Estimating the Effective Population Size of *Crassostrea virginica*

Honors Thesis
Presented to the College of Agriculture and Life Sciences
Ecology and Evolutionary Biology
Cornell University

In Partial Fulfillment of the Requirements for the
Biology Honors Program

By
Ashley Brown
May 2012

Supervisor: Dr. Matthew Hare

Abstract

Effective population size (N_e) is an important parameter used to estimate the magnitude of genetic drift, a major evolutionary force, that a population experiences. Across many taxa N_e is one tenth smaller on average than the census population size because of overlapping generations, uneven sex ratios, and demographic factors that increase the variance in reproductive success. Organisms that experience high fecundity, external fertilization, and larval dispersal, like the eastern oyster (*Crassostrea virginica*), are hypothesized to experience "sweepstakes reproduction", or a high stochastic variance in reproductive success among individuals, resulting in a much lower N_e relative to N_{census}. Previous tests utilizing direct genetic measures of N_e have supported sweepstakes in oysters but reported a wide range of local contemporary N_e from 20 to 1500. Here the objective was to further test the sweepstakes hypothesis by estimating N_e in oyster populations using two distinct, direct measures. Temporal samples of *C. virginica* from two eastern Florida localities were genotyped for nine microsatellite loci to estimate N_e. For several of the loci, populations were out of Hardy-Weinberg equilibrium due to heterozygote deficiencies, predictably due to null alleles. Effective population size was estimated using two methods, a two sample moments-based temporal estimation and a single sample linkage disequilibrium estimation. Despite broad variation in the finite estimates of N_e and high uncertainty due to infinite upper confidence limits, results from both methods are consistent with N_e between 500 and 10,000. These values are similar to past studies indicating a sweepstakes reproduction hypothesis, although not as extreme of a sweepstakes reproduction as some past work has suggested. However, predicted differences for allelic diversity and N_e were not detected between cohorts of spat

(early juveniles) and local adults, preventing a strong conclusion that *C. virginica* is in fact experiencing sweepstakes reproduction based upon the tested indicators in this study.

Introduction

To an uninformed eye, the homogenizing ocean currents and long distance dispersal potential of pelagic larvae would be suspected to keep marine populations well connected and genetically homogeneous. However, years of research (Hellberg 2009) have indicated that marine populations are often not homogeneous and even at the smallest spatial scales larval cohorts differ between seasons and from the local adults. This stochastic variance is surprising, not only because of the large dispersal potential of marine species, but also because population abundances are very large. Given that this spatial heterogeneity, sometimes referred to as 'chaotic patchiness' (Hellberg 2009), is observed more in some species than others, these patterns raise interesting questions related to the distribution and population structure in marine populations. What factors or life strategies cause marine populations to be so unevenly and poorly connected (Marshall 2010) when their planktonic larval duration allows them to disperse so far?

One of the most important parameters that have been used to describe the potential for population differentiation is effective population size (N_e). N_e describes the rate of evolution of a population through genetic drift. The strength of genetic drift, or of chance factors in allele frequency change, depends primarily on population size. It allows us to measure an important component of evolution, drift or the stochastic variation in a population, rather than being able to only concentrate on selection or adaptive forces. Large populations will experience a smaller amount of random change in allele frequencies between generations

because there are a larger number of individuals to pass on alleles. In addition, as with statistical sampling error, reproductive sampling error (or the effect of the unequal pairing up of mates) will be smaller when the next generation is drawn from a larger parental population. For contrasting reasons, small populations experience a larger amount of genetic drift. Populations experiencing strong genetic drift are relatively less able to adapt to environmental changes as alleles and heterozygosity may move in and out of the population rapidly. The N_e parameter is therefore useful in determining the vulnerability of a population to genetic drift for conservation and management planning (Hauser 2002, Hoarau 2005). For marine populations, this genetic drift is predicted to be a large component of the previously mentioned 'chaotic patchiness'. Because this stochastic heterogeneity exists among these species, populations that seem quite large on the outside may actually be a series of smaller populations once the genetic composition is taken into consideration. Therefore, marine populations might have smaller population sizes and experience higher genetic drift than we suspect. Effective population size parameter can be useful in beginning to characterize these populations that experience this 'chaotic patchiness'.

Census numbers do not inform us about genetic drift because demographic factors inherent to the species studied alter the severity of genetic drift. These demographic factors (such as variance in reproductive success or skewed sex ratio) reduce the genetic contribution of the parents in one generation to the offspring in the next. For example, a population with 99 males and 1 female will not have 100 effective breeders, but rather a much lower number as not all males will get to reproduce with the female in a single mating season. Therefore, the genetic composition will drift drastically compared to another population that has a more even sex ratio. The variation of these demographic factors across populations and species makes it

difficult to compare the strength of genetic drift as an evolutionary force (Charlesworth 2009). The concept of effective population size provides a way around this non-comparability by relating observations in real populations to a theoretical ideal population (Pollack 1983, Waples 1989, Jorde & Ryman 1995). In the simplest definition, effective population size is an estimate of the size of a theoretical population that would experience the same amount of genetic drift as the population being studied (Wright 1931). The N_e parameter characterizes many different aspects of the population, making it useful in comparison between species with different life strategies. A demographic approach to estimating effective population size is possible by adjusting the census number to account for known stochastic variation in reproductive success, skewed sex ratio, overlapping generations, and other factors influencing the effective number of breeders in a specific species (Nunney 1994, Hedgecock 1994, Hauser 2002, Turner 2002, Hoarau 2005, Charlesworth 2009). A variety of demographic factors could be predicted and have been observed to affect the effective population size, with a few examples given below. Within highly polygynous mating systems, N_e increases with increasing generation time because the variance among male reproductive success is equalized as there are more opportunities to mate (Nunney 1993). Effective population size can also be impacted by demographies after fertilization, including high fecundity, larval dispersal, and mortality. Both considerable temporal and spatial variation has been shown to exist among pelagic larvae (Li & Hedgecock 1998). Depending on any family specific dispersal or mortality, only a small number of individuals that contributed to the high fecundity of a population could end up producing the next cohort (Waples 2002). After consideration of all the possible demographic factors, the resulting population size estimate is formed that describes the number of individuals that will contribute genetically to the next

generation even though genetics has not been sampled directly. In an evolutionary genetic sense, this is the genetically *effective* population size.

Estimating N_e using the wide variety of potentially changing demographic factors makes the calculation extremely multi-faceted and laborious. While this is difficult in terrestrial species with generally “easy” life strategies, it is nearly impossible to estimate with marine species that normally have more complex life histories (for example: multiple life stages). However, molecular methods have made it possible to estimate this parameter more directly through the impact of drift across one or more generations. Genetic drift has predictable impacts on allelic correlations (linkage disequilibrium, LD) within cohorts and on the magnitude of allele frequency change between generations. Therefore, by measuring linkage disequilibrium and temporal changes in allele frequency, and accounting for sampling error, the observed strength of genetic drift can be related to the size of an ideal population with an equivalent amount of drift.

Genetic approaches to estimating effective population size differ primarily by time scale. Evolutionary effective population size can be estimated as a long term harmonic mean over many generations which characterize the historic population size of a species, but that is not a relevant framework here because most of the ‘chaotic patchiness’ seen in marine populations are observed over small spatial and temporal scales. A more contemporary estimate of N_e can be based on measures of genetic drift during a period of population sampling (Waples 1991). Contemporary N_e has been used to examine sweepstakes reproduction in past studies because it focuses on the most relevant near-term scale. This contemporary N_e can be estimated using the previously mentioned genetic patterns from a single sample or based on a comparison across temporal samples. Of the single sample

methods, the linkage disequilibrium (LD) method uses association among alleles at different loci to provide an estimate of N_e. In highly polymorphic unlinked loci, this method has been shown to be particularly useful to calculate contemporary N_e because linkage disequilibrium is more closely associated with recent mating (e.g. variance in reproductive success) rather than historical population structure (Hill 1981). Assuming that the sampled species experiences random mating and their genes are neutral and freely recombining at each locus, the alleles are expected to be randomly associated in large populations. However, reduced numbers of contributing parents will generate more allelic correlations as a consequence of genetic drift (Hill 1981). Thus, allelic correlations in a cohort sample inform us about the effective number of breeders in the parental generation. The temporal approach requires genetic comparisons across at least one generation and has been widely applied using an assortment of different methods (Palstra and Ruzzante 2008). The moment-based method uses two samples (Krimbas and Tsakas 1971, Nei and Tajima 1981, Pollak 1983, Waples 1989, Jorde & Ryman 2005) to estimate the temporal variance in allele frequencies. After accounting for sampling error, the variance provides an estimate of N_e for the time interval between samples. This method has been shown to be best suited with organisms having high juvenile mortality and limited effective population size (Waples 1989).

Contemporary genetic methods for estimating N_e require knowledge and caution to apply and interpret them properly. In large N_e populations with minimal drift, sampling error may swamp out the signal of genetic drift in these methods. Therefore, large populations showing essentially no measurable drift are indistinguishable from an infinitely large N_e (Waples 1989). Both methods also require more than a few loci for accurate estimates because

there is a large amount of evolutionary variance across loci. Similarly, large samples are needed to increase the precision of estimates (Waples 2006).

The assumptions of the LD method are that the population studied is stable, panmictic, and has no selection, migration, or mutation (England 2006). Beyond these assumptions, this method has also been shown to have particular sensitivities. Small sample sizes relative to the known N_e were shown to underestimate the actual effective population size value (England 2006). This small sample to N_e ratio causes the sampling error to affect the estimate much more than drift leading to a significant underestimation (Waples 2006). The LD method has also been shown to be sensitive to low-frequency alleles (England 2006). However, estimates with the LD method do not seem to be affected much by removal or addition of loci (although precision is reduced) or from population mixing and continuous migrants (Waples 2006). For populations with non-overlapping generations, the LD in a sample estimates the N_e in the parental generation. In age-structured populations with overlapping generations however, the LD estimate approximates the number of breeders for that cohort rather than the effective size of the population (Waples 2006).

The temporal method assumes constant N_e within the sampling interval, neutral alleles, non-overlapping generations, and no selection, mutation, or migration (Krimbas & Tsakas 1971, Nei & Tajima 1981). Overlapping generations can cause the estimate to be bias downwards (Waples 1989). Migration can skew the estimate either upwards or downwards depending on the source population but usually skews downward depending on the size of the temporal interval between the samples (Wang & Whitlock 2003, Fraser 2007). Temporal estimates benefit from both increasing the number of generations between samples, to about

3-5 generations (Waples 2006), and increasing the ratio of sample size to effective size (Nei & Tajima 1981, Jorde & Ryman 1995).

Characteristically in marine species, census population sizes can be enormous yet contemporary N_e is often much lower than the census population size (Crow 1955, Frankham 1995, Hauser & Carvalho 2008). One hypothesis used to explain this in species with high fecundity and larval dispersal is ‘sweepstakes reproduction’. Sweepstakes reproduction means that, as a result of high variance in reproductive success and stochastic juvenile mortality, only a small, random number of parents contributes to the next generation. Sweepstakes reproduction has been inferred in past studies by finding 1) reduced genetic variation in discrete cohorts compared to aggregate samples of adults from the local population (Hedgecock 2007) and 2) higher genetic heterogeneity over time in one location than seen spatially at relevant scales in relation to the expanse of larval dispersal (Li & Hedgecock 1998, Moberg & Burton 2000, Planes & Lenfant 2002, Rhodes 2003, Pujolar 2006, Hedgecock 2007, Lallias 2010). Nonetheless, other studies with strong statistical power to detect these patterns have not supported sweepstakes reproduction (Flowers 2002; Bernal-Ramirez 2003, Poulsen 2006, Calderon 2009, Taris 2009). Thus, the generality and magnitude of sweepstakes reproduction is largely still unknown among high fecundity marine invertebrates.

This study compares the effective population size of the eastern oyster (*Crassostrea virginica*) over several generations, in two locations, and with two estimation methods. *Crassostrea virginica* is a keystone member of the estuarine community across a broad species range in the western North Atlantic (Jackson et al 2001). *Crassostrea virginica* is a sessile organism that spawns gametes for external fertilization and has a subsequent planktonic larval stage lasting two to three weeks (Kennedy 1996). *Crassostrea virginica*

also experiences a relatively short generation time as a result of a one to two year span from fertilization to reproductive maturity (Hayes and Menzel 1981) and longevity of only a few years. Many studies with the goal of understanding ‘chaotic patchiness’ of genetic variation have concentrated on the larval dispersal and survivorship patterns of marine invertebrates and fish (Johnson & Black 1982; Avise 1994, Palumbi 1996) to explain the scale of genetic drift and population stochasticity that the populations experience. Through the use of N_e, this study has the potential to increase understanding of population structure and reproduction in high fecundity marine species like oysters (Buston 2009), including information supporting the suspected sweepstakes reproduction (Hedgecock 1989, Hedgecock 1994, Waples 2002, Hendrick 2005). Because of the large amount of previously accumulated data on the topic of N_e and sweepstakes reproduction, *Crassostrea virginica* is a particularly useful organism for this study.

Crassostrea virginica also presents challenges as a study organism though. First, this study will be affected by the overlapping generations of *C. virginica.* Secondly, *C. virginica* has an overlapping population structure distributed continuously along the eastern coast of Florida in a series of lagoons and estuaries (personal observation). While the oysters sampled belong to two largely separated populations (spatially and temporally), the extent of the migration and disruption of unique cohorts is difficult to calculate. The northern population has a more genetic homogeneity with its surrounding populations compared to the southern population (Hare & Avise 1996). Beyond those descriptions however, potential migration is hard to characterize. Depending on if the migration is temporally or spatial episodic, there may be a variance in N_e over time. Effects on the N_e estimations by migrants would depend on the characteristics of the source population—whether it be large or small, with high or low

allelic diversity relative to the population or cohort under consideration, and if the source population is continuously providing migrants (Fraser 2007, Palstra & Ruzzante 2008).

Recently, studies have used genetic methods to examine the N_e of *C. virginica* and similar species. Hedgecock 1992 was the first of these studies and estimated the N_e of *C. virginica* to be extremely low, between 10 and 20 based on temporal analysis of allozyme data in the Chesapeake Bay. Using microsatellite markers in Chesapeake Bay populations Rose et al 2006 made a temporal N_e estimate between 500 and 1500. Most recently, Hedgecock 2007 and He et al (In press) have shown lower allelic diversity and N_e in spat samples relative to adults within single estuaries. These studies have reported a wide variety of estimates and hypotheses that point towards a low N_e/N_c ratio and sweepstakes reproduction, but the body of knowledge still lacks precision in relation to how extreme sweepstakes reproduction is generally and how consistently this reproduction strategy manifests. The present study will contribute by comparing across generations, locations, and two different N_e estimation methods using *C. virginica* from the eastern coast of Florida. I will examine both the N_e and allelic diversity in several different samplings of populations over time. The decrease in N_e relative to census size from sweepstakes reproduction is expected to be large (e.g., an N_e/N_c ratio estimated to be $1{,}517/10^9$ in Chesapeake oysters [Rose 2006], as opposed to an average ratio of 0.1 across diverse taxa [Frankham 1995]). However, census numbers in Florida assumed to be very large ($>10^6$) but unknown, so sweepstakes reproduction will be evaluated with respect to previous N_e estimates in C. virginica ($N_e = 20 – 1500$) (Hedgecock 1992, Rose 2006, He In Press). According to the previous work done on *C. virginica* and similar marine species, I am expecting indication of sweepstakes reproduction to be expressed as (i) a small N_e, (ii) a reduction in N_e in juvenile

cohorts relative to adults, and (iii) a decrease in allelic diversity in juvenile cohorts relative to local adults. Through testing for these defining characteristics of sweepstakes reproduction, this study will contribute to the growing body of knowledge on the sweepstakes reproduction hypothesis of *C. virginica*.

Methods

Juvenile and Adult Sampling

Juvenile samples (also called spat) were allowed to settle on oyster shells in mesh bags set out along a transect parallel to the eastern coast of Florida. Adult samples were collected on the intertidal shore near the juvenile collection locations. Samples were collected at both the University of Florida Whitney Laboratory at 29°40.209' N, 81°12.940' W (WHL) and Deerfield Beach at 26°17.750' N, 80°4.846' W (DFB) (Figure 1). Adult samples were collected in May 2007 and spat samples were collected in both June 2007, June 2009 (WHL only), and July 2009 (DFB only) by Hare Lab members. DNA was extracted using DNeasy 96 Blood & Tissue Kit following the manufacturer's protocol.

DNA Amplification

Each sample was amplified at nine polymorphic unlinked (Ximing Guo, Rutgers Univ., unpublished data) microsatellite loci; CVi9, CVi2i23, CVi2k14, CVi12, RUCV045, RUCV060, RUCV063, RUCV131, and RUCV374. Further description of the loci and primers are located in Table 1. CVi9 amplification was performed with 1.33ng/ul of template, 1X of buffer, 3uM BSA, 0.1mM dNTPS, 1.75mM $MgCl_2$, 0.2uM of forward and reverse primer (CVi9F diluted with non-fluorescent primer 1:5, CVi9R-PIG), and 0.375 units of taq. CVi9

was the only primer that required pig-tailing (PIG) in an effort to reduce microsatellite stutter. CVi2i23 amplification was performed with 1.33ng/ul of template, 1X of buffer, 3uM BSA, 0.1mM dNTPS, 1.5mM $MgCl_2$, 0.2uM of forward and reverse primer (2i23F 1:5, 2i23R), and 0.25 units of taq. CVi2k14 amplification was performed with 1.33ng/ul of template, 1X of buffer, 3uM BSA, 0.1mM dNTPS, 1.75mM $MgCl_2$, 0.2uM of forward and reverse primer (2k14F 1:5, 2k14R), and 0.25 units of taq. CVi12 amplification was performed with 1.33ng/ul of template, 1X of buffer, 3uM BSA, 0.1mM dNTPS, 1.5mM $MgCl_2$, 0.2uM of forward and reverse primer (primer F 1:5, primer R), and 0.375 units of taq. RUCV060, RUCV063, and RUCV374 were amplified using a multiplex (Multiplex A) containing 1.33ng/ul of template, 3uM BSA, 1X Qiagen multiplex kit, and 0.2uM of each forward and reverse primer. RUCV045 and RUCV131 were amplified in using a multiplex (Multiplex B) containing 1.33ng/ul template, 3uM BSA, 1X Qiagen multiplex kit, and 0.2uM of each forward and reverse primer. All reactions were performed in 7.5uL volume. PCR amplification was performed on MJ Research PTC-225 Thermal Cycler using five different PCR programs. CVi9 and CVi12 used the same program with an initial denaturization step (95C for 1 min), denaturization-annealing-polymerization step (95C, 30s; 62C -1C/cycle, 30s; 1C/sec, 72s; 72C, 30s; repeat previous nine times; 95C, 30s; 52c, 30s; 1C/sec to 72; 7C, 1min; go to beginning 29 times) and final elongation step (72C, 5.5min; 24C, 2min). CVi2i23 used a program with an initial denaturization step (95C, 1min), denaturization-annealing-polymerization step (95C, 30s; 51.5C, 30s; 1C/sec to 72; 72C, 1min; repeated 30 times) and final elongation step (70C, 5.5min; 24C, 2min). CVi2k14 used a program with an initial denaturization step (95C, 1min), denaturization-annealing-polymerization step (95C, 30s; 52C, 60s; 72C, 1min; repeated 30 times) and final elongation step (72C, 5.5min; 24C, 2min).

Multiplex A used a program with an initial denaturization step (95C, 15min), denaturization-annealing-polymerization step (94C, 30s; 53C, 90s; 72C, 60s, repeated 24 times), and final elongation step (72C, 30min). Multiplex B used the same program as Multiplex A, but with an annealing temperature of 57C.

Microsatellite Genotyping

PCR products were prepared for fragment analysis using 1mL of HiDi to 14ul (24ul when coloading CVi9, CVi2i23, and CVi2k14) of LIZ GeneScan™ 500 LIZ™ Size Standard. 1ul of PCR product (1.5uL for 2i23) was added to 9ul of HiDi/LIZ mixture. Fragment Analysis was performed using an Applied BioSystems 3730xl DNA Analyzer at Cornell University Life Sciences Core Laboratories Center. Results were scored using Genemapper v4.0. Binsets used for each locus were previously designed by binning alleles into length classes by eye by Hare Lab members over several different projects. All automated allele assignments were confirmed by eye. Samples were redone if the highest peak was under 100 (200 for CVi9 and CVi2i23), there were more than 2 alleles, or alleles were not clear. For CVi12 the threshold for allele calls was changed to 20 percent. Alleles were usually called if they were more than 50 percent the height of the called homozygous allele or deleted if they were less than 50 percent the height of the smallest heterozygous allele in the sample. When adenylation was encountered, the larger peak was always called. Individuals that had missing data were not included in results.

Data Quality Control

Genotyping error was estimated from replicate genotypes by calculating the number of allele calls that were mismatched, divided by the total number of allele calls (two per genotype). MICRO-CHECKER 2.2.3 was used to test for genotypic patterns potentially indicative of scoring error, including allele dropout (Wattier 1998), stuttering (Shinde 2003), and null alleles (Brookfield 1996).

Population Genetics

GenePop v4 (Raymond & Rousset 1995) was used to test Hardy-Weinberg equilibrium (HWE) using Markov chain estimation of exact p-values with 5000 iterations per batch (Guo & Thompson 1992). In addition to conventional Hardy-Weinberg (HW) tests of population samples, the R script described in Morin et al 2009 was used to implement leave-one-out Hardy-Weinberg tests, testing individuals to determine if any caused, by themselves, a significant deviation from Hardy-Weinberg expectations. GenePop was also used to estimate F_{is}, a measure of inbreeding or Hardy-Weinberg deviation that indicates heterozygote deficiency when positive. Significance of heterozygote deficiency was tested using one-sided randomization tests and evaluated with an adjusted alpha of 0.00093 to account for multiple tests. FSTAT 2.9.3.2 (Goudet 2001) was used to determine heterozygote deficits based on expectations from null alleles, allelic richness values, and the significance of allelic richness differences between groups of samples by permutation.

Effective Population Size estimation

Oyster generation time in Florida is one to several years, depending on the population (Hare unpublished data). In this study, both populations are assumed to have a two year generation time. Single sample effective population size based on linkage disequilibrium was estimated using LDNe v1.31 (Waples 2006) using the random mating option and 0.02 critical value to correct for known biases from rare alleles. Temporal-based Ne was estimated using the moments based approach in NeEstimator v1.3. (Peel 2004).

Results

Data Quality Control

A total of 534 oysters were each genotyped at nine microsatellite loci; 254 from DFB (99 from June 2007, 100 from July 2009, and 55 adults from May 2007) and 280 from WHL (108 from June 2007, 112 from June 2009, and 60 adults from May 2007). Each locus was duplicate genotyped in 5.0% - 37.6% of samples to estimate locus specific errors ranging from 0% to 5.77% (Table 2). Although citations of error rates are uncommon in microsatellite papers, reasonable error rates seem to fall within a range around 2.0% (Bonin 2004).

The presence of null alleles was identified by MICRO-CHECKER in all populations at CVi9 and CVi12 and in some populations at RUCV063, RUCV374, and RUCV045. Stuttering error in CVi9 was suggested for two populations and large allele dropout was suggested for RUCV374 in one population. Analysis of HWE showed significant deviations in some loci and populations (Table 3). All populations were significantly out of HWE for CVi9 and CVi12, and 5 out of 6 populations for RUCV374. Further analysis showed that this deviation was most likely due to heterozygote deficiency (Table 4). CVi9, CVi12, RUCV060,

and RUCV063 had all positive F_{is} values that were significantly positive, which reinforced the finding of null alleles presented by Micro-checker.

The leave-one-out HWE analysis (Morin 2009) showed that 8 individuals in the WHL adult population were significantly skewing the HW genotype proportions (Morin 2010). The individuals seemed to be significantly affected by allele drop out at the RUCV060 and RUCV063 loci. These individuals were removed so that the presented data for population differentiation and effective population size estimates were not skewed by allele drop out effects (all tables other than 1 and 2). Additionally, the programs used for estimation of effective population size excluded alleles that were less than 2% frequency, reducing the effect of the higher than average error rates.

Population Genetics

The nine loci had a wide range of allelic diversity (Table 1), but the most diverse loci had many rare alleles. Thus, in terms of allelic richness (Table 5), RUCV045 had the highest A_r with 30.3 and 2k14 with the lowest of 5.0. For each locus, A_r was fairly consistent among all populations. Allelic richness, H_o, H_s, and F_{is} were compared between the 2007 adults (WHL57a and DFB57a samples), all spat (DFB67s, DFB79s, WHL67s, and WHL69s samples), and 2007 spat (DFB67s and WHL67s samples) (Table 6). None of the comparisons showed significant changes in allelic richness, observed heterozygosity, expected heterozygosity, of F_{is} values between spat and adults.

Effective Population Size Estimation

After HWE, F_{is}, and A_r characterization (Tables 3, 4, 5, 6), CVi9, 2k14, and CVi12 showed to be the least informative and most problematic loci, so population N_e estimations were done with and without these loci to see if there was any effect on the results. Single-sample estimates of N_e (Tables 7, 8, 9) were often unattainable with the current data using the LD method. Several estimates produced by LDNe were negative values. This was as a result of the sample size to effective population size ratio being too small and therefore the sampling error overpowered the effect of drift within the sample. Those were assumed to be infinity for ease of conceptualization (Laurie-Ahlberg & Weir 1979, Hill 1981) but marked as a ? in the table to distinguish between the true infinite estimations of the temporal samples.

Finite effective population size estimates from LDNe ranged from 400 to 8900. The lower bounds of the confidence intervals fell mostly in the low hundreds although one ranged as high as 33224. Removal of low quality loci created more finite estimates, but finite estimates and lower confidence limits generally decreased as more loci were removed.

For both the DFB and WHL populations, effective population size estimated from LD in adults was 400 – 500 (lower limits of 130 – 200). The spat for DFB seemed to have a higher effective population size in the thousands, with a decrease from 2007 to 2009 in both the estimates and lower confidence bounds. No finite point estimates were achieved for WHL spat. Focusing on the lower confidence limits for the WHL spat, the effective population size was never smaller than 600 and showed a four to six-fold increase from 2007 to 2009. However, most of the estimates and lower confidence limits for both populations indicate effective population sizes between 500 and 10,000.

Finite effective population size estimates based on temporal comparisons (Table 10, 11, 12) ranged from 143 to 4644. The lower bounds of the confidence intervals were mostly under 500 and all were under 2000. Removal of low quality loci changed most point estimates to infinity, while finite estimates and lower confidence limits increased as more loci were removed.

For both populations, the estimates (determined either by the finite estimate values or the lower confidence intervals) increased with an enlargement in the sampling interval. While the set of estimates that used all the loci produced a greater number of finite estimates, the lower bounds were fairly constant throughout all estimates. However for simplicity, the estimates including all loci were considered accurate and were used in analysis (Table 10). DFB had a slightly smaller effective population size than WHL (lower limit means for DFB and WHL were 120 vs. 604, respectively) Finite estimates predicted an effective population size of between 100 and 700. However, estimates and lower confidence limits for both populations indicate effective population sizes between 500 and 3000.

Discussion

Past studies have suggested that oyster life history, like that of many high-fecundity species with larval dispersal, entails a reproduction sweepstakes. This dramatically elevates variance in reproductive success, thereby lowering N_e (Li & Hedgecock 1998, Moberg & Burton 2000, Planes & Lenfant 2002, Rhodes 2003, Pujolar 2006, Hedgecock 2007, Lallias 2010). Populations that undergo sweepstakes reproduction were predicted to display three patterns as a consequence: (i) very low effective population size, (ii) lower effective population size of spat cohorts compared to the population size of local adults, and (iii) lower

allelic richness in spat cohorts compared to the allelic richness in local adults. This study tested these three predictions in Florida populations where temporal and spatial sampling of adults and spat enabled estimation of contemporary N_e in two different populations. Analysis of microsatellite genotypes, using both single-sample and temporal estimation methods, produced population size estimates with low precision and mostly upper confidence limits of infinity. Despite this, all point estimates obtained do support sweepstakes reproduction based on the first prediction: estimated effective population size was quite small similar to the estimates in other studies that hypothesized sweepstakes reproduction in *C. virginica*. In contrast, neither effective population size nor allelic richness was lower in spat cohorts relative to local adults.

The effective population size estimates of both LDNe and NeEstimator were critical to the analysis of the reproduction patterns of these *C. virginica* populations. Using both the single sample and temporal approach induced some variation into the estimates although both methods produced similar estimates and similar amounts of variation. Many of the estimates were potentially unreliable because of the lack of point estimates and/or infinite upper bounds on the confidence intervals. However, even when the point estimate was indeterminate because the genetic drift signal was low relative to sampling error, lower confidence limits were generally estimated. Comparing lower confidence interval bounds which reflected the minimum N_e (Fraser 2007), allowed me at the very least to compare the N_e values to some of the very low previously discovered values (Hedgecock 1992). With an infinite upper confidence bound however, the true estimate could be very large. Therefore caution should be exercised in interpreting these effective population size estimates.

The methods used assume HWE for the samples analyzed which many of the samples did not follow for several loci. Most of the Hardy-Weinberg deviation seemed to be due to null alleles and resulted in large heterozygote deficiencies at some loci. Null alleles are a common problem reported in microsatellite studies with oysters (Rose et al 2006, He et al) As expected, removal of these loci did not change LD-based population size estimates much but reduced the precision. However it is surprising that these loci were determined to be poor quality and yet still did not affect the estimations much. This appears to suggest that the null alleles were equally distributed across samples (Jehle 2001, Zeller 2008). The temporal analysis did show an increase in estimation values as loci were removed. As a result of this observed trend, the estimates using all the loci were considered the most accurate and were the estimates used for analysis, as they utilized the most data and were the most precise (Waples 1989).

Furthermore, comparing adult and spat populations introduced variability into the results. While the spat populations are assumed (to the best of the samplers abilities) to have belonged to a single cohort, adult samples likely included individuals from multiple cohorts across several generations. The LD method attributes all linkage disequilibrium to genetic drift which may not have been the case if multiple cohorts were used in a single sample. In this case, the LD could have been due to factors such recent selection or migration as well as genetic drift, although past analysis of this method shows that migration does not overpower drift effects in most cases (Waples 2006). Overlapping generations, on the other hand, do tend to cause the LD method to estimate the number of breeders of that cohort instead of the effective population size (Waples 2006). The temporal analysis also attributes all change in allele frequency to genetic drift which also may not have been if multiple cohorts were used

in a single sample or migration occurred between temporal samples. This method has been shown to be less severely impacted than the LD method by multiple cohorts in a sample (as long as they are separated by multiple generations), however these mixed samples still cause a depression in N_e as a result of overlapping generations (Waples 1989). This temporal analysis requires at least one generation to be between samples and also assumes discrete generation time (Waples 1991). In species that have overlapping generations, the number of generations between a pair of samples is difficult to estimate, but it has been shown that inaccurate values of generations do not cause huge biases in the estimate when small (Hare 2011). On the other hand, the small generation number between the temporal samples predictably had an effect on the temporal estimates, as this method is shown to depressed estimates as a result of small generation number separating samples (Waples 2006). The sample sizes for adults were also only ~50% of that for the spat causing potential skews of the accuracy of the estimates. Analysis of the LD method has shown that reasonable precision of estimates require that the sample size should be greater than or equal to the predicted Ne. (England 2006). Assuming that the effective population size from the calculated estimates is correct, samples sizes of 60 and 55 adults oysters are significantly smaller than the N_e, giving us probable evidence that the estimates have less than desirable precision.

Despite these reservations and violations of assumptions, the estimated effective population sizes of DFB and WHL are approximately between 500 and 10,000 oysters and are in agreement with estimates (of about 1,500) by Rose et al. (2006) for oysters in a Chesapeake tributary. Additionally, He et al found *C. virginica* N_e estimates (of upwards of 50,000) to be even higher than the Rose et al paper, which are consistent with the range of my study's findings.

While effective population size estimates with the single sample LD and temporal methods in many cases provided infinite values, the few finite estimates provide some insight into the range of possible effective population sizes given by the data. These finite estimates put the effective population size between 100 and 10,000 oysters. This caused the effective population size of the populations to seem to still fall at a number much lower than the assumed census size magnitude of $10^6 – 10^8$ (Ross & Luchenbach 2009). However, these rough Ne estimations and predictions of a sweepstakes reproduction were supported by previous studies done on *Crassostrea* virginica and other similar species. While effective population sizes varied considerably among and within previous studies (Hedgecock 1992, Rose 2006, Hedgecock 2007, He In Press) many of the estimates seem to agree with the general magnitude of values obtained for the N_e estimates, as well as the variability seen in the estimates and the confidence intervals. While this is not strong evidence for sweepstakes because of the uncertainty of the estimates, it does give some data into supporting and gauging the severity of the sweepstakes reproduction in *C. virginica*.

Allelic richness was also hypothesized to be an indicator of sweepstakes reproduction with an expectation of reduced allelic richness in spat cohorts as compared to adult populations (Hedgecock 1994). No significant differences in allelic richness were found, and in many cases the allelic richness of the spat was larger than that of the adults. As a result, the sweepstakes hypothesis cannot be confirmed with this indicator. The samples themselves may have skewed this result with respect to the mixed cohort nature of the adult samples. The adults probably belonged to several cohorts of oysters while the spat only belonged to a single cohort, impacting the diversity of alleles in the sample to an unknown degree depending on the diversity and number of cohorts contained in a sample.

Finally, seeing a smaller N_e in the spat populations as compared to adults was the third test of sweepstakes reproduction. Here, for the majority of the estimated population size estimates, the adults had much lower N_e estimates than the spat. However, as discussed earlier, mixed cohort samples from overlapping generations may have depressed the N_e estimates of these samples, especially in the temporal samples where there are small numbers of generations separating the samples (Waples 2006). Additionally, migration among differentiated populations could have caused LD that is not caused simply by genetic drift of a population as these populations have the potential to be highly continuous (personal observation). Therefore, by looking at the data given and the suspicions of violated assumptions, conclusions or inferences that the effective population size of the spat is less than that of the adults cannot be confirmed.

In summary, the results suggest that the effective population sizes of *Crassostrea virginica* are small, similar to previous estimates that suggested a sweepstakes reproduction hypothesis. Effective population size estimates of these Florida populations were found somewhere between a few hundred and 10,000, which are consistent with the conclusions made by Rose et al (2006) and He et al (In Press). Furthermore this is also consistent with Rose and He's conclusion that the very extreme sweepstakes reproduction presented by Hedgecock (1992) is probably not the case in wild populations. While conclusions cannot be made definitely about sweepstakes reproduction processes in Florida oysters, as a result of large variation among estimates and violation of assumptions by the samples which may have affected the results, evidence presented in this study aids us in the ability to distinguish the severity of a possible sweepstakes reproduction life strategy along with a growing body of other studies.

Acknowledgements

I thank Dr. Matthew Hare for being my thesis advisor and helping me extensively with writing, software utilization, and guidance in sample processing. Also, special thanks to Harmony Borchardt-Wier for assistance in sample processing. Finally thanks to Laura Eierman, Dana Hills, Katrina Krebs, and Ellen Meyer for reviewing and editing drafts of my manuscript

References

Árnason, V. (2004). Coding and consent: Moral challenges of the database project in Iceland. *Bioethics, 18*(1), 27-49.

Avise, J. C. (1994). *Molecular markers, natural history and evolution. chapman and hall, new york*. New York: Chapman and Hall.

Bernal-Ramírez, J. H., Adcock, G. J., Hauser, L., Carvalho, G. R., & Smith, P. J. (2003). Temporal stability of genetic population structure in the new Zealand snapper, pagrus auratus, and relationship to coastal currents. *Marine Biology, 142*(3), 567-574.

Bonin, A., Bellemain, E., Bronken, E., Pompanon, F., Brochmann, C., & Taberlet, P. (2004). How to track and assess genotyping errors in population genetics studies. *Molecular Ecology, 13*(11), 3261-3273.

Brookfield , J. F. (1996). A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology, 5*(3), 453-455.

Brown, B. L., Franklin, D. E., Gaffney, P. M., Hong, M., Dendanto, D., & Kornfield, I. (2000). Characterization of microsatellite loci in the eastern oyster, crassostrea virginica. *Molecular Ecology, 9*(12), 2216.

Buston, P. M., Fauvelot, C., Wong, M. Y. L., & Planes, S. (2009). Genetic relatedness in groups of the humbug damselfish dascyllus aruanus: Small, similar-sized individuals may be close kin. *Molecular Ecology,18*(22), 4707-4715.

Calderon, I., Palacin, C., & Turon, X. (2009). Microsatellite markers reveal shallow genetic differentiation between cohorts of the common sea urchin paracentrotus lividus (lamarck) in northwest mediterranean.*Molecular Ecology, 18*(14), 3036-3049.

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews. Genetics, 10*(3), 195-205.

Crow, J. F., & Newton E. Morton. (1955). Measurement of gene frequency drift in small populations. *Evolution, 9*(2), pp. 202-214.

Dewoody, J., Nason, J. D., & Hipkins, V. D. (2006). Mitigating scoring errors in microsatellite data from wild populations. *Molecular Ecology Notes, 6*(4), 951-957.

England, P., Cornuet, J., Berthier, P., Tallmon, D., & Luikart, G. (2006). Estimating effective population size from linkage disequilibrium: Severe bias. *Conservation Genetics, 7*(2), 303-308.

Flowers, J. M., Schroeter, S. C., & Burton, R. S. (2002). The recruitment sweepstakes has many winners: Genetic evidence from the sea urchin strongylocentrotus purpuratus. *Evolution, 56*(7), 1445-1453.

Frankham, R. (1995). Effective population size/adult population size ratios in wildlife: A review. *Genetic Research, 66*, 95-107.

Fraser, D. J., Hansen, M. M., Østergaard, S., Tessier, N., Legault, M., & Bernatchez, L. (2007). Comparative estimation of effective population sizes and temporal gene flow in two contrasting population systems.*Molecular Ecology, 16*(18), 3866-3889.

Goudet, J. (1995). FSTAT (version 1.2): A computer program to calculate F-statistics. *Journal of Heredity, 86*(6), 485-486.

Guo, X. *Unpublished material.* Unpublished manuscript.

Guo, S. W., & Thompson, E. A. (1992). Performing the exact test of hardy-weinberg proportion for multiple alleles. *Biometrics, 48*(2), pp. 361-372.

Hare, M. P., Nunney, L., Schwartz, M. K., Ruzzante, D. E., Burford, M., Waples, R. S., et al. (2011). Understanding and estimating effective population size for practical application in marine species management; entendiendo y estimando el tamaño poblacional efectivo para aplicación práctica en el manejo de especies marinas. *Conservation Biology, 25*(3), 438-449.

Hare, M. P., & Avise, J. C. (1996). Molecular genetic analysis of a stepped multilocus cline in the american oyster (crassostrea virginica). *Evolution, 50*(6), pp. 2305-2315.

Hauser, L., Adcock, G. J., Smith, P. J., Bernal Ramírez, J. H., & Carvalho, G. R. (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, 99*(18), 11742-11747.

Hauser, L., & Carvalho, G. R. (2008). Paradigm shifts in marine fisheries genetics: Ugly hypotheses slain by beautiful facts. *Fish and Fisheries, 9*(4), 333-362.

Hayes, P. F., & Menzel, R. W. (1981). The reproductive cycle of early setting crassostrea virginica (gmelin) in the northern gulf of mexico, and its implications for population recruitment. *The Biological Bulletin,160*(1), 80-88.

He, Y., Ford, S., Bushek, D., Powell, E., Bao, Z., & Guo, Z. (In Press). Effective population sizes of the eastern oyster *crassostrea virginica* (gmelin) populations in delaware bay, USA.

Hedgecock, D. (1994). Does variance in reproductive success limit effective population sizes of marine organisms. In A. R. Beaumont (Ed.), *Genetics and evolution of aquatic organisms* (pp. 122-134). London: Chapman & Hall.

Hedgecock, D., Launey, S., Pudovkin, A., Naciri, Y., Lapègue, S., & Bonhomme, F. (2007). Small effective number of parents (N_b) inferred for a naturally spawned cohort of juvenile european flat oysters *ostrea edulis*. *Marine Biology, 150*(6), 1173-1182.

Hedgecock, D. (1986). Is gene flow from pelagic larval dispersal important in the adaptation and evolution of marine invertebrates? *Bulletin of Marine Science, 39*(2), 550-564.

Hedgecock, D., Chow, V., & Waples, R. S. (1992). Effective population numbers of shellfish broodstocks estimated from temporal variance in allelic frequencies. *Aquaculture, 108*(3–4), 215-232.

Hedgecock, D., & Pudovkin, A. I. (2011). Sweepstakes reproductive success in highly fecund marine fish and shellfish: A review and commentary. *Bulletin of Marine Science, 87*(4), 971-1002.

Hedrick, P. (2005). LARGE VARIANCE IN REPRODUCTIVE SUCCESS AND THE Ne/N RATIO. *Evolution, 59*(7), 1596-1599.

Hellberg, M. E. (2009). Gene flow and isolation among populations of marine animals. *Annual Review of Ecology and Systematics, 40*, 291.

Hill, W. G. (1981). Estimation of effective population size from data on linkage disequilibrium. *Genetic Research, 38*, 209-215.

Hoarau, G., Boon, E., Jongma, D. N., Ferber, S., Palsson, J., Van der Veer, H. W., et al. (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*(1562), 497-503.

Jackson, J. B. C., Kirby, M. X., Berger, W. H., Bjorndal, K. A., Botsford, L. W., Bourque, B. J., et al. (2001). Historical overfishing and the recent collapse of coastal ecosystems. *Science, 293*(5530), 629-637.

Jehle, R., Arntzen, J. W., Burke, T., Krupa, A. P., & Hödl, W. (2001). The annual number of breeding adults and the effective population size of syntopic newts (triturus cristatus, T. marmoratus). *Molecular Ecology,10*(4), 839-850.

Johnson, M. S., & Black, R. (1984). Pattern beneath the chaos: The effect of recruitment on genetic patchiness in an intertidal limpet. *Evolution, 38*(6), pp. 1371-1383.

Jorde, P. E., & Ryman, N. (1995). Temporal allele frequency change and estimation of effective size in populations with overlapping generations. *Genetics, 139*(2), 1077-1090.

Kennedy, V. S. (1996). The biology of larvae and spat. In V. S. Kennedy, Newell R.I.E & A. F. Eble (Eds.), *The eastern oyster: Crassostrea virginica* (pp. 371-422). College Park, MD: Maryland Sea Grand.

Krimbas, C. B., & Tsakas, S. (1971). The genetics of dacus oleae. V. changes of esterase polymorphism in a natural population following insecticide control-selection or drift? *Evolution, 25*(3), pp. 454-460.

Lallias, D., Taris, N., Boudry, P., Bonhomme, F., & Lapegue, S. (2010). Variance in the reproductive success of flat oyster *ostrea edulis* L. assessed by parentage analyses in natural and experimental conditions.*Genetics Research, 92*(03), 175.

Laurie-Ahlberg, C. C., & Weir, B. S. (1979). Allozyme variation and linkage disequilibrium in some laboratory populations of drosophila melanogaster. *Genetics, 92*, 1295-1314.

Li, G., & Hedgecock, D. (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*(4), 1025-1033.

Marshall, D. J., Monro, K., Bode, M., Keough, M. J., & Swearer, S. (2010). Phenotype?environment mismatches reduce connectivity in the sea. *Ecology Letters, 13*(1), 128-140.

Moberg, P. E., & Burton, R. S. (2000). Genetic heterogeneity among adult and recruit red sea urchins, *strongylocentrotus franciscanus*. *Marine Biology, 136*(5), 773-784.

Morin, P. A., Leduc, R. G., Archer, F. I., Martien, K. K., Huebinger, R., Bickham, J. W., et al. (2009). Significant deviations from hardy?weinberg equilibrium caused by low levels of microsatellite genotyping errors.*Molecular Ecology Resources, 9*(2), 498-504.

Morin, P. A., Martien, K. K., Archer, F. I., Cipriano, F., Steel, D., Jackson, J., et al. (2010). Applied conservation genetics and the need for quality control and reporting of genetic data used in fisheries and wildlife management. *Journal of Heredity, 101*(1), 1-10.

Nei, M., & Tajima, F. (1981). Genetic drift and estimation of effective population size. *Genetics, 98*(3), 625-640.

Nunney, L. (1993). The influence of mating system and overlapping generations on effective population size. *Evolution, 47*(5), pp. 1329-1341.

Nunney, L., & Elam, D. R. (1994). Estimating the effective population size of conserved populations; estimacion del tamaño poblacional efectivo de poblaciones protejidas. *Conservation Biology, 8*(1), 175-184.

Ovenden, J. R., Peel, D., Street, R., Courtney, A. J., Hoyle, S. D., Peel, S. L., et al. (2007). The genetic effective and adult census size of an australian population of tiger prawns (penaeus esculentus). *Molecular Ecology, 16*(1), 127-138.

Palstra, F. P., & Ruzzante, D. E. (2008). Genetic estimates of contemporary effective population size: What can they tell us about the importance of genetic stochasticity for wild population persistence? *Molecular Ecology, 17*(15), 3428-3447.

Palumbi, S. R. (1996). Macrospatial genetic structure and speciation in marine taxa with high dispersal abilities. In J. D. Ferrais, & S. R. Palumbi (Eds.), *Molecular zoology: Advances, strategies and protocols* (pp. 101-117). New York: Wiley-Liss.

Planes, S., & Lenfant, P. (2002). Temporal change in the genetic structure between and within cohorts of a marine fish, diplodus sargus, induced by a large variance in individual reproductive success. *Molecular Ecology, 11*(8), 1515-1524.

Pollak, E. (1983). A new method for estimating the effective population size from allele frequency changes. *Genetics, 104*(3), 531-548.

Poulsen, N. A., Nielsen, E. E., Schierup, M. H., Loeschcke, V., & Gronkjaer, P. (2006). Long-term stability and effective population size in north sea and baltic sea cod (gadus morhua). *Molecular Ecology, 15*(2), 321-331.

Pudovkin, A. I., Zaykin, D. V., & Hedgecock, D. (1996). On the potential for estimating the effective number of breeders from heterozygote-excess in progeny. *Genetics, 144*(1), 383-387.

Pujolar, J. M., Maes, G. E., & Volckaert, F. A. M. (2006). Genetic patchiness among recruits in the european eel anguilla anguilla. *Marine Ecology Progress Series, 307*, 209-217.

Raymond, M., & Rousset, F. (1995). GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity, 86*(3), 248-249.

Reece, K. S., Ribeiro, W. L., Gaffney, P. M., Carnegie, R. B., & Allen, S. K. (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*(4), 346-352.

Rhodes, K. L., Lewis, R. I., Chapman, R. W., & Sadovy, Y. (2003). Genetic structure of camouflage grouper, epinephelus polyphekadion (pisces: Serranidae), in the western central pacific. *Marine Biology, 142*(4), 771-776.

Rose, C. G., Paynter, K. T., & Hare, M. P. (March/April 2006). Isolation by distance in the eastern oyster, crassostrea virginica, in chesapeake bay. *Journal of Heredity, 97*(2), 158-170.

Ross, P. G., & Luckenbach, M. W. (2009). *Population assessment of eastern oysters (crassostrea virginica) in the seaside coastal bays*. Wachapreague, VA: Eastern Shore Laboratory. Retrieved from http://www.deq.virginia.gov/Portals/0/DEQ/CoastalZoneManagement/task10-02-07.pdf

Shinde, D., Lai, Y., Sun, F., & Arnheim, N. (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*(3), 974-980.

Taris, N., Boudry, P., Bonhomme, F., Camara, M. D., & Lapègue, S. (December 2009). Mitochondrial and nuclear DNA analysis of genetic heterogeneity among recruitment cohorts of the european flat oyster ostrea edulis. *The Biological Bulletin, 217*(3), 233-241.

Turner, T. F., Wares, J. P., & Gold, J. R. (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*(3), 1329-1339.

Wang, J., & Whitlock, M. C. (2003). Estimating effective population size and migration rates from genetic samples over space and time. *Genetics, 163*(1), 429-446.

Wang, Y., Wang, X., Wang, A., & Guo, X. (2010). A 16-microsatellite multiplex assay for parentage assignment in the eastern oyster (crassostrea virginica gmelin). *Aquaculture, 308, Supplement 1*(0), S28-S33.

Wang, Y., & Guo, X. (2007). Development and characterization of EST-SSR markers in the eastern oyster *crassostrea virginica*. *Marine Biotechnology, 9*(4), 500-511.

Wang, Y., Shi, Y., & Guo, X. (2009). Identification and characterization of 66 EST-SSR markers in the eastern oyster crassostrea virginica (gmelin). *Journal of Shellfish Research, 28*(2), 227-234.

Waples, R. S. (1989). A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics, 121*(2), 379-391.

Waples, R. S. (1991). *Genetic methods for estimating the effective size of cetacean populations* No. 13)

Waples, R. S., & Do, C. (2008). LDNE: A program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources, 8*(4), 753-756.

Waples, R. S. (2002). Evaluating the effect of stage-specific survivorship on the Ne/N ratio. *Molecular Ecology, 11*(6), 1029-1037.

Waples, R. S. (2006). A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci*. *Conservation Genetics, 7*(2), 167-184.

Waples, R. S., & Do, C. (2010). Linkage disequilibrium estimates of contemporary ne using highly variable genetic markers: A largely untapped resource for applied conservation and evolution. *Evolutionary Applications, 3*(3), 244-262.

Waples, R. S., & Yokota, M. (January 2007). Temporal estimates of effective population size in species with overlapping generations. *Genetics, 175*(1), 219-233.

Wattier, Engel, Samitou-Laprade, & 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*(11), 1569-1573.

Wright, S. (1931). Evolution in mendelian populations. *Genetics, 16*, 97-159.

Zeller, M., Reusch, T. B. H., & Lampert, W. (2008). Small effective population sizes in two planktonic freshwater copepod species (eudiaptomus) with apparently large census sizes. *Journal of Evolutionary Biology,21*(6), 1755-1762.

Tables and Figures

Figure 1. Locations of Collection Sites

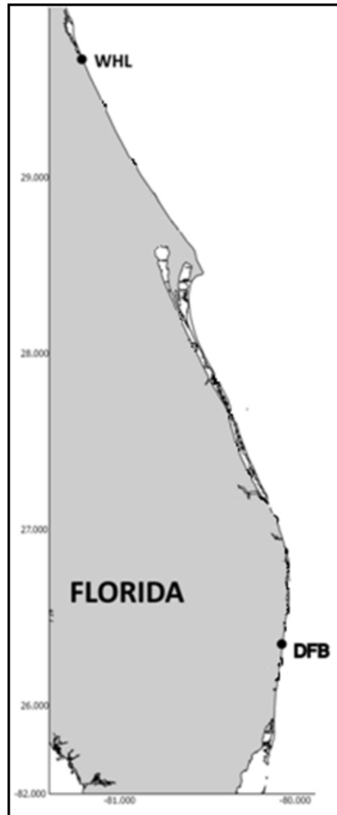

Table 1. Locus names (citation), descriptions and PCR conditions. Ta, annealing temperature centigrade.

Loci	Multiplex	Fluorescent Label	T_a (C)	Repeat Unit	Observed Alleles
CVi9 (Brown 2000)	--	Green (HEX)	62	Tri-nucleotide	21
CVi2i23 (Reece 2004)	--	Blue (6FAM)	51.5	Tetra-nucleotide	36
CVi2k14 (Reece 2004)	--	Yellow (NED)	52	Tri-nucleotide	6
CVi12 (Brown 2000)	--	Green (HEX)	62	Tetra-nucleotide	48
RUCV045 (Wang & Guo 2007)	B	Blue (6FAM)	57	Di-nucleotide	10
RUCV060 (Wang & Guo 2007)	A	Green (HEX)	53	Di-nucleotide	41
RUCV063 (Wang & Guo 2007)	A	Yellow (NED)	53	Di-nucleotide	11
RUCV131 (Wang 2009)	B	Yellow (NED)	57	Tri-nucleotide	44
RUCV374 (Wang 2010)	A	Blue (6FAM)	53	Tri-nucleotide	10

Table 2. Genotyping Error Estimates

Loci	**Samples Duplicated (Percent)**	**Error Rate per Allele**
Cvi9	29 (5.5%)	0.00%
CV2i23	31 (5.9%)	1.61%
CV2k14	32 (6.1%)	4.69%
CVi12	28 (5.3%)	7.14%
RUCV060	51 (9.7%)	2.94%
RUCV063	26 (5.0%)	5.77%
RUCV374	95 (37.6%)	0.00%
RUCV045	42 (16.7%)	3.57%
RUCV131	37 (14.7%)	4.05%

Table 3. HWE P-values for populations and loci

	Loci								
Population	**CVi9**	**2i23**	**2k14**	**CVi12**	**RU060**	**RU063**	**RU374**	**RU045**	**RU131**
DFB67s	0.0000*	0.0411*	0.4336	0.0000*	0.3484	0.5833	0.0006*	0.2034	0.3648
DFB79s	0.0000*	0.4213	0.1544	0.0000*	0.9301	0.0310	0.0005*	0.9335	0.0004*
WHL67s	0.0245*	0.5088	0.8187	0.0000*	0.0780	0.1077	0.0015*	0.8239	0.8508
WHL69s	0.0000*	0.9305	1.0000	0.0000*	0.3173	0.0789	0.0115*	0.8675	0.2535
DFB57a	0.0000*	0.5689	0.3239	0.0000*	0.8813	0.3159	0.1027	0.3172	0.8571
WHL57a	0.0000*	0.9850	0.2799	0.0000*	0.2743	0.0000*	0.0148*	0.0160*	0.8911

Table 4. Listed for each locus and population are observed and expected number of heterozygotes (H_o and H_e), F_{is} values, and probability (P_r) of a significantly positive Fis (heterozygote deficiency).

	Loci																		
Population	**CVi9**		**2i23**		**2k14**		**CVi12**		**R060**		**R063**		**R374**		**R045**		**R131**		**Total F_{is}**
	H_o/H_e P_r	F_{is}	H_o/H_e P_r	F_{is}	H_o/H_e P_r	F_{is}	H_o/H_e P_r	F_{is}	H_o/H_e P_r	F_{is}	H_o/H_e P_r	F_{is}	H_o/H_e P_r	F_{is}	H_o/H_e P_r	F_{is}	H_o/H_e P_r	F_{is}	
DFB67s (n=99)	63/90 **0.000**	0.301	91/89 0.758	-0.018	53/50 0.842	-0.056	63/85 **0.000**	0.262	49/52 0.176	0.066	92/95 0.089	0.033	60/73 **0.002**	0.183	91/94 0.091	0.036	70/75 0.098	0.072	**0.105**
DFB79s (n=100)	67/91 **0.000**	0.261	86/90 0.111	0.044	57/53 0.952	-0.084	66/86 **0.000**	0.261	55/52 0.869	-0.054	90/96 **0.007**	0.062	53/72 **0.000**	0.270	96/96 0.474	0.004	65/72 0.032	0.100	**0.108**
WHL67s (n=108)	81/98 **0.000**	0.173	101/101 0.563	0.000	34/32 0.756	-0.043	69/93 **0.000**	0.255	64/61 0.796	-0.044	102/104 0.193	0.020	59/77 **0.000**	0.233	102/103 0.305	0.014	77/81 0.202	0.049	**0.083**
WHL69s (n=112)	64/103 **0.000**	0.380	103/106 0.160	0.027	37/35 0.824	-0.060	63/100 **0.000**	0.372	53/58 0.095	0.084	100/108 **0.001**	0.073	64/71 0.052	0.105	108/107 0.772	-0.012	79/80 0.398	0.018	**0.127**
DFB57a (n=55)	37/49 **0.000**	0.254	49/48 0.764	-0.024	36/30 0.999	-0.218	32/50 **0.000**	0.359	27/25 0.868	-0.082	52/53 0.350	0.016	31/39 0.012	0.206	52/53 0.373	0.016	41/41 0.580	-0.001	**0.079**
WHL57a (n=60)	25/47 **0.000**	0.471	49/48 0.780	-0.019	12/13 0.243	0.099	31/46 **0.000**	0.328	29/27 0.868	-0.075	46/49 0.035	0.070	34/36 0.249	0.068	46/50 0.026	0.076	41/39 0.862	-0.064	**0.121**
Total F_{is}	**0.296**		**0.007**		**-0.075**		**0.301**		**-0.005**		**0.047**		**0.187**		**0.017**		**0.040**		

Table 5. Allelic Richness per locus and population

	Loci								
Population	**CVi9**	**2i23**	**2k14**	**CVi12**	**R060**	**R063**	**R374**	**R045**	**R131**
DFB67s	16.7	20.2	5.0	18.8	6.9	27.9	5.5	26.4	7.0
DFB79s	14.5	22.1	5.5	20.9	7.8	29.1	5.0	33.6	7.5
WHL67s	15.2	24.8	3.5	25.2	6.6	28.6	7.6	30.2	6.7
WHL69s	17.5	25.1	4.2	24.7	6.4	29.9	7.3	29.0	6.2
DFB57a	15.8	19.6	5.0	22.6	5.9	26.7	5.0	30.6	6.9
WHL57a	15.0	25.0	3.0	25.0	7.0	29.0	6.0	28.0	8.0
Total	16.1	26.4	5.0	24.8	7.1	28.8	6.8	30.3	7.4

Table 6. Comparisons of Allelic Richness (A_r) and Observed and Expected Heterozygosity (H_o and H_o) between spat and adults.

Populations	Allelic Richness	P-values Adults>Spat	Ho	P-values Adults>Spat	Hs	P-values Adults>Spat	Fis	P-values Adults>Spat
2007 Adults	15.786		0.696		0.772		0.099	
2007 Spat	15.708	0.474	0.709	0.694	0.782	0.738	0.093	0.403
All Spat	16.083	0.660	0.697	0.520	0.779	0.750	0.106	0.543

Table 7. Single sample Ne estimates using all loci and parametric 95% confidence limits.

Population	Estimate	95% CI
DFB57a	417.4	[179.6 - ∞]
DFB67s	?	[993.8 - ∞]
DFB79s	3008.5	[583.9 - ∞]
WHL57a	450.8	[179.8 - ∞]
WHL67s	?	[1056.9 - ∞]
WHL69s	?	[5309.1 - ∞]

Table 8. Single Sample Ne Estimation without CVi12 Locus.

Population	Estimate	95% CI
DFB57a	413.2	[162.6 - ∞]
DFB67s	?	[709.9 - ∞]
DFB79s	7236.3	[592.0 - ∞]
WHL57a	499.7	[198.5 - ∞]
WHL67s	?	[849.6 - ∞]
WHL69s	?	[33224.3 - ∞]

Table 9. Single Sample Ne Estimation without CVi9, 2k14, or CVi12 Loci.

Population	Estimate	95% CI
DFB57a	478.1	[142.9 - ∞]
DFB67s	8908.4	[462.4 - ∞]
DFB79s	2614.6	[415.6 - ∞]
WHL57a	399.9	[133.3 - ∞]
WHL67s	?	[599.8 - ∞]
WHL69s	?	[3949.5 - ∞]

Table 10. Temporal Ne estimates using all loci.

Populations	Sampling Interval (Generations)	Estimate [95% CI]
DFB57a & DFB67s	1	142.9 [62.2 - 2198.3]
DFB57a & DFB69s	2	651.7 [183.7 - ∞]
DFB67s & DFB69s	1	330.4 [114.5 - ∞]
WHL57a & WHL67s	1	∞ [508.9 - ∞]
WHL57a & WHL69s	2	∞ [1165.6 - ∞]
WHL67s & WHL69s	1	404.5 [137.2 - ∞]

Table 11. Temporal Ne estimates without CVi12 Locus.

Populations	Sampling Interval (Generations)	Estimate [95% CI]
DFB57a & DFB67s	1	∞ [353.9 - ∞]
DFB57a & DFB69s	2	∞ [1826.6 - ∞]
DFB67s & DFB69s	1	∞ [390.7 - ∞]
WHL57a & WHL67s	1	∞ [298.5 - ∞]
WHL57a & WHL69s	2	∞ [604.5 - ∞]
WHL67s & WHL69s	1	1653.8 [342.4 - ∞]

Table 12. Temporal Ne estimates without CVi9, 2k14, or CVi12 Loci.

Populations	Sampling Interval (Generations)	Estimate [95% CI]
DFB57a & DFB67s	1	∞ [326.6 - ∞]
DFB57a & DFB79s	2	∞ [1553.2 - ∞]
DFB67s & DFB79s	1	4644.3 [314.8 - ∞]
WHL57a & WHL67s	1	∞ [407.3 - ∞]
WHL57a & WHL69s	2	∞ [1079.4 - ∞]
WHL67s & WHL69s	1	3106.2 [357.1 - ∞]