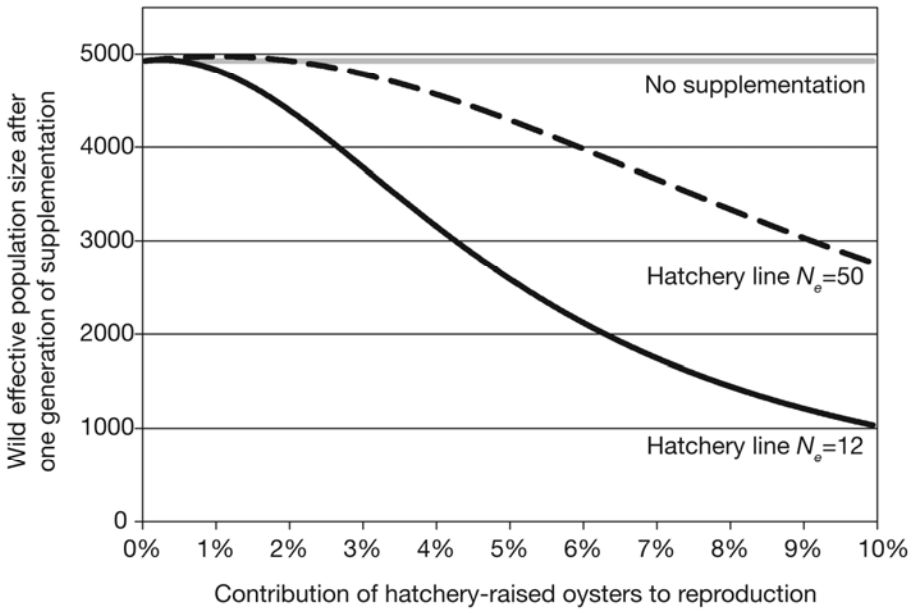
Wild effective population size after one generation of supplementation
5000
4000
3000
2000
1000
0
No supplementation
Hatchery line N_e=50
Hatchery line N_e=12
0%
1%
2%
3%
4%
5%
6%
7%
8%
9%
10%
Contribution of hatchery-raised oysters to reproduction

Figure 4-3

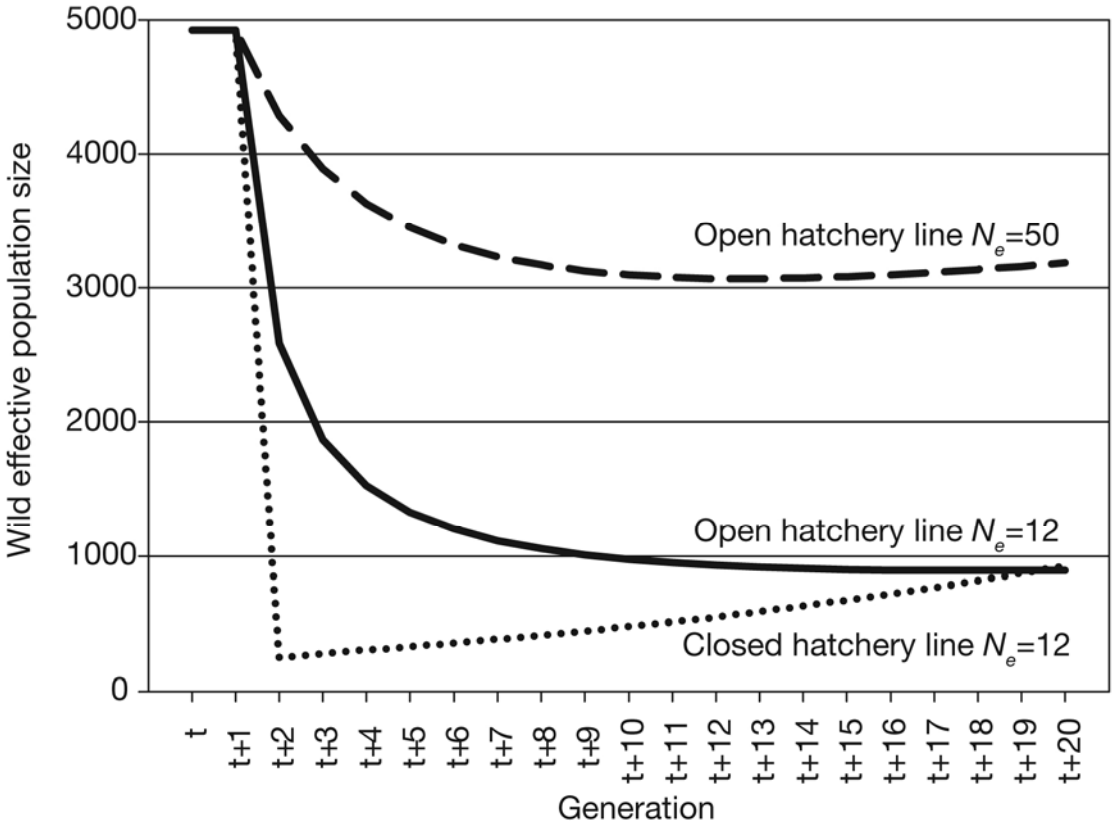
5000
4000
3000
2000
1000
0
Wild effective population size
Open hatchery line N_e=50
Open hatchery line N_e=12
Closed hatchery line N_e=12
t
t+1
t+2
t+3
t+4
t+5
t+6
t+7
t+8
t+9
t+10
t+11
t+12
t+13
t+14
t+15
t+16
t+17
t+18
t+19
t+20
Generation

Figure 4-4

Chapter 5: General Discussion

Effective Population Size

Three different estimates of effective size were calculated for the Chesapeake Bay oyster population. Temporally spaced samples in the James River led to an estimate of variance effective size of approximately 1,500. Allelic correlations of microsatellites from oysters in the Great Wicomico River indicated that of the number of breeding adults was approximately 4,500 and 11,500 in the years sampled; given two to five year generation time, this suggests that effective size was on the order of 10,000. Coalescent analysis of DNA sequences collected in the Choptank River indicated that the effective population size was 160,000. These estimates varied from one another by orders of magnitude, but two important patterns emerged. First, all estimates of effective size suggested that the oyster population was larger than what would be expected with extreme sweepstakes reproduction. While the low N_e/N ratio suggested that sweepstakes variation may play a role in oyster evolution, it did not appear to have affected the population as dramatically as Hedgecock (1994) originally suggested.

Second, estimates of inbreeding and variance effective size were smaller than the coalescent effective size. This is consistent with the expectation that the effective population size of a metapopulation is greater than the demes that make it up (Maruyama and Kimura 1980; Nei and Takahata 1993). The estimate in Chapter 2 (N_e = 1,516)

reflected the variance effective size only for the James River population between 1990 and 2003; similarly the estimates of effective size from allelic correlations in Chapter 4 spoke only to the number of breeders estimated in James River during in 2002 and 2005 (N_e = 4,925 and 11,677 respectively). On the other hand, the coalescent estimate of effective size in Chapter 3 (N_e = 160,000) reflected a Bay-wide value because it was calculated from a sample that contains ancestral levels of polymorphism. This pattern supported the notion that there are limits to migration between subpopulations in Chesapeake Bay.

Natural Selection

The large effective size of the oyster population means that natural selection is likely to be an important force shaping patterns of genetic variation, especially in regions of low recombination like mitochondrial DNA. Bazin et al. (2006b) argued that compared to vertebrates, invertebrates tend to have large effective population sizes and therefore that positive selection on mitochondrial DNA is likely to be common. The signal of mitochondrial selection indicated that genetic draft has indeed been an important force determining levels of variation for oysters. The issue of selective neutrality is critical because mitochondrial DNA under selection does not provide information about demography, but instead reflects patterns of the natural selection itself. Since the goal of many population genetic studies is the determination of demographic history, it is critical to confirm the selective neutrality of mitochondrial DNA before making demographic inferences.

Population Structure

The pattern of isolation by distance demonstrated in Chapter 2 shows that, despite the long duration of the oyster larval stage, dispersal distance tends to be restricted. Isolation by distance is an equilibrium pattern, meaning that the differentiation among rivers or subestuaries is a historical process, and probably not due to recent habitat fragmentation. After the findings in Chapter 2 were published by Rose et al. (2006), the conclusions that some rivers are trap-like and that oyster behavior may reduce larval transport were bolstered by the findings of North et al. (2008). North et al. used a hydrodynamic model of Chesapeake Bay to predict the movement and transport of oyster larvae and found that large tributary systems tended to have circulation patterns conducive to trapping *C. virginica* larvae, and that populations found in the northern parts of the Bay were at least one generation removed from populations in the southern Bay. They concluded that hydrodynamic and behavioral forces restricting larval movement could contribute to a pattern of isolation by distance. Genetic structure in the Chesapeake Bay oyster population shows that long larval duration does not necessarily lead to long dispersal distances. Despite the lack of obvious barriers to movement, it appears that behavior and hydrography interact in a way that may promote local recruitment.

Supportive Breeding

Supplementation of wild Chesapeake Bay oysters with disease tolerant hatchery stock is a recent but increasingly important restoration tool. The risks of supplementation, however, may outweigh any benefits if the wild effective population size drops precipitously. Data indicated that hatchery and wild effective population sizes differed by several orders of

magnitude, and that hatchery oysters can have potentially harmful effects on the wild population. The maintenance of high wild effective population size after three years of supplementation indicated that hatchery oysters have not and perhaps will not contribute to wild reproduction. Changes in restoration policy that emphasize increasing the hatchery effective population size may mitigate the risk of inbreeding depression in the wild.

Prospects for Future Work

The conclusions of this study suggest several directions for future research. First, my estimates of relatively large effective population size can be compared with those from populations of *C. virginica* outside of Chesapeake Bay, other oyster species, and other non-crassostreid marine invertebrates. Additional future work focusing on interannual variation in effective size may be informative about the role that variance in reproductive success plays in shaping the composition of cohorts. Understanding how these patterns compare among marine invertebrates—particularly in relation to census size—may prove to be an effective way of evaluating the sweepstakes hypothesis.

Next, the presence of isolation by distance raises questions about larval dispersal in Chesapeake Bay. An investigation of genetic differentiation in other Chesapeake Bay invertebrate species will provide insight into the relative importance that behavior and hydrography play in determining population structure. Studies of other oyster populations may indicate whether the relatively small-scale differentiation found in Chesapeake Bay is typical in other estuarine or coastal habitats.

Another exciting prospect for future work is to investigate whether mitochondrial selection is the cause of non-neutral patterns observed in other oyster populations. The description by Boom et al. (1994) of skewed mitochondrial haplotype frequency in *C. gigas* was very similar to the work described here. The data that Boom et al. collected have been used applied to the multiple-mergers coalescent theory, which relies on the assumption of selective neutrality. Tests of selection on *C. gigas* mitochondrial DNA might provide information about whether the multiple-mergers coalescent is an appropriate model for the data. Selection could also explain the mitochondrial skew found in the data collected by Reeb and Avise (1990); if so, positive selection may have played a role in the evolution of Floridian populations of *C. virginica*. The steep genetic cline found in mitochondrial markers along Florida's Atlantic coast could be explained by a selective gradient, perhaps related to water temperature or to stress related to hypoxic or hyperosmotic conditions. Determining whether natural selection has played a role in the mitochondrial differentiation of "Atlantic" and "Gulf of Mexico" haplotypes along Florida's coast could prove informative about the forces determining the introgression of genes in a secondary contact zone.

Finally, data presented here regarding the effects of supportive breeding provide several clear directions for the future of oyster conservation. Research into raising the effective size of hatchery lines and determining the effective population size of supplemented populations should be top priorities. If supportive breeding is to continue in Chesapeake Bay, managers should ensure that the risk of long-term harm is minimized. Populations

that have already been supplemented with closed hatchery lines should be monitored for declining effective population size due to reproductive success of hatchery offspring.

List of References

Adams, R. I., K. M. Brown, and M. B. Hamilton. 2004. The impact of microsatellite electromorph size homoplasy on multilocus population structure estimates in a tropical tree (*Corythophora alta*) and an anadromous fish (*Morone saxatilis*). Molecular Ecology 13:2579-2588.

Addison, J. A., and M. W. Hart. 2004. Analysis of population genetic structure of the green sea urchin (*Strongylocentrotus droebachiensis*) using microsatellites. Marine Biology 144:243-251.

Akashi, H. 1999. Inferring the fitness effects of DNA mutations from polymorphism and divergence data: Statistical power to detect directional selection under stationarity and free recombination. Genetics 151:221-238.

Allen, S. K., R. D. Brumbaugh, and D. Schulte. 2003. Terraforming Chesapeake Bay. Virginia Marine Resource Bulletin 35:2-8.

Allendorf, F. W. 1983. Linkage disequilibrium generated by selection against null alleles at duplicate loci. The American Naturalist 121:588-592.

Andrews, J. D. 1979. Pelecypoda: Ostreidae. Pp. 293-341 *in* A. C. Giese, and J. S. Pearse, eds. Reproduction of Marine Invertebrates. Academic Press, New York.

Appleyard, S. A., and R. D. Ward. 2006. Genetic diversity and effective population size in mass selection lines of Pacific oyster (*Crassostrea gigas*). Aquaculture 254:148-159.

Araki, H., B. Cooper, and M. S. Blouin. 2007. Genetic effects of captive breeding cause a rapid, cumulative fitness decline in the wild. Science 318:100-103.

Árnason, E. 2004. Mitochondrial cytochrome b DNA variation in the high-fecundity Atlantic cod: Trans-atlantic clines and shallow gene genealogy. Genetics 166:1871-1885.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics 18:489-522.

Avise, J. C., R. M. Ball, and J. Arnold. 1988. Current versus historical population sizes in vertebrate species with high gene flow: A comparison based on mitochondrial DNA lineages and inbreeding theory for neutral mutations. Mol Biol Evol 5:331-344.

Baker, P., and R. Mann. 2003. Late stage bivalve larvae in a well-mixed estuary are not inert particles. Estuaries 26:837-845.

Ballard, J. W. O., and M. Kreitman. 1995. Is mitochondrial DNA a strictly neutral marker? Trends in Ecology & Evolution 10:485-488.

Ballard, J. W. O., and D. M. Rand. 2005. The population biology of mitochondrial DNA and its phylogenetic implications. Annual Review of Ecology Evolution and Systematics 36:621-642.

Balloux, F., H. Brunner, N. Lugon-Moulin, J. Hausser, and J. Goudet. 2000. Microsatellites can be misleading: an empirical and simulation study. Evolution 54:1414-1422.

Balloux, F., and J. Goudet. 2002. Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. Molecular Ecology 11:771-783.

Barrier, M., C. D. Bustamante, J. Y. Yu, and M. D. Purugganan. 2003. Selection on rapidly evolving proteins in the *Arabidopsis* genome. Genetics 163:723-733.

Bartley, D., M. Bagley, G. Gall, and B. Bentley. 1992. Use of linkage disequilibrium data to estimate effective size of hatchery and natural fish populations. Conservation Biology 6:365-375.

Bazin, E., S. Glemin, and N. Galtier. 2006a. Mitochondrial DNA and population size - Response. Science 314:1390-1390.

Bazin, E., S. Glemin, and N. Galtier. 2006b. Population size does not influence mitochondrial genetic diversity in animals. Science 312:570-572.

Beaumont, M. A., W. Zhang, and D. J. Balding. 2002. Approximate Bayesian computation in population genetics. Genetics 162:2025-2035.

Beckenbach, A. T. 1994. Mitochondrial haplotype frequencies in oysters: neutral alternatives to selection models. Pp. 188-198 *in* G. B. Golding, ed. Non-Neutral Evolution: Theories and Molecular Data. Chapman & Hall, New York.

Beerli, P., and J. Felsenstein. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. Genetics 152:763-773.

Beerli, P., and J. Felsenstein. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proceedings of the National Academy of Sciences of the United States of America 98:4563-4568.

Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste, and F. Bonhomme. 2001. GENETIX 4.02, logiciel sous Windows^{TM} pour la génétique des populations. Laboratoire Génome, Populations, Interactions CNRS UMR 5000, Université de Montpellier II, Montpellier (France).

Bertness, M. D., and S. D. Gaines. 1993. Larval dispersal and local adaptation in acorn barnacles. Evolution 47:316-320.

Bierne, N., S. Launey, Y. Naciri-Graven, and F. Bonhomme. 1998. Early effect of inbreeding as revealed by microsatellite analyses on *Ostrea edulis* larvae. Genetics 148:1893-1906.

Bierne, N., A. Tsitrone, and P. David. 2000. An inbreeding model of associative overdominance during a population bottleneck. Genetics 155:1981-1990.

Boesch, D., E. Burreson, W. Dennison, E. Houde, M. Kemp, V. Kennedy, R. Newell, K. Paynter, R. Orth, and R. Ulanowicz. 2001a. Factors in the decline of coastal ecosystems. Science 293:1589-1590.

Boesch, D. F., R. B. Brinsfield, and R. E. Magnien. 2001b. Chesapeake Bay eutrophication: scientific understanding, ecosystem restoration, and challenges for agriculture. Journal of Environmental Quality 30:303-320.

Bohonak, A. J. 1999. Dispersal, gene flow, and population structure. Quarterly Review of Biology 74:21-45.

Boom, J. D. G., E. G. Boulding, and A. T. Beckenbach. 1994. Mitochondrial DNA variation in introduced populations of Pacific oyster, *Crassostrea gigas*, in British Columbia. Canadian Journal of Fisheries and Aquatic Sciences 51:1608-1614.

Botsford, L. W., F. Micheli, and A. Hastings. 2003. Principles for the design of marine reserves. Ecological Applications 13:S25-S31.

Boudry, P., B. Collet, F. Cornette, V. Hervouet, and F. Bonhomme. 2002. High variance in reproductive success of the Pacific oyster (*Crassostrea gigas*, Thunberg) revealed by microsatellite-based parentage analysis of multifactorial crosses. Aquaculture 204:283-296.

Braverman, J. M., R. R. Hudson, N. L. Kaplan, C. H. Langley, and W. Stephan. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics 140:783-796.

Breitburg, D. L., L. D. Coen, M. W. Luckenbach, R. Mann, M. Posey, and J. A. Wesson. 2000. Oyster reef restoration: convergence of harvest and conservation strategies. Journal of Shellfish Research 19:371-377.

Brookfield, J. F. Y. 1996. A simple new method for estimating null allele frequency from heterozygote deficiency. Molecular Ecology 5:453-455.

Brown, B. L., D. E. Franklin, P. Gaffney, M. Hong, D. Dendanto, and I. Kornfield. 2000. Characterization of microsatellite loci in the eastern oyster, *Crassostrea virginica*. Molecular Ecology 9:2155-2234.

Brown, B. L., and K. T. Paynter. 1991. Mitochondrial DNA analysis of native and selectively inbred Chesapeake Bay oysters, *Crassostrea virginica*. Marine Biology 110:343-352.

Brown, W. M. 1983. Evolution of animal mitochondrial DNA. Pp. 95-130 *in* M. Nei, and R. K. Koehn, eds. Evolution of Genes and Proteins. Sinauer, Sunderland, MA.

Brumbaugh, R. D., L. A. Sorabella, C. O. Garcia, W. J. Goldsborough, and J. A. Wesson. 2000. Making a case for community-based oyster restoration: an example from Hampton Roads, Virginia, USA. Journal of Shellfish Research 19:467-472.

Buonaccorsi, V. P., M. Westerman, J. Stannard, C. Kimbrell, E. Lynn, and R. D. Vetter. 2004. Molecular genetic structure suggests limited larval dispersal in grass rockfish, Sebastes rastrelliger. Marine Biology 145:779-788.

Buroker, N. E. 1983a. Genetic differentiation and population structure of the American oyster *Crassostrea virginica* (Gmelin) in Chesapeake Bay. Journal of Shellfish Research 3:153-167.

Buroker, N. E. 1983b. Population genetics of the American oyster *Crassostrea virginica* along the Atlantic coast and the Gulf of Mexico. Marine Biology 75:99-112.

Buroker, N. E., W. K. Hershberger, and K. K. Chew. 1979. Population genetics of the family Ostreidae. II. Interspecific studies of the genera *Crassostrea* and *Saccostrea*. Marine Biology 54:171-184.

Burreson, E. M., and L. 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.

Butler, P. A. 1949. Gametogenesis in the oyster under conditions of depressed salinity. Biol Bull 96:263-269.

Caley, M. J., M. H. Carr, M. A. Hixon, T. P. Hughes, G. P. Jones, and B. A. Menge. 1996. Recruitment and the local dynamics of open marine populations. Annual Review of Ecology and Systematics 27:477-500.

Carlsson, J., C. L. Morrison, and K. S. Reece. 2006. Wild and aquaculture populations of the eastern oyster compared using microsatellites. Journal of Heredity 97:595-598.

Carlton, J. T., and R. Mann. 1996. Transfers and world-wide introductions. Pp. 691-706 *in* V. Kennedy, R. Newell, and A. F. Eble, eds. The Eastern Oyster *Crassostrea virginica*. Maryland Sea Grant, College Park.

Castric, V., and L. Bernatchez. 2003. The rise and fall of isolation by distance in the anadromous brook charr (*Salvelinus fontinalis* Mitchill). Genetics 163:983-996.

Chakraborty, R., M. Deandrade, S. P. Daiger, and B. Budowle. 1992. Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. Annals of Human Genetics 56:45-57.

Charlesworth, B., M. Nordborg, and D. Charlesworth. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genetical Research 70:155-174.

Clement, M., D. Posada, and K. Crandall. 2000. TCS: A computer program to estimate gene genealogies. Molecular Ecology 9:1657-1660.

Cockerham, C. C., and B. S. Weir. 1993. Estimation of Gene Flow from F-Statistics. Evolution 47:855-863.

Cornuet, J. M., S. Piry, G. Luikart, A. Estoup, and M. Solignac. 1999. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genetics 153:1989-2000.

Cowen, R. K., K. M. M. Lwiza, S. Sponaugle, C. B. Paris, and D. B. Olson. 2000. Connectivity of marine populations: open or closed? Science 287:857-859.

Crow, J. F. 1954. Breeding structure of populations. II. Effective population number. Pp. 543-560 *in* O. Kempthorne, T. A. Bancroft, J. W. Gowen, and L. Lush, eds. Statistics and Mathematics in Biology. Iowa State College Press, Ames.

Crow, J. F., and K. Aoki. 1984. Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences 81:6073-6077.

Dekshenieks, M. M., E. E. Hofmann, J. M. Klinck, and E. N. Powell. 1996. Modeling the vertical distribution of oyster larvae in response to environmental conditions. Marine Ecology-Progress Series 136:97-110.

Depaulis, F., S. Mousset, and M. Veuille. 2003. Power of neutrality tests to detect bottlenecks and hitchhiking. Journal of Molecular Evolution 57:S190-S200.

Ehrlich, P. R., and P. H. Raven. 1969. Differentiation of Populations. Science 165:1228-1232.

Eldon, B., and J. Wakeley. 2006. Coalescent processes when the distribution of offspring number among individuals is highly skewed. Genetics 172:2621-2633.

Ellis, L. L., J. M. Burcham, and K. T. Paynter. 1985. Amino acid metabolism in euryhaline bivalves: regulation of glycine accumulation in ribbed muscle gills. Journal of Experimental Zoology 233:347-358.

Elson, J. L., D. M. Turnbull, and N. Howell. 2004. Comparative Genomics and the evolution of human mitochondrial DNA: Assessing the effects of selection. American Journal of Human Genetics 74:229-238.

Estoup, A., P. Jarne, and J. M. Cornuet. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Molecular Ecology 11:1591-1604.

Ewens, W. J. 1972. Sampling theory of selectively neutral alleles. Theoretical Population Biology 3:87-112.

Ewens, W. J. 1982. On the concept of the effective population size. Theoretical Population Biology 21:373-378.

Fay, J. C., G. J. Wyckoff, and C. I. Wu. 2002. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. Nature 415:1024-1026.

Felsenstein, J. 1992. Estimating effective population size from samples of sequences: A bootstrap Monte Carlo integration method. Genetical Research 60:209-220.

Finelli, C. M., and D. S. Wethey. 2003. Behavior of oyster (*Crassostrea virginica*) larvae in flume boundary layer flows. Marine Biology 143:703-711.

Flowers, J. M., S. C. Schroeter, and R. S. Burton. 2002. The recruitment sweepstakes has many winners: genetic evidence from the sea urchin *Strongylocentrotus purpuratus*. Evolution 56:1445-1453.

Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3:294-299.

Foltz, D. W. 1986. Null alleles as a possible cause of heterozygote deficiencies in the oyster *Crassostrea virginica* and other bivalves. Evolution 40:869-870.

Ford, M. J. 2002. Selection in captivity during supportive breeding mayreduce fitness in the wild. Conservation Biology 16:815-825.

Frankham, R. 1995a. Effective population size/adult population size ratios in wildlife: a review. Genetical Research 66:95-107.

Frankham, R. 1995b. Inbreeding and extinction: a threshold effect. Conservation Biology 9:792-799.

Franklin, I. R. 1980. Evolutionary changes in small populations. Pp. 135-149 *in* M. E. Soulé, and B. A. Wilcox, eds. Conservation Biology: An Evolutionary-Ecological Perspective. Sinauer Associates, Sunderland, MA.

Fu, Y. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915-925.

Fu, Y. X., and W. H. Li. 1993. Statistical Tests of Neutrality of Mutations. Genetics 133:693-709.

Funk, D. J., and K. E. Omland. 2003. Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annual Review of Ecology Evolution and Systematics 34:397-423.

Galtier, N., F. Depaulis, and N. H. Barton. 2000. Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. Genetics 155:981-987.

Galtsoff, P. S. 1964. The American oyster. U.S. Government Printing Office, Washington, DC.

Gilg, M. R., and T. J. Hilbish. 2003. The geography of marine larval dispersal: Coupling genetics with fine-scale physical oceanography. Ecology 84:2989-2998.

Gillespie, J. H. 2000. Genetic drift in an infinite population: The pseudohitchhiking model. Genetics 155:909-919.

Gillespie, J. H. 2001. Is the population size of a species relevant to its evolution? Evolution 55:2161-2169.

Gillespie, J. H. 2004. Population Genetics: A Concise Guide. The Johns Hopkins University Press, Baltimore and London.

Gold, J. R., C. P. Burridge, and T. F. Turner. 2001. A modified stepping-stone model of population structure in red drum, *Sciaenops ocellatus* (Sciaenidae), from the northern Gulf of Mexico. Genetica 111:305-317.

Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Institut de Zoologie et d'Ecologie Animale, Universite de Lausane, Lucerne.

Goudet, J., M. Raymond, T. deMeeus, and F. Rousset. 1996. Testing differentiation in diploid populations. Genetics 144:1933-1940.

Grumet, R. S. 2000. Bay, Plain, and Piedmont: A Landscape History of the Chesapeake Heartland from 1.3 Billion Years Ago to 2000. U.S. Department of the Interior, National Park Service, Annapolis, MD.

Hahn, M. W. 2008. Toward a selection theory of molecular evolution. Evolution 62:255-265.

Hahn, M. W., M. D. Rausher, and C. W. Cunningham. 2002. Distinguishing between selection and population expansion in an experimental lineage of bacteriophage T7. Genetics 161:11-20.

Hare, M. P., S. K. Allen, P. Bloomer, M. D. Camara, R. B. Carnegie, J. Murfree, M. Luckenbach, D. Meritt, C. Morrison, K. Paynter, K. S. Reece, and C. G. Rose. 2006. A genetic test for recruitment enhancement in Chesapeake Bay oysters, *Crassostrea virginica*, after population supplementation with a disease tolerant strain. Conservation Genetics 7:717-734.

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.

Hare, M. P., S. A. Karl, and J. C. Avise. 1996. Anonymous nuclear DNA markers in the American oyster and their implications for the heterozygote deficiency phenomenon in marine bivalves. Molecular Biology and Evolution 13:334-345.

Hartl, D. L., and A. G. Clark. 1997. Principles of population genetics. Sinauer Associates, Sunderland, MA.

Hauser, L., G. J. Adcock, P. J. Smith, J. H. B. Ramirez, and G. R. Carvalho. 2002. Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). Proceedings of the National Academy of Sciences of the United States of America 99:11742-11747.

Hedgecock, D. 1994. Does variance in reproductive success limit effective population size of marine organisms? Pp. 122-134 *in* A. R. Beaumont, ed. Genetics and Evolution of Aquatic Organisms. Chapman & Hall, London.

Hedgecock, D., V. Chow, and R. S. Waples. 1992. Effective population numbers of shellfish broodstocks estimated from temporal variance in allelic frequencies. Aquaculture 108:215-232.

Hedgecock, D., G. Li, S. Hubert, K. Bucklin, and V. Ribes. 2004. Widespread null alleles and poor cross-species amplification of microsatellite DNA loci cloned from the Pacific oyster, *Crassostrea gigas*. Journal of Shellfish Research 23:379-385.

Hedrick, P. 2005. Large variance in reproductive success and the Ne/N ratio. Evolution 59:1596-1599.

Hedrick, P. W. 1999. Perspective: highly variable loci and their interpretation in evolution and conservation. Evolution 53:313-318.

Hedrick, P. W., D. Hedgecock, and S. Hamelberg. 1995. Effective population size in winter-run chinook salmon. Conservation Biology 9:615-624.

Hedrick, P. W., and S. T. Kalinowski. 2000. Inbreeding depression in conservation biology. Annual Review of Ecology and Systematics 31:139-162.

Hey, J., and R. Nielsen. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167:747-760.

Hilbish, T. J. 1996. Population genetics of marine species: the interaction of natural selection and historically differentiated populations. Journal of Experimental Marine Biology and Ecology 200:67-83.

Hill, A. E. 1991. A mechanism for horizontal zooplankton transport by vertical migration in tidal currents. Marine Biology 111:485-492.

Hill, W. G. 1981. Estimation of effective population size from data on linkage disequilibrium. Genetical Research 38:209-216.

Ho, M. S., and P. L. Zubkoff. 1982. Anaerobic metabolism of the ribbed mussel, *Geukensia demissa*. Comparative Biochemistry and Physiology 73B:931-936.

Ho, M. S., and P. L. Zubkoff. 1983. Volatile fatty acids of the ribbed mussel, *Geukensia demissa*. Comparative Biochemistry and Physiology 74B:539-542.

Hoarau, G., E. Boon, D. N. Jongma, S. Ferber, J. Palsson, H. W. Van der Veer, A. D. Rijnsdorp, W. T. Stam, and J. L. Olsen. 2005. Low effective population size and evidence for inbreeding in an overexploited flatfish, plaice (*Pleuronectes platessa* L.). Proceedings of the Royal Society B-Biological Sciences 272:497-503.

Hobbs, C. H. 2004. Geological history of Chesapeake Bay, USA. Quaternary Science Reviews 23:641-661.

Holm, S. 1979. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6:65-70.

Hood, G. M. 2004. PopTools version 2.6.4.

Hoover, C. A., and P. M. Gaffney. 2005. Geographic variation in nuclear genes of the eastern oyster, *Crassostrea virginica* Gmelin. Journal of Shellfish Research 24:103-112.

Hudson, K. L., M. Kreitman, and M. Aguadé. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics 116:153-159.

Hunt, H. L., and R. E. Scheibling. 1997. Role of early post-settlement mortality in recruitment of benthic marine invertebrates. Marine Ecology-Progress Series 155:269-301.

Hunter, R. L., and C. L. Markert. 1957. Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels. Science 125:1294-a-1295.

Jackson, J. B. C. 2001. What was natural in the coastal oceans? Proceedings of the National Academy of Sciences of the United States of America 98:5411-5418.

Jackson, J. B. C., M. X. Kirby, W. H. Berger, K. A. Bjorndal, L. W. Botsford, B. J. Bourque, R. H. Bradbury, R. Cooke, J. Erlandson, J. A. Estes, T. P. Hughes, S. Kidwell, C. B. Lange, H. S. Lenihan, J. M. Pandolfi, C. H. Peterson, R. S. Steneck, M. J. Tegner, and R. R. Warner. 2001. Historical overfishing and the recent collapse of coastal ecosystems. Science 293:629-638.

Johannesson, K. 1988. The paradox of Rockall: why is a brooding gastropod (*Littorina saxatilis*) more widespread than one which has a planktonic dispersal stage (*L. littorea*)? Mar. Biol. 99:507-513.

Johannesson, K., J. Lundberg, C. Andre, and P. G. Nilsson. 2004. Island isolation and habitat heterogeneity correlate with DNA variation in a marine snail (Littorina saxatilis). Biological Journal of the Linnean Society 82:377-384.

Johnson, M. S., and R. Black. 1984. Pattern beneath the chaos: the effect of recruitment on genetic patchiness in an intertidal limpet. Evolution 38:1371-1383.

Jordan, S. J., and J. M. Coakley. 2004. Long-term projections of eastern oyster populations under various management scenarios. Journal of Shellfish Research 23:63-72.

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

Karl, S. A., and J. C. Avise. 1992. Balancing selection at allozyme loci in oysters: Implications from nuclear RFLPs. Science 256:100-102.

Kelly, F. P. 1977. Exact results for the Moran neutral allele model. Advances in Applied Probability 9:197-201.

Kennedy, V. S. 1996. The biology of larvae and spat. Pp. 371-422 *in* V. S. Kennedy, R. I. E. Newell, and A. F. Eble, eds. The Eastern Oyster: *Crassostrea virginica*. Maryland Sea Grant, College Park, MD.

Kimura, M. 1979. Neutral Theory of Molecular Evolution. Scientific American 241:98-127.

Kimura, M., and J. F. Crow. 1963. The measurement of effective population numbers. Evolution 17:279-288.

Kimura, M., and G. H. Weiss. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 49:561-576.

Kingman, J. F. C. 1982a. The coalescent. Stochastic Processes and Their Applications 13:235-248.

Kingman, J. F. C. 1982b. On the genealogy of large populations *in* J. Gani, and E. J. Hannan, eds. Essays in Statistical Science. Applied Probability Trust, London.

Kirby, M. X. 2004. Fishing down the coast: Historical expansion and collapse of oyster fisheries along continental margins. Proceedings of the National Academy of Sciences of the United States of America 101:13096-13099.

Kishino, H., and M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J. Mol. Evol. 29:170-179.

Kluytmans, J. H., M. van Graft, J. Janus, and H. Pieters. 1978. Production and excretion of volatile fatty acids in the sea mussel *Mytilus edulis* L. Journal of Comparative Physiology B123:163-167.

Koehn, R. K., R. I. E. Newell, and F. Immermann. 1980. Maintenance of an aminopeptidase allele frequency cline by natural selection. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences 77:5385-5389.

Kuhner, M. K. 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. Bioinformatics 22:768-770.

Lande, R. 1995. Mutation and conservation. Conservation Biology 9:782-791.

Lande, R., and S. Shannon. 1996. The role of genetic variation in adaptation and population persistence in a changing environment. Evolution 50:434-437.

Lapègue, S., A. Beaumont, P. Boudry, and P. Goulletquer. 2007. European flat oyster: *Ostrea edulis*. Pp. 70-75 *in* T. Svåsand, D. Crosetti, E. García-Vázquez, and E. Verspoor, eds. Genetic impact of aquaculture activities on native populations. Genimpact final scientific report.

Launey, S., M. Barre, A. Gerard, and Y. Naciri-Graven. 2001. Population bottleneck and effective size in *Bonamia ostreae*-resistant populations of *Ostrea edulis* as inferred by microsatellite markers. Genetical Research 78:259-270.

Launey, S., and D. Hedgecock. 2001. High genetic load in the Pacific oyster *Crassostrea gigas*. Genetics 159:255-265.

Launey, S., C. Ledu, P. Boudry, F. Bonhomme, and Y. Naciri-Graven. 2002. Geographic structure in the European flat oyster (*Ostrea edulis* L.) as revealed by microsatellite polymorphism. Journal of Heredity 93:331-338.

Laurie-Ahlberg, C. C., and B. S. Weir. 1979. Allozymic variation and linkage disequilibrium in some laboratory populations of *Drosophila melanogaster*. Genetics 92:1295-1314.

Li, G., and D. Hedgecock. 1998. Genetic heterogeneity, detected by PCR-SSCP, among samples of larval Pacific oysters (*Crassostrea gigas*) supports the hypothesis of large variance in reproductive success. Canadian Journal of Fisheries and Aquatic Sciences 55:1025-1033.

Luikart, G., F. W. Allendorf, J. M. Cornuet, and W. B. Sherwin. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. Journal of Heredity 89:238-247.

Lynch, M., J. Conery, and R. Bürger. 1995. Mutation Accumulation and the Extinction of Small Populations. American Naturalist 146:489-518.

Lynch, M., and M. O'Hely. 2001. Captive breeding and the genetic fitness of natural populations. Conservation Genetics 2:363-378.

MacKenzie, C. 1996. Management of Natural Populations. Pp. 707-721 *in* V. Kennedy, R. Newell, and A. Eble, eds. The Eastern Oyster *Crassostrea virginica*. Maryland Sea Grant College, College Park.

Maes, G. E., and F. A. M. Volckaert. 2002. Clinal genetic variation and isolation by distance in the European eel *Anguilla anguilla* (L.). Biological Journal of the Linnean Society 77:509-521.

Mann, R. 1988. Distribution of bivalve larvae at a frontal system in the James River, Virginia. Marine Ecology-Progress Series 50:29-44.

Mann, R., and D. A. Evans. 2004. Site selection for oyster habitat rehabilitation in the Virginia portion of the Chesapeake Bay: a commentary. Journal of Shellfish Research 23:41-49.

Mann, R., and E. N. Powell. 2007. Why oyster restoration goals in the Chesapeake Bay are not and probably cannot be achieved. Journal of Shellfish Research 26:905-917.

Mantel, N. 1967. Detection of disease clustering and a generalized regression approach. Cancer Research 27:209-220.

Maruyama, T., and M. Kimura. 1980. Genetic variability and effective population size when local extinction and recolonization of subpopulations are frequent. Proceedings of the National Academy of Sciences of the United States of America 77:6710-6714.

McDonald, J. H., and M. Kreitman. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 351:652-654.

McDonald, J. H., B. C. Verrelli, and L. B. Geyer. 1996. Lack of geographic variation in anonymous nuclear polymorphisms in the American oyster, *Crassostrea virginica*. Molecular Biology and Evolution 13:1114-1118.

McGoldrick, D. J., and D. Hedgecock. 1997. Fixation, segregation and linkage of allozyme loci in inbred families of the Pacific oyster *Crassostrea gigas* (Thunberg): implications for the causes of inbreeding depression. Genetics 146:321-334.

McGoldrick, D. J., D. Hedgecock, L. J. English, P. Baoprasertkul, and R. D. Ward. 2000. The transmission of microsatellite alleles in Australian and North American stocks of the Pacific oyster (*Crassostrea gigas*): selection and null alleles. Journal of Shellfish Research 19:779-788.

Meiklejohn, C. D., K. L. Montooth, and D. M. Rand. 2007. Positive and negative selection on the mitochondrial genome. Trends in Genetics 23:259-263.

Milbury, C. 2007. The mitochondrial genome of the eastern oyster, *Crassostrea virginica*: the complete DNA sequence and its application in local restoration efforts. College of Marine Studies. University of Delaware, Lewes.

Milbury, C. A., D. W. Meritt, R. I. E. Newell, and P. M. Gaffney. 2004. Mitochondrial DNA markers allow monitoring of oyster stock enhancement in the Chesapeake Bay. Marine Biology 145:351-359.

Moran, P. A. P. 1958. Random processes in genetics. Proceedings of the Cambridge Philosophical Society 54:60-72.

Moritz, C., T. E. Dowling, and W. M. Brown. 1987. Evolution of animal mitochondrial DNA: Relevance for population biology and systematics. Annual Review of Ecology and Systematics 18:269-292.

Mulligan, C. J., A. Kitchen, and M. M. Miyamoto. 2006. Comment on "Population size does not influence mitochondrial genetic diversity in animals". Science 314:1390a.

Murray, M. C., and M. P. Hare. 2006. A genomic scan for divergent selection in a secondary contact zone between Atlantic and Gulf of Mexico oysters, *Crassostrea virginica*. Molecular Ecology 15:4229-4242.

Nei, M. 1987. Molecular Evolutionary Genetics. Pp. 164-165. Columbia University Press, New York.

Nei, M., and N. Takahata. 1993. Effective population size, genetic diversity, and coalescence time in subdivided populations. J. Mol. Evol. 37:240-244.

Neigel, J. E. 1996. Estimation of effective population size and migration parameters from genetic data. Pp. 329-346 *in* T. B. Smith, and R. K. Wayne, eds. Molecular Genetic Approaches in Conservation. Oxford University Press, New York.

Neigel, J. E., and J. C. Avise. 1993. Application of a random walk model to geographic distributions of animal mitochondrial DNA variation. Genetics 135:1209-1220.

Neigel, J. E., R. M. Ball, and J. C. Avise. 1991. Estimation of Single Generation Migration Distances from Geographic-Variation in Animal Mitochondrial-DNA. Evolution 45:423-432.

Newell, R. I. E. 1988a. Ecological changes in Chesapeake Bay: are they the result of overharvesting the eastern oyster (*Crassostrea virginica*)? Pp. 536-546 *in* M. P. Lynch, and E. C. Krome, eds. Understanding the Estuary: Advances in Chesapeake Bay Research. Chesapeake Research Consortium Publication 129 (CPB/TRS 24/88), Gloucester Point, VA.

Newell, R. I. E. 1988b. Understanding the Estuary: Advances in Chesapeake Bay Research *in* M. P. Lynch, and E. C. Krome, eds. Chesapeake Research Consortium.

Nielsen, R. 2001. Statistical tests of selective neutrality in the age of genomics. Heredity 86:641-647.

Nielsen, R. 2005. Molecular signatures of natural selection. Annual Review of Genetics 39:197-218.

Nordborg, M., and S. M. Krone. 2002. Separation of time scales and convergence to the coalescent *in* M. Slatkin, and M. Veuille, eds. Modern Developments in Theoretical Population Genetics: The Legacy of Gustave Malécot. Oxford University Press, Oxford.

North, E. W., Z. Schlag, R. R. Hood, M. Li, L. Zhong, T. Gross, and V. S. Kennedy. 2008. Vertical swimming behavior influences the dispersal of simulated oyster larvae in a coupled particle-tracking and hydrodynamic model of Chesapeake Bay. Marine Ecology-Progress Series 359:99-115.

O'Reilly, P. T., M. F. Canino, K. M. Bailey, and P. Bentzen. 2004. Inverse relationship between Fst and microsatellite polymorphism in the marine fish, walleye pollock (*Theragra chalcogramma*): implications for resolving weak population structure. Molecular Ecology 13:1799-1814.

Ohta, T. 1992. The nearly neutral theory of molecular evolution. Annual Review of Ecology and Systematics 23:263-286.

Palumbi, S. R. 1994. Genetic divergence, reproductive isolation, and marine speciation. Annual Review of Ecology and Systematics 25:547-572.

Palumbi, S. R. 2003. Population genetics, demographic connectivity, and the design of marine reserves. Ecological Applications 13:S146-S158.

Palumbi, S. R., F. Cipriano, and M. P. Hare. 2001. Predicting nuclear gene coalescence from mitochondrial data: The three-times rule. Evolution 55:859-868.

Palumbi, S. R., G. Grabowsky, T. Duda, L. Geyer, and N. Tachino. 1997. Speciation and population genetic structure in tropical Pacific sea urchins. Evolution 51:1506-1517.

Pamilo, P., S. Palsson, and O. Savolainen. 1999. Deleterious mutations can reduce differentiation in small, subdivided populations. Hereditas 130:257-264.

Paynter, K. T., and L. DiMichele. 1990. Growth rate of tray-cultured oysters (*Crassostrea virginica* Gmelin) in Chesapeake Bay. Aquaculture 87:289-297.

Paynter, K. T., L. L. Ellis, and S. H. Bishop. 1984. Cellular location and partial purification of the alanine aminotransferase from ribbed musssel gill tissue. Journal of Experimental Zoology 232:51-58.

Paynter, K. T., G. A. Karam, L. L. Ellis, and S. H. Bishop. 1985. Pyruvate dehydrogenase complex from ribbed mussel gill mitochondria. Journal of Experimental Zoology 236:251-257.

Peatman, E., H. Kucuktas, P. Li, C. He, J. Feng, X. Wei, and Z. Liu. 2003. Differentially expressed oyster (*Crassostrea virginica*) genes after exposure to mercury. Auburn University.

Peel, D., J. R. Ovenden, and S. L. Peel. 2004. NeEstimator: software for estimating effective population size, Version 1.3. Queensland Government, Department of Primary Industries and Fisheries.

Pesole, G., C. Gissi, A. De Chirico, and C. Saccone. 1999. Nucleotide substitution rate of mammalian mitochondrial genomes. J. Mol. Evol. 48:427-434.

Petersen, J. L., A. M. Ibarra, J. L. Ramirez, and B. May. 2008. An induced mass spawn of the hermaphroditic lion-paw scallop, *Nodipecten subnodosus*: Genetic assignment of maternal and paternal parentage. Journal of Heredity 99:337-348.

Peterson, M. A., and R. F. Denno. 1998. The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects. American Naturalist 152:428-446.

Piry, S., A. Alapetite, J.-M. Cornuet, D. Paetkau, L. Baudouin, and A. Estoup. 2004. GENECLASS2: a software for genetic assignment and first-generation migrant detection. J Hered 95:536-539.

Planes, S., and C. Fauvelot. 2002. Isolation by distance and vicariance drive genetic structure of a coral reef fish in the Pacific Ocean. Evolution 56:378-399.

Pogson, G. H., C. T. Taggart, K. A. Mesa, and R. G. Boutilier. 2001. Isolation by distance in the Atlantic cod, *Gadus morhua*, at large and small geographic scales. Evolution 55:131-146.

Pollak, E. 1983. A new method for estimating the effective population size from allele frequency changes. Genetics 104:531-548.

Powell, E. N., M. Kasschau, E. Chen, M. Koenig, and J. Pecon. 1982. Changes in the free amino acid pool during environmental stress in the gill tissue of the oyster, *Crassostrea virginica*. Comparative Biochemistry and Physiology 71A:591-598.

Pray, L. A., C. J. Goodnight, L. Stevens, J. M. Schwartz, and G. Y. Yan. 1996. The effect of population size on effective population size: An empirical study in the red flour beetle *Tribolium castaneum*. Genetical Research 68:151-155.

Ragone Calvo, L., 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.

Ragone Calvo, L., V. Harmon, and E. M. Burreson. 1997. Selection of oysters for resistance to two protozoan parasites. Journal of Shellfish Research 16:327-328.

Rand, D. M. 2001. The units of selection on mitochondrial DNA. Annual Review of Ecology and Systematics 32:415-448.

Rand, D. M., and L. M. Kann. 1996. Excess amino acid polymorphism in mitochondrial DNA: Contrasts among genes from *Drosophila*, mice, and humans. Molecular Biology and Evolution 13:735-748.

Rannala, B., and J. L. Mountain. 1997. Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences of the United States of America 94:9197-9201.

Raymond, M., and F. Rousset. 1995a. An exact test for population differentiation. Evolution 49:1280-1283.

Raymond, M., and F. Rousset. 1995b. Genepop (version 1.2) - Population genetics software for exact tests and ecumenicism. Journal of Heredity 86:248-249.

Reeb, C. A., and J. C. Avise. 1990. A genetic discontinuity in a continuously distributed species: mitochondrial DNA in the American oyster, *Crassostrea virginica*. Genetics 124:397-406.

Reece, K. S., W. L. Ribeiro, P. M. Gaffney, R. B. Carnegie, and S. K. Allen. 2004. Microsatellite marker development and analysis in the eastern oyster (*Crassostrea*

virginica): confirmation of null alleles and non-Mendelian segregation ratios. Journal of Heredity 95:346-352.

Riginos, C., and M. W. Nachman. 2001. Population subdivision in marine environments: the contributions of biogeography, geographical distance and discontinuous habitat to genetic differentiation in a blennioid fish, *Axoclinus nigricaudus*. Molecular Ecology 10:1439-1453.

Rose, C. G., K. T. Paynter, and M. P. Hare. 2006. Isolation by distance in the eastern oyster, *Crassostrea virginica*, in Chesapeake Bay. Journal of Heredity 97:158-170.

Rose, R. L. 1984. Genetic variation in the oyster, *Crassostrea virginica* (Gmelin), in relation to environmental variation. Estuaries 7:128-132.

Rosenberg, N. A., and A. E. Hirsh. 2003. On the use of star-shaped genealogies in inference of coalescence times. Genetics 164:1677-1682.

Rothschild, B. J., J. S. Ault, P. Goulletquer, and M. Héral. 1994. Decline of the Chesapeake Bay oyster population: a century of habitat destruction and overfishing. Marine Ecology Progress Series 111:29-39.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145:1219-1228.

Rousset, F. 2003. Inferences from spatial population genetics. Pp. 681-712 *in* D. J. Balding, M. Bishop, and C. Cannings, eds. Handbook of Statistical Genetics, 2nd ed. John Wiley & Sons, Chichester, West Sussex, England; Hoboken, NJ.

Rozas, J., J. C. Sánchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496-2497.

Ruzzante, D. E. 1998. A comparison of several measures of genetic distance and population structure with microsatellite data: bias and sampling variance. Canadian Journal of Fisheries and Aquatic Sciences 55:1-14.

Ruzzante, D. E., C. T. Taggart, and D. Cook. 1996. Spatial and temporal variation in the genetic composition of a larval cod (*Gadus morhua*) aggregation: cohort contribution and genetic stability. Canadian Journal of Fisheries and Aquatic Sciences 53:2695-2705.

Ryman, N., and L. Laikre. 1991. Effects of supportive breeding on the genetically effective population size. Conservation Biology 5:325-329.

Saavedra, C. 1997. Low effective sizes in hatchery populations of the European oyster (*Ostrea edulis*): Implications for the management of genetic resources. J. Shellfish Res. 16:441-446.

Sargsyan, O., and J. Wakeley. 2008. A coalescent process with simultaneous multiple mergers for approximating the gene genealogies of many marine organisms. Theoretical Population Biology 74:104-114.

Schmidt, P. S., and D. M. Rand. 2001. Adaptive maintenance of genetic polymorphism in an intertidal barnacle: Habitat- and life-stage-specific survivorship of Mpi genotypes. Evolution 55:1336-1344.

Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin: A software for population genetics data analysis. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.

Shanks, A. L. 1995. Mechanisms of cross-shelf dispersal of larval invertebrates and fish. Pp. 323-367 *in* L. R. McEdward, ed. Ecology of marine invertebrates. CRC, Boca Raton, Florida.

Shanks, A. L., B. A. Grantham, and M. H. Carr. 2003. Propagule dispersal distance and the size and spacing of marine reserves. Ecological Applications 13:S159-S169.

Shinde, D., Y. L. Lai, F. Z. Sun, and N. Arnheim. 2003. *Taq* DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)n and (A/T)n microsatellites. Nucleic Acids Research 31:974-980.

Shumway, S. E. 1996. Natural Environmental Factors. Pp. 467-513 *in* V. S. Kennedy, R. I. E. Newell, and A. F. Eble, eds. The Eastern Oyster *Crassostrea virginica*. Maryland Sea Grant, College Park, MD.

Siegel, D. A., B. P. Kinlan, B. Gaylord, and S. D. Gaines. 2003. Lagrangian descriptions of marine larval dispersion. Marine Ecology-Progress Series 260:83-96.

Simonsen, K. L., G. A. Churchill, and C. F. Aquadro. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. Genetics 141:413-429.

Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. Science 236:787-792.

Slatkin, M. 1993. Isolation by distance in equilibrium and nonequilibrium populations. Evolution 47:264-279.

Sorabella, L., and M. Luckenbach. 2003. A comparison of two oyster (*Crassostrea virginica*) stocks to determine suitability for use in oyster reef restoration in Virginia. Journal of Shellfish Research 22:355.

Soulé, M. E. 1980. Thresholds for survival: maintaining fitness and evolutionary potential. Pp. 151-170 *in* M. E. Soulé, and B. A. Wilcox, eds. Conservation Biology: An Evolutionary-Ecological Perspective. Sinauer Associates, Sunderland, MA.

Southworth, M., and R. Mann. 1998. Oyster reef broodstock enhancement in the Great Wicomico River, Virginia. Journal of Shellfish Research 17:1101-1114.

Spencer, C. C., J. E. Neigel, and P. L. Leberg. 2000. Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. Molecular Ecology 9:1517-1528.

Stajich, J. E., and M. W. Hahn. 2005. Disentangling the effects of demography and selection in human history. Molecular Biology and Evolution 22:63-73.

Storey, J. D. 2002. A direct approach to false discovery rates. Journal of the Royal Statistical Society Series B-Statistical Methodology 64:479-498.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595.

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.

Tankersley, R. A., L. M. McKelvey, and R. B. Forward. 1995. Responses of estuarine crab megalopae to pressure, salinity and light: implications for flood tide transport. Marine Biology 122:391-400.

Templeton, A. R., K. A. Crandall, and C. F. Sing. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132:619-633.

Templeton, A. R., and B. Read. 1983. The elimination of inbreeding depression in a captive herd of Speke's gazelle. Pp. 241-261 *in* C. M. Schonewald-Cox, S. M. Chambers, B. MacBryde, and L. Thomas, eds. Genetics and conservation: A reference for managing wild animal and plant populations. Benjamin/Cummings, Menlo Park, CA.

Turner, T. F., L. A. Salter, and J. R. Gold. 2001. Temporal-method estimates of *Ne* from highly polymorphic loci. Conservation Genetics 2:297-308.

Turner, T. F., J. P. Wares, and J. R. Gold. 2002. Genetic effective size is three orders of magnitude smaller than adult census size in an abundant, estuarine-dependent marine fish (Sciaenops ocellatus). Genetics 162:1329-1339.

Van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4:535-538.

van Oosterhout, C., W. G. Zijlstra, M. K. van Heuven, and P. M. Brakefield. 2000. Inbreeding depression and genetic load in laboratory metapopulations of the butterfly *Bicyclus anynana*. Evolution 54:218-225.

VIMS. Chesapeake Bay Oyster Population Estimation. Last modified Aug 31, 2003. http://www.vims.edu/mollusc/cbope/index.htm.

Vrijenhoek, R. C., S. E. Ford, and H. H. Haskin. 1990. Maintenance of heterozygosity during selective breeding of oysters for resistance to MSX disease. Journal of Heredity 81:418-423.

Vucetich, J. A., and T. A. Waite. 1998. Number of censuses required for demographic estimation of effective population size. Conservation Biology 12:1023-1030.

Wang, J. L. 2001. A pseudo-likelihood method for estimating effective population size from temporally spaced samples. Genetical Research 78:243-257.

Wang, J. L. 2005. MLNE version 1.1.

Wang, J. L., and N. Ryman. 2001. Genetic effects of multiple generations of supportive breeding. Conservation Biology 15:1619-1631.

Waples, R. S. 1989. A generalized approach for estimating effective population size from temporal changes in allele frequency. Genetics 121:379-391.

Waples, R. S. 1998. Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. Journal of Heredity 89:438-450.

Waples, R. S. 2005. Genetic estimates of contemporary effective population size: to what time periods do the estimates apply? Molecular Ecology 14:3335-3352.

Waples, R. S. 2006. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. Conservation Genetics 7:167-184.

Waples, R. S., and C. Do. 1994. Genetic risk associated with supplementation of Pacific Salmonids: captive broodstock programs. Canadian Journal of Fisheries and Aquatic Sciences 51:310-329.

Waples, R. S., and C. Do. 2008. LDNE: a program for estimating effective population size from data on linkage disequilibrium. Molecular Ecology Resources 8:753-756.

Wares, J. P., P. H. Barber, J. Ross-Ibarra, E. E. Sotka, and R. J. Toonen. 2006. Mitochondrial DNA and population size. Science 314:1388-1389.

Watterson, G. A. 1975. Number of segregating sites in genetic models without recombination. Theoretical Population Biology 7:256-276.

Wattier, R., C. R. Engel, P. Saumitou-Laprade, and M. Valero. 1998. Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). Molecular Ecology 7:1569-1573.

Wayne, M. L., and K. L. Simonsen. 1998. Statistical tests of neutrality in the age of weak selection. Trends in Ecology & Evolution 13:236-240.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358-1370.

Wirth, T., and L. Bernatchez. 2001. Genetic evidence against panmixia in the European eel. Nature 409:1037-1040.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97-159.

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

Wright, S. 1951. The genetical structure of populations. Annals of Eugenics 15:323-354.

Yang, H.-P., A. Y. Tanikawa, and A. S. Kondrashov. 2001. Molecular nature of 11 spontaneous *de novo* mutations in *Drosophila melanogaster*. Genetics 157:1285-1292.

Yu, Z., and X. Guo. 2003. Genetic linkage map of the eastern oyster *Crassostrea virginica* Gmelin. Biol. Bull. 204:327-338.

Zouros, E., S. M. Singh, and H. E. Miles. 1980. Growth rate in oysters: an overdominant phenotype and its possible explanations. Evolution 34:856-867.

Zurburg, W., and A. De Zwaan. 1981. The role of amino acids in anaerobiosis and osmoregulation in bivalves. Journal of Experimental Zoology 215:315-325.