# ABSTRACT

Title of dissertation:                    POPULATION GENOMICS OF SELECTION IN
THE EASTERN OYSTER CONTACT ZONE

                                         Maria C. Murray, Doctor of Philosophy, 2012

Dissertation directed by:           Professor Thomas D. Kocher
Department of Biology

Intraspecific clinal systems are ideal for investigating the how divergence occurs in the presence of gene flow because they represent a balance between selection and gene flow prior to speciation. High dispersal marine species with clinal variation are particularly informative to test for divergent selection because selection likely is strong enough to counteract high gene flow. The degree of population structure varies considerably among loci, such that the genome acts as a sieve allowing gene flow at neutral loci and impeding it at selected loci, creating a genomic mosaic of differentiation. In this study, I examine genomic and geographic patterns of differentiation among parapatric populations of the eastern oyster (*Crassostrea virginica*) along their contact zone in Florida estuaries. The planktotrophic larval phase of this species gives it the potential for regular long-distance dispersal and genetically homogeneous populations. However Florida populations at the center of its range exhibit a sharp step cline at some loci, suggesting a role for divergent selection. Using 217 AFLP loci, including seven candidate loci for differential selection

between the two populations, I genotyped 1,011 spat over two seasons and 274 adults at sites along the contact zone. I examined: (1) whether genome scans can detect divergent selection in a clinal system, (2) the genomic and geographic patterns of differentiation along the cline at neutral and selected loci, and (3) regional patterns of differentiation and genotypic distributions among the life stages. Results demonstrated: (1) candidate loci for regionally divergent selection, (2) a genomic and geographic mosaic of differentiation, (3) regional and localized selection at a non-trivial portion of loci, (4) lower recruitment and some mortality in the center of the cline, and (5) strong exogenous, post-settlement viability selection against intermediate and non-native-like genotypes. While a combination of neutral and adaptive processes likely shape genomic and geographic patterns of differentiation, this study revealed evidence for divergent selection in an estuarine species with high potential for gene flow. Overall, these results point to a major role for post-zygotic, environment-dependent selection in the maintenance of the contact zone between Atlantic and Gulf-type oyster populations.

POPULATION GENOMICS OF SELECTION
IN THE EASTERN OYSTER
CONTACT ZONE


by

Maria C. Murray




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
2012




Advisory Committee:

Professor Thomas D. Kocher, Chair
Professor Matthew P. Hare, Cornell University
Professor David. J. Hawthorne, Dean's Representative
Professor Anson H. Hines
Professor Marjorie L. Reaka
Professor Sara Via

# DEDICATION

For my parents Elodie and Dwight,
who unfailingly prioritized my education and pursuit of knowledge

and

For my loving and supportive husband Woody

# ACKNOWLEDGEMENTS

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# GENERAL INTRODUCTION

Describing the process of divergence has been a major goal of evolutionary biologists. It has been more challenging to explain how adaptive divergence occurs in the presence of gene flow (i.e., sympatry and parapatry) because divergent selection must be strong enough to overcome just one successful migrant per generation (Wright 1931). Examining divergence at different points along the continuum leading to speciation can help illuminate the underlying process and forces involved in developing reproductive barriers. Intraspecific clinal systems, or species whose distribution includes a geographic shift between populations differentiated in phenotype or genotype frequencies, are ideal for investigating the process of divergence because they represent a balance between selection and gene flow at a stage prior to incipient speciation.

Hybrid zones are natural laboratories for investigating how selection and dispersal interact to maintain divergence (Harrison 1990). They are often the result of secondary contact and are frequently located in regions of environmental transition, which provide opportunities for selection gradients to act (Arnold 1997). Selection can act through two pathways to maintain divergence in such zones of contact between differentiated populations: endogenous selection against genetic incompatibilities in hybrids and exogenous selection for genotypes conferring higher fitness in their respective habitats (Arnold 1997). These two pathways are the basis for hybrid zone maintenance models. In the former, a balance between dispersal and intrinsic selection against hybrids

produces an environment-independent, clinal "tension zone" (Barton & Hewitt 1985). In contrast, the latter model states that environmental selection gradients maintain hybrid zones (Slatkin 1973; Moore 1977; Endler 1977). Under this environment-dependent model, genotypes are not necessarily distributed in clines but loosely follow the distribution of critical environmental features.

Both of these models have difficulty explaining sharp marine clines in high gene flow species without assuming very strong selection or barriers to dispersal. For example, cohort analyses indicated that a narrow *Mytilus edulis-M. galloprovincialis* hybrid zone is maintained by exogenous viability selection through differential vulnerability to wave action (Wilhelm & Hilbish 1998). Nonrandom hydrographic dispersal barriers also seem to play a role (Gilg & Hilbish 2003). The endogenous and exogenous models are not mutually exclusive (Arnold 1992). A combination of intrinsic and extrinsic selection agents has been demonstrated in both terrestrial (reviewed in Arnold 1997) and marine (e.g., *Mercenaria* in Bert & Arnold 1995) hybrid zones.

High dispersal marine species with clinal variation offer particularly informative systems to test for divergent selection because selection is likely to be strong to counteract high gene flow. For example, viability selection maintains leucine aminopeptidase allele frequency differences in the face of high dispersal along a narrow, estuarine salinity gradient in *Mytilus edulis* (Koehn et al. 1980). Indeed, stressful environmental conditions found in intertidal (e.g., extreme temperature fluctuations, desiccation) and estuarine (e.g., extreme salinity fluctuations, limited dissolved oxygen) habitats impose strong

2

selection on inhabitants (Gardner 1997).  Selection gradients producing clinal variation

across broad species distributions offer opportunities to study geographic patterns of

differentiation (e.g., latitudinal physiological races in oyster reproductive timing in

Loosanoff & Nomejko 1951; Barber et al. 1991).  Where historical and/or contemporary

processes amplify latitudinal clines, investigation into those processes becomes more

tractable and potentially relevant to parapatric speciation.

Both historical and contemporary forces must be considered to explain population

structure (Hilbish 1996).  Secondary contact is a particularly interesting historical legacy

that could result in a contemporary signal of strong population divergence in species with

high dispersive potential (Avise 1992).  During vicariance, isolated subpopulations

escape the homogenizing effects of gene flow and differentiate through drift or

differential selection (Coyne & Orr 2004).  If vicariance occurs within an environmental

gradient, divergent adaptations can arise where gene flow previously prevented them

(Conner & Hartl 2004).  After secondary contact forms a zone of transition between

subpopulations, differential introgression erodes the accumulated neutral differentiation

while selection maintains differences over evolutionary time (Barton & Gale 1993).  If

sufficient time has passed to allow homogenization through gene flow at non-adaptive

genes, genomic patterns can illuminate the filtering effects of divergent selection.

Disentangling the effects of historical legacy and contemporary selection continues to

present a major challenge, especially along ecotones.  Therefore, the genome will reflect

strong differentiation in regions subject to strong divergent selection across environments

or between genetic incompatibilities and will appear as a single structured population in

neutral regions (Beaumont 2005). These selected regions are the unit of adaptation. As the populations diverge, the portion of the genome experiencing restricted gene flow expands. This theory has led to a 'genic view of speciation' (Wu 2001), where the process of adaptive speciation can be informed by examining genomic patterns resulting from differentiation in the face of gene flow.

Through empirical studies of model organisms, we are only beginning to uncover the genetic basis of adaptation (Orr 2005). Experiments looking for quantitative trait loci, examining the roles of candidate genes, and using association mapping have led to specific discoveries of the genetic basis of adaptive traits (e.g., Szabo & Burr 1996; Albertson et al. 2003; Nachman et al. 2003; Caicedo et al. 2004; Colosimo et al. 2005; Tishkoff et al. 2007). These approaches, however, require a great deal of prior knowledge of the genome and/or gene functions, typical of model organisms. In addition, the limitations to model organisms makes it difficult to assess how applicable these discoveries are to all organisms.

## Population Genomics

Separating genome-wide neutral patterns from locus-specific patterns of selection can increase understanding of how evolutionary forces work to create and maintain differentiation in the presence of gene flow. Neutral theory has gone far to explain and emphasize the role of random drift in generating genetic polymorphism and differentiation (e.g., Kimura 1968). Recently, however, the argument for selection

playing a prominent role in generating and maintaining divergence has strengthened

through theoretical (Doebeli & Dieckmann 2003; Doebeli et al. 2005; Gavrilets & Vose

2005) and empirical (e.g., Wilhelm & Hilbish 1998; Schneider et al. 1999; Wilding et al.

2001; Ogden & Thorpe 2002; Riginos & Cunningham 2005) lines of evidence.

Capitalizing on the advances in neutral theory as a null model, evolutionary genomics can

now address outstanding questions in evolutionary genetics of how and why divergence

varies across the genome (Orr 2005).

Population genomics is defined as the simultaneous sampling of many loci across a

genome to understand how evolutionary processes shape genetic variation across

genomes and populations (Black et al. 2001; Luikart et al. 2003).  It combines genomic

concepts and technical approaches with population genetic theory and statistics to

separate genome-wide effects (i.e., random genetic drift, gene flow, and inbreeding) from

locus-specific effects (i.e., selection, mutation, assortative mating, and recombination).

Of interest here, is the ability of this contrast to identify loci that indicate selective

divergence despite gene flow at other loci.  This genome scan process can be summarized

in four basic steps (Luikart et al. 2003).  (1) Sample many individuals from populations

of interest.  (2) Genotype many loci, ideally tens to hundreds.  (3) Test for outlier loci

using neutral theory of molecular evolution as a basis for the null hypothesis (Lewontin

& Krakauer 1973).  For this test, simulations create a null distribution of neutral

divergence due to mutation against which to compare the empirical genetic variation

(Beaumont & Nichols 1996; Beaumont & Balding 2004).  Outlier loci that do not

conform statistically to the expected null distribution are indicators of non-neutral

processes, whereas neutrally behaving loci are informative regarding demography and phylogenetic history.  This test assumes that: genes evolve independently due to pervasive recombination; enough loci have been sampled to accurately represent the genomic variation; and genome-wide effects influence all parts of the genome equally (Black et al. 2001).  (4) Use the outlier loci putatively under selection for further study to confirm and investigate selection.

For non-model organisms, genome scans offer an alternative approach to address questions about the genetic basis of adaptive divergence (Luikart et al. 2003; Beaumont & Nichols 1996; Beaumont 2005; Hedrick 2006; Butlin 2010).  By comparing differentiation measured by hundreds of anonymous amplified fragment length polymorphisms (AFLP) to a neutral model of divergence, an investigator can: (1) identify candidate regions in the genome that are under the influence of divergent selection, either directly or through tight physical linkage, (2) assess the role selection plays in contributing to population differentiation, (3) evaluate how the effects of selection vary across the genome, and (4) compare patterns of selection on a geographic scale.  An increasing number of studies have used the genome scan approach to detect candidate loci under the influence of selection (Wilding et al. 2001; Emelianov et al. 2004; Campbell & Bernatchez 2004; Bonin et al. 2006; Savolainen et al. 2006; Egan et al. 2008; Eveno et al. 2008; Nosil et al. 2008; Mariac et al. 2011).

Although genome scans have been applied to a wide variety of organisms including various plants (e.g., Savolainen et al. 2006; Oetjen & Reusch 2007; Minder & Widmer

2008; Mariac et al. 2011), invertebrates (e.g., Wilding et al. 2001; Emelianov et al. 2004; Murray & Hare 2006; Egan et al. 2008; Nosil et al. 2008), and vertebrates (e.g., Campbell & Bernatchez 2004; Bonin et al. 2006), the range in methodologies utilized to study divergent populations has not been consistent. Different applications have included: a single pair of populations; multiple, evolutionarily replicated pairs of populations analyzed separately; a single analysis incorporating multiple populations; varying arbitrary significance levels; and adjustments to address multiple comparisons. The selection of any given approach often proceeds with little justification, but can have a large effect on the conclusions that can be drawn. Therefore, these studies should be interpreted with caution, particularly when performing meta-analyses (e.g., Nosil et al. 2009).

Some generalizations are possible, however, that provide a window on the genomic landscape of divergent selection. In a 2009 review of 20 genome-scan studies, Nosil and colleagues report an average of 8.5% of loci were putatively under the influence of selection (range: 0.4-24.5%). This suggests that a considerable proportion of the genome may be affected by divergent selection. A study not utilizing a genome scan supports this finding that selective divergence is heterogeneous across the genome (Payseur et al. 2004). For studies examining the genomic distribution of outlier loci via genetic maps and quantification of linkage disequilibruim, Nosil and colleagues found mixed evidence supporting genomic dispersion and genomic clustering among outliers. In addition, nearly 50% of outliers were detected for multiple population pairs (Nosil et al. 2009). This suggests the same adaptive allele often either is selected in parallel repetition or

spreads through migratory gene flow. In fact, accumulating examples demonstrate parallel divergence when independent origins of replicated ecotypes are known (Wilding et al. 2001; Campbell & Bernatchez 2004; Nosil 2008). Another pattern emerging from genome scans is one of genealogical discordance between population trees based on non-outlier loci and those based on outlier loci (Wilding et al. 2001; Bonin et al. 2006; Egan et al. 2008). This corroborates theory of divergence with gene flow, where neutral gene trees reflect geographic patterns of demography, gene flow, and drift while gene trees of adaptive loci reflect the ecologically and genomically adaptive constraints. Finally, to lend credence to the genome scan approach, QTL and outlier loci do collocate on genetic maps in some cases (Rogers & Bernatchez 2007; Via & West 2008).

Amplified fragment length polymorphism (AFLP) assays combine the replicability of RFLPs with the power of PCR to offer a genomic alternative for non-model organisms that is useful for genome scans. The assay generates many dominant loci with random genomic coverage, mostly from non-coding nuclear DNA (Vos et al. 1995). AFLPs are used to evaluate many aspects of population genetics including: gene flow, dispersal, outcrossing, introgression, hybridization, DNA fingerprinting, parentage, genome-wide variation, and selection (Mueller & Wolfenbarger 1999; Meudt & Clarke 2007). While AFLPs have been extensively reviewed (Mueller & Wolfenbarger 1999; Bonin et al. 2007; Meudt & Clarke 2007), their limitations and advantages are of interest here. As dominant markers, individual loci are less informative, but the large number of loci generated can easily overcome this (Bonin et al. 2007). Because Hardy-Weinberg equilibrium is not testable directly in dominant loci, estimating allele frequencies requires

an assumption of equilibrium based on prior information or that Bayesian procedures be

used (Zhivotovsky 1999). Although this requires larger sample sizes, AFLPs are cheap,

easy, fast, and reliable, with a reproducibility of 2-5% (Meudt & Clarke 2007; Zhang &

Hare 2012). They require little tissue and no prior knowledge of genomic composition or

sequence, making them ideal for non-model and small organisms. Very efficient for a

given assay, AFLPs generate a large number of markers and high polymorphism. Their

advantages make them ideal for studies investigating genomic patterns of adaptive

divergence between populations in non-model organisms.


## Study System


The eastern oyster, *Crassostrea virginica,* is a high dispersal species that can be

especially informative because it has a sharp intraspecific cline located within a relatively

narrow ecotonal transition between temperate and subtropical communities. Several

features of the oyster system facilitate a test for divergent selection in this high gene flow

species. First, the genetic cline along eastern Florida is sharp relative to its dispersal

potential (Hare and Avise 1996). Second, the sessile nature of the oyster facilitates

investigation of local adaptation. Third, phylogeographic concordance among broadly

distributed estuarine species in eastern Florida strongly suggests that divergence at this

biogeographic province boundary is the result of historical vicariance and secondary

contact but is now maintained by contemporary selection gradients (Avise 1992; Avise

2004). Thus, the oyster cline involves a contact zone between historically and genetically

distinct subpopulations and presents an intriguing system in which to study the role of selection in maintaining sharp differentiation in a high gene flow species.

**Oyster Biology**

In the intertidal and subtidal estuarine environment, sessile organisms like the oyster encounter strong environmental and biological gradients spatially and temporally (Pennings & Bertness 2001). Under a scenario where the spatial scale of gene flow surpasses the scale at which selection varies, generalist traits or phenotypic plasticity are expected (Slatkin 1973; Grosberg & Cunningham 2001). However, in some marine species with extensive dispersal potential, population genetics show new recruits to be well mixed over local scales at most loci while other loci associated with physiological tolerances diverge among environmental patches due to recurrent post-settlement diversifying selection (Koehn et al. 1980; Schmidt & Rand 1999).

The eastern oyster is a broadly distributed, economically and ecologically important estuarine bivalve. As a reef-builder, *C. virginica* is a keystone species in estuaries across the American Atlantic and Gulf of Mexico coasts. It plays a critical role in the ecology of these systems by providing food, habitat, substrate, and shelter to a variety of organisms (e.g., worms, crabs, tunicates, small fish, other bivalves, amphipods, and algae; personal observation; Galtsoff 1964; White & Wilson 1996). The eastern oyster also reduces turbidity and deposits nutrients to the bottom as it filters about 10 L of water per hour (Newell & Langdon 1996) to clean it of silt and phytoplankton. Oyster reef structures slow water currents and prevent erosion ashore (Galtsoff 1964). While it provides proper

habitat conditions for other organisms, the eastern oyster as a species lives in a range of salinities (5-40 ppt; Shumway 1996), temperatures (-2 - 36$^{o}$C; Shumway 1996), and latitudes (18-49$^{o}$N).

The geographic range of the eastern oyster stretches from the Gulf of St. Lawrence in Canada to the Yucatan Peninsula in Mexico (Carriker & Gaffney 1996). The distribution of the oyster is controlled by synergistic tolerances of salinity and temperature and by biotic factors like predators, parasites, and disease (Shumway 1996). Various studies have demonstrated a range of salinity tolerance and the control temperature has over oyster growth, development, and reproduction (reviewed in Shumway 1996). Spanning a latitudinal temperature gradient and biogeographic province boundaries, the oyster has been called a generalist characterized by phenotypic plasticity that enables it to tolerate and thrive in a variety of environments. Not mutually exclusive to its generalist characteristics, local to regional adaptation may contribute to the ability of the eastern oyster to inhabit such a broad geographic range. Indeed, physiological races exist between geographically separated populations, sometimes in relatively close proximity (Barber et al. 1991; Dittman 1997; Dittman et al. 1998). The life cycle of *C. virginica* also supports long distance dispersal and large genetic variation, enabling oysters to reach and be fit in many places.

The oyster life cycle begins with adult oysters synchronously spawning gametes into the water column. Over the next two to three weeks, planktonic fertilized eggs develop into swimming veliger larvae that eventually develop an eye spot and foot. These

11

pediveligers locate and explore the bottom environment and settle onto hard substrate, preferably existing adult oyster shell, and metamorphose into the juvenile spat stage. Over the next one to three years, these spat develop into reproductively competent adults, usually protandrously beginning as males and changing to females when they reach a large enough size (Thompson et al. 1996). The age at which oysters reach reproductive competency varies geographically and with growth rate. For oysters in more temperate environments like those found in the Northeast, first reproduction occurs a year after settlement. In contrast, the warmer climate of the Southeast supports faster growth and first reproduction within the spat's first season. Like the timing of reproductive competency, the number of broods per season varies latitudinally with only one brood in the Northeast and two in the Southeast. As with many organisms, fecundity increases with size such that females can produce a range of approximately 2-100 million eggs per season (Shumway 1996).

Through sexual recombination, the great fecundity of oysters produces a wide variety of genotypes on which selection can act. For organisms like many trees and bivalves that face intense attrition in early phases of the life cycle, this genetic variation is highly advantageous (Williams 1975). For the oyster, Williams estimated that 90% of progeny perish in the plankton and that 90% of the remaining individuals die from non-selective factors (1975). Settling at a much higher density that the area can support, competition occurs among adjacent individuals. Slight disadvantages in viability and fitness-related traits suffered over extended periods in some leads to the best genotypes surviving. The high fecundity and resulting high genetic diversity maximizes the probability that at least

12

one offspring will survive.  Those that survive are slightly more efficient at using

resources and can tolerate local conditions better.


In the natural mode of dispersal for this sessile benthic bivalve, the two to three weeks

spent as planktotrophic larvae gives the eastern oyster a great potential for long-distance

dispersal and gene flow.  While practical challenges exist for directly tracking actual

dispersal distance in many planktonic larvae, population genetics present alternative

approaches to studying planktonic dispersal on evolutionary time scales.  In a meta-

analysis of marine larval dispersal, clear positive correlation between dispersal distance

and larval period translates to 20-120 km dispersal distance for a 14-21 day planktonic

period (Kinlan & Gaines 2003; Shanks et al. 2003; Siegel et al. 2003; Shanks 2009).  In

the Chesapeake Bay, the average squared dispersal distance was 479 km$^2$ for the eastern

oyster specifically (Rose et al. 2006), empirically demonstrating effective gene flow over

moderate distances in this species.


### The Cline at Cape Canaveral, Florida

For an organism considered a generalist with long distance dispersal living in an aquatic

environment with no physical barriers to dispersal, there might be an expectation of

genetic homogeneity among populations (Grosberg & Cunningham 2001).  However, *C.*

*virginica* is composed of two genetically distinct populations in the Atlantic and Gulf of

Mexico (Reeb & Avise 1990).  Genetic homogeneity observed within each region is

consistent with high gene flow (Reeb & Avise 1990; Hoover & Gaffney 2005).  Along

the eastern coast of Florida, a zone of rapid transition in allele frequency connects

'Atlantic' and 'Gulf' populations (Reeb & Avise 1990; Karl & Avise 1992). Within the transition zone hybrids are formed; however, it remains unclear at what frequency. When examined at a finer scale, the cline in allele frequency was quite sharp in comparison to potential dispersal distance (Hare & Avise 1996). Allele frequencies changed by 50-75% over a span of only 20 km near Cape Canaveral (Hare & Avise 1996), suggesting a strong physical barrier to dispersal or a role for regional selection.

In contrast to mitochondrial (mt)DNA, which shows a clear pattern of reciprocal monophyly when distant allopatric oyster populations are compared (Reeb & Avise 1990), alternate fixation has not been found for any nuclear locus (Hare & Avise 1998). In fact, the degree of differentiation varies dramatically among presumed neutral nuclear loci (Buroker 1983; Karl & Avise 1992; McDonald et al. 1996). This variation has spurred controversial hypotheses to explain the variance among loci.

In the first study of population structure across the oyster range, populations had high genetic similarity (96.2 – 99.7%), interpreted as a result of high gene flow in presumed neutral allozyme markers (Buroker 1983). When DNA markers were examined, they began to reveal a very different pattern of genetic structure. A restriction fragment length polymorphism (RFLP) analysis of mtDNA showed a distinct phylogeographic break in eastern Florida and homogeneity on either side of the break (Reeb & Avise 1990). The mtDNA results were corroborated by RFLPs at four single-copy nuclear (n)DNA loci (Karl & Avise 1992). Karl and Avise suggested that the disagreement between marker

classes was due to uniform balancing selection on allozyme loci, such that they were unaffected by the vicariant history registered in DNA polymorphisms.

Sparked by this controversial hypothesis, Cunningham and Collins (1994) reanalyzed the allozyme data using phylogeographic methods. A subtle geographic break was found, distinguishing Northern and Southern lineages, but it was located on the Gulf side of Florida. To test whether marker classes truly differ or if the Karl and Avise sampled loci were non-representative of DNA variation, McDonald and colleagues (1996) sampled polymorphisms at six additional anonymous nDNA loci in two populations (Charleston, SC and Panacea, FL) and demonstrated allele frequency differentiation no higher than observed in the original allozyme data. From these data, it was evident that the small Karl and Avise sample may have been biased toward differentiated loci. Yet, no explanation was offered for the large variance in differentiation measured across loci ($F_{ST}$ range: $0 - 0.757$; McDonald et al. 1996). Four additional nDNA loci showed intermediate levels of Atlantic-Gulf differentiation (Hoover & Gaffney 2005). Whereas Karl and Avise (1992) concluded that different evolutionary forces must shape different patterns across marker classes, the same logic now applies to different nuclear loci, but it must be tested explicitly.

**Patterns Geographically Concordant with the Oyster Cline**

Along eastern Florida, oysters inhabit a system of lagoons joined to the ocean at 11 narrow inlets and to each other through a stabilized intracoastal waterway. The extent to which larval transport occurs outside the lagoons, via inlet hopping, remains unknown.

15

However, for species that disperse as larvae, the shallow lagoons may create a common dispersal barrier because tidal mixing is weak, except near inlets (mean tide heights within the lagoons is ± 5 cm versus ± 26 cm along the oceanic coast; Smith 1987; Smith 1993). In non-tidal lagoons, wind shear can drive water movement, but the resulting larval dispersal is unknown. Other species with similar modes of dispersal might be affected by a physical barrier, such as when hydrodynamic circulation is weak as in the Florida lagoons or when oceanographic currents prevent migrants from random dispersal (e.g., southward migration across the Point Conception biogeographic boundary in Wares et al. 2001; and unidirectional migration of *M. edulis* alleles from the hybrid zone to *M. galloprovincialis* populations in Gilg & Hilbish 2003).

Although the genetic cline is strikingly sharp for a species with such a high potential for dispersal and dispersal barriers may play a role, the question of what evolutionary forces maintain the cline becomes more interesting in the biogeographic and environmental context of eastern Florida. The eastern oyster is not the only estuarine species in this region exhibiting clines (e.g., ribbed mussel in Sarver et al. 1992; killifish in Duggins et al. 1995; hard clams in Dillon & Manzi 1989; Bert & Arnold 1995), but these clines are centered in different locations in northeastern Florida. The Atlantic coast of Florida also contains a well-recognized biogeographic province boundary defined by clusters of species range limits between the Georgia-Florida border and Palm Beach, Florida (Briggs 1974; Engle & Summers 1999). The variation in cline locations and the diffuse nature of the province boundary implicate species-specific differential responses to selection gradients and/or more than a single, strong dispersal barrier. The transition in species

composition across the boundary, coupled with a particularly steep temperature gradient generates a dramatic ecotone.  With these properties overlaid onto the step cline, the possibility is clear that the environmental gradients (abiotic and biotic) could play a role in maintaining differentiation through divergent selection.

The oyster cline likely originated from genetic differences generated by vicariant processes that subsequently have come into secondary contact (Reeb & Avise 1990). Hybrid zones frequently collocate with regions of secondary contact between previously allopatric populations (Hewitt 1989).  Intraspecific phylogeographic concordance among multiple marine species (*Limulus polyphemus, Malaclemys terrapin, Centropristis striata, Ammodramus maritimus*) provides powerful evidence for this secondary contact hypothesis (Avise 1992).  During at least ten Pleistocene glacial advances and retreats, which enlarged the Florida peninsula, Atlantic and Gulf of Mexico populations were isolated and able to diverge (Reeb & Avise 1990).  These lines of evidence offer strong indications of cline origin and generate a need to understand its maintenance after secondary contact.

## Possible Explanations for Cline Maintenance

One potential explanation for the sharpness of this cline is that of recent secondary contact, implying that cline shape is not stable.  Recent secondary contact is consistent with the ephemeral-zone hypothesis, which predicts the temporary nature of hybrid zones to result in either speciation or fusion (Dobzhansky 1940).  Under this hypothesis, the sharpness of the oyster cline should erode at a speed that is proportional to gene flow

(Barton & Gale 1993; Endler 1977). However, numerous examples of stable hybrid zones exist (Barton & Hewitt 1985), demonstrating that the zones are not always short-lived. Indeed, when the oyster populations spanning Cape Canaveral were revisited 11 years after the documentation of this genetic break, the sharpness of the cline remained (Murray & Hare 2006). These findings suggest that the cline is somewhat stable, motivating investigation into the contemporary evolutionary processes maintaining this cline over time.

Factors that may be maintaining the oyster cline include: reduced dispersal, exogenous differential selection along the Florida coast, endogenous selection against hybrids, or a combination thereof. Selection is a particularly attractive evolutionary explanation when the ecotone along eastern Florida is considered. It is very possible that the low tidal flux (Bert & Arnold 1995) and weak currents (Smith 1987; Smith 1993) in the Florida lagoons are operating in concert with selection for local adaptation and/or against hybrids to maintain and even sharpen the oyster step cline. Tests for physical barriers to gene flow are outside the scope of this study. The strength of selection must be greater than the migration rate to prevent homogenization (Conner & Hartl 2004). While both processes must be quantified to fully understand cline maintenance, this study focuses on the initial step of testing whether divergent selection is acting in this system. To explore the role of selective forces through the following questions, this study uses a combination of genomic AFLP data, geographic and temporal sampling, and simulations.

- In Chapter 1:

    o Is the expected neutral distribution of differentiation sensitive to the details of a secondary contact history?

    o With an increase in the number of loci used to estimate the genomic mean $F_{ST}$ between Atlantic and Gulf of Mexico oyster populations, can neutral migration-drift processes explain the among-locus heterogeneity observed in previous studies?

- In Chapter 2:

    o Does the population structure near the cline center reflect clinal expectations of two differentiated populations in parapatry?

    o How does selection shape the genomic pattern of differentiation?

    o Is there a clinal geographic pattern of selection along the ecotone?

- In Chapter 3:

    o Is there successful recruitment in the Cape Canaveral area?

    o To what extent does hybridization occur between Atlantic and Gulf types?

    o Does selection occur pre-settlement or post-settlement?

    o Is selection endogenous or exogenous?  That is, does selection vary geographically across the zone of contact?

# CHAPTER 1: A Genomic Scan for Divergent Selection in a Secondary Contact Zone Between Atlantic and Gulf of Mexico Oysters: *Crassostrea virginica*

## Abstract

The degree of population structure within species often varies considerably among loci. This makes it difficult to determine whether observed variance reflects neutral-drift stochasticity or locus-specific selection at one or more loci. This uncertainty is exacerbated when evolutionary equilibrium cannot be assumed and/or admixture potentially inflates genomic variance. Thus, the value of a "genome scan", where locus-specific summary statistics are compared with a simulated neutral distribution among loci, may be limited in secondary contact zones if the null distribution is sensitive to the timing of secondary contact. Of particular interest here is the wide variance previously observed in locus-specific patterns of population structure between Atlantic and Gulf of

Mexico populations of eastern oyster, *Crassostrea virginica*. To test the robustness of an equilibrium null model, we compared among-locus distributions of $F_{ST}$ simulated under migration-drift equilibrium and several nonequilibrium secondary contact histories. We then tested for evidence of divergent selection between two oyster populations on either side of a secondary contact zone using 215 amplified fragment length polymorphism (AFLP) loci. Constant-migration equilibrium and nonequilibrium secondary-contact simulations produced equivalent distributions of $F_{ST}$ when anchored by the global mean $F_{ST}$ observed in oysters, 0.0917. The 99th quantile of simulated neutral $F_{ST}$ encompassed most of the variation among oyster loci. Three AFLP loci exhibited $F_{ST}$ values higher than this threshold. Although no locus was significant after correcting for multiple tests, our results show in geographically clinal organisms: AFLPs can efficiently characterize the genomic distribution of $F_{ST}$; equilibrium models can be used to evaluate outliers; these procedures help focus research on genomic regions of interest.

## Introduction

Determining the evolutionary forces that act to generate and maintain genetic diversity in natural populations is a fundamental goal of population genetics (Hartl & Clark 1997). Molecular markers provide an ever more powerful tool for this endeavor as types of polymorphic markers diversify and analytical methods improve (Hey & Machado 2003; Avise 2004; Manel et al. 2005). However, genomic sampling error remains a serious limitation to accurate inference of evolutionary processes from genetic patterns in natural populations. In a typically small, random sample of loci compared among populations, there easily can be one or two loci that appear to have extreme patterns of population differentiation (biship pine, Millar et al. 1988; ponderosa pine, Latta & Mitton 1999; cod fish, Pogson et al. 1995; eastern oysters, Karl & Avise 1992; blue mussels, Riginos et al. 2002). Demography will have generalized effects across the genome, whereas selection will influence variation at localized sections of tightly linked loci in the genome. Thus, a locus-specific pattern of variation over and above the genomic pattern implicates selection, provided an accurate estimate of the neutral genomic variance and corresponding central moment can be made. In a small sample of loci, one gene with especially high differentiation may indicate divergent selection, or it could reflect a broad (and poorly sampled) distribution of neutral variation. Demographic and historical inferences can be very different depending on the set of loci chosen to represent neutral variation (Ford 2002). Proper interpretation of variation among loci requires two things: (1) accurate assessment of the neutral drift variance among loci, and (2) an objective

basis for discerning whether selection is influencing either or both ends of the distribution (balancing or purifying selection).

It has become well known that a large amount of stochastic variation in the evolution of independent loci is generated by neutral drift (Knowles & Maddison 2002; Rosenberg & Nordberg 2002; Rosenberg 2003). It is also expected that hybridization and secondary contact zones, or any contrast in phylogeographic structure among populations, will generate larger among-locus variation in levels of differentiation than expected within a set of populations with homogeneous structure (Robertson 1975; Harrison 1990; Latta 2003). Thus, surprisingly large heterogeneity in $F_{ST}$ among loci is often consistent with neutral drift expectations under these evolutionary scenarios (Latta 2003). In other cases, more detailed investigations of outlier loci have strongly supported a causative role for selection (e.g., the pantophysin gene in cod; Pogson 2001; Pogson & Mesa 2004; Case et al. 2005). In two cases where genomic sampling error was suspected as the cause of an apparent distinction between relatively low-$F_{ST}$ allozymes and high-$F_{ST}$ DNA markers, follow-up studies added more loci to the data set but did not contrast the empirical results to an estimate of the expected neutral drift variance (oysters, McDonald et al. 1996; mussels, Bierne et al. 2003). Instead, after estimating a statistic describing locus-specific differentiation, a nonparametric test was used to reject homogeneity between marker classes (allozyme versus DNA; McDonald et al. 1996; Bierne et al. 2003).

Even when classes of marker show patterns that are not significantly different, the variance observed among loci may be elevated above neutral expectations because of

selection on individual loci. To address this, an estimate of the drift variance among loci is needed to rigorously identify candidate loci potentially under selection (or linked to targets of selection). A genome scan is an efficient approach for accomplishing this. The procedure involves collecting low-resolution polymorphism data on a large genomic scale to more completely sample genomic diversity, identifying "outlier" loci based on statistical criteria or by comparison to a null model, and ultimately testing observed data against alternative nonequilibrium models of historical demography/selection (Kauer et al. 2003; Akey et al. 2004; Storz et al. 2004; Haddrill et al. 2005). However, genome scans may be unreliable if an equilibrium model is used to identify outlier loci in nonequilibrium populations (Nei & Maruyama 1975; Robertson 1975; Luikart et al. 2003). Of more specific interest here, Bierne et al. (2003) argued that in a secondary contact zone, a combination of pre-contact differentiation variance and post-contact differential introgression can inflate the $F_{ST}$ variance over equilibrium expectations. Unfortunately, it is rarely known with any certainty whether population differentiation originated in situ or resulted from allopatric divergence and secondary contact at some time before present (Endler 1977), or even whether the populations are at drift-migration equilibrium. If the among locus variance is a function of these unknowns, then there is little clarification provided by comparison of observed patterns to an equilibrium model and little basis for determining the proper nonequilibrium model to apply to clinal populations.

Some resolution has been provided for the opposing interpretations of allozyme and DNA marker variation in the Scandinavian mussel hybrid zone. This was accomplished in two

ways.  First, because some markers were not alternatively fixed between the parental

populations (*Mytilus edulis* and *M. trossulus*), it was informative to examine allele

frequency differentiation across the narrow hybrid zone relative to that between allopatric

parental populations (Riginos & Cunningham 2005).  When scaled by divergence

between parental species, DNA markers showed less differentiation across the hybrid

zone as compared to allozyme loci.  While this result may still suffer from biased

sampling of DNA markers and did not account for the drift variance among loci, it was

coupled with a second criterion, environmental correlations, to support selection as the

cause of allele frequency patterns at several allozyme loci (Riginos & Cunningham

2005).

Here, we focus on the other example of high among-locus variance in $F_{ST}$ between

Atlantic and Gulf of Mexico populations of the eastern oyster, *Crassostrea virginica*, to

revisit the question of whether neutral migration-drift processes can explain the genomic

heterogeneity.  This species has a broad range distribution spanning from New

Brunswick, Canada to Brazil (Carriker & Gaffney 1996).  Reproduction is by

synchronous broadcast spawning and external fertilization, followed by a planktonic

larval stage lasting two to three weeks before permanent settlement on hard estuarine

substrate (Thompson et al. 1996).  The lengthy planktonic larval stage of eastern oysters

confers a great potential for long distance dispersal (Kinlan & Gaines 2003).  As a result,

genetic differentiation found between Atlantic and Gulf of Mexico oyster populations

was both surprising and intriguing, given their continuous distribution around Florida.

Reciprocal monophyly was observed for restriction fragment length polymorphisms

(RFLPs) of mitochondrial (mt)DNA (Reeb & Avise 1990).  Strong Atlantic-Gulf of

Mexico allele frequency differentiation was also found in single copy nuclear (scn)DNA

markers (Karl & Avise 1992).  Based on these macrogeographic studies, relative

homogeneity of allele frequencies *within* Atlantic and Gulf of Mexico regions suggested

that high gene flow indeed was maintained across populations in large parts of the species

range.  The Atlantic-Gulf of Mexico differentiation, localized primarily along eastern

Florida, reflects a combination of vicariant history and contemporary barriers to gene

flow (Reeb & Avise 1990).  Curiously, allozyme loci in the oysters showed a very

different pattern.  A broad survey examining 21 polymorphic allozyme loci in Atlantic

and Gulf of Mexico oysters uncovered subtle differentiation between these regions at

only two loci (Buroker 1983; Cunningham & Collins 1994).

The contrast between allozyme and DNA patterns led Karl & Avise (1992) to

hypothesize that geographically uniform balancing selection maintained homogeneity of

allele frequencies at allozyme loci, while population differentiation of DNA

polymorphisms reflected a history of vicariance.  Because Karl & Avise (1992) reported

on only four scnDNA markers plus mtDNA, McDonald et al. (1996) examined whether

genomic sampling error might have created a false dichotomy between patterns of

differentiation at allozyme and DNA markers.  To accomplish this, McDonald et al.

(1996) compared the previously published allozyme data with the Karl & Avise (1992)

markers and seven new scnDNA RFLPs, all analyzed in single representative populations

from both the Atlantic and Gulf of Mexico.  Strong differentiation was found with the

Karl & Avise (1992) markers, replicating their results, but the new scnDNA markers

exhibited low differentiation comparable to levels found in allozymes (McDonald et al. 1996). McDonald et al. (1996) concluded that patterns of Atlantic – Gulf differentiation did not differ between marker classes. Additional studies have contributed data on Atlantic – Gulf oyster population structure from yet more DNA markers (Hoover & Gaffney 2005), but the hypothesis that genetic drift can explain the among-locus heterogeneity in differentiation has never been formally tested.

With more intensive sampling in eastern Florida, it was shown that the transition in oyster mtDNA and scnDNA allele frequencies occurred as coincident, sharp step clines near Cape Canaveral (Hare & Avise 1996). Several facts support the hypothesis that this is a secondary contact zone. First, if introgression is ignored by comparing populations far removed from the cline, mitochondrial haplotypes in the Atlantic and Gulf of Mexico are reciprocally monophyletic with average sequence divergence of 2.6 % between the clades (Reeb & Avise 1990). Similar Atlantic-Gulf of Mexico intraspecific phylogeographic breaks in multiple estuarine and marine species provide comparative data implicating shared vicariance (Avise 1992; Avise 2000; but see Lee & Ó Foighil 2004). Hybridization and clinal variation are also common features between parapatric estuarine species in eastern Florida (Duggins et al. 1995; Sarver et al. 1992; reviewed in Gardner 1997). Finally, a well-recognized boundary zone between temperate and subtropical faunal provinces occurs along the eastern Florida coast (Briggs 1974). Although it is impossible to confirm that the oyster cline originated with secondary contact, this likely scenario would produce a large variance among loci in levels of differentiation (Bierne et al. 2003). The narrowness of some oyster clines suggests that

they may be maintained by selection (Barton & Hewitt 1985), further adding to the among-locus variance in differentiation.

A cline resulting from secondary contact should deteriorate rapidly due to gene flow and introgression of neutral markers, unless it is stabilized by biotic or abiotic factors (Endler 1977; Barton & Hewitt 1985). Atlantic and Gulf of Mexico oysters are routinely interbred for the purposes of aquaculture (X. Guo and S. Allen pers. comm.), indicating that any intrinsic genomic incompatibilities may be minimal at the $F_1$ generation and beyond. Local samples along the oyster cline mostly exhibited equilibrium genotype frequencies within and among loci, also consistent with a lack of strong reproductive barriers between populations on either side of the cline and/or relatively local recruitment (Hare & Avise 1996). These considerations suggest two possible explanations for sharp oyster clines: (1) the clines are decaying every generation, but their narrowness reflects the recency of secondary contact or the strength of a dispersal barrier at Cape Canaveral; (2) divergent selection maintains differentiation at some loci, while gene flow dissipates the differentiation at others. The second process would increase among-locus variance in $F_{ST}$, but to detect this in a genome scan would require a suitable null model for neutral divergence. In this report, we supply data on the temporal stability of the cline and increase the genomic sampling of polymorphisms ten-fold to test the second hypothesis using a genome scan.

Our first goal was to determine whether the neutral distribution of $F_{ST}$ was sensitive to details of a secondary contact history. Improving on a test introduced by Lewontin &

Krakauer (1973), Beaumont & Nichols (1996) found that simulated distributions of neutral $F_{ST}$, anchored on the observed genomic mean $F_{ST}$ and plotted against locus heterozygosity, provided expectations robust to many features of population structure and demography as well as to variation in mutation rates among loci.  Specifically, neutral genetic drift produced equivalent distributions of $F_{ST}$ in colonization and island models with the same overall mean $F_{ST}$.  However, as predicted by Robertson (1975) and Nei & Maruyama (1975), models in which more than two subpopulations had different divergence times or migration rates showed a broader distribution of $F_{ST}$ relative to a symmetric island model (Beaumont & Nichols 1996).  Fortunately, restricting analyses to pairwise population comparisons nearly eliminates the risk of simulating an overly simplistic and narrow null distribution of $F_{ST}$ in which outliers would be incorrectly identified as non-neutral (Tsakas & Krimbas 1976; Vitalis et al. 2001; Beaumont 2005). The neutral distribution of $F_{ST}$ may not be as robust in nonequilibrium secondary contact populations (Bierne et al. 2003).  Therefore, we utilized simulations to compare the neutral distributions of $F_{ST}$ in two populations generated under migration-drift equilibrium versus different stages of nonequilibrium after secondary contact, using the same genomic mean $F_{ST}$ in each case.

Our second goal was to increase the number of loci used to estimate the genomic mean $F_{ST}$ between Atlantic and Gulf of Mexico oyster populations and determine whether neutral migration-drift processes can explain the among-locus heterogeneity.  We surveyed amplified fragment length polymorphisms (AFLP) in one Atlantic population just north of the step cline and one population at the opposite end of the cline in

southwestern Florida, just inside the Gulf of Mexico (Hare & Avise 1996).  AFLP

analysis is ideal for rapidly screening many loci that are randomly distributed throughout

the genome, allowing a more accurate estimate of the genomic differentiation among

populations and a more thorough test for portions of the genome under divergent

selection.

## Materials and Methods

### Sampling and DNA Extraction

Adult oysters were sampled in two locations along the Florida coast (Fig. 1), New

Smyrna Beach in 2002 (NSB; 29.0795° N, 80.9559° W) and Port Charlotte in 2004

(PCH; 26.4286° N, 81.8668° W).  These sample sites were selected as representatives of

divergent populations in the Atlantic and Gulf of Mexico based on Karl & Avise 1992

(see Fig. 1).  For 30 individuals from each site, gill or mantle tissue was dissected and

preserved in 95% ethanol for subsequent DNA extraction. Approximately 20 mg of tissue

was used for DNA extraction using the DNeasy 96 Tissue kit (Qiagen Inc., Valencia,

CA) following the manufacturer protocol for animal tissues.  Genomic DNA was diluted

to 50 ng/μl based on spectrophotometry.

### RFLP Analysis

To test the temporal stability of the oyster allele frequencies in the sampled locations, one

mitochondrial (mtDNA) and two nuclear loci were amplified with the polymerase chain

reaction (PCR) and digested with restriction enzymes to score RFLPs previously shown

to be differentiated between Atlantic and Gulf of Mexico populations.  Procedures

followed Hare et al. (1996) for the CV-32.4 and CV-7.7 loci and Hare & Avise (1996) for

the mtDNA RFLP in the NADH dehydrogenase 4 gene.  To test for significant deviations

from Hardy-Weinberg equilibrium, we used permutation tests as implemented in FSTAT

ver. 2.9.3 (Goudet 2001).

**AFLP Analysis**

AFLP fragments were generated according to a modified version of the procedure outlined by Vos et al. (1995). Primer and adaptor sequences are provided in Table 1. For each sample, 50 ng of genomic DNA were added to a 50 µl total volume solution composed of: 1X enzyme buffer, 0.1 mg/ml BSA, 0.2 mM ATP, 1.8 mM dithiothreitol, 0.05 µM *Mse*I adaptor, 0.05 µM *Eco*RI adaptor, 25 U *Mse*I, 10 U *Eco*RI, and 100 U DNA ligase. To allow for complete digestion and ligation, this reaction was incubated at 37°C for 5 hours and then diluted 1:10 with dH$_2$O.

A pre-selective PCR was then performed on each sample in 10 µl reactions containing 2.0 µl 1:10 digestion-ligation product, 1.5 mM MgCl$_2$, 0.125 mM dNTP, 1X PCR buffer, 0.5 µM primer *Eco*RI-A, 0.5 µM primer *Mse*I-C, and 0.5 U *Taq* polymerase (Invitrogen). Pre-selective PCR conditions were: 95°C for 1.5 min, then 22 cycles stepping from 95°C for 0.5 min to 56°C for 1 min and 72°C for 1 min. The selective PCR reactions contained 2.0 µl of pre-selective PCR product diluted 1:40, 1.5 mM MgCl$_2$, 0.125 mM dNTP, 1X PCR buffer, 0.5 µM fluorescently labeled primer *Eco*RI-ACT, 0.5 µM primer *Mse*I-CXX, and 0.5 U *Taq* polymerase. Four different *Mse*I-CAX primers were used (Table 1). The selective PCR conditions began with 95°C for 1.5 min; then 9 touch-down cycles of 95°C for 0.5 min, 65°C for 0.5 min, 72°C for 1 min where the annealing temperature decreased by 1°C per cycle; followed by 22 cycles of 95°C for 0.5 min, 56°C for 0.5 min and 72°C for 1 min.

AFLP fragments were electrophoresed with an ABI 3100 genetic analyzer (Applied Biosystems). Using GeneScan 3.7 and Genotyper 2.5 software (Applied Biosystems), individuals were scored for fragment presence/absence at well-resolved loci with sizes between 75 and 490 bp. AFLP loci were accepted as scorable if, across all gels and all individuals (n = 60), the distribution of fragment lengths was unimodal. Then, a ±0.5 bp bin was centered on the mean fragment size for each locus. Because this bin size contained the entire distribution of fragment sizes for each scored locus, and excluded fragments from loci of similar size, miscalling fragments of similar size was avoided. Triplicate assays on the genomic DNA of several individuals from each population demonstrated repeatability with all four primer pairs. Loci with similar fragment profiles across individuals, potentially indicating linkage, were identified using the 'loci filtering' option of AFLPOP (Duchesne & Bernatchez 2002).

### Population Differentiation

Because most AFLPs are dominant markers, in which heterozygotes cannot be distinguished from band-present homozygotes, frequencies of the presence and absence alleles for each locus were calculated assuming Hardy-Weinberg equilibrium. Monomorphic loci were excluded from analyses. For each remaining locus, mean allele frequency was calculated as the arithmetic average of the band-absent frequency in the two populations. For each locus, genetic differentiation between populations was measured in terms of $F_{ST}$, calculated as $(1-(H_S / H_T))$, where $H_S$ is the mean locus-specific heterozygosity within populations and $H_T$ is the locus-specific total heterozygosity (Nei 1973). In addition, a correction for finite sample size was applied as suggested by Nei &

33

Chesser (1983). Negative values of $F_{ST}$ were converted to zero. Then global $F_{ST}$, a measure of central tendency for the distribution of $F_{ST}$ across loci, was calculated as (1- (mean $H_S$ / mean $H_T$)), where mean $H_S$ and mean $H_T$ are the arithmetic averages across loci of $H_S$ and $H_T$ (Nei & Chesser 1983). Standard error of $F_{ST}$ for individual loci was estimated by calculating the $F_{ST}$ from a sample of 30 individuals from each of two synthetic populations (N=1000 for each population) in Hardy-Weinberg equilibrium, where the allele frequencies of these populations were equal to those observed for the corresponding AFLP locus, and permuting this process 1000 times.

Previously published codominant allele frequencies for 27 loci were available for samples from South Carolina and western Florida, representing Atlantic and Gulf of Mexico oysters, respectively. These included 19 allozyme loci from Buroker (1983) with the frequency of the most common allele $\leq 0.99$ (n = 90-100 individuals per sample), four scnDNA RFLPs initially reported by Karl & Avise (1992; n = 21-38), and four scnDNA polymorphisms (indel or RFLP) from McDonald et al. (1996; n = 60). Allele frequencies for two of the Karl & Avise (1992) loci were obtained from the reanalysis by Hare et al. (1996). Mean $F_{ST}$ between these populations was calculated among loci, weighted by their heterozygosities, using the program DATACAL distributed with fdist2 (Beaumont & Nichols 1996; http://www.rubic.rdg.ac.uk/~mab/software.html).

### AFLP Isolation and Characterization

For one AFLP locus with high $F_{ST}$, a DNA fragment was isolated from a band-present individual on a polyacrylamide gel and sequenced directly using the AFLP primers.

Gene-walking PCR was used to amplify segments of genomic DNA flanking the AFLP

fragment following the recommended procedures in the GenomeWalker kit (BD

Biosciences Clontech). Gene-walk amplifications were cloned or directly sequenced, and

PCR primers were designed to amplify a short segment of DNA containing the

polymorphic *Eco*RI site that had generated the AFLP (oyCAA273-R3, 5'-

CCCAAATTCAAAGCTTCA ACG A-3'; oyCAA273-L4, 5'-

ACCCCGTGTGCGACAATTATAC-3').


## Comparison of Equilibrium and Nonequilibrium Secondary Contact $F_{ST}$ Distributions

Simulations were conducted to generate the null distributions of $F_{ST}$ expected among

neutral loci under two conditions; an equilibrium between gene flow and drift versus

nonequilibrium secondary contact between two populations. The equilibrium simulation

was performed using the computer program WINK (Wilding et al. 2001). Drawing

initially from a uniform distribution of allele frequencies, 500 bi-allelic loci were

simulated with a mutation rate $\mu$ in two diploid populations of size N sharing migrants at

a constant symmetrical rate m. At generation 10N (i.e., at equilibrium), 30 individuals

were sampled from each population and their genotypes stored. Mean allele frequency

and $F_{ST}$ were calculated for each locus in the same manner as for the AFLP data (band

present is dominant to band-absent). Repeating this simulation numerous times provided

a total of 144,331 independent loci. Monomorphic loci were excluded, and negative

values of $F_{ST}$ were converted to zero.

Because the simulated $F_{ST}$ distributions are robust to different combinations of population size and mutation rate (Beaumont & Nichols 1996; Wilding et al. 2001), these parameters were held constant at a 10:1 ratio, calculated based on N$\mu$ estimated from AFLP homozygosity (1 – mean $H_S$) using equation 2.7 in Kimura (1968). Migration rate was adjusted in the simulation to produce a global $F_{ST}$ similar to, but never below, the observed global $F_{ST}$. As a conservative measure, maintaining a simulated global $F_{ST}$ above the empirical value reduced the likelihood of false positives, i.e., neutral oyster loci with high $F_{ST}$ misidentified as targets of divergent selection.

Simulations were also conducted under a non-equilibrium neutral model of secondary contact for comparison to the equilibrium null distribution of $F_{ST}$. A two-stage simulation using EASYPOP v1.8 software (Balloux 2001) generated data for two populations of 1000 diploid individuals with a constant mutation rate $\mu$. Approximately 1,000 bi-allelic codominant loci were initiated with a random assignment of alleles using the 'maximum variability' option. During the first 10,000 generations, m = 0 between the populations. Subsequent generations had a symmetrical constant m > 0 to simulate secondary contact. In separate simulations using identical starting conditions, genotypes for 30 individuals were sampled at t generations following secondary contact (t = 100, 200, 400, 800, 2500) to examine different stages of introgression. Codominant genotypes were converted to dominant presence/absence data. Simulated secondary contact data were analyzed as described above for the equilibrium model, including adjustment of migration rate m for each simulation so that at time t, the simulated global $F_{ST}$ was similar to that observed in the AFLP data. Binned distributions of $F_{ST}$ under different

equilibrium and nonequilibrium conditions were compared using the Kolmogorov-Smirnov non-parametric test with a strict Bonferroni correction for multiple comparisons (Sokal & Rolf 1995, pp. 434-439).

**Outlier Test for Selection**

For populations diverging by genetic drift, the expected $F_{ST}$ and its variance among loci depend on the heterozygosity of the loci. In order to test for selection, the joint distribution of $F_{ST}$ and heterozygosity is derived by simulation under a null hypothesis of neutral genetic drift with migration (Beaumont & Nichols 1996). To test for selection among the previously published codominant loci, coalescent simulations using Fdist2 (Beaumont & Nichols 1996) under the infinite-allele model were performed for 50,000 independent loci in two ways: 50 populations or 2 populations simulated. In either case n = 50 individuals were sampled from each of two populations. The simulation program adjusts mutation rate to produce a nearly uniform distribution of heterozygosity and sets migration rate to reproduce the empirical weighted mean $F_{ST}$ at equilibrium. Another module of the Fdist2 package (cplot) was used to calculate 0.99 quantiles, based on many overlapping bins, for the two sets of simulated $F_{ST}$ values. The probability of extreme $F_{ST}$ observations relative to a simulated distribution was calculated as in Beaumont and Nichols (1996; program pv distributed with Fdist2) and compared to a conventional Bonferroni correction for an experiment-wide alpha of 0.05. The $F_{ST}$ observed for locus CAA-270, the AFLP locus converted to a codominant marker, was then compared to the distribution simulated above, which was based on the weighted mean of the previously

published loci.  The probability of the CAA-270 $F_{ST}$ was calculated relative to the same distribution.

For the AFLP loci, we analyzed $F_{ST}$ conditioned on the mean frequency of the absent allele.  To identify empirical AFLP loci for which the neutral drift model could be rejected, the 0.99 quantile of the simulated $F_{ST}$ data was plotted against the mean allele frequency for the recessive band-absent allele.  The 99th quantile and probability of empirical $F_{ST}$ was calculated as above.  The null model of neutral divergence was rejected for AFLP loci with p-values less than the Bonferroni-corrected 0.05 alpha.

# Results

## Temporal Stability of Allele Frequencies

Allele frequencies were not significantly different when compared across eleven years (4-5 generations) at New Smyrna Beach and across 14 years (4-7 generations) at Port Charlotte (Fig. 1). Genotype proportions at the nuclear loci were also consistently very close to Hardy-Weinberg expectations (Hare et al. 1996). This suggests that Hardy-Weinberg deviations in neutral AFLP markers, if present, would have mostly technical rather than biological causes, and therefore relatively small effects.

## AFLP Differentiation in Oysters

A total of 226 AFLP bands (loci) were reliably scorable for polymorphism. Of these, 11 (4.9%) were monomorphic and excluded from further analysis. The number of scorable AFLP loci varied somewhat among the four primer pairs, but the proportion of polymorphic loci within a population was always greater than 80% (Table 2). A total of 27 loci had allele frequency differences of at least 0.35 between the Atlantic and Gulf samples. Among these high-differentiation loci, no pair exhibited identical patterns of fragment presence and absence across individuals (no tight linkage). Over all loci, the global $F_{ST}$ was 0.0917, but varied among primer pairs (Table 2).

**Comparison of Simulated $F_{ST}$ Distribution under Two Evolutionary Histories**

When two populations diverge in allopatry ($m = 0$), the magnitude of neutral $F_{ST}$ at

equilibrium depends on the balance between drift ($N_e$) and mutation ($\mu$) (Maruyama

1970). In our simulations, this equilibrium global $F_{ST}$ preceding secondary contact was

0.38, obtained after approximately 3000 generations. For any given level of constant

gene flow after secondary contact, as expected, the global mean $F_{ST}$ decreased

monotonically over time until a drift-migration-mutation equilibrium was reached (e.g.,

$Nm = 2$ in Fig. 2a). Also, the among-locus variance in $F_{ST}$ decreased in parallel with $F_{ST}$

after secondary contact (Fig. 2a). We were interested in the equilibrium and

nonequilibrium conditions that could produce the observed global $F_{ST} = 0.092$, and the

among-locus variance in neutral $F_{ST}$ associated with those conditions. For the parameters

in our simulation ($N = 1000$ and $\mu = 0.0001$), only $Nm = 1.48$ produced an equilibrium

$F_{ST}$ of 0.092, obtained after 2500 generations (Fig. 2b). Higher levels of constant gene

flow after secondary contact produce $F_{ST} = 0.092$ transiently during the approach to

equilibrium (Fig. 2b). For example, with post-contact $Nm = 9.0$, the global $F_{ST}$ quickly

decreased from 0.38 to 0.092 in $t + 50$ generations (Fig. 2b). Lower rates of constant

gene flow after secondary contact increased the number of generations before $F_{ST}$

transiently reached 0.092 on the approach to equilibrium (Fig. 2b). It is usually difficult

to determine whether populations are at equilibrium or not. Therefore, the important

result from these simulations is that every time point and gene flow combination that

yielded the empirical estimate of global $F_{ST} = 0.092$ also produced the same variance in

$F_{ST}$ that was observed at equilibrium, 0.013 (Fig. 2b). This result is reinforced by a

comparison of frequency distributions for binned $F_{ST}$ values from different simulations

(Fig. 3). There were no significant differences in pairwise tests among distributions from two nonequilibrium and two equilibrium simulations that produced a global $F_{ST}$ of 0.092 (six tests, Bonferroni corrected alpha = 0.008; Fig. 3). Thus, *for a given $F_{ST}$*, there was no period after secondary contact when neutral processes inflated the among locus variance in $F_{ST}$ above that expected at equilibrium.

### Tests of Selection

The weighted mean $F_{ST}$ among 27 previously published codominant markers was 0.1142 (range 0 to 0.6647). In all simulations, the weighted mean $F_{ST}$ for simulated data was never more than 0.003 higher than the empirical mean. Compared with the 0.99 quantile of simulated values, only one (3.7 %) previously published locus, CV-7.7, had a higher $F_{ST}$ value than expected under neutral evolution (Fig. 4a). The probability of the $F_{ST}$ value at CV-7.7 was 0.0026 or 0.0045, depending on the details of the simulation defining the neutral distribution. Neither p-value was significant after Bonferroni correction with an experiment-wise alpha of 0.05 (for 27 loci, corrected alpha = 0.0019).

Because distributions generated under the equilibrium and nonequilibrium models did not differ, an equilibrium island model was used to produce the null distribution of $F_{ST}$ against which the empirical AFLP values could be tested. The mean homozygosity for polymorphic AFLPs, 0.745, was used to estimate $N_e\mu = 0.13$. This population-mutation parameter value was assumed for all simulations by setting $N = 1000$ and $\mu = 0.0001$. After removal of monomorphic loci (13.3 %), a total of 125,091 simulated loci remained. The global $F_{ST}$ among polymorphic simulated loci was 1.1% higher than the empirical

estimate from AFLPs, providing a slightly conservative test for non-neutral $F_{ST}$ values. The test revealed three AFLP loci (1.4 %) with $F_{ST}$ values greater than the 0.99 quantile for $F_{ST}$ simulated under neutral drift-migration (Fig. 4b). These three loci are the strongest candidates for divergent selection. However, no locus remained significant after Bonferroni correction with an experiment-wise alpha of 0.05 (for 215 loci, corrected alpha = 0.0002).

The homology of same-sized AFLP bands was confirmed for locus CAA-270 based on DNA sequence comparisons between fragments isolated from the New Smyrna Beach and Port Charlotte samples. Gene walking and sequence analysis revealed that the AFLP was caused by a nucleotide substitution in the *Eco*RI site at one end of the fragment (data not shown). Codominant analysis of this *Eco*RI RFLP produced allele frequency estimates very similar to those calculated from AFLP data. Also, RFLP genotype frequencies did not deviate significantly from Hardy-Weinberg expectations ($F_{IS}$ = -0.033 and -0.018, $P \geq 0.70$ for each sample). Finally, $F_{ST}$ calculated from codominant CAA-270 data was more extreme than the 0.99 quantile based on previously published codominant data (Fig. 4a).

**Discussion**

**Genomic Variance after Secondary Contact**

Genome scans have been heralded as an efficient and robust method of evaluating among-locus variance in a population genetic statistic, such as $F_{ST}$, relative to the variance expected under neutrality. A major constraint on the use of genome scans, however, appeared to be the appropriateness of an equilibrium null model when applied to clinal populations that might represent nonequilibrium stages of hybridization or secondary contact (Luikart et al. 2003; Bierne et al. 2003). Yet, these are particular contexts in which the statistical evaluation of among-locus variance would be especially valuable (e.g., Payseur et al. 2004). Therefore, we used simulations to test whether a history of allopatric divergence and recent secondary contact elevates the neutral variance of $F_{ST}$ over what would be expected from an equilibrium model for a given genomic global $F_{ST}$. We found no statistically significant distinction between the $F_{ST}$ distributions simulated under these contrasting models. This result may not match intuitive expectations (Bierne et al. 2003) because it is well known that the distribution of $F_{ST}$ has a high variance immediately after secondary contact compared with the variance obtained once neutral migration-drift equilibrium is reached. However, our simulations demonstrate that non-equilibrium changes in the distribution of $F_{ST}$ parallel the reduction in global $F_{ST}$ that occurs as neutral gene flow homogenizes secondary contact populations. Regardless of how recently secondary contact occurred, whether the oyster populations are at equilibrium, or how much constant gene flow is occurring to produce the empirical global $F_{ST} = 0.092$, in each case this magnitude of global $F_{ST}$ is expected to

43

have the same distribution among loci.  Thus, a reliable test for outliers, when applied

pairwise to populations in secondary contact, does not hinge on equilibrium status or the

age of secondary contact.

Another goal of this study was to increase the genomic sampling of oyster loci to

determine how much among locus variance in $F_{ST}$ can be explained by neutral migration-

drift processes.  The genome scan results make it clear that genetic drift and migration

can explain most of the among-locus variance in $F_{ST}$ observed between Atlantic and Gulf

of Mexico oyster populations to date.  This conclusion is consistent with that of

McDonald et al. (1996), but is now based on a ten-fold increase in the number of loci

examined and a comparison to the distribution of $F_{ST}$ expected under the null hypothesis

of neutrality.

In contrast to our conclusions about the large neutral variance of $F_{ST}$ and its robustness to

equilibrium assumptions, our interpretation of individual outlier loci is more equivocal

until additional independent criteria can be applied to evaluate marginally significant

non-neutral patterns.  Three aspects of the methods and results contribute to this

uncertainty: (1) the probability of false positives with multiple testing, (2) small

population sample sizes and Hardy-Weinberg assumptions with AFLPs, and (3)

assumptions underlying the genome scan.  A total of four loci (three AFLP, one

codominant) were significant at the $P < 0.01$ level.  However, with 215 independent tests

we expect two loci to fall above the $99^{th}$ quantile by chance (type I errors), and the

binomial probability of finding three or more loci above this critical value is $P = 0.36$.  A

standard Bonferroni correction for multiple tests applied to the conventional alpha of 0.05 yields a critical value of 0.0002 for all 215 AFLPs and 0.0019 for 27 codominant loci. Using the procedures of Beaumont & Nichols (1996; pv program) to estimate the probability of each empirical $F_{ST}$ value relative to the simulated distribution, no locus had a p-value below the Bonferroni-corrected alpha (Fig. 4b). Bonferroni corrections have been criticized as too conservative for applications in ecology and genomics (Moran 2003). As the number of loci increases, a Bonferroni correction over the entire data set becomes prohibitive for an exploratory study aimed at identifying candidate loci that exhibit patterns of selection. Alternative measures of significance in terms of false discovery rate have been proposed (Storey & Tibshirani 2003) but do not alter our conclusions here.

Ultimately, the biological interpretation of outlier loci in a genome scan is not clarified by statistics, but by more focused data collection and analysis of these candidate loci. In this study in particular, the small population sample sizes and low-resolution dominant variation created large sampling error around locus-specific estimates of allele frequency and $F_{ST}$. Sampling error was also evident in the widely fluctuating quantiles on the simulated null distribution. These sampling errors were presumably random. In addition, Hardy-Weinberg equilibrium was assumed for each AFLP locus to estimate allele frequencies. Although this assumption is justified by genotype proportions at several codominant markers and was confirmed for one AFLP locus, bivalves in general and oysters in particular have a propensity for heterozygote deficits in natural populations (Hare et al. 1996; Gaffney 1996). We simulated a range of Hardy-Weinberg deviations

to explore their potential to affect our results. Even if the codominant polymorphisms underlying oyster AFLPs have a systematic bias toward heterozygote deficits, the impact on $F_{ST}$ estimated from dominant band presence/absence data depends on the distribution of Hardy-Weinberg deviations across the populations. For example, a downward bias in $F_{ST}$ results when both populations are deficient in heterozygotes, whereas an upward bias is created if only one population is deficient (data not shown). These considerations suggest that our general conclusions regarding the breadth of the neutral variance in $F_{ST}$ are robust, but strong conclusions about individual loci are not warranted at this time. The marginally significant outlier loci provide candidates for more in-depth analyses that will be capable of more definitively rejecting neutral evolution should this hypothesis be false.

Finally, interpreting outlier loci in a genome scan in terms of selection assumes that the observed genomic mean $F_{ST}$ is a good approximation of the average level of differentiation among neutral loci. Loci under divergent selection would upwardly bias the initial estimate of global $F_{ST}$, creating a more conservative test for this form of selection, whereas balancing selection acting on some loci in the sample would have the opposite effect. It is possible, even probable, that balancing selection is influencing patterns of differentiation at some oyster loci, as posited by Karl & Avise (1992). We did not test this hypothesis because $F_{ST}$ distributions provide low statistical power for detecting balancing selection (Beaumont & Balding 2004). To account for such potential biases, genome scans for divergent selection sometimes remove statistically significant outliers from the data set to calculate a less upwardly biased estimate of the global mean

and rerun simulations to iteratively refine estimates of the neutral variance (Wilding et al. 2001).  Without being able to test each end of the distribution with equal power, however, this one-sided 'refinement' of statistical results might simply shift bias rather than remove it.

## Factors Maintaining the Oyster Step Cline

This study was partly motivated by the stability in allele frequencies observed over several generations of dispersal and reproduction among oysters in eastern Florida (Fig. 1).  Temporal stability was found at New Smyrna Beach, only 40 km north of the oyster step cline, where typical patterns of oyster gene flow (Rose et al. 2006) should have generated visible introgression under the null hypothesis of neutrality.  Instead, like many other hybrid zones (Barton & Hewitt 1985), this temporal stability suggests that biotic and/or abiotic processes are maintaining strong population differentiation at some loci.

Hare & Avise (1996) speculated that the forces maintaining this stepped cline included a low density of oysters in the Cape Canaveral lagoons compared to the north and south. The low density in this area may reflect relatively poor habitat in these lagoons or physical impediments to larval dispersal, both of which would create a barrier to gene flow.  This barrier hypothesis has not been tested, but even if these mechanisms are responsible for sharpening the cline (Endler 1977) and limiting introgression, their effect would be to maintain latitudinal adaptations evolved in allopatry.  A generalist species such as the oyster maintains flexibility in its response to environmental heterogeneity by generating an abundance of genetically diverse progeny (Williams 1975), each of which

is phenotypically plastic (Warner 1997). With high gene flow among populations continuously distributed along a latitudinal gradient, this life history strategy would produce 'local' adaptation only on very large geographic scales equivalent to climatic provinces (Conover & Schultz 1995; Dittman et al. 1998; Whitehead & Crawford 2006). If gene flow between climatic provinces is terminated by a vicariant event, then the balance shifts from the homogenizing influences of gene flow toward diversifying selection for adaptations to regional climate. Spatially variable selection pressures could emerge from any factors likely to co-vary with climate such as species richness, community composition, and parasite load. Secondary contact would bring together populations with more latitudinal specialization than would occur elsewhere in the species range (where gene flow was never terminated). In the eastern oyster, this model predicts that maintenance of these latitudinal adaptations over a relatively small geographic scale depends on dispersal barriers, relative fitness of hybrids, and relative fitness of 'specialist' genotypes along the environmental and ecological gradients in eastern Florida.

The oyster genetic cline is located within a broad ecological gradient where many selective agents potentially reinforce population differences that evolved in allopatry. Under these circumstances, the exploratory aspect of our genome scan was a desirable feature because it identified potentially non-neutral candidate loci without reference to phenotypes, coding versus noncoding sequence, or specific models of selection.

Our results are a starting point for population genetic and functional analyses, which are needed to detect selection and to determine how it is operating. Previous genome scan studies have used parallel testing to substantiate selection acting on outlier loci by showing that they have higher than expected $F_{ST}$ between two morphotypes in several geographic replicates (Wilding et al. 2001; Campbell & Bernatchez 2004). Such geographic replication is not possible with the oyster cline, but cline shape across the secondary contact zone can be estimated for each AFLP locus to further test whether $F_{ST}$ outliers are spurious. Under the divergent selection hypothesis, loci with high Atlantic – Gulf of Mexico $F_{ST}$ should be associated with exceptionally steep clines that have a step at Cape Canaveral. In contrast, under the null hypothesis clines should be gradual and wide or might show smaller-scale patchy differentiation if genetic variation is shaped by habitat heterogeneities within regions. Conversion of AFLPs into codominant markers will permit a more robust analysis of clinal variation to estimate the strength of selection opposing introgression (Barton & Gale 1993).

Candidate loci in oysters can be tested for an association with quantitative, fitness-related traits in an experimental reciprocal transplant of Atlantic and Gulf of Mexico populations and ultimately in the wild hybrid population (Rieseberg & Buerkle 2002). Where such associations exist, we can begin to infer the genetic basis for adaptive variation across the oyster cline. It will be important to supplement these inferences from field populations with a more conventional QTL mapping approach using experimental crosses to determine whether candidate loci are clustered in regions of low recombination and

whether they correspond to QTL for adaptive traits such as growth rate or disease resistance (Campbell & Bernatchez 2004).

Differences among chromosomes in ploidy, inheritance patterns, and recombination rates can elevate the genomic variance in $F_{ST}$ and generate false positive signals of selection relative to a null model that does not account for these factors (Charlesworth 1998; Kauer et al. 2003; Emelianov et al. 2004). We are not able to evaluate the impact of these factors on our results at this time. Interestingly, mitochondrial DNA is the only known locus that is reciprocally monophyletic between Atlantic and Gulf of Mexico populations of eastern oyster (Reeb & Avise 1990; Hare & Avise 1998). This level of differentiation is more extreme than at any nuclear locus identified as an outlier here and has been attributed to vicariance and drift (Reeb & Avise 1990). The smaller effective population size of the haploid mitochondrial genome could have caused lineage sorting by neutral drift in allopatry without similar consequences at autosomal nuclear loci (Palumbi et al. 2001; Hudson & Turelli 2003; Rosenberg 2003). However, this hypothesis does not explain why mitochondrial DNA shows one of the steepest allele frequency clines along eastern Florida (Hare & Avise 1996). Also, it is common for populations of *Crassostrea* spp. to have a skewed mitochondrial haplotype frequency spectrum (Beckenbach 1994; MPH unpublished data), consistent with a genome subject to recurrent selective sweeps (but see Eldon & Wakeley 2006). Thus, efforts to understand how selection shapes genetic diversity in oysters will need to reconcile patterns observed in the mitochondrial and nuclear genomes.

## Conclusions

Evolutionary geneticists working on non-model systems are usually sampling loci in the dark, unaware of the genomic heterogeneity in locus-specific patterns that must be characterized to properly interpret results. Under these conditions, AFLPs provide a large, random sample of loci that can illuminate the genomic landscape. By providing a more accurate estimate of genomic differentiation, AFLP comparison to a null model helps interpret the among-locus variance. Genome scans provide a statistical framework for testing neutrality, but their utility is largely exploratory because underlying assumptions are difficult to evaluate. For the case of pairwise population comparisons across a secondary contact zone, we have shown that the genomic distribution of $F_{ST}$ is not dependent on migration-drift equilibrium, nor on the details of secondary contact.

Most of the among locus variation in $F_{ST}$ between Atlantic and Gulf of Mexico *C. virginica* can be explained by neutral migration-drift processes. Because the oyster cline is spatially coincident with phylogeographic breaks in many other estuarine taxa, secondary contact is the most likely explanation for a high global mean $F_{ST}$ between Atlantic and Gulf of Mexico oysters and for a high genomic variance in $F_{ST}$. This history makes it more difficult to detect spatially variable selection based solely on the distribution of $F_{ST}$, and strong statistical support was absent for divergent selection here. Nonetheless, three AFLPs and one previously published codominant marker had $F_{ST}$ values above the 99th quantile for the neutral distribution of $F_{ST}$. We regard these as

candidates for divergent selection that help focus future tests using more definitive

methods.

# Tables

Table 1: Sequences of adaptors and primers used for AFLP ligation and PCR.  Sequences are provided 5' to 3'.

| Name | Function | AFLP stage | Sequence |
|------|----------|------------|----------|
| EcoRI-F | Adaptor | Digestion-ligation | CTC GTA GAC TGC GTA CC |
| EcoRI-R | Adaptor | Digestion-ligation | AAT TGG TAC GCA GTC TAC |
| MseI-F | Adaptor | Digestion-ligation | GAC GAT GAG TCC TGA G |
| MseI-R | Adaptor | Digestion-ligation | TAC TCA GGA CTC AT |
| EcoRI-A | Primer | Pre-Selective PCR | GAC TGC GTA CCA ATT CA |
| EcoRI-ACT | Primer | Selective PCR | Fam - GAC TGC GTA CCA ATT CAC T |
| MseI-C | Primer | Pre-Selective PCR | GAT GAG TCC TGA GTA AC |
| MseI-CAA | Primer | Selective PCR | GAT GAG TCC TGA GTA ACA A |
| MseI-CAC | Primer | Selective PCR | GAT GAG TCC TGA GTA ACA C |
| MseI-CAG | Primer | Selective PCR | GAT GAG TCC TGA GTA ACA G |
| MseI-CAT | Primer | Selective PCR | GAT GAG TCC TGA GTA ACA T |

Table 2: Summary statistics for the level of polymorphism and differentiation observed with four AFLP primer pairs. The number of AFLP loci with $F_{ST}$ greater than the 0.99 quantile of the neutral simulated distribution is reported.

| | EcoRI-MseI primers | | | | |
| --- | --- | --- | --- | --- | --- |
| | ACT-CAA | ACT-CAC | ACT-CAG | ACT-CAT | Overall |
| Number of loci | 55 | 72 | 51 | 48 | 226 |
| Percent polymorphic (Total) | 92.7 | 98.6 | 92.2 | 95.8 | 95.1 |
| (NSB) | 72.7 | 88.9 | 90.2 | 83.3 | 84.1 |
| (PCH) | 85.5 | 86.1 | 80.4 | 85.4 | 84.5 |
| Number of $F_{ST}$ > 0.99 quantile | 1 | 2 | 0 | 0 | 3 |
| Maximum $F_{ST}$ | 0.9008 | 0.8688 | 0.6561 | 0.381 | 0.9008 |
| Global $F_{ST}$ | 0.1246 | 0.0664 | 0.0882 | 0.0914 | 0.0917 |

**Figures**
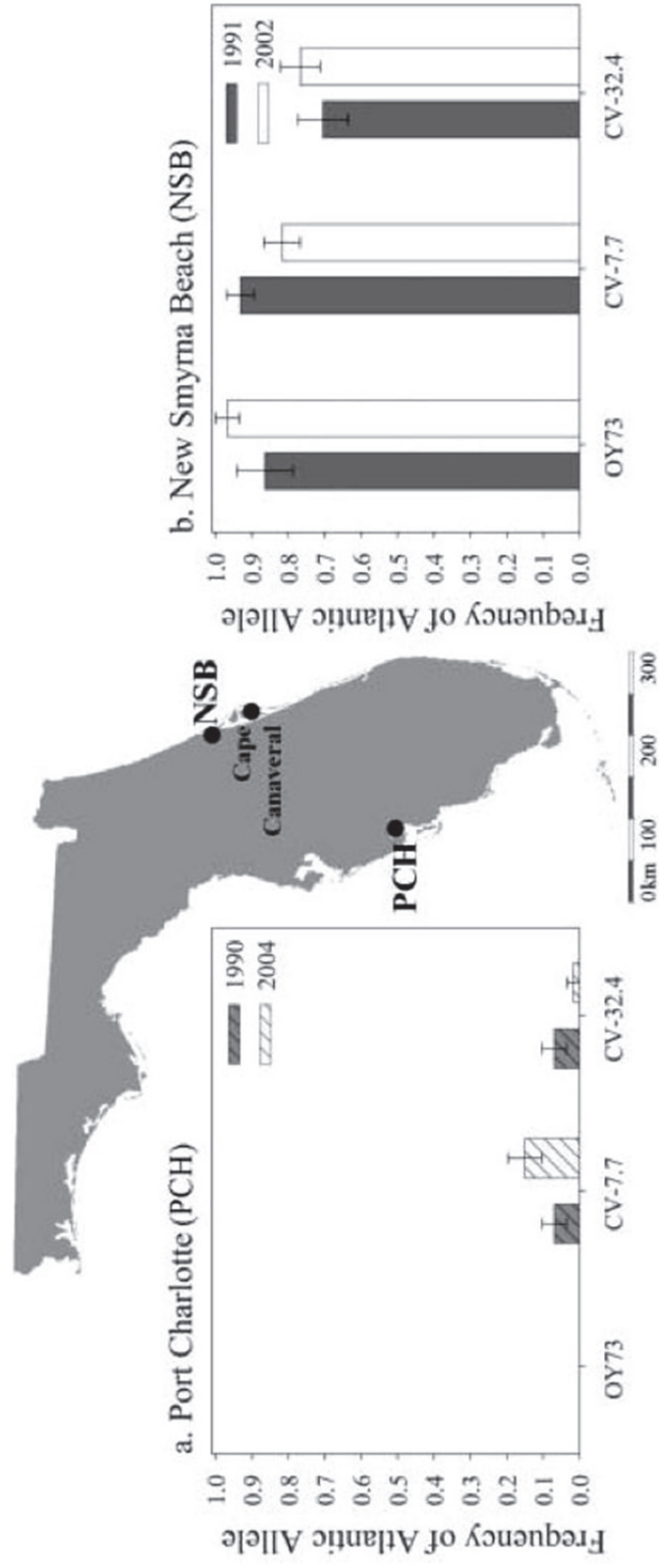


Fig. 1. Allelic frequencies at two time points for one mitochondrial and two nuclear loci in samples from Port Charlotte (a) and New Smyrna Beach (b), Florida. Error bars are one binomial standard error. Recent samples are the same as used for AFLP analysis, n = 30. Port Charlotte 1990 time point from a Karl & Avise (1992) sample (n = 29) reanalyzed by Hare et al. (1996) and Hare & Avise (1996). New Smyrna Beach 1991 time point from Hare et al. (1996; n = 22) with mitochondrial data from Hare & Avise (1996). No significant deviations from Hardy-Weinberg genotype proportions occurred at the nuclear loci (P > 0.05).

55
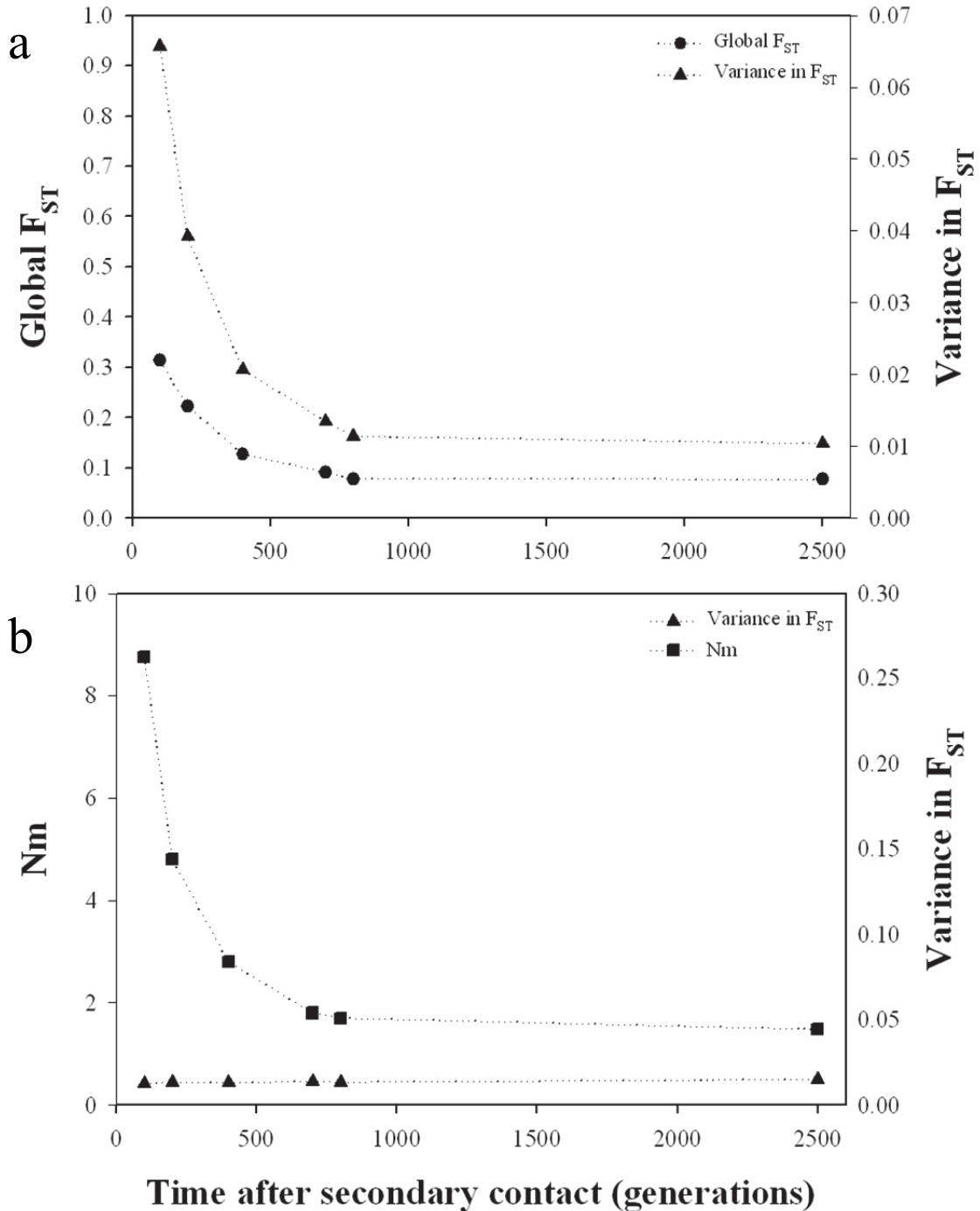
Fig. 2: Relationship among gene flow, $F_{ST}$, and variance of $F_{ST}$ over time after secondary contact in simulations under neutrality. (a) Results of six simulations using the same starting conditions and parameter values, $N = 1000$, $\mu = 0.0001$ and $Nm = 2$, but with samples analyzed at different times after secondary contact. (b) Results of six simulations as above but using different constant postcontact $Nm$ values required to reproduce the empirical global $F_{ST}$ of 0.092 either transiently ($Nm = 8.75, 4.80, 2.80, 1.80, 1.70$ for $t = 100, 200, 400, 700, 800$, respectively) or at equilibrium ($Nm = 1.48$ for $t = 2500$).

56
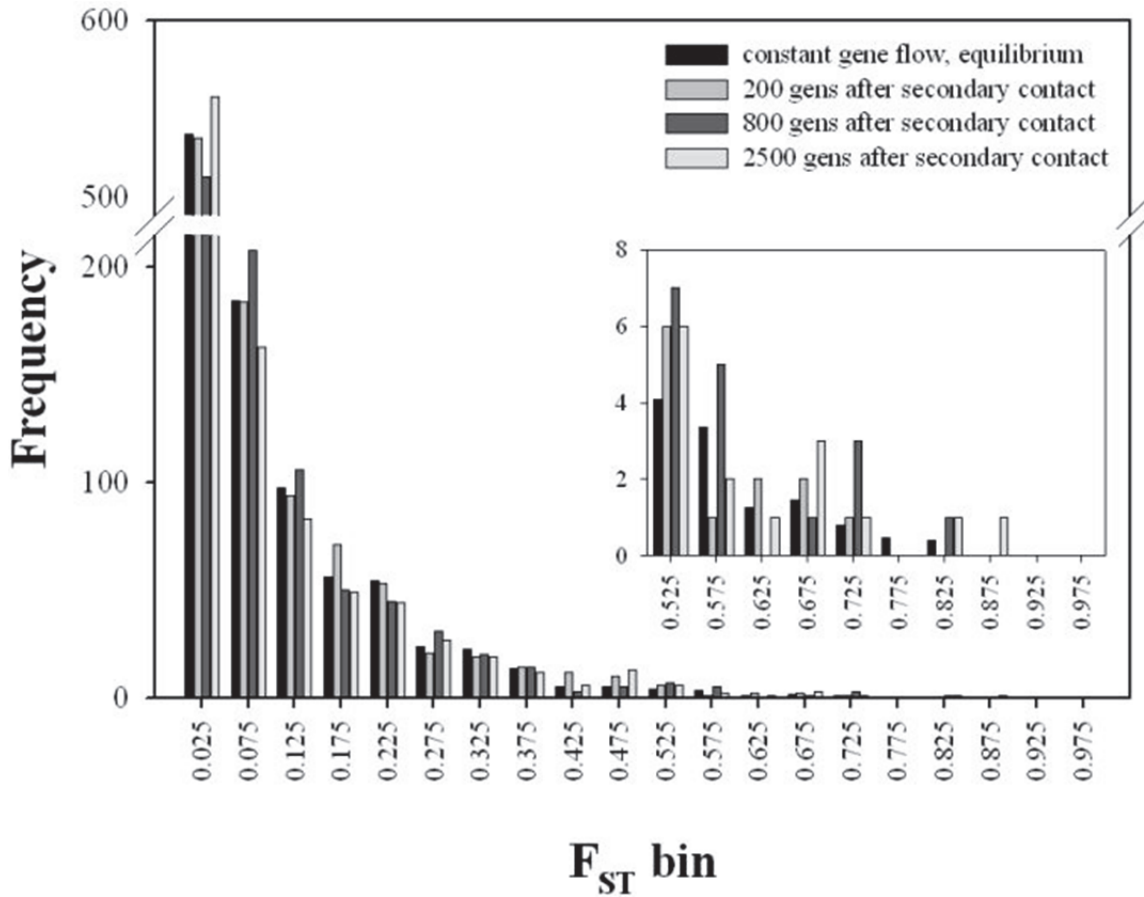
Fig. 3: Frequency distributions of loci in 20 $F_{ST}$ bins for four simulations. Two bars represent migration-drift equilibria; without vicariance in black and the equilibrium reached after vicariance and secondary contact in white. The other two bars show results for nonequilibrium time points after secondary contact. No significant differences were found in pairwise comparisons between the four distributions (Kolmogorov-Smirnov p < 0.05, Bonferroni corrected). Note broken y-axis scale.
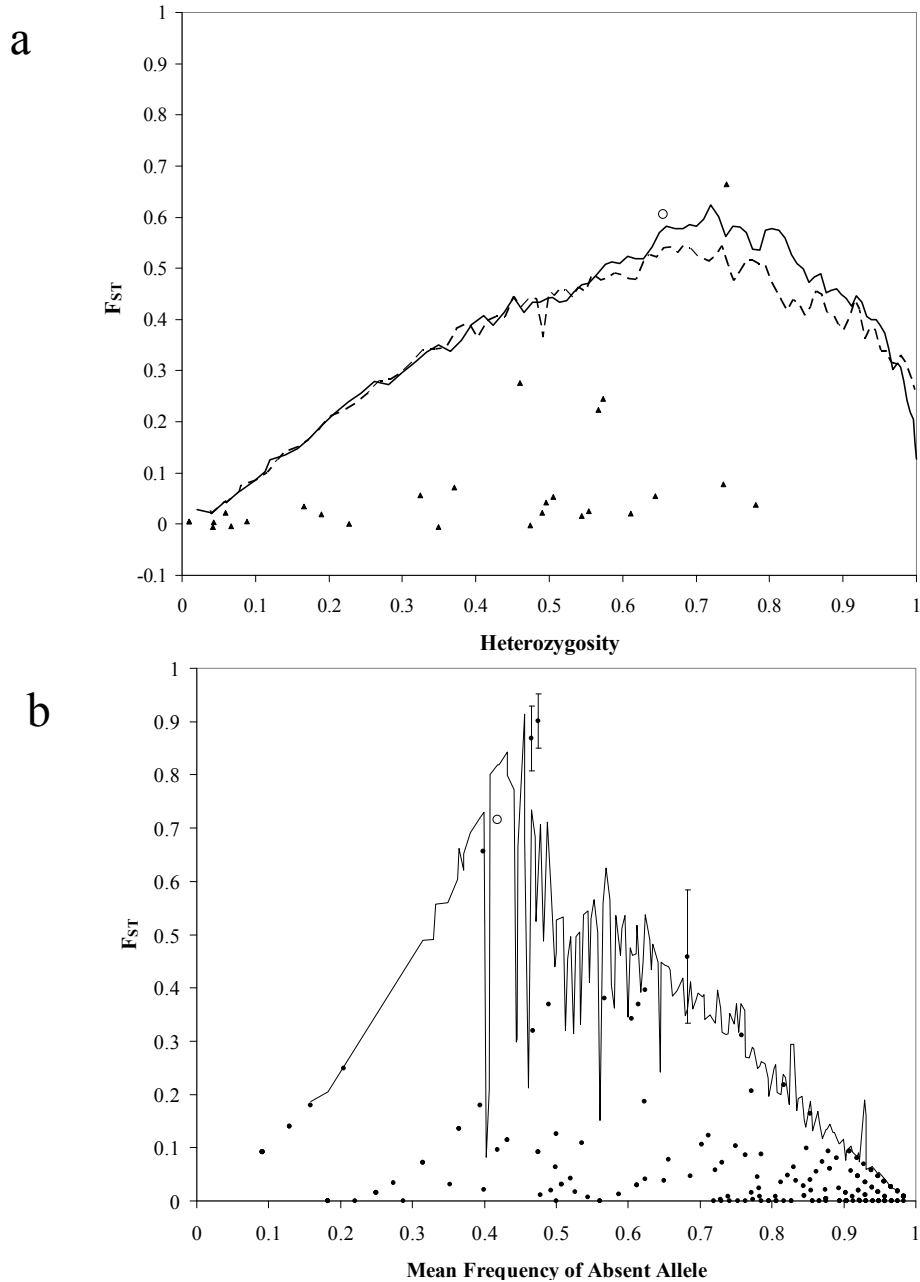
Fig. 4: Empirical and simulated $F_{ST}$ distributions for previously published codominant markers (a) and AFLP loci (b). Empirical $F_{ST}$ values are denoted by triangles (a) and dots (b). Lines depict the 0.99 quantiles of the neutral model based on many overlapping bins of simulation data. P-values denote the probability of observing the $F_{ST}$ of the outlying loci by chance relative to the simulated neutral distributions. Results of two simulation methods are depicted in (a), where the dashed line corresponds to 50 demes simulated and the bold line to two demes simulated, with two demes ultimately sampled in both cases (see Methods). The jagged quantile in (b) reflects sampling error with n = 30. For outlying AFLP loci, the error bars indicate one standard error. The AFLP that was converted to a codominant marker, CAA-270, is shown as an open circle in (a) and (b).

# CHAPTER 2: Geographic and Genomic Patterns of Neutral and Adaptive Differentiation along the Eastern Oyster Contact Zone

## Abstract

When populations undergo gene flow while maintaining differential adaptations, their genomes act as sieves. At the genic level, gene flow allows free exchange of alleles at neutral and uniformly selected genes. Simultaneously, selection impedes gene flow at target genes and closely linked loci. The resulting genomic mosaic of differentiation illuminates the underlying evolutionary processes governing population differentiation, particularly in parapatric contact zones. The geographic pattern of differentiation at neutral and selected genes can expose these processes further. In this study, I examine genomic and geographic patterns of differentiation among parapatric populations of the eastern oyster (*Crassostrea virginica*) along their contact zone in Florida estuaries. The planktotrophic larval phase of this species gives it the potential for regular long-distance dispersal and genetically homogeneous populations. However Florida populations at the center of its range exhibit a sharp step cline at some loci, suggesting a role for divergent selection. Here, I genotyped 217 AFLP loci at 274 adult oysters from seven locations along the cline to ask: (1) whether this cline is generalized across the genome, (2) what the genomic pattern of differentiation is at neutral and selected loci, and (3) whether the clines observed at some loci are maintained by differential selection. I used genome scans to separate classes of loci that behave neutrally and as though they are under the

influence of locally and regionally divergent selection. I compared levels and geographic patterns of differentiation among these locus classes using $F_{ST}$, neighbor-joining trees, assignment tests, principal coordinates analysis, and AMOVA. I also tested for higher differentiation between sites spanning the cline over sites on the same side of the cline.

Results demonstrated a mosaic of differentiation among loci and along the transect, where differentiation was low at putatively neutral loci and high at loci influenced by divergent selection ($F_{ST}$ = 0.0263 and 0.1841, respectively). While, the vast majority of loci were not clinal, differentiation was significantly higher for comparisons of populations on opposite sides of the cline than those on the same side only for repeated outliers. At a 0.01 $\alpha$ level, genome scans identified 24 outlier loci in multiple population comparisons. Four of these putatively selected loci were detected only in comparisons of populations on opposite sides of the cline, supporting a role for regional clinal selection at some loci despite gene flow at others. Patchy or localized selection gradients likely drive divergence at the remaining 20 outliers. In general, the data supported a two-population parapatric model of differentiation, regardless of locus class. However, the strength of the geographic patterns varied among locus classes, being weakest in putatively neutral loci. Some sites south of the step cline (Fort Pierce, Titusville, and Port Charlotte) showed patterns of differentiation that contrasted with other sites in the same region and suggest a complex substructure based on local selection and/or reduced gene flow. The observed selection-based mosaicism raises questions about the various genomic and geographic scales on which environmental divergent selection may be acting and how these interact with local gene flow patterns, particularly south of Cape

Canaveral.  Selection is certainly occurring on a smaller scale than the dispersal potential of planktonic oyster larvae because divergent selection on local and regional scales maintains differentiation at some loci despite gene flow over the remainder of the genome.  While a combination of neutral and adaptive processes, both historical and contemporary, likely shape genomic and geographic patterns of differentiation, this study revealed evidence for divergent selection at local and regional scales in an estuarine species with high potential for gene flow.

# Introduction

Elucidating the process of divergence has been a major goal of evolutionary biologists. Determining the genetic basis of adaptation can facilitate deeper understanding of some evolutionary processes that generate and maintain biological diversity. The selection that generates and maintains divergence can be based on genetic incompatibilities between divergent genomes and/or ecological differences in phenotypes. Either form of selection has a genetic basis and can be measured using population genetic tools.

Given the variety of ecological conditions many species encounter within their geographic range, locally adaptive differentiation can evolve among populations and progress toward reproductive isolation, even in the presence of gene flow (Schilthuizen 2000; Wu 2001; Ogden & Thorpe 2002; Doebeli & Dieckmann 2003; Rosenblum 2006). Environmental gradients can facilitate this branching process, especially those of intermediate slope, and can generate geographic patterns of parapatry between adaptively diverging populations (Doebeli & Dieckmann 2003). Further evidence suggests that biogeographic province boundaries, range edges, and hybrid zones are strongly associated with ecological transition zones, or ecotones (Remington 1968; Endler 1977; Arnold 1997; Swenson & Howard 2005). Indeed, Schilthuizen (2000) proposed that ecotones are "speciation-prone" because divergent habitat adaptation can lead to parapatric speciation. Such regions of ecological transition are ideal laboratories in which to investigate geographic and genetic patterns of adaptive divergence.

Traditionally by Wright's (1931) rule of thumb, only one migrant per generation on average is necessary to eliminate or prevent differentiation caused by neutral genetic drift. In order for differentiation to proceed or be maintained in the presence of gene flow, the strength of selection must be greater than the migration rate (Haldane 1930; Wright 1931; Bulmer 1972). From a multilocus point of view, the necessary strength of selection on a given locus asymptotically diminishes as the number of loci under divergent selection increases (Fry 2003). As evidence mounts for the theoretical probability and empirical reality of divergence in sympatry and parapatry, where gene flow is more likely to be greater than one migrant per generation, it is clear that divergent selection plays an undeniable role in many instances of biological diversification (e.g., Raffaelli 1978; Grant & Grant 1989; Feder 1998; Hatfield & Schluter 1999; Schneider et al. 1999; Berlocher 2000; Ogden & Thorpe 2002; Irwin et al. 2005; Grahame et al. 2006). One resulting question is: what is the role of selection in shaping the genomic landscape, particularly in cases where divergence exists in the presence of gene flow?

**Hybrid Zones**

Examining the genomic patterns of hybrid zones, where divergent populations meet and experience gene flow, can help identify genes that matter to ecology and reproductive isolation (Schmidt et al. 2008; Nosil et al. 2009). The majority of theoretical work in hybrid zones has been in clinal tension zones (Ross & Harrison 2002). Allele frequency or phenotypic clinal patterns in hybrid zones are the result of the interaction between selection and gene flow between divergent populations. Several conditions attract and/or produce clines, such as environmental gradients, ecotones, secondary contact zones, low-

density areas, and dispersal barriers (Endler 1977; Barton & Hewitt 1985). Therefore, the presence of a cline does not necessarily implicate a role for selection, as allele frequency clines can be generated neutrally by chance, history, etc. (Nei & Maruyama 1975; Robertson 1975). If neutral processes generate a cline, it will degrade in the presence of gene flow at a rate proportional to the dispersal distance, provided there is no physical/genetic barrier to prevent or divergent selection to counteract genetic introgression (Endler 1977; Barton & Hewitt 1989). Therefore, strong clines found in hybrid zones of sufficient age and magnitude of gene flow may be strong indicators of divergent selection, either in the form of selection against hybrids due to genomic incompatibilities or of local adaptation.

Theory states that, regardless of the type of selection, the genomic landscape of divergent groups experiencing gene flow, like those found in hybrid zones, is a porous one (Key 1968; Hodges & Arnold 1994; Feder 1998; Wu 2001; Nosil et al. 2009). While gene flow causes neutral alleles to pass freely between populations, divergent selection impedes target alleles from passing between populations (Wu 2001; Beaumont 2005; Gavrilets & Vose 2005; Bonin et al. 2006). Over generations, recombination disassociates neutral loci from selected loci. As a result, loci under divergent selection or causing reproductive isolation and those that are closely linked are more strongly differentiated than other loci. Only a few loci will show such a pattern during the early stages of divergence, those loci that are directly involved in local adaptation, mate choice, or genetic incompatibilities between populations (Butlin 2010). Over time, these

selection-targeted loci affect increasing parts of the genome through coadapted gene complexes (Wu 2001).

Comparing geographic patterns of allele frequencies among loci makes it possible to detect divergent selection, even when the specific phenotypes underlying ecological differences and reproductive isolation are unknown (Bonin et al. 2006; Teeter et al. 2008). Along an environmental gradient, ecologically important loci will display a shift in allele frequency in response to the gradient, and neutral loci will show relatively small changes in allele frequency as a result of genetic drift (Schmidt et al. 2008). If insufficient time has passed to allow for complete breakdown of neutral differentiation in a recent secondary contact zone, then neutral alleles might still display weak clines (Endler 1977; Barton & Hewitt 1985). However, divergent selection and reproductive barriers will maintain strong clines in target and closely linked loci despite the passage of time and gene flow. Furthermore, the targets of balancing or uniform selection will be even more homogeneous between populations than neutral loci (Lewontin & Krakauer 1973).

Hybrid zones are natural laboratories to make such geographic comparisons of allele frequency because neutral loci will introgress between populations more than loci involved in divergent selection or reproductive isolation (Rieseberg et al. 1999; Buerkle & Rieseberg 2001). Indeed, hybrid zones have been called windows on the evolutionary process because they elucidate the processes governing the continuum of divergence (Harrison 1990). In particular, they offer opportunities to decipher the balance between

gene flow and selection (Barton & Hewitt 1985; Hewitt 1988; Harrison 1990; Harrison 1998). However, multiple models exist to describe this balance, each distinguished by the type of selection and geographic pattern observed in the system (Arnold 1997). Tension zones are clinal and maintained by reduced hybrid fitness due to endogenous (environmentally independent) selection against genomic incompatibilities (Barton & Hewitt 1985). Ecological gradient hybrid zones also are clinal, but in contrast are maintained by exogenous selection along an environmental gradient or ecotone (Endler 1977). Mosaic hybrid zones are patchy in structure and maintained by spatially variable selection due to patchy environments (Harrison 1990). Depending on the spatial scale of sampling, the same hybrid zone may have clinal or mosaic structure, as observed in *Gryllus* crickets (Ross & Harrison 2002) and *Chorthippus* grasshoppers (Bridle et al. 2002). Therefore, the geographic sampling scheme can affect the inferred model of hybrid zones.

## Genomic Analysis of Divergence with Gene Flow

As genomic tools begin to illuminate the evolutionary processes governing contact zones at the genetic level, empirical results continue to reflect theory. Genomes undergoing divergence and gene flow are indeed sieves that retain alleles important to ecological differences and reproductive isolation while permitting introgression of neutral and positively selected alleles (e.g., Rieseberg et al. 1999). The growing collection of genomic studies reveals the complex architecture of divergence in hybrid zones (e.g., Teeter et al. 2008). At divergently selected loci, some studies have demonstrated geographic discordance of clines with a wide variation in cline width (e.g., Marshall &

Sites 2001; Nikula et al. 2008; Teeter et al. 2008). As a result, the set of loci chosen to examine contact zones can have a large effect on inferences from their geographic structure (Ross & Harrison 2002). Indeed, a multilocus geographic approach in contact zones permits discernment among potential evolutionary mechanisms maintaining locus-specific levels of differentiation.

The genome scan is one genomic approach that has lead to deeper understanding of the balance of selection and gene flow during the process of divergence and reproductive isolation (Butlin 2010). The theoretical basis of this statistical approach is that neutral expectations for divergence offer a null model against which to test for such adaptive loci (Lewontin & Krakauer 1973; Beaumont & Nichols 1996; Luikart et al. 2003; Beaumont 2005). This method compares locus-specific empirical estimates of differentiation to a simulated distribution of the expected neutral differentiation (Beaumont & Nichols 1996; Luikart et al. 2003). The genome scan is extremely robust to different demographic histories (Beaumont and Nichols 1996; Beaumont 2005) and is unaffected by a history of secondary contact (Murray & Hare 2006). Population genomic methods such as this one enable researchers to distinguish genome-wide effects from locus-specific effects (Luikart et al. 2003). Processes such as drift, inbreeding, and gene flow are expected to affect all neutral variation in genome in the same way (Black et al. 2001). In contrast, locus-specific processes like selection affect only a few loci at a time. Although most loci will fall within the neutral distribution, loci that exhibit particularly strong differentiation will fall outside because they are either under divergent selection themselves or tightly linked to a gene that is. For loci exhibiting significantly greater

differentiation than the neutral distribution, the null hypothesis of neutral differentiation is rejected and divergent selection inferred.

Over the years of application of the genome scan method, authors have assembled an informal list of best practices (reviewed in Butlin 2010).  For example, when examining multiple sites, pairwise population comparisons should be done to avoid effects of differences in gene frequency among subpopulations in the metapopulation (Beaumont 2005).  Because this can lead to an increased number of loci identified as outliers from the neutral distribution, the best candidate loci for divergent selection are those with the smallest p-value and identified as outliers in multiple pairwise population comparisons (Emelianov et al. 2004; Bonin et al. 2006).

While genome scans have increased in use (e.g., Wilding et al. 2001; Emelianov et al. 2004; Campbell & Bernatchez 2004; Vasemagi et al. 2005; Bonin et al. 2006; Murray & Hare 2006; Savolainen et al. 2006; Yatabe et al. 2007; Egan et al. 2008; Nosil et al. 2008; Via & West 2008; Galindo et al. 2009; Gagnaire et al. 2011), they are most often applied to species characterized by effective assortative mating, low dispersal, and clear phenotypic differences.  However, the method also has the power to detect divergent selection in non-model systems where the association between divergent selection and target traits is complex and/or cryptic (Bonin et al. 2006).  The vast majority of genome scans focus interpretation and conclusions on outliers, though the approach can be used to study the behavior of different classes of loci (outliers versus non-outliers) and tease apart the various evolutionary forces that govern distinct behaviors.  Such a novel

approach to genome scans would be particularly useful in contact zones where divergent selection interfaces with gene flow.

More genomic studies of marine hybrid zones are needed because little is known about the forces facilitating divergence in marine animals, where the nature of the barriers between water masses is largely unknown or speculative, especially for marine organisms with planktonic dispersal (Mayr 2001).  As a result of simulations of planktonic larval dispersal, Cowen et al. (2006) asserted that a combination of small-scale dispersal, distribution of suitable habitat, and oceanographic patterns can create ecologically isolated subregions or genetically differentiated biogeographic regions.  Although environmental and habitat variability along the species distribution should result in local adaptation or phenotypic plasticity, empirical evidence for divergent selection in marine bivalves is somewhat limited (Drent 2002).  Limited geographic and genomic sampling can oversimplify the spatial scales over which populations are selectively and evolutionarily divergent, especially in marine populations (Sotka et al. 2004).  Therefore, it is even more critical to have good genomic and geographic sampling in marine animals with planktonic larval dispersal.  However, genome scans rarely have been applied to such organisms (but see Gagnaire et al. 2011).

**The System: The *Crassostrea virginica* Contact Zone in Florida**

Like many marine species that disperse via a planktonic larval phase, the eastern oyster, *Crassostrea virginica*, has a high potential for long distance dispersal.  In the Chesapeake Bay, dispersal distances have been estimated at 497 km$^2$ over a two to three week larval

period (Rose et al. 2006).  Although it is not known if these dispersal distances translate throughout this oyster's range from Canada to the Yucatan Peninsula, the potential exists for high levels of gene flow.  Despite this high potential for long distance gene flow, several loci exhibit coincident and surprisingly steep step clines near Cape Canaveral, Florida, dividing the oyster's range between two morphologically cryptic genetic groups termed "Atlantic" and "Gulf" genotypes (Reeb & Avise 1990; Karl & Avise 1992; Hare & Avise 1996).  Such genetic homogeneity within broad regions found in juxtaposition to a relatively sharp cline indicates that the extent of gene flow in the clinal region is not a species-specific characteristic, but rather a reflection of local environmental conditions and/or historical events (Sotka et al. 2004).  The oyster step cline is coincident with a biogeographic province boundary (Avise 2000; Avise 2004), an ecotone between temperate and subtropical communities (Briggs 1974), a narrow canal connecting linear lagoons, and a trough in oyster density (Hare & Avise 1996).  Either through ecological selection or as incomplete dispersal barriers, all of these characteristics potentially attract and sharpen clines (Endler 1977; Barton & Hewitt 1985).  Given the gene flow potential, it is unclear what mechanisms maintain the sharp differentiation observed in this system (Hare & Avise 1996; Avise 2004).

Although mtDNA shows alternately fixed differences between the centers of the two regions (Reeb & Avise 1990), allele frequencies at a few clinal nuclear loci are nondiagnostic.  These loci are generally homogeneous among Atlantic-type populations between Massachusetts and Cape Canaveral, but are heterogeneous in the Gulf-type region between Cape Canaveral and Louisiana (Karl & Avise 1992; Hare & Avise 1996).

However, not all nuclear loci show a strong differentiation between Atlantic and Gulf regions (Buroker 1983; Karl & Avise 1992; McDonald et al. 1996; Hoover & Gaffney 2005; Murray & Hare 2006). A genomic study examining a pair of Florida populations (one north of Cape Canaveral, one on the Gulf of Mexico) documented a genetic signal of divergent selection at some loci over and above the neutral level of differentiation (Murray & Hare 2006). Currently, this is the strongest evidence for divergent selection operating at a few loci in this system. Due to the great distance between these populations, it is unlikely they exchange migrants, making it difficult to assess whether divergence is maintained in the face of gene flow. Although the pair of populations was chosen to be representative of the two regions, it is also unclear whether the genomic pattern documented between them can be translated to the regions as a whole, particularly near the zone of contact where cross dispersal is likely.

As stated above, the genomic and geographic scale of sampling can affect interpretation of evolutionary forces (Sotka et al. 2004). For both the full oyster range and the mesogeographic scale of eastern Florida, departures from the classic sigmoid cline shape have been observed (Hare & Avise 1996). Of particular interest are the localized allele frequency reversals that occur south of Cape Canaveral because such a pattern can indicate mosaic ecological selection (Harrison 1990; Arnold 1997) or contact zone movement (Barton & Hewitt 1985). However, this geographic pattern has only been documented in two nuclear and one mtDNA loci (Hare & Avise 1996). Because these few loci were also biased for strong differentiation between Atlantic and Gulf types (McDonald et al. 1996), it remains unknown whether the geographic pattern they

displayed reflects a broader genomic pattern of localized genetic reversals or anomalies in an otherwise clinal system.  In addition, the moderate frequency of Atlantic-type alleles along southeastern Florida may indicate high levels of hybridization (a unimodal genotypic distribution) or highly mixed locales of pure Atlantic and Gulf types (bimodal genotypic distribution) (Jiggins & Mallet 2000).  Hare & Avise (1996) tested for disproportionate levels of parental types at a few loci and found nonsignificant excesses and deficits of parental genotypes.  More loci may need to be sampled to clarify genotypic distribution of populations exhibiting moderate allele frequencies. Having an accurate picture of the genomic landscape of population differentiation in the contact zone is vital to deciphering the evolutionary processes maintaining it.

In this chapter, I aim to address the following questions by leveraging both genomic and geographic approaches.

1) Does the population structure near the cline center reflect clinal expectations of two differentiated populations in parapatry?

2) How does selection shape the genomic pattern of differentiation?

3) Is there a clinal geographic pattern of selection along the ecotone?

## Methods and Materials

### Sample Collection

A total of 274 adult oysters were collected from six locations in lagoons along Florida's

Atlantic coast and one location on Florida's Gulf of Mexico coast (Table 2-1, Fig. 2-1).

Although collection dates ranged between 2002-2006, previous data indicate allele

frequencies do not vary significantly in adult oysters over a span of 11-14 years (Murray

& Hare 2006).  Therefore, the four-year span of collections should not bias results.

Oysters were put on ice and brought to the lab, where gill tissue was dissected and

preserved in 95% ethanol.  Tissue samples were refrigerated or frozen until DNA

extraction.  Whole genomic DNA was extracted from approximately 20 mg of preserved

gill tissue following the DNeasy animal tissue extraction protocol (QIAGEN Inc.).  To

equalize DNA concentrations among individuals for further analysis, all samples of DNA

were diluted to 50 ng/$\mu$L based on spectrophotometry.

### AFLP Data Collection

To generate AFLP fragments, I used a modified version of the assay outlined by Vos and

colleagues (1995) (See Murray & Hare 2006 for protocol details).  The fluorescently

labeled AFLP products of the selective amplification were electrophoresed on an ABI

Prism 3100 Genetic Analyzer or an ABI 3730*xl* DNA Analyzer (Applied Biosystems

Inc.).  I used GeneMapper 4.0 (Applied Biosystems Inc.) to view samples and create bins

around fragments for further analysis.  Because slight size shifts occur among

electrophoretic runs, I overlaid chromatographs of several samples from every run to create bins.  Fragments smaller than 75 bp and larger than 490 bp were excluded from analysis.  The average bin width was 0.85 bp, and each bin was centered on the average peak size.  The ABI 3100 and ABI 3730*xl* exhibited a difference in peak-sizing of about 1.2 bp.  Therefore, duplicate samples (see Data Reliability) were used to measure the shifts among runs and create correspondingly shifted bins for identical peaks.  After creating all bins for each primer pair, GeneMapper generated data tables containing binned peak heights for all individuals.

For each primer pair, peak heights were then automatically scored using the program AFLPScore (Whitlock et al. 2008).  This software first normalizes peak heights to compensate for variance in PCR product concentrations among samples.  It then allows the user to optimize the error rate by removing bins that do not meet mean peak height thresholds and by creating a genotyping threshold of peak height for the presence allele. Using the filter and absolute threshold options that minimized error rate for each primer pair, I calculated the overall error rate based on duplicate samples and created a table of 0/1 (fragment absence / fragment presence) scores.  I manually checked all these automated scores by viewing samples in GeneMapper.  I corrected any erroneous peak calls and entered blank genotypes whenever the presence or absence of a binned peak was ambiguous.

**Data Reliability**

In order to establish the reliability of this data set, I estimated error rates for each locus and the full data set. For every plate run through electrophoresis, I included four positive controls and one negative control. In addition, I ran 16 samples from the PCH population twice for the ACT-CAT primer pair. I used all these replicates to calculate a locus error rate as the ratio of the total number of mismatches divided by the total number of replicates (following Pompanon et al. 2005). To ensure a minimum of 90% repeatability for each locus, I removed all loci with error rates higher than 10%. Then, I used the retained loci to estimate the error rate for the entire data set as the ratio of the total number of mismatches over all loci to the product of the numbers of replicates and loci (Pompanon et al. 2005).

I also removed loci that were not informative or provided similar results due to linkage. First I removed all monomorphic loci, which provided no additional information about genetic variation. Next I culled all loci whose allele frequency was less than 0.05 in all populations. These loci do not provide significant information on population differences and may have low frequency due to null alleles. Therefore, what little genetic information they do provide may be misleading due to experimental difficulties in amplification. Because AFLP are random genomic markers, some loci may be allelic or tightly physically linked and therefore produce redundant, dependent results (Wilding et al. 2001). To avoid as much redundancy and potential resulting bias as possible, I calculated a similarity index for every possible pairwise combination of loci (following

Gaudeul et al. 2000). For each locus pair (k, l), the pairwise linkage index was calculated as:

$$I_{kl} = \frac{1}{n} \sum_i |m_{ki} - m_{li}|$$

where $n$ is the number of individuals, $m_{ki}$ is the score (0/1) for individual $i$ at locus $k$, and $m_{li}$ is the score (0/1) for individual $i$ at locus $l$.

Because this task created many pairwise locus comparisons, I wrote a program to automate the task using MatLab 7.0 (MathWorks, Inc.). This program calculated the pairwise linkage index for the whole data set. Based on these results, I removed all loci with more than 20 pairwise indices greater than 0.95 (those where the same allele appeared in both loci very frequently) or less than 0.05 (those where opposite alleles appeared at each locus very frequently). I assumed that these locus pairs were so similar as to produce redundant results, possibly skewing the data set (Gaudeul et al. 2000). Then I ran the remaining loci through the program again and removed one locus from each pair that surpassed the similarity thresholds above, thus producing a final set of loci with linkage indices between 0.05 and 0.95.

### Tests for Genetic Signals of Selection

When populations diverge under neutral conditions, the differentiation measured by individual loci across the genome has an expected variance and distribution due to the stochastic nature of drift (Lewontin & Krakauer 1973). Simulations can model neutral

divergence and provide a null distribution against which to test an empirical distribution of differentiation among loci sampled across the genome (Beaumont & Nichols 1996). Loci falling within the neutral distribution are likely diverging neutrally. In contrast, loci exhibiting differentiation significantly greater than that expected by the neutral distribution are putatively subject to divergent selection, either directly or through physical linkage to a selected locus.

To statistically test for significantly differentiated loci, I used the Dfdist package, which employs methods developed by Beaumont & Nichols (1996) and Beaumont & Balding (2004) and applies them to dominant markers. This program first uses Bayesian methods to estimate allele frequencies for the sampled populations in the empirical data set and then calculate the distribution of $F_{ST}$ in the data set (Zhivotovsky 1999). Next, it uses simulations to create a null, neutral distribution of differentiation ($F_{ST}$) that has the same mean $F_{ST}$ as the "trimmed mean" of the empirical data. The "trimmed mean" is calculated from the middle 40% of the empirical data, which should avoid loci under the influence of divergent (upper 30%) and uniform selection (lower 30%) (Beaumont 2005). The simulations calculate $F_{ST}$ conditioned on heterozygosity using an island model of subdivided populations through a hierarchical Bayesian approach. I generated 50,000 simulated loci for the null distribution. Finally, the empirical and neutral distributions are compared to identify outliers. I chose to use two $\alpha$ values (0.05 and 0.01), which for a one-tailed test equate to the 97.5 and 99.5 quantiles of the null distribution, and the default smoothing parameter of 0.04 for graphing purposes. Empirical loci falling above these thresholds were considered outliers and those below were non-outliers. This

approach is robust to a variety of demographic and equilibrium scenarios (Beaumont &
Nichols 1996; Murray & Hare 2006) and has substantial power to detect loci subject to
selection (reviewed by Beaumont 2005).

A total of 21 possible pairwise population comparisons were performed in triplicate. I
used an in-house program written in LabView 8.5 (National Instruments Corporation) to
automate the steps of the Dfdist package for the given set of populations. Then I wrote a
program in MatLab to process and summarize the "pv" output files, which report the
heterozygosity, $F_{ST}$, and p-value for each locus. Specifically, this program calculated the
mean $F_{ST}$ and p-value across each triplicate for each locus. Because Dfdist is based on
Bayesian simulations of population divergence, some variation in results may occur from
run to run. By calculating mean statistics among triplicates, I reduced the possibility of
reporting an aberrant outlier that would not be an outlier in subsequent runs. This
MatLab program also summarized into a single table the outliers with their mean
statistics and corresponding population pair. I then divided loci into four classes: all loci,
non-outliers, outliers identified in multiple independent comparisons (MI outliers; where
'multiple independent' indicates at least two comparisons that do not have any
populations in common), and other outliers (SD outliers; loci appearing as outliers in only
one comparison or in multiple comparisons that all share one population).

It is important to note that the label for 'multiple independent' (MI) comparisons does not
refer to evolutionary independence of contact zones among the population pairs (e.g.,
Wilding et al. 2001). Replicated contact zones are not possible in the single oyster cline

system.  Rather, it indicates a statistical sampling independence where at least two

comparisons that do not share any sites or sampled individuals returned the same locus as

an outlier.  Under these criteria, the independent comparisons can include population

pairs on the same side of the step cline and/or on opposite sides of the cline.  Therefore,

MI outliers are not necessarily clinal in allele frequency or of higher differentiation

among regions than within regions.

When dealing with multiple comparisons, it is important to consider the possibility of

accumulating type I error (Excoffier et al. 2009; Galindo et al. 2009).  The probability

($p_m$) of detecting a single locus in $x$ comparisons due to type I error is simply:

$$p_m = \left(\frac{\alpha}{2}\right)^x$$

where $\alpha$ is the alpha value and $x$ is the number of population comparisons.  To evaluate

the strength of statistical evidence for a role of regional selection between Atlantic and

Gulf types, I calculated the probability ($p_r$) of detecting a single locus in $o$ comparisons of

population pairs on different sides of the step cline and no comparisons on the same side

as the following (adapted from Nosil et al. 2008):

$$p_r = \left(\frac{\alpha}{2}\right)^o \times \left[\frac{\left(\frac{d!}{o!(d-o)!}\right)}{\left(\frac{n!}{o!(n-o)!}\right)}\right]$$

where $\alpha$ is the alpha value, $o$ is the number of comparisons of population pairs on different sides of the cline in which the locus was observed as an outlier, $d$ is the total number of comparisons of population pairs on different sides of the cline, and $n$ is the total number of population comparisons. The number of loci expected as outliers in $o$ comparisons of population pairs on different sides of the step cline and no comparisons on the same side ($l_e$) is simply:

$$l_e = p_r \ \text{x} \ L$$

where $L$ is the total number of loci.

## Population Structure

I used the program AFLP-SURV 1.0 (Vekemans 2002), which deals specifically with dominant data (Lynch & Milligan 1994), to estimate locus allele frequencies and overall and pairwise population genetic differentiation ($F_{ST}$). I also estimated the pairwise population genetic distances as Nei's $D$ (following Lynch & Milligan 1994). I used a Bayesian method of allele frequency estimation with a non-uniform prior distribution because this method delivers the most accurate results and is the most generally applicable (Zhivotovsky 1999). Differences in allele frequencies between adjacent population pairs were calculated from the output. Because prior examination has not found Hardy-Weinberg deviations within Florida oyster populations (Hare & Avise 1996), I chose the option to assume Hardy-Weinberg genotypic proportions. I repeated these analyses separately for each class of loci. To determine whether genetic

differentiation was higher among population pairs on opposite sides of the cline step versus on the same side, I performed one-sided t-tests on pairwise $F_{ST}$ for each class of loci and for all 217 loci with a Bonferroni correction for four tests ($\alpha = 0.0125$; JMP version 9).

In order to visualize these genetic relationships between populations, I constructed population trees based on 10,000 bootstrapped matrices of Nei's *D* generated by AFLP-SURV. Using the programs NEIGHBOR and CONSENSE in PHYLIP 3.68 (Felsenstein 2005), each category of loci was used to construct separate neighbor-joining consensus trees with a 50% majority rule. I then used the program Dendroscope to plot consensus trees, rotate tree branches, and denote bootstrap values (Huson et al. 2007).

In addition to population-level measures of genetic differentiation, the genetic similarity between individuals across loci can be used in assignment tests, which remove the assumption that a sample site is equivalent to a sampled admixed population. Such individual-based approaches can not only delineate admixed populations, but also identify migrants and individuals of mixed ancestry. The software STRUCTURE v. 2.2.3 utilizes a Markov Chain Monte Carlo (MCMC) algorithm, with an extension to deal with dominant data, to assign individuals into clusters based on their multilocus genotypes (Falush et al. 2007). To determine the number of genetic populations, I ran the model for $K = 1$-$7$, with 50 runs for each value of K. Options included the use of the admixture model with correlated allele frequencies and no prior information. Each run incorporated a burn-in period of $10^5$ MCMC iterations and a data collection period of $10^6$ iterations.

For each K value, results across runs were summarized using the software CLUMPP v. 1.1.2 (Jakobsson & Rosenberg 2007), which assesses the similarity among runs and integrates over them to provide a summary solution. I then used the software Distruct v. 1.1 (Rosenberg 2004) to graphically display the final summary solution. I determined the most likely number of K clusters for the full data set using two methods: the maximum log probability of the data ($Pr(X|K)$, Pritchard et al. 2000), and the mode of the rate of change in log probability for consecutive K values (Evanno et al. 2005). I repeated this procedure for each of the locus classes and visually interpreted differences in geographic assignment patterns between locus classes.

Using all loci, I evaluated the ancestry proportions generated by CLUMPP under the most likely number of clusters to determine which sites exhibited the greatest levels of admixture. For each individual, I determined which cluster contributed the most to its ancestry as estimated by the ancestry proportions for each cluster. These highest ancestry proportion values for each individual were the maximum ancestry proportions. I then calculated the mean of the maximum ancestry proportions across individuals within a site and tested for significant differences in these means across sites using pairwise t-tests with Tukey adjustment for all pairwise tests (JMP 9.0).

An alternative individual-based approach to visually identify genetically similar populations is the Principal Coordinates Analysis (PCoA), which identifies similarities among individuals across loci using multidimensional scaling. For this analysis, I used the Excel macro GenAlEx 6.2 (Peakall & Smouse 2006). First, I calculated the pairwise

linear genetic distances between individuals (following Huff et al. 1993).  The PCoA was

then conducted on the covariances of these distances using the data standardization

option.  This analysis measured the percent variation explained by the three major axes,

both individually and cumulatively.  Individuals that group together are likely to be from

the same genetic population (Reeves & Richards 2009).  By labeling individuals with

their collection site, population clusters, hybrids, and migrant individuals can be

identified (Young et al. 2001).  I performed a separate PCoA on each of the locus classes.

To test for geographic patterns in genetic variation, an analysis of molecular variance

(AMOVA) was performed in GenAlEx 6.2 for each locus class.  Specifically, I tested for

three patterns of geographically hierarchical partitioning of variance: (1) a two-region

model to represent Atlantic and Gulf of Mexico (Gulf) populations, where Atlantic

populations are north of Cape Canaveral and Gulf populations are in Cape Canaveral and

to the south (Reeb & Avise 1990); (2) a three-region model of North populations

(populations north of Cape Canaveral), Cape Canaveral populations, and South

populations (populations south of Cape Canaveral); (3) the hierarchical grouping

corresponding to the best K clusters identified by STRUCTURE.  Model 1 corresponds to

a parapatric genetic distribution.  In contrast, model 2 represents a geographic distribution

expected by the bounded hybrid superiority hypothesis, in which hybrids are more

successful than parental types in transitional ecotone habitats (Moore 1977).  Model 3

tests for an alternatively supported geographic structure, which may be indicative of a

mosaic zone pattern.  For each geographic pattern, the AMOVA partitioned the variance

among the geographic regions, between populations within regions, and within

populations.

# Results

## Data Reliability

Locus error rates ranged from 0 to 30% before removing any loci. However, average locus and data-wide error rates were low (Table 2-2; 6.53% and 5.96%, respectively). For the ACT-CAT primer pair, error rates were lower (4.55% and 4.52%, respectively) and resulted in retaining a higher proportion of loci. For this primer pair more individuals were tested in duplicate to calculate error rate, which represented a larger portion of the total individuals examined in this study. Thus, the proportion of effect on error rate that each individual could have was reduced. Removing the 97 loci with error rates above 10% reduced the average locus error rate and the data-wide error rate to 3.26% and 3.33%, respectively. This error rate exclusion process left 291 loci with 90% or higher repeatability. After removing highly similar loci (see below), the final set of loci resulted in an average locus error rate of 4.03% and a data-wide error rate of 3.96%. These error rates are comparable to, or lower than, those found in other population genetic studies using AFLPs (e.g., Bonin et al. 2006; Egan et al. 2008; Nosil et al. 2008; Galindo et al. 2009). Blank/missing genotypes comprised 2.58% of the final data set and were randomly distributed among individuals.

After removing 17 monomorphic loci (5.84% of the 291 loci) and 29 loci with no population allele frequency greater than 0.05 (9.96% of the 291 loci), the pairwise linkage index was calculated for the remaining 245 polymorphic loci. The mean pairwise linkage index was $0.4098 \pm 0.2412$. Because indices near the extremes (0 and 1) indicate

high similarity between loci, the magnitude of the absolute difference from 0.5 provides a more direct estimate of similarity, i.e., low absolute differences would indicate low similarity and chance of linkage. The mean absolute difference between $I_{kl}$ values and 0.5 was $0.2218 \pm 0.1309$. This first set of $I_{kl}$ values resulted in the removal of 26 loci that surpassed the $I_{kl}$ thresholds in greater than 20 pairwise comparisons. The second calculation run identified two additional loci for removal, leaving the final data set of 217 remaining loci. The presence allele was at very low frequency for the majority of loci that were removed. Removing loci made very little difference in the Bayesian allele frequency estimates for the final 217 loci. The mean difference in allele frequency estimated for each locus when 291 loci and 217 loci were included in the data set was minimal (-0.0005, which corresponds to a median change of -1.45%). Furthermore, retaining these loci would have shifted downward the global and trimmed mean $F_{ST}$ values used for the null distribution in simulations, resulting in a less stringent test for outliers. For this final set of loci, the means for $I_{kl}$ and the absolute difference between $I_{kl}$ and 0.5 were $0.4299 \pm 0.2202$ and $0.1973 \pm 0.1202$, respectively.

**Tests for Genetic Signals of Selection**

At the 0.05 (0.01) $\alpha$ levels, the 21 pairwise population comparisons revealed 85 (49) loci as outliers. Of these outliers, 33 (16) were detected in multiple independent comparisons (MI outliers). The remaining 52 (33) outliers (SD outliers) were divided between 28 (15) single-comparison outliers and 24 (18) loci detected in multiple dependent comparisons. Results for each population comparison are summarized in Table 2-3 and depicted in Figure 2-2. All population comparisons detected at least one outlier at the 0.05 $\alpha$

threshold. However, while the LPA-HOB comparison exhibited only one SD outlier, all other comparisons ranged between nine and 29 outliers. At the 0.01 $\alpha$ level, the lone LPA-HOB outlier was no longer detected, but the number of outliers in the remaining comparisons ranged from two to 21.

Although conducting multiple comparisons raises the possibility of type I error, Dfdist corrects for this possibility through Bayesian methods that use the prior distribution (Beaumont & Balding 2004). The number of loci observed in multiple comparisons further illustrates that their outlier status is less likely to be due to chance alone. For example, the probability of observing a given locus in two comparisons is 6.25 $\times 10^{-4}$ (2.50 $\times 10^{-5}$) and for eight comparisons is 1.53 $\times 10^{-13}$ (3.91 $\times 10^{-19}$). Furthermore, there is an even greater statistical argument for those outliers detected in multiple comparisons of only populations on different sides of the cline. The probability of detecting one of these outliers in eight different-cline-side comparisons is 1.30 $\times 10^{-17}$ (8.64 $\times 10^{-23}$). However, three (one) such loci were observed.

The ratio of the percents of MI outliers to SD outliers detected in each comparison was significantly different between different-cline-side and same-cline-side comparisons ($p =$ 0.00158 (0.00337), t-test). Within each comparison type, the average ratio was 1.255 (1.067) for same-cline-side comparisons, but was 5.487 (4.117) for different cline side comparisons. Within sides, selective differentiation appears to involve different loci. The tests indicate population pairs on different sides of the cline are more likely to share the same genetic basis of adaptive differentiation than those on the same side because a

greater proportion of outliers were detected in MI comparisons.  For the $\alpha$ threshold of 0.05, the percentage of MI outliers was not different between the two pair types ($p =$ 0.34240), while the percentage of SD outliers was significantly greater in same-cline-side comparisons than in different-side comparisons ($p = 0.00033$, t-test).  In contrast, the 0.01 $\alpha$ threshold percentage of MI outliers was marginally significantly greater in different-cline-side comparisons ($p = 0.04779$, t-test), while the percentage of SD outliers was marginally significantly greater in same-side-cline comparisons ($p = 0.04340$, t-test).

A handful of loci were detected as outliers in at least four comparisons of population pairs on different cline sides but in no same-cline-side comparisons (DSO outliers).  Because the statistical argument for these loci being under the influence of regionally divergent selection is strong, it is helpful to identify them, all of which were MI outliers at their given $\alpha$ threshold.  At the $\alpha$ threshold of 0.05, locus 80 was an outlier in four different-cline-side comparisons.  This event has a $1.37 \times 10^{-8}$ probability of occurring and is expected by chance in $2.97 \times 10^{-6}$ loci.  Loci 13, 31, and 51 were outliers in eight such comparisons.  This event has a $3.37 \times 10^{-17}$ probability of occurring and is expected by chance in $7.32 \times 10^{-15}$ loci.  For the $\alpha$ threshold of 0.01, locus 31 was an outlier in four different cline side comparisons.  This event has a $2.19 \times 10^{-11}$ probability of occurring and is expected by chance in $4.76 \times 10^{-9}$ loci.  Loci 13, and 51 were outliers in five such comparisons.  This event has a $3.87 \times 10^{-14}$ probability of occurring and is expected by chance in $8.40 \times 10^{-12}$ loci.  Finally, locus 64 was an outlier in eight different cline side comparisons and no same side comparisons.  This event has an $8.64 \times 10^{-23}$ probability of occurring and is expected by chance in $1.87 \times 10^{-20}$ loci.

The locus classes determined by the genome scans using the 0.05 α value were the focus of the remaining analyses to create a more neutral non-outlier data set for comparison to the outlier classes. This allowed for comparison of population genetic patterns among locus classes using a non-outlier data set that would be less likely to include loci under the influence of divergent selection.

## Population Structure

Through examination of loci with large allele frequency changes/shifts between adjacent populations, regional relationships were further revealed (Fig. 2-3). For loci with only one allele frequency shift greater than 0.3 along the transect, the majority of the shifts (14 loci) occurred at the step cline interval between NSB and TTV. In addition nearly 20% occurred each at the TTV-LPA (5 loci) and HOB-PCH (6 loci) intervals. The proportion of loci with one shift became even more skewed toward the Cape Canaveral region, especially the NSB-TTV interval, as the magnitude of the shift increases. All allele frequency shifts greater than 0.6 (2 loci) were located within the NSB-TTV step cline interval. These two loci (loci 13 and 31) were also in the group of loci with only one shift greater than 0.3 and were detected in multiple comparisons between population pairs on different sides of the step cline and no comparisons of pairs on the same side (see above), making them strong candidates for being under the influence of regionally divergent selection.

Overall population differentiation across all populations for all 217 loci was moderate ($F_{ST}$ = 0.0756, Table 2-4). No genetic differentiation was observed between TTV and HOB ($F_{ST}$ = 0.0000, whereas the greatest differentiation was between two adjacent sites, NSB and TTV ($F_{ST}$ = 0.1294), which are only 60 km apart and span the interval of the previously described step cline (Hare & Avise 1996). Among locus classes, genetic differentiation was greatest for MI outliers and least for non-outliers ($F_{ST}$ = 0.1841 and 0.0263, respectively). As compared to population pairs on opposite sides of the cline, those on the same side of the cline showed remarkably lower differentiation for all 217 loci and significantly lower differentiation for MI outliers (p = 0.0224 and 0.0027, respectively), although the average geographic distance between population pairs was similar. In contrast, within both non-outliers and SD outliers, the average pairwise differentiation was nearly equivalent between the two comparison types (p = 0.0991 and 0.4398, respectively). Taken together, these results underscore the role divergent selection plays in the effects regional geographic patterning have on loci across the genome. Regardless of the distance between populations, the variability in differentiation among locus classes demonstrated restricted gene flow at some loci, particularly those subject to regional selection, and freer gene flow at others. Population pairs including TTV had the highest $F_{ST}$ values within cline side groups for all loci, neutral loci, and SD outliers, but not for MI outliers. These results reveal the distinctiveness of TTV from other populations at most loci.

Phylogenetic trees of population genetic distances exhibited different topologies among locus classes (Fig. 2-4). The topologies for all 217 loci and MI outliers were identical

with similarly strong bootstrap support. Remarkably, this topology was not completely consistent with cline divisions, grouping southern FTP between the northern clade (WHI and NSB) and the remaining populations. All populations south of the step cline and the two northern populations formed well-supported monophyletic groups. SD outliers produced a similar topology with generally weaker bootstrap support and one polytomy among TTV, LPA, and HOB. In contrast, the tree resulting from analysis of non-outlier loci was weakly resolved, only showing a high bootstrap value for the (WHI, NSB) node. The remainder of the tree consisted of one polytomy and two nodes with less than 60% bootstrap support. All trees exhibited a comparatively longer branch length for PCH. These trees provide further support for the $F_{ST}$ results by underscoring the greater resolution found in the MI outlier data set over the non-outliers and SD outliers and strong distinction between populations at the step cline. The latter is particularly true for the non-outlier tree.

Individual-based assignments performed by STRUCTURE and summarized by CLUMPP created cluster patterns that differed by locus class (Fig. 2-5). Sample sites for all 217 loci and MI outliers clustered similarly to the tree topologies (Fig. 2-5.A and C, respectively). The optimum number of clusters (Fig. 2-6) was either two (one cluster grouping WHI, NSB, most of FTP, and some component of PCH; another grouping TTV, LPA, HOB, and most of PCH) or four (one cluster grouping WHI, NSB, and most of FTP, a second grouped LPA and HOB, and the last two consisted of TTV and PCH as one site each). Three clusters were most likely for the SD outlier data set (Fig. 2-6), but mixed ancestry was limited only in WHI, NSB, TTV, and LPA (Fig. 2-5.D). The WHI-

91

NSB cluster dominated FTP and represented about half the individuals in PCH. TTV was its own cluster. The LPA-HOB cluster dominated these sites and represented the other half of PCH individuals. For non-outlier loci, the optimum number of clusters was difficult to evaluate (Fig. 2-6). The Evanno method (Evanno et al. 2005) cannot evaluate K=1 and provided the best support for K=5 and secondary support for K=2. The high proportion of cluster mixing within each site for K=5 suggests this was too many cluster divisions (Fig. 2-5.B). As a result, the likely number of clusters for neutral loci is either one or two. In the case of two clusters, all individuals grouped in a single cluster except for most individuals from TTV, which grouped with some individuals of LPA, FTP, and HOB. In this clustering and the higher optimum clusters for the other locus classes, TTV formed its own cluster in each case. In addition, FTP began to form its own cluster at K=5 for all 217 loci and for MI outliers. TTV and FTP likely have some unique, but weakly, distinguishing genetic features in the MI data set. While the clusters created by CLUMPP provide additional detail on substructure, focusing on the best K value may ignore important information revealed at other K values, particularly if there is an alternative K value that has considerable statistical support.

Ancestry proportions generated by CLUMPP using all 217 loci differed between the most likely cluster models. Under the K=2 model, six of the seven sites had very low levels of mixed ancestry (mean maximum ancestry > 0.955; Fig. 2-7). The mean maximum ancestry at PCH was significantly lower than all other sites (0.8466). Although a limited number of individuals had mixed ancestry in each site, FTP showed the most mixture of clusters in the K=4 model (mean maximum ancestry = 0.8163). The next lowest mean

maximum ancestry was 0.0857 at TTV. All other sites had mean maximum ancestry proportions above 0.930. FTP had significantly higher levels of mixed ancestry than all other sites except TTV, which was significantly more mixed than WHI, LPA, HOB, and PCH. Combined with the visual bar plots of ancestry assignment (Fig. 2-5.A), these results suggest that: FTP exhibits complex ancestry patterns dominated by Atlantic-type alleles; TTV ancestry is largely Gulf-like with some unique differentiation; and PCH features unique multilocus ancestry patterns that cannot be captured by the K=2 model.

The PCoA revealed differing population patterns of individual-based genetic distances among locus classes (Fig. 2-8). It is important to note that individual axes are not comparable between locus-class data sets and necessarily should not be interpreted as describing similar aspects of the underlying data. Mean eigen vector values for each site exhibited the greatest spread for the entire 217 loci data set, followed by the MI outliers. This suggests that genetic differences are most discernable and contain the largest variance among individuals using the entire 217 loci data set, but particularly for the MI outlier loci. The proportion of variation explained by all three axes was highest for MI outliers (0.7308) and lowest for non-outlier loci (0.5611; Fig. 2-9). Indeed, the error bars demonstrate clear differences among sites for the full data set and the MI outliers. Groupings from these two analyses are generally in agreement with the STRUCTURE K=4 clusters. For the entire data set, WHI and NSB clustered closely, with FTP overlapping their distribution. Along axis 1, these three populations all differed from (i.e., did not overlap with) TTV, LPA, and HOB, with WHI additionally differing from PCH. A similar pattern was found along axis 1 for MI outliers, except FTP and TTV

93

overlapped partially.  Additionally, TTV differed in its distribution from LPA along axis

2 and from PCH along axis 3 using the entire data set, further corroborating the

distinctiveness of this population from others.   Meanwhile, no differences existed among

sites for non-outlier loci.  Only one population differed for SD outliers, where TTV

differs from WHI, NSB, and FTP along axis 1.  This result strengthens the argument

against K=5 found in STRUCTURE analyses of non-outlier loci and supports a one-

cluster designation for non-outlier loci, which could not be evaluated above.  For all locus

classes, WHI, NSB, and FTP had positive mean eigen vector values along axis 1, whereas

TTV, LPA, HOB, and PCH all had negative values (Fig. 2-8).  The population

distributions along axis 1 are highly reflective of the groups and bootstrap support

detected in each phylogenetic analysis.


Due to the magnitude of within-site variance, migrants and hybrids were difficult to

identify (data not reported here).  This was especially true for non-outlier loci and SD

outliers, which lacked differential clustering among populations.  Using all loci, only one

migrant could be identified as having a genetic background similar to TTV at the HOB

site.  This result corroborated the TTV cluster identity of the HOB individual in the

STRUCTURE K=4 analysis.


Geographic patterns of genetic variation were similar among locus classes.  For each

class, the three-region model explained the least amount of variance after AMOVA

testing (Table 2-5).  The K=4 model had the highest among-region differentiation ($\phi_{RT}$),

likely because designating TTV and PCH as individual clusters moved some of the

among-population differentiation to among-region differentiation. The two-region model explained similar amounts of variance for among-region and among-populations for all data sets except the SD outliers. Nearly all of the variation was contained within populations for non-outlier loci under all models (93-95%), hence the weakest hierarchical differentiation was found in non-outlier loci. In contrast, the MI outliers had the highest $\phi_{RT}$ and $\phi_{PR}$ values, which were an order of magnitude greater than those found in non-outliers and at least twice those of the SD outliers. Given the tree and AMOVA results, the clusters created by the K=4 value in STRUCTURE may be too refined to describe the overarching pattern of parapatry between two differentiated groups.

## Discussion

Here, I used hundreds of anonymous genetic loci randomly spread across the genome to examine levels and geographic patterns of divergence among eastern oyster populations along an ecotone. This study supports a general geographic pattern of parapatry across putatively neutral loci and those influenced by regionally and localized divergent selection. However, the geographic patterns at the genome level revealed a complex substructure of differentiation featuring localized reversals and localized distinctiveness. Data from the 217 loci support divergence with gene flow expectations of a porous genome, where differentiation is low at putatively neutral loci (suggesting higher gene flow) and high at loci influenced by divergent selection (suggesting gene flow is much restricted in locus-specific patterns due to selection acting at or near these loci). My results further suggest that regionally divergent selection is playing a role at some of these loci in maintaining strong divergence, despite evidence for gene flow at putatively neutral loci. Finally, this study revealed evidence for divergent selection occurring on local scales within regions and possibly at the population level.

The patchy genetic and spatial structure found in the eastern oyster Florida populations sampled here are two supporting lines of evidence that at least some of this selection is exogenous (Arnold 1997). However, it appears that a combination of neutral and non-neutral processes, both historical and contemporary, have influenced the geographic and genomic patterns of differentiation observed in Florida eastern oysters. This study supports and builds on findings from previous investigations (Buroker 1983; Reeb &

Avise 1990; Karl & Avise 1992; Cunningham & Collins 1994; Hare & Avise 1996; Hoover & Gaffney 2005; Murray & Hare 2006). It also highlights the geographically complex role divergent selection can play on the genome in the presence of gene flow and how selection can produce a genomic landscape featuring strong differentiation at some loci in an estuarine species with high potential for gene flow.

### Geographic Patterns of Population Structure

The first goal of this study was to determine the geographic pattern of population structure in a zone of secondary contact using genomic tools. Given the clinal results of a previous transect study (Hare & Avise 1996) and lack of clinal decay over time (Murray & Hare 2006), I expected two populations, Atlantic and Gulf of Mexico, to meet at Cape Canaveral with potential integradation between them. In general this pattern was supported. However, the genomic population structure was not completely parapatric, but rather more geographically mosaic with definite substructure. While the northern region (WHI and NSB) showed high genetic similarity ($F_{ST} = 0.0139$), the structure within the previously described Gulf group was highly heterogeneous ($F_{ST}$ range: 0.0000 – 0.1040). The strong genetic distinctiveness of TTV and PCH from the other southern populations and the similarity between FTP and the northern populations contrasted starkly with the near identity between non-adjacent populations separated by 92 km, LPA and HOB ($F_{ST} = 0.0000$). It is unclear what factors drive the population genetic heterogeneity observed in the south (Hare & Avise 1996), but the mosaic geographic structure implicates a role for exogenous selection (Arnold 1997; Ross & Harrison 2002).

Homogeneity in Atlantic-type populations and heterogeneity in Gulf-type populations has been observed previously (Hare & Avise 1996). There are a few of possible explanations for the strong genetic similarity among populations north of Cape Canaveral. Temperate communities and their genetic diversity were much more strongly affected by the glacial cycles of the Pliocene and Pleistocene (Hewitt 1996). As glaciers retreated and species ranges advanced poleward, newly founded populations and expanding refuges often exhibit comparatively genetically depauperate populations toward the poles due to bottleneck and expansion effects. Higher gene flow, either among localities or swamping from larger populations further north, is another possible explanation for the low differentiation among Atlantic-type populations (Garcia-Ramos & Kirkpatrick 1997; Bridle & Vines 2007). Finally, the selection regime may be stronger and/or more homogeneous north of Cape Canaveral, where winter temperatures reach freezing more often above the frost line (Hull & the Federal Geodetic Control Committee 1989; (Southeast Regional Climate Center, http://www.sercc.com/climateinfo/historical/historical_fl.html)). However, results from this study alone are insufficient to draw conclusions regarding Atlantic homogeneity because only two sites in this region are represented.

The relative distinctiveness of PCH from the other Gulf-type populations sampled here was an unexpected discovery. Using one mitochondrial and two nuclear loci, previous allele frequency plots along a transect from Massachusetts to Louisiana showed this location to be the overall most Gulf-like (Hare & Avise 1996). Based on this characterization, this population was used to represent the Gulf-type in a genome scan for

divergent selection between an Atlantic and a Gulf population (Murray & Hare 2006).

However, the present study reveals that PCH is somewhat unique among the sampled

populations. Indeed, allele frequency plots along the present transect show a reversal to

Altantic-type frequencies at PCH in an otherwise generally clinal pattern for a handful of

loci (data not shown). These results suggest that PCH may not be truly representative of

the Gulf-type and underscore the importance of sampling many loci before drawing

conclusions about population structure at the genome level.

Despite the heterogeneity in the south, the phylogenetic population trees indicate an

incomplete signal of parapatry between Atlantic and Gulf type populations. Except for

FTP being placed between Atlantic and Gulf type populations, the division between these

two types was strong. This division may reflect the system's history of vicariance. The

non-outlier population tree exemplifies this demographic legacy, providing the only

strong bootstrap support at the node dividing the populations north of the step cline from

those to the south. Whether or not the allopatric evolution of Atlantic and Gulf

populations involved divergent selection in different environments, it appears a neutral

legacy of vicariance remains (Reeb & Avise 1990; Avise 2004). Indeed, the

biogeographic province boundary between range edges of temperate and subtropical

assemblages and the concordant boundary zones within species along the Florida

coastline (e.g., *Ammodramus maritimus, Cicindela dorsalis, Geukensia demissa, Limulus

polyphemus, Malaclemys terrapin*) are the strongest evidence for a history of vicariance

(Briggs 1974; Avise 2000; Avise 2004). The genetic fingerprints of local adaptations and

genomic incompatibilities that likely arose while populations were allopatric may be

retained due to history and/or due to contemporary selection (Avise 2004). The general parapatric pattern of differentiation observed in this study may have arisen through secondary contact and be sustained by regionally divergent selection along an ecotone, genomic incompatibilities, and/or contemporary barriers to dispersal (Reeb & Avise 1990; Bert & Arnold 1995; Hare & Avise 1996; Arnold 1997).

The similarity of FTP to the two northern populations was unexpected, but not inconsistent with the previous finding that this location exhibited some of the highest Atlantic allele frequencies of sites south of the Cape Canaveral region (Hare & Avise 1996). Additional support for a northern genetic signature in FTP comes from two phylogenetic studies (Cunningham & Collins 1994; Hoover & Gaffney 2005). The predominantly northern genetic signature of this population could be a relict of a northward moving hybrid zone (Barton & Hewitt 1985) or past contact events during glacial cycles (Hewitt 1996). If the hybrid zone moved, it may have left remnants of the Atlantic type along its path until stabilizing its position at Cape Canaveral. The higher mixed ancestry and intermediate genetic distance between northern and other southern populations may be the result of gene flow and introgression from neighboring southern populations into a relict Atlantic population or more effective southward gene flow than northward. There is an inlet at Fort Pierce, which would allow localized immigration. In consideration of the possibility of plate effects, FTP may have had a more northern signature because a third of these samples were run on the same experimental plates as northern samples. However, this explanation does not seem valid because a plate effect pattern was not observed in the other FTP samples or in other populations run with

samples from across the cline (See Appendix Fig. A-1).  Furthermore, positive controls

on each plate minimized potential plate biases.  It is unknown whether the FTP

population is a localized anomaly or indicative of a yet unsampled pattern.

According to tension zone theory, cline centers will move until dispersal barriers and

density troughs entrap and stabilize their location (Barton & Hewitt 1985; Endler 1977).

Differentiation is highest in Cape Canaveral, the cline center.  This is true even for

neutral loci, indicating that neutral processes have a stronger influence in this region than

other areas.  Aside from the neutral contribution of historical vicariance legacy, strong

dispersal barriers in Cape Canaveral are consistent with the physical attributes of the

lagoons in this region.  Cape Canaveral is the region in eastern Florida most isolated from

the Atlantic Ocean.  The nearest inlets to the north and south are at distances of

approximately 70 and 80 km, respectively.  As a result of the isolation, tidal flow in Cape

Canaveral lagoons is very weak (5 cm tidal height, Smith 1987) and would limit dispersal

of planktonic larvae.  Indeed, tidal mixing extends only 1.6 km from an inlet (Smith

1987), and residence times in the region range from 0.04 to 0.62 yr (Florida Department

of Environmental Protection).  Furthermore, the Mosquito lagoon is connected to

northern lagoons by only a narrow canal.  Land separates this lagoon from Banana River

lagoon, located on the eastern side of the Cape.

In addition, there is a comparatively low density of oysters in the Cape Canaveral region

(Hare & Avise 1996; pers. obs.).  This density trough may limit the number of potential

migrants, thereby easing the gene flow – selection balance and reducing the rate of

neutral cline decay.  A severely low density in oysters potentially can produce an allee effect as a byproduct (Stephens & Sutherland 1999; Liermann & Hilborn 2001; Berec et al. 2007).  Oyster sperm and eggs have a limited life span of 1-5 hours in seawater (Galtsoff 1964).  If the distance between potential mates is too great to allow adequate travel of gametes, fertilization will not succeed.  The resulting dearth of progeny would also result in a lower number of potential recruits, only continuing the density trough through a negative feedback loop.

In accord with the above highlighted physical and biological conditions that attract hybrid zones, this region exhibits elevated population differentiation among all locus classes and is the location where the majority of loci with single allele frequency shifts greater than 0.3 and all shifts greater than 0.5 occur.  Clearly, oysters in the Cape Canaveral region experience reduced gene flow.  The physical geography and species density of Cape Canaveral lagoons likely facilitates the genetic distinctiveness of the TTV site by enabling it to evolve in more isolation.  It is important to note that for non-outliers $F_{ST}$ between TTV and the adjacent sample sites was about 0.05, which indicates reduced gene flow relative to other adjacent population pairs but does not support strong restrictions to gene flow.  It is also possible that TTV samples exhibited distinctiveness because they may not have amplified as well in the laboratory as other samples for biological or experimental reasons.  This site had the lowest average band-present allele frequency, but the difference was not significant.  Another experimental difference was that TTV was the only set of samples analyzed on the ABI 3730 because these samples were acquired late in the study, but positive controls helped reduce any run differences.  However, if the

later sampling of this site made TTV less comparable to other samples, it may indicate less stability in allele frequencies along the cline than previously thought (Murray & Hare 2006). This seems like an unlikely explanation because the interval between samples in this study is much less than the one Murray & Hare examined. These potential experimental differences in TTV samples did not result in a consistent effect across primer pairs or across loci within primer pairs, and therefore should not be driving the overall patterns observed.

It appears that a combination of neutral and selection-related processes, both historical and contemporary, play a role in shaping the patterns of geographic differentiation observed in this study. Many of the maritime contact zones in Florida do not have cline centers in the same location (e.g., ribbed mussel in Sarver et al. 1992; weakfish in Cordes & Graves 2003; killifish in Duggins et al. 1995; hard clams in Bert & Arnold 1995). Indeed, the broader, well-recognized biogeographic province boundary spans clusters of species range limits between the Georgia-Florida border and Palm Beach, FL (Briggs 1974; Engle and Summers 1999). Neutral processes limiting dispersal likely have similar effects on species with similar life cycles. The differences in contact zone locations imply that neutral processes are not the only factors determining cline shape and center in these species. A combination of neutral forces, endogenous selection, and exogenous selection has been demonstrated in another concordant bivalve hybrid zone (hard clams *Mercenaria spp.*: Dillon & Menzi 1989; Bert & Arnold 1995).

**Divergent Selection Drives the Genomic Pattern of Differentiation**

The second goal of this study was to determine if selection shapes the genomic landscape of differentiation, using a null hypothesis of genome-wide neutral differentiation processes. This null hypothesis was rejected, as it was in a study by Emelianov and colleagues (2004). There was clear evidence here for a porous genome, as theory predicts in a system of divergence with gene flow (Hodges & Arnold 1994; Feder 1998; Wu 2001; Nosil et al. 2009). In this study, approximately 39% of loci were detected as highly diverged $F_{ST}$ outliers in at least one population comparison putatively due to the effects of divergent selection, leaving nearly 61% as putatively neutral across all populations. This large percentage of outlier loci is higher than those found in most other genome scans (e.g., Wilding et al. 2001; Bonin et al 2006; Egan et al. 2008; reviewed in Nosil et al. 2009), possibly due to the larger number of population comparisons performed to detect outliers here. Sampling sites may include additional sources of environmental heterogeneity produced by strong selective parameters found in intertidal estuaries. This is consistent with the large majority of outliers being site-specific (SD outliers). Nonetheless, the percentages should be cautiously interpreted with the realization that they likely include false positives at the $\alpha = 0.05$ level and less so at the $\alpha = 0.01$ threshold (see below for further discussion). Indeed, the percentage of outlier loci drops to 22.6% when $\alpha = 0.01$. The average proportion of loci detected as outliers within individual comparisons was 6.7% for $\alpha = 0.05$ and 3.1% for $\alpha = 0.01$. These per-comparison outlier proportions are consistent with those found in other studies (see above references), including the previous Florida oyster study (Murray & Hare 2006).

Although it is sometimes done in genome scan studies (e.g., Galindo et al. 2009; Nosil et al. 2008; Nosil et al. 2009), the percentage of loci found to be outliers must not be inferred to represent the proportion of the genome under selection. Without a genetic map of the markers, the anonymous nature of AFLPs precludes an understanding of their genomic distribution and the magnitude of hitchhiking effects. As the strength of selection increases and/or recombination decreases, the genomic region experiencing indirect hitchhiking effects of selection broadens (Maynard Smith & Haigh 1974; reviewed in Barton 2000; and Andolfatto 2001). Although studies have demonstrated general genomic dispersion of AFLP loci (reviewed in Black et al. 2001), the likelihood of a differentiated region including more than one AFLP locus increases with the size of the hitchhiking region and number of loci. Therefore, the possibility exists that more than one locus sample the effects of divergent selection on a single genetic target, though the linkage index threshold minimized this potential.

In addition, various strategies commonly employed in multiple-population genome scans can lead to increased false positives (Bonin et al. 2006; Excoffier 2009). Significance criteria are rarely corrected for multiple comparisons (except Eveno et al. 2008; Galindo et al. 2009), yet interpreted as though there is no increased possibility of false positives. Excoffier et al. (2009) recently pointed out that ignoring population substructure produces high levels of false positives when comparing populations in aggregate, but that performing pairwise comparisons is one way of addressing this. However, the software Dfdist does implement algorithms to reduce type I error (Beaumont & Balding 2004; Nosil et al. 2008). Necessarily, I cautiously interpret the identity of any given locus and

the percentage of outlier loci as putatively being under the influence of selection. Furthermore, non-outliers quite possibly represent a mixture of truly neutral loci and loci under the influence of weak selection that is either uniform, balancing, or weak and divergent. It is important to note, though, that the probability of a given locus truly being under the influence of divergent selection increases with the number of independent comparisons in which it was identified as an outlier. In this way, larger patterns across multiple comparisons can be more reliable indicators of broader divergent selection regimes.

## Multiple Independent Outliers

Likewise, the statistical likelihood of repeatedly identifying a given locus by chance drops dramatically as the number of comparisons detecting the locus increases. Other genome scan studies have used repeated outliers as their most reliable indicators for patterns of divergent selection (Bonin et al. 2006; Nosil et al. 2008; Egan et al. 2008). Indeed, false-positive outliers are not expected to exhibit parallel behavior in several comparisons (Bonin et al. 2006). This study goes one step further by distinguishing between outliers detected in multiple independent population comparisons and those that are only repeated when a population is shared in all detecting comparisons. Such multiple independent (MI) outliers represented 15% of all loci at the $\alpha = 0.05$ level and 7% of loci when $\alpha = 0.01$, percentages consistent with those found in other studies (e.g., Wilding et al 2001; Egan et al. 2008). However, having at least two population comparisons that do not share any sites is purely a statistical argument here and does not imply replicated, independent divergence among population pairs. While some contact

zones have this replication in nature, the eastern oyster does not. Furthermore, this locus

class contained outliers detected in population pairs on the same side of the cline and on

different sides. In this study, continued selection regimes in nature, whether patchy or

monotonic, have non-negligible effects on the genomic structure of genetic

differentiation.

Loci under the influence of divergent selection appear to drive population structure, as

the striking similarities between results from all 217 loci and MI outliers indicate.

Although population structure was assessed using a variety of methods, results

consistently documented weak population differentiation for non-outliers and SD outliers

distinct from the stronger pattern observed in all loci and MI outliers in particular. The

reduced differentiation found in non-outlier loci is not a surprise because $F_{ST}$-based

criteria were used to distinguish outliers from non-outliers. However, population-level

$F_{ST}$ values of non-outlier and MI outliers were not correlated (correlation analysis not

shown), suggesting differentiation at the two sets of loci is governed by a different

balance of evolutionary forces.

Cautions considered, the MI outliers represent the strongest candidates for divergent

selection (Bonin et al. 2006; Nosil et al. 2008). These outliers are putative markers for

genomic regions that either are targets for divergent selection or facilitate reproductive

isolation along the speciation continuum. These regions are selection-based barriers to

the gene flow that occurs freely across the remainder of the genome. The high number of

these outliers may indicate that the divergent populations are farther along on the

speciation continuum compared to other outlier studies (Wu 2001). Because they were outliers in at least two independent comparisons, they show that divergent selection has the same genetic basis in multiple populations along the transect. Although the contact zone is not a case of parallel evolution driven by the same selective regimes (e.g., whitefish: Pigeon et al. 1997; sticklebacks: Rundle et al. 2000; Colosimo et al. 2005; *Littorina saxatilis*: Johannesson et al. 2009), selective differences within regions or that may or may not have arisen in allopatry are being maintained consistently in the face of gene flow at other parts of the genome.

## Non-outliers

Non-outlier loci indicate the presence of gene flow through their extremely low genetic differentiation, low resolution of their phylogenetic tree, and the inability of assignment tests to identify robust clusters. The scale of gene flow is unclear, but previous estimates of dispersal in Chesapeake Bay eastern oysters indicate the evolutionary scale of dispersal is 479 km$^2$ (Rose et al. 2006). In the Florida lagoons, dispersal distances are likely much less given the longer retention times in this shallow system with weak tidal influences (Smith 1987).

Completely neutral explanations may not explain the extremely low differentiation observed in some non-outliers. Instead, some of these loci may be under the influence of balancing or uniform selection between population pairs (Lewontin & Krakauer 1973; Beaumont & Nichols 1996). Genome scans are most often performed as one-tailed tests to search for extraordinarily high levels of differentiation. The method is not as powerful

for detecting uniform selection (Beaumont & Nichols 1996).  Ignoring the likelihood of

uniform selection affecting a data set potentially biases the genomic $F_{ST}$ downward,

which is used to create the simulated neutral data set.  However, Dfdist does account for

this possibility by trimming extremely low and high values of $F_{ST}$ before calculating the

mean (here, upper and lower 30% trimmed).  Though this reduces the probability of false

positives due to a downwardly biased estimate of genomic differentiation, the one-tailed

nature of the test still limits the ability to distinguish between truly neutral loci and those

under uniform or balancing selection (Beaumont & Nichols 1996; Beaumont & Balding

2004).  Therefore, non-outliers should be interpreted as just that, not necessarily neutral

loci.

Although the vast majority of genetic variation observed in non-outliers was contained

within populations, the subtle structure that was found reflected the pronounced structure

found in MI outliers.  Differentially adapted loci can potentially drive differentiation at

unlinked neutral loci by creating general barriers to gene flow when divergent selection

against immigrants is strong, thereby facilitating drift (Barton & Bengtsson 1986; Nosil

et al. 2009).  Though gene flow at neutral loci breaks down much of the adaptive

genomic associations found, it may not be strong enough to counteract this phenomenon

of isolation by adaptation completely (Nosil et al. 2009).  Alternatively, the weak

structure observed in non-outliers could be the result of statistical categorization of

selection, which varies continuously.  Although most analyses here focused on the locus

classes produced by $\alpha = 0.05$ criterion in order to create a more neutral data set, non-

outliers likely still contain loci under the influence of weak divergent selection and loci

that are more loosely linked to target loci.  Under these circumstances, the effects of selection are too weak to elevate the loci to outlier status.  In such cases, some loci are false negatives because they carry a faint signal of differentiation based on divergent selection.  The hitchhiking effect depends greatly on the ratio of recombination frequency and strength of selection (r/s), such that higher recombination rates and weaker selection produces smaller hitchhiking effects on loci linked to fitness-related loci (Barton 2000; Andolfatto 2001; Ortiz-Barrientos et al. 2002; Butlin 2005; Via & West 2008).  Other possible explanations for the weak but present structure at putatively neutral loci include the strengths and balance of gene flow and selection and insufficient time since secondary contact.  Given the porous nature of the genomic structure of differentiation and the subtle signals of differentiation in non-outliers, gene flow is most likely large enough to break down the bulk of differentiation through recombination but may be too weak to overcome the effects of divergent selection entirely.

<div align="center">Singleton and Dependent Outliers</div>

The moderate differentiation observed in SD outliers has several potential explanations.  These outliers may indicate localized selection, specific to individual populations or fine-scale habitat differences not repeatedly sampled here.  Alternatively, the influence of divergent selection may be relatively weak at these loci, only elevating them to outlier status in a limited number of comparisons.  Finally, multiple comparisons elevate the chances of type I error, or false positives.  It is likely that some SD outliers are statistical artifacts, particularly in the case of outliers detected in only one comparison (Bonin et al. 2006; Nosil et al. 2008).  Considering only outliers detected by the 0.01 $\alpha$ level reduces

the possibility of type I error in inferring selection at particular loci. However, the purpose of this study was not to identify specific loci influenced by divergent selection, but rather to characterize genomic and geographic patterns of differentiation through the behavior of multiple locus classes.

## Mixed Ancestry

Contact zones are sometimes categorized by the distribution of genotypic classes (Harrison & Bogdanowicz 1997; Howard 1993). Hybrid swarms, where intermediate genotypes dominate, are called 'unimodal' (e.g., *Bombina* frogs: Szymura & Barton 1991). 'Flat' hybrid zones have a relatively even representation of genotypic classes (e.g. crickets, Howard & Waring 1991; Cyprinid fishes: Meagher & Dowling 1991; leopard frogs: Sage & Selander 1979; and oaks: Howard et al. 1997). When parental genotypes dominate and intermediates are rare, the hybrid zone is termed 'bimodal' (e.g., *Heliconius* butterflies: Jiggins et al. 1997; Louisiana irises: Cruzan & Arnold 1994). This continuum of genotypic distributions ends with speciation. Bimodal contact zones are strongly associated with prezygotic barriers, assortative fertilization or assortative mating (Jiggins & Mallet 2000). Despite the low ratio of hybrids, bimodal zones show no association with postzygotic incompatibilities, ranging from hybrid breakdown to hybrid vigor (reviewed in Jiggins & Mallet 2000). As a result, endogenous postzygotic selection can be weak in these systems, suggesting a prominent role for exogenous postzygotic selection and/or other premating barriers to gene flow in bimodal zone maintenance.

Past studies on the oyster step cline all used population-level statistics to describe geographic and genomic patterns of differentiation (Buroker 1983; Reeb & Avise 1990; Karl & Avise 1992; Hare & Avise 1996; Murray & Hare 2006). Individual assignment approaches can answer additional questions regarding the role divergent selection plays in shaping genomic differentiation. The strong barriers to gene flow found at the population level for MI outliers are consistent with reduced levels of mixed ancestry at the individual level found in this class of loci compared to other classes. Comparing bar plot results from DISTRUCT analyses for the different locus classes further supports a porous genomic landscape, whereby individuals appear to have greater mixed ancestry at non-outlier and SD outlier loci than at MI outliers.

The low levels of mixed ancestry in most populations suggest the oyster contact zone is genotypically bimodal. Because bimodality is associated with prezygotic barriers and exogenous selection, the few migrants detected here may be explained by reduced migrant fitness, but may simply be the result of weak hydrodynamics. These inferences of low levels of mixed ancestry and migration should be interpreted with caution because these analyses of population structure were not designed to test ancestry specifically or vigorously. Nonetheless, the possibility of pre- and post-zygotic barriers should be investigated. Though behaviorally based assortative mating is unlikely in a broadcast spawner, a recent fertilization trial between two populations located on opposite sides of the step cline failed to demonstrate any strong reproductive barriers for embryo survival, fertilization success, and paternity (Zhang et al. 2010). The authors of this study suggest that prezygotic and early postzygotic genetic incompatibilities do not maintain the step

cline, leaving physical dispersal barriers and exogenous selection as possible mechanisms for cline maintenance.  However, endogenous selection may play a role at a life stage later than the embryo and cannot be ruled out.

## Geographic Pattern of Selection along the Ecotone

The third goal of this study was to determine whether selection produces a clinal geographic pattern along the ecotone.  In general, the geographic pattern along the transect was not smoothly clinal, even for most loci putatively under the influence of divergent selection.  The broad range of allele frequency patterns along the transect reflect the mosaicism of genomic regions under geographically variable balances of evolutionary forces.  In parapatric contact zones, clinal loci are of particular interest because they provide insights into the evolutionary dynamics of gene flow and selection (Barton & Hewitt 1985; Harrison 1990).  In this AFLP study, some loci did show overall clinal patterns in allele frequency with occasional local reversals, particularly involving FTP.  Although the vast majority of such loci are classified as MI outliers, several are classified as SD outliers and even as non-outliers.  However, not all MI outliers were clinal.  Many MI loci exhibited multiple large allele frequency shifts along the transect, providing further indication that ecotonal selection may not be the only form of divergent selection operating in this system.

### Clinal Selection

Although the hypothesis of two parapatric populations corresponding to Atlantic and Gulf clades (Reeb & Avise 1990) proved incomplete across the genome, it can still be

used as a framework to examine divergent selection on regional scales. By comparing

trends in outliers for population pairs on opposite sides of the previously described step

cline versus those on the same side, it appears that divergent selection between these

regions is affecting genetic differentiation at some loci. The nearly four-fold more MI

outliers versus SD outliers observed in population pairs across regions demonstrate that

selection differences between regions have a strong influence over the patterns of

geographic differentiation observed at many loci. Furthermore, examination of outliers

detected in multiple comparisons only on opposite cline sides and no comparisons on the

same side revealed several candidate loci for regional selection. Two of the three outlier

loci identified by Murray & Hare (2006) at the $\alpha = 0.01$ level were MI outliers in this

study using the same significance criteria. In fact, these loci were identified in at least

five comparisons of populations on opposite sides of the cline and zero comparisons on

the same side. The third locus was not examined in this study. These loci are also clinal,

demonstrating that these criteria are a promising way to identify clinal loci involved in

regionally divergent selection (Wilding et al. 2001; Campbell & Bernatchez 2004; Bonin

et al. 2006; Egan et al. 2008; Nosil et al. 2008).

For those MI outliers that are generally clinal along the transect, the possible processes

governing their geographic pattern of differentiation are most likely ecotonal or

endogenous selection. Barriers to dispersal between two historically vicariant

populations, the Atlantic and Gulf types, may play some role (Avise 2004). However, the

broad admixture occurring at most non-outlier loci indicates that gene flow and time have

been sufficient to erase nearly all neutral divergence that drift may have generated in

vicariance.  Clinal outliers may be tracking latitudinal ecotone transitions in biotic composition or the latitudinal gradient in temperature.

Hybrid zone theory offers multiple explanations for the geographic location of clines, depending on whether selection is predominantly exogenous or endogenous.  For exogenous selection, sharp clines are generated and maintained by steep environmental changes and low gene flow between differentially adapted populations (Endler 1977). The geography, hydrography, and low oyster density of Cape Canaveral lagoons serve as a physical backdrop that likely limits planktonic larval dispersal, as mentioned above. Indeed, the Cape Canaveral region experiences extremely low recruitment success and high mortality in the sporadic recruitment events that do occur (pers. obs., see Chapter 3). Neutral processes physically limiting dispersal are probably working in concert with a strong selective regime in this area (Bert & Arnold 1995) and ease cline maintenance at selected loci.  Populations nearest to dispersal barriers are usually the most differentiated (Endler 1977).  The comparatively higher differentiation observed in TTV non-outliers and SD outliers exemplify this.  Furthermore, TTV has most of the large, single allele frequency shifts.  Genetic isolation in this population can facilitate local adaptation at this site and regional adaptation beyond it.

Tension zone theory predicts that reduced hybrid fitness due to intrinsic barriers, or genomic incompatibilities between divergent parental populations, produces a pattern of linkage disequilibrium (LD) in the region of transition (Barton & Hewitt 1985). Although intrinsic barriers to gene flow have not been explicitly tested here, Hare &

Avise (1996) tested LD in codominant loci and found low to no LD and Hardy-Weinberg equilibrium genotypic proportions within sites. They determined that there was free interbreeding within sites and predominantly local recruitment. Furthermore, Zhang and colleagues recently found no early postzygotic barriers existed between Atlantic and Gulf-type samples (2010). A similar lack of LD was found in a study of the co-distributed *Fundulus* hybrid zone, leading the authors to conclude that this was not a tension zone (Duggins et al. 1995). Increasing evidence shows that hybrid zones do not conform completely to the tension zone cline model of dispersal balanced by endogeneous selection against genomic incompatibilities (reviewed in Arnold 1997).

Though not examined in Florida populations, several studies show evidence of local adaptation in the form of physiological races in the eastern oyster. Newkirk (1977) observed races adapted for salinity tolerance. In a common garden environment, Long Island Sound (LIS) oysters exhibited earlier gametogenesis and spawning than Delaware Bay (DB) oysters, allowing them to begin reproduction in colder conditions and closer to the time of their southern counterparts (Barber et al. 1991). LIS oysters also had higher gill ciliary activity at lower temperatures than DB oysters, enabling them to feed at the lower temperatures found more northward (Dittman 1997). Although growth rate is predominantly considered highly phenotypically plastic in the species, LIS oysters exhibited countergradient or physiological compensation through higher growth rates than DB and Chesapeake Bay (CB) oysters in common gardens (Conover & Schultz 1995; Dittman et al. 1998). The results of these studies are more profound given the lack of physical or known hydrographic barriers between CB, DB, and LIS. This

demonstration of local adaptation in this broadcast spawner suggests the lagoon

geography and hydrography in Florida should provide greater facilitation for local

adaptation in this contact zone.  Certainly these environmental parameters vary in Florida

estuaries (Smith 1993; Southeast Regional Climate Center,

http://www.sercc.com/climateinfo/historical/historical_fl.html) and could be acting as

selective agents for countergradient and local adaptation.


Broad variation in geographic differentiation patterns among loci is not unique to this

system.  Different selection regimes can cause differential introgression among loci

within the same contact zone (Stuckus et al. 2009).  Differentiation was strongly

heterogeneous among chromosomes of larch moth ecotypes when comparing different

host populations but not within the same host (Emelianov et al. 2004).  Introgression was

highly variable across the genome in a mouse hybrid zone and produced significant

differences in cline widths and centers, despite purposeful selection for loci that were

alternately fixed between parental subspecies (Teeter et al. 2008).  For a marine clam,

*Macoma*, the general clinal pattern between hybridizing populations had locus-specific

deviations, including local reversals in allele frequency and strong similarities among

distant populations (Nikula et al. 2008).  These patterns were particularly similar to that

found in this study.  Nikula and colleagues implicated hydrographic patterns and salinity

gradients in the broad clinal pattern, but did not offer an explanation for the smaller scale

inconsistencies.

<u>Within-region Heterogeneous Selection</u>

Indeed, spatially heterogeneous selection has been implicated as a possible explanation

for the southern heterogeneity observed in the oyster system (Hare & Avise 1996). In the

Hare & Avise study, the region from Cape Canaveral to Miami showed allele frequency

heterogeneity for a few markers that were strongly differentiated between Atlantic and

Gulf clades. The authors offered additional potential explanations for this southern

heterogeneity. Through the resulting sea level changes, repeated cycles of glaciation and

recession of glaciers during the Pleistocene might have provided the opportunity for

multiple secondary contact events, each of which could leave relict populations.

Idiosyncratic demographic events like occasional long distance dispersal could also

generate this geographic pattern. Finally, the authors suggested spatially heterogeneous

selection also could play a role. A similar pattern of southern heterogeneity was found in

the collocated *Fundulus* hybrid zone (Duggins et al. 1995).


In contrast to generally clinal loci, some MI outliers exhibited multiple allele frequency

shifts along the transect. These loci are likely targets or linked to targets of finer-scale

environmental selection. The characteristic population-specific differentiation observed

in SD outliers, particularly those detected in multiple dependent comparisons, also is

evidence for localized selection. Estuarine intertidal environments experience broad

variation in numerous abiotic and biotic parameters over small and large temporal and

spatial scales. Key parameters affecting survival in oysters and other intertidal sedentary

organisms include: salinity, temperature, dissolved oxygen, tidal height affecting

exposure to desiccation, rate of hydrographic flows, community interactions,

sedimentation, and turbidity (Shumway 1996; Pennings & Bertness 2001). Because the sampling scheme here did not attempt to hold any of these parameters constant, there is potential for one or more of them to act as divergently selective agents among sample sites. Given that local recruitment is likely for Florida oysters (Hare & Avise 1996), populations could be locally adapted to their population-specific environments. The mosaic model of hybrid zones has several characteristics: (1) geographically mosaic genotype frequencies, (2) parental types adapted to different environments that are patchily distributed, (3) the presence of exogenous selection, and (4) reduced hybrid fitness (Howard 1986; Harrison 1990). This study of adult oysters along the contact zone in Florida supports for the first two of these characteristics for some loci and hints at the last two, indicating that this contact zone has some characteristics of a mosaic hybrid zone in addition to those of an ecotonal zone at different parts of the genome.

Although the formerly mentioned neutral scenarios could contribute to the southern heterogeneity and cannot be excluded, this study implicates spatially heterogeneous selection in southern genetic differentiation for outlier loci. This is particularly exemplified by the northern genetic signature observed at FTP, which groups between Atlantic clade populations and other Gulf sites in population trees for MI outliers and SD outliers and has mixed ancestry among MI outliers. Fort Pierce exhibited relatively more northern genetic signatures in other studies as well (Cunningham & Collins 1994; Hare & Avise 1996; Hoover & Gaffney 2005). The fact that genome scans between Fort Pierce and its adjacent sample sites returned the greatest number of outliers over all other comparisons, with MI outliers outnumbering SD outliers two to one at $\alpha = 0.05$,

demonstrates the uniqueness of the Fort Pierce selective regime from other southern sites sampled here. Exogenous selection may be driving these outlier characteristics. Pure parental genotypes may be selected against in the Fort Pierce area. Under a bounded hybrid superiority model, hybrids are more fit in intermediate environments than parental genotypes (Moore 1977). The Fort Pierce site may have more northern-like physical and/or biotic regime than its southern counterparts. Patterns of mixed ancestry depending on locus class may be shared at other populations beyond Fort Pierce, but were not sampled elsewhere in this study. It is possible that the anomalous patterns observed in FTP may be representative of fine-scale, patchy, or mosaic selection regimes.

Though previous population genetic analysis of a few loci across the majority of the oyster range revealed a clinal transition from Atlantic to Gulf types (Karl & Avise 1992), more focused analyses at these loci (Hare & Avise 1996) and at hundreds of anonymous loci (here) reveal mosaically patterned large allele frequency reversals within the zone of contact, even at most generally clinal loci. Thus, there is good evidence in Florida oysters for selection driving and/or maintaining differentiation in the presence of gene flow on various spatial scales. While some loci most likely represent historical genomic incompatibilities or are responding to ecotonal and/or latitudinal gradients, others are likely responding to selective pressures that operate at finer scales like those often found in intertidal estuaries.

**Future Prospects**

Some genome scan studies have benefitted from geographically replicated contact zones, enabling comparison among evolutionarily independent events of divergent selection (e.g., whitefish: Pigeon et al. 1997; sticklebacks: Rundle et al. 2000; Colosimo et al. 2005; *Littorina saxatilis*: Johannesson et al. 2009).  The oyster contact zone involves only one geographic zone of historically vicariant populations.  As a result, population comparisons are not evolutionarily independent laboratories to test for similar regimes of divergent selection.  This population genomic study of differentiation along a previously reported cline revealed complex geographic patterns at regional and local scales. Although this study examines more sites than most genome scan studies, some of the patterns observed will require examination of additional sites to eliminate possible explanations.  With denser sampling along the transect, it will be easier to identify behaviors of some populations (e.g., FTP, TTV) as site-specific, anomalous, or representative of a yet undiscovered pattern.  Including multiple samples at each locale will illuminate the lower limit of the geographic scale of selection in this intertidal estuarine system.

A more powerful geographic sampling scheme would be greatly enhanced by incorporation of environmental data.  As of yet, both the targets and agents of selection remain unknown in this system.  When no documented phenotypic differences exist, one way to investigate selective agents is by looking for genetic associations with environmental factors.  In particular, landscape genetic approaches can address whether there are geographic relationships between environmental parameters and outlier allele

frequencies (Manel et al. 2003; Storfer et al. 2007). Some features of the environment may explain a clinal allele frequency pattern, while others may explain some of the localized frequency reversal patterns observed. For the previously described clinal pattern and collocated biogeographic province boundaries, several studies implicate the role of a latitudinal temperature gradient. In this case, the best candidates for genetic associations with temperature or biological community are outliers that were identified in multiple independent comparisons only between populations on opposite sides of the step cline. Other MI outliers and those from multiple dependent comparisons are well suited for studies examining environmental parameters that vary patchily on smaller scales, such as dissolved oxygen, salinity, and turbidity.

An association between genetic differentiation and exogenous variables should help to identify the phenotypic characters that are under selection. Such associations with phenotype have identified adaptive variation in several systems (e.g., lizards: Rosenblum 2006; pocket mice: Nachman et al. 2003). In the intertidal zone, local selection gradients for desiccation risk, predation, and wave action have all produced thicker shells in several mollusks (e.g., littorinid snails: Hull et al. 1996; *Mytilus*: Bierne et al. 2002). Along the latitudinal temperature gradient in Florida, regional adaptation might manifest in temperature tolerance thresholds, particularly for seasonal extremes like freezing conditions during winter temperature minimums. Outside of Florida populations, there is already evidence of phenotypes that show local adaptation along a latitudinal temperature gradient. An investigation into such physiological tolerance thresholds might yield interesting results that could link genotype, phenotype, and environment. To identify loci

underlying adaptive traits, it would be helpful to have loci mapped to a sequenced and annotated genome for this commercially important species.

The genomic mosaic caused by the opposing forces of gene flow and selection during adaptive divergence has revealed a great deal about the genic view of speciation (Wu 2001). This genome scan and study of locus class behavior is a step toward determining whether divergent selection plays a role in maintenance of the oyster step cline. However, these statistical arguments must be verified through additional study. Reciprocal transplant field studies could conclusively evaluate if exogenous and/or endogenous selection reduces fitness of immigrants and hybrids, thus maintaining the cline. Such environmentally based local adaptations can cause strong differentiation observed at outlier loci. Through cohort analyses of spat, allele frequency changes over time would indicate whether post-settlement selection plays a role in the variable patterns of differentiation across space and would help distinguish when that selection occurs in the life cycle (e.g. Koehn et al. 1980; Bert & Arnold 1995). Newly settled spat also will reveal temporal and geographic patterns in dispersal and how much gene flow must be overcome by selection.

# Conclusions

Although transect studies of the oyster contact zone conducted by Avise and colleagues demonstrate dramatic genetic clines (Reeb & Avise 1990; Karl & Avise 1992; Hare & Avise 1996), this study builds on a previous two-population genomic study (Murray & Hare 2006) to clearly demonstrate a genomic and geographic mosaic of differentiation in these populations. Divergent selection plays a large role in patterning these mosaics and is countered by gene flow at neutral loci. In contrast to the two-population approach (McDonald et al. 1996; Murray & Hare 2006), combining genomic and geographic approaches enabled the identification of specific loci that had clinal frequency patterns echoing those documented by Avise and colleagues and highlighted candidate loci for regional and fine-scale selection for further examination.

Because the geographically mosaic population structure does not completely agree with a simple cline, using the mitochondrial dichotomy of "Atlantic" and "Gulf" types may not be universally appropriate at the nuclear genome scale. Indeed, this population classification for the oyster contact zone must be reconsidered for contemporary population structure in light of the complex substructure observed among and within locus classes, but is informative regarding the vicariant history of the zone and its clinal selection. The selection-based mosaicism raises questions about the various scales on which environmental divergent selection may be acting and how these interact with local gene flow patterns, particularly south of Cape Canaveral. Selection likely is occurring on a smaller scale than the dispersal potential of planktonic oyster larvae because divergent

selection on local and regional scales maintains differentiation at some loci despite gene flow over the remainder of the genome. The identification of adaptively distinct genetic groups made here will be important in formulating conservation strategies for this economically important keystone species.

**Tables**

Table 2-1: Adult transect sampling scheme.

| Site Abbreviation | Number of Individuals | Year Collected | Location | Latitude (°N) | Longitude (°W) | Distance from WHI (km) |
|---|---|---|---|---|---|---|
| WHI | 46 | 2004 | Whitney Lab Dock, Marineland | 29.670 | 81.216 | 0.0 |
| NSB | 30 | 2002 | New Smyrna Beach | 29.080 | 80.956 | 71.6 |
| TTV | 32 | 2006 | Titusville | 28.618 | 80.803 | 131.1 |
| LPA | 45 | 2004 | Launch Pad A, Cape Canaveral | 28.437 | 80.582 | 222.6 |
| FTP | 45 | 2004 | Fort Pierce | 27.491 | 80.315 | 266.7 |
| HOB | 46 | 2004 | Hobe Sound | 27.100 | 80.141 | 314.1 |
| PCH | 30 | 2002 | Port Charlotte | 26.429 | 81.867 | 779.0 |
| **Total** | **274** | | | | | |

Table 2-2: Error rates by primer pair for each level of data refinement. Primer pair corresponds to the selective amplification for the MseI primer, all of which were paired with a selective EcoRI primer extension of ACT. Number of replicates equals the number of DNA samples used in replicate analyses for error rates. Average locus error rate equals the mean error rate per locus. Data-wide error rate equals the error rate based on summed mismatches across loci. All loci corresponds to the 388 loci that were scored before culling error-prone loci. Of these, the 291 retained loci met the locus error threshold. After removing loci with high similarity using the pairwise linkage index, 217 final loci remained.

| Primer pair | # replicates | Average locus error rates | | | Data-wide error rates | | |
|---|---|---|---|---|---|---|---|
| | | All loci | 291 retained loci | 217 final loci | All loci | 291 retained loci | 217 final loci |
| CAA | 30 | 6.36% | 2.63% | 4.49% | 6.34% | 2.63% | 4.49% |
| CAC | 29 | 8.16% | 3.64% | 3.65% | 7.99% | 3.61% | 3.62% |
| CAG | 31 | 6.80% | 3.62% | 4.20% | 7.41% | 3.59% | 4.15% |
| CAT | 91 | 4.56% | 3.55% | 3.88% | 4.52% | 3.52% | 3.86% |
| **All** | **181** | **6.53%** | **3.26%** | **4.03%** | **5.96%** | **3.33%** | **3.96%** |

Table 2-3: Number and percent of loci in each locus class by population pair. Population pairs are categorized by whether the included populations are on the 'same' or 'different' sides of the previously documented step cline in northern Cape Canaveral. Results from the 0.05 and 0.01 α levels are reported as the values adjacent to and inside parentheses, respectively. Overall number and percent of loci in each locus class are the total unique loci summed across all population comparisons. Multiple independent (MI) outliers are outliers detected in two or more independent population comparisons. Single and multiple dependent (SD) outliers are outliers detected in only one population comparison or multiple population comparisons with one population in common.

| Cline side | Population 1 | Population 2 | Total # loci | # Non-outliers | # Outliers | # MI outliers | # SD outliers |
|---|---|---|---|---|---|---|---|
| Same | NSB | WHI | 213 | 193 (210) | 20 (3) | 10 (2) | 10 (1) |
| Same | LPA | TTV | 208 | 193 (202) | 15 (6) | 6 (2) | 9 (4) |
| Same | FTP | TTV | 216 | 203 (211) | 13 (5) | 6 (1) | 7 (4) |
| Same | HOB | TTV | 204 | 195 (199) | 9 (5) | 4 (2) | 5 (3) |
| Same | PCH | TTV | 213 | 204 (211) | 9 (2) | 5 (1) | 4 (1) |
| Same | FTP | LPA | 212 | 183 (191) | 29 (21) | 19 (7) | 10 (14) |
| Same | LPA | HOB | 193 | 192 (193) | 1 (0) | 0 (0) | 1 (0) |
| Same | LPA | PCH | 211 | 194 (203) | 17 (8) | 12 (5) | 5 (3) |
| Same | FTP | HOB | 212 | 186 (198) | 26 (14) | 18 (6) | 8 (8) |
| Same | FTP | PCH | 214 | 196 (207) | 18 (7) | 11 (4) | 7 (3) |
| Same | PCH | HOB | 209 | 190 (203) | 19 (6) | 10 (4) | 9 (2) |
| **Average** | | | **209.5** | **193.5 (202.5)** | **16.0 (7.0)** | **9.2 (3.1)** | **6.8 (3.9)** |
| **Std. Dev.** | | | **6.4** | **6.2 (6.9)** | **7.9 (5.9)** | **5.8 (2.3)** | **2.8 (4.0)** |
| Different | TTV | WHI | 216 | 204 (213) | 12 (3) | 8 (1) | 4 (2) |
| Different | LPA | WHI | 212 | 202 (208) | 10 (4) | 9 (4) | 1 (0) |
| Different | FTP | WHI | 212 | 203 (208) | 9 (4) | 7 (3) | 2 (1) |
| Different | HOB | WHI | 214 | 203 (210) | 11 (4) | 10 (4) | 1 (0) |
| Different | PCH | WHI | 213 | 195 (205) | 18 (8) | 14 (5) | 4 (3) |
| Different | NSB | TTV | 213 | 202 (209) | 11 (4) | 8 (3) | 3 (1) |
| Different | LPA | NSB | 211 | 200 (203) | 11 (8) | 11 (8) | 0 (0) |
| Different | FTP | NSB | 211 | 195 (205) | 16 (6) | 11 (6) | 5 (0) |
| Different | NSB | HOB | 211 | 196 (201) | 15 (10) | 13 (9) | 2 (1) |
| Different | NSB | PCH | 212 | 201 (203) | 11 (9) | 9 (6) | 2 (3) |
| **Average** | | | **212.5** | **200.1 (206.5)** | **12.4 (6.0)** | **10.0 (4.9)** | **2.4 (1.1)** |
| **Std. Dev.** | | | **1.6** | **3.5 (3.7)** | **2.9 (2.5)** | **2.3 (2.4)** | **1.6 (1.2)** |
| **Overall** | | | **217** | **132 (168)** | **85 (49)** | **33 (16)** | **52 (33)** |

128

Table 2-3 continued.

| Cline side | Population 1 | Population 2 | Total # loci | % Non-outliers | % Outliers | % MI outliers | % SD outliers |
|---|---|---|---|---|---|---|---|
| Same | NSB | WHI | 213 | 90.6 (98.6) | 9.4 (1.4) | 4.7 (0.9) | 4.7 (0.5) |
| Same | LPA | TTV | 208 | 92.8 (97.1) | 7.2 (2.9) | 2.9 (1.0) | 4.3 (1.9) |
| Same | FTP | TTV | 216 | 94.0 (97.7) | 6.0 (2.3) | 2.8 (0.5) | 3.2 (1.9) |
| Same | HOB | TTV | 204 | 95.6 (97.5) | 4.4 (2.5) | 2.0 (1.0) | 2.5 (1.5) |
| Same | PCH | TTV | 213 | 95.8 (99.1) | 4.2 (0.9) | 2.3 (0.5) | 1.9 (0.5) |
| Same | FTP | LPA | 212 | 86.3 (90.1) | 13.7 (9.9) | 9.0 (3.3) | 4.7 (6.6) |
| Same | LPA | HOB | 193 | 99.5 (100.0) | 0.5 (0.0) | 0.0 (0.0) | 0.5 (0.0) |
| Same | LPA | PCH | 211 | 91.9 (96.2) | 8.1 (3.8) | 5.7 (2.4) | 2.4 (1.4) |
| Same | FTP | HOB | 212 | 87.7 (93.4) | 12.3 (6.6) | 8.5 (2.8) | 3.8 (3.8) |
| Same | FTP | PCH | 214 | 91.6 (96.7) | 8.4 (3.3) | 5.1 (1.9) | 3.3 (1.4) |
| Same | PCH | HOB | 209 | 90.9 (97.1) | 9.1 (2.9) | 4.8 (1.9) | 4.3 (1.0) |
| **Average** | | | **209.5** | **92.4 (96.7)** | **7.6 (3.3)** | **4.3 (1.5)** | **3.2 (1.8)** |
| **Std. Dev.** | | | **6.4** | **3.7 (2.8)** | **3.7 (2.8)** | **2.7 (1.1)** | **1.3 (1.9)** |
| | | | | | | | |
| Different | TTV | WHI | 216 | 94.4 (98.6) | 5.6 (1.4) | 3.7 (0.5) | 1.9 (0.9) |
| Different | LPA | WHI | 212 | 95.3 (98.1) | 4.7 (1.9) | 4.2 (1.9) | 0.5 (0.0) |
| Different | FTP | WHI | 212 | 95.8 (98.1) | 4.2 (1.9) | 3.3 (1.4) | 0.9 (0.5) |
| Different | HOB | WHI | 214 | 94.9 (98.1) | 5.1 (1.9) | 4.7 (1.9) | 0.5 (0.0) |
| Different | PCH | WHI | 213 | 91.5 (96.2) | 8.5 (3.8) | 6.6 (2.3) | 1.9 (1.4) |
| Different | NSB | TTV | 213 | 94.8 (98.1) | 5.2 (1.9) | 3.8 (1.4) | 1.4 (0.5) |
| Different | LPA | NSB | 211 | 94.8 (96.2) | 5.2 (3.8) | 5.2 (3.8) | 0.0 (0.0) |
| Different | FTP | NSB | 211 | 92.4 (97.2) | 7.6 (2.8) | 5.2 (2.8) | 2.4 (0.0) |
| Different | NSB | HOB | 211 | 92.9 (95.3) | 7.1 (4.7) | 6.2 (4.3) | 0.9 (0.5) |
| Different | NSB | PCH | 212 | 94.8 (95.8) | 5.2 (4.2) | 4.2 (2.8) | 0.9 (1.4) |
| **Average** | | | **212.5** | **94.2 (97.2)** | **5.8 (2.8)** | **4.7 (2.3)** | **1.1 (0.5)** |
| **Std. Dev.** | | | **1.6** | **1.4 (1.2)** | **1.4 (1.2)** | **1.1 (1.2)** | **0.7 (0.6)** |
| | | | | | | | |
| **Overall** | | | **217** | **60.8 (77.4)** | **39.2 (22.6)** | **15.2 (7.4)** | **24.0 (15.2)** |

Table 2-4: Pairwise $F_{ST}$ for all locus classes. Same cline side describes population pairs on the same side of the previously described step cline at northern Cape Canaveral. Different cline side describes population pairs on opposite sides of the step cline. Distance is between populations 1 and 2 in km. $F_{ST}$ values for each population pair are provided for each locus class: all 217 loci, non-outliers, multiple independent (MI) outliers, and single and multiple dependent (SD) outliers. The average and standard deviation (Std. Dev.) distances and $F_{ST}$ values are for each population pair type and $F_{ST}$ for the overall data set, which consisted of all seven populations.

| Cline Side | Population 1 | Population 2 | Distance (km) | All 217 | Non-outliers | MI outliers | SD outliers |
|---|---|---|---|---|---|---|---|
| Same | NSB | WHI | 71.6 | 0.0139 | 0.0000 | 0.0426 | 0.0275 |
| Same | LPA | TTV | 91.4 | 0.0902 | 0.0479 | 0.1482 | 0.1461 |
| Same | FTP | TTV | 135.6 | 0.1040 | 0.0428 | 0.1637 | 0.1751 |
| Same | HOB | TTV | 183.0 | 0.0834 | 0.0424 | 0.1318 | 0.1404 |
| Same | PCH | TTV | 647.8 | 0.0973 | 0.0512 | 0.1632 | 0.1523 |
| Same | FTP | LPA | 44.1 | 0.0621 | 0.0030 | 0.1951 | 0.0694 |
| Same | LPA | HOB | 91.5 | 0.0000 | 0.0000 | 0.0000 | 0.0035 |
| Same | LPA | PCH | 556.4 | 0.0495 | 0.0124 | 0.1649 | 0.0524 |
| Same | FTP | HOB | 47.4 | 0.0542 | 0.0062 | 0.1685 | 0.0534 |
| Same | FTP | PCH | 512.3 | 0.0506 | 0.0137 | 0.1382 | 0.0576 |
| Same | PCH | HOB | 464.9 | 0.0506 | 0.0174 | 0.1496 | 0.0557 |
| **Same** | **Average** | | **258.7** | **0.0596** | **0.0215** | **0.1333** | **0.0849** |
| **Same** | **Std. Dev.** | | **234.4** | **0.0329** | **0.0203** | **0.0586** | **0.0577** |
| Different | TTV | WHI | 131.1 | 0.1291 | 0.0551 | 0.2303 | 0.1843 |
| Different | LPA | WHI | 222.6 | 0.0997 | 0.0299 | 0.2733 | 0.0794 |
| Different | FTP | WHI | 266.7 | 0.0297 | 0.0115 | 0.0649 | 0.0356 |
| Different | HOB | WHI | 314.1 | 0.0946 | 0.0364 | 0.2562 | 0.0583 |
| Different | PCH | WHI | 779.0 | 0.1032 | 0.0319 | 0.2717 | 0.0958 |
| Different | NSB | TTV | 59.5 | 0.1294 | 0.0523 | 0.2607 | 0.1712 |
| Different | LPA | NSB | 151.0 | 0.0945 | 0.0299 | 0.2679 | 0.0800 |
| Different | FTP | NSB | 195.1 | 0.0397 | 0.0125 | 0.1142 | 0.0300 |
| Different | NSB | HOB | 242.5 | 0.0898 | 0.0311 | 0.2585 | 0.0626 |
| Different | NSB | PCH | 707.4 | 0.0940 | 0.0263 | 0.2698 | 0.0880 |
| **Different** | **Average** | | **306.9** | **0.0904** | **0.0317** | **0.2268** | **0.0885** |
| **Different** | **Std. Dev.** | | **241.5** | **0.0326** | **0.0142** | **0.0743** | **0.0517** |
| **Overall** | | | | **0.0756** | **0.0263** | **0.1841** | **0.0912** |

Table 2-5: AMOVA results for each locus class by hierarchical model. The two-region model corresponds to Atlantic and Gulf parapatric patterns. The three-region model corresponds to the geographic pattern expected by the bounded hybrid zone hypothesis. The K=4 model corresponds to cluster delineations prescribed by CLUMPP for K=4 clusters.

| | | AMOVA Phi-statistics | | | Percent variance explained | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | Locus Class | PhiRT | PhiPR | PhiPT | Among Regions | Among Pops | Within Pops |
| **Two Region Model** | 217 Loci | 0.065 | 0.087 | 0.147 | 7% | 8% | 85% |
| | Non-outliers | 0.032 | 0.027 | 0.058 | 3% | 3% | 94% |
| | MI outliers | 0.177 | 0.212 | 0.351 | 18% | 17% | 65% |
| | SD outliers | 0.025 | 0.120 | 0.142 | 2% | 12% | 86% |
| **Three Region Model** | 217 Loci | 0.041 | 0.088 | 0.125 | 4% | 8% | 88% |
| | Non-outliers | 0.017 | 0.029 | 0.045 | 2% | 3% | 95% |
| | MI outliers | 0.120 | 0.210 | 0.305 | 12% | 18% | 70% |
| | SD outliers | 0.018 | 0.118 | 0.134 | 2% | 12% | 87% |
| **K=4 Model** | 217 Loci | 0.107 | 0.033 | 0.137 | 11% | 3% | 86% |
| | Non-outliers | 0.041 | 0.009 | 0.050 | 4% | 1% | 95% |
| | MI outliers | 0.260 | 0.087 | 0.324 | 26% | 6% | 68% |
| | SD outliers | 0.109 | 0.047 | 0.151 | 11% | 4% | 85% |

**Figures**
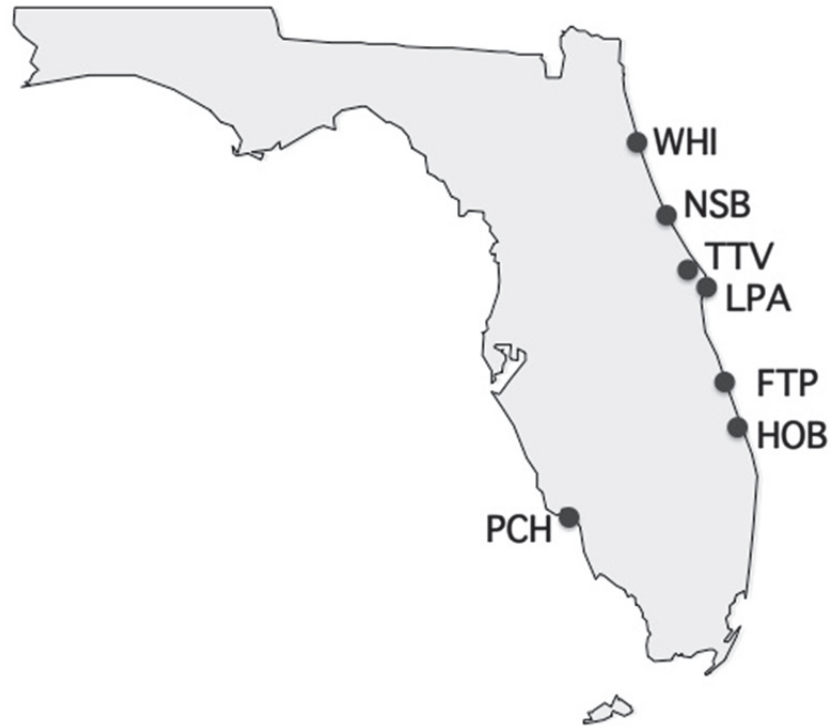


Fig. 2-1:  Map of adult transect sites.  Site abbreviations are as follows: WHI, Whitney Lab Dock, Marineland; NSB, New Smyrna Beach; TTV, Titusville; LPA, Launch Pad A, Cape Canaveral; FTP, Fort Pierce; HOB, Hobe Sound; PCH, Port Charlotte.
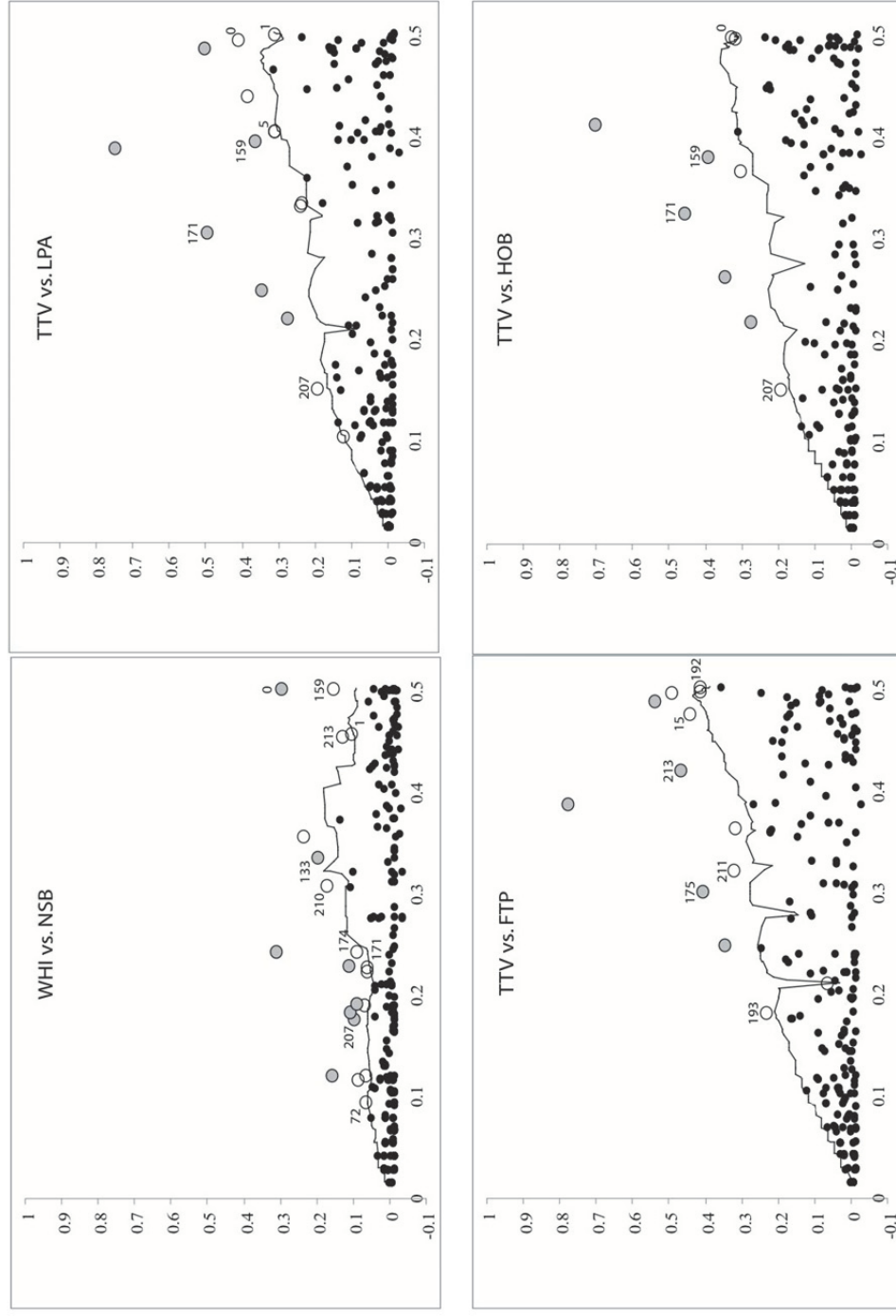
Fig. 2-2: Plots of Dfdist results for each pairwise population comparison with heterozygosity and $F_{ST}$ as the x and y axes, respectively. The line depicts the 97.5 percentile threshold of the simulated neutral distribution of $F_{ST}$. Solid black circles are empirical non-outlier loci. Open white circles are empirical outliers at the $\alpha = 0.05$ level, but not at the $\alpha = 0.01$ level. Gray circles are empirical outliers at the $\alpha = 0.01$ level. Numbered circles correspond to the locus number for loci in the multiple independent (MI) outlier class. (A) The population comparisons comprised of populations on the same side of the cline. Comparisons are ordered by increasing mean distance from WHI.

133

Fig. 2-2.A continued.

134

Fig. 2-2.A continued.

135
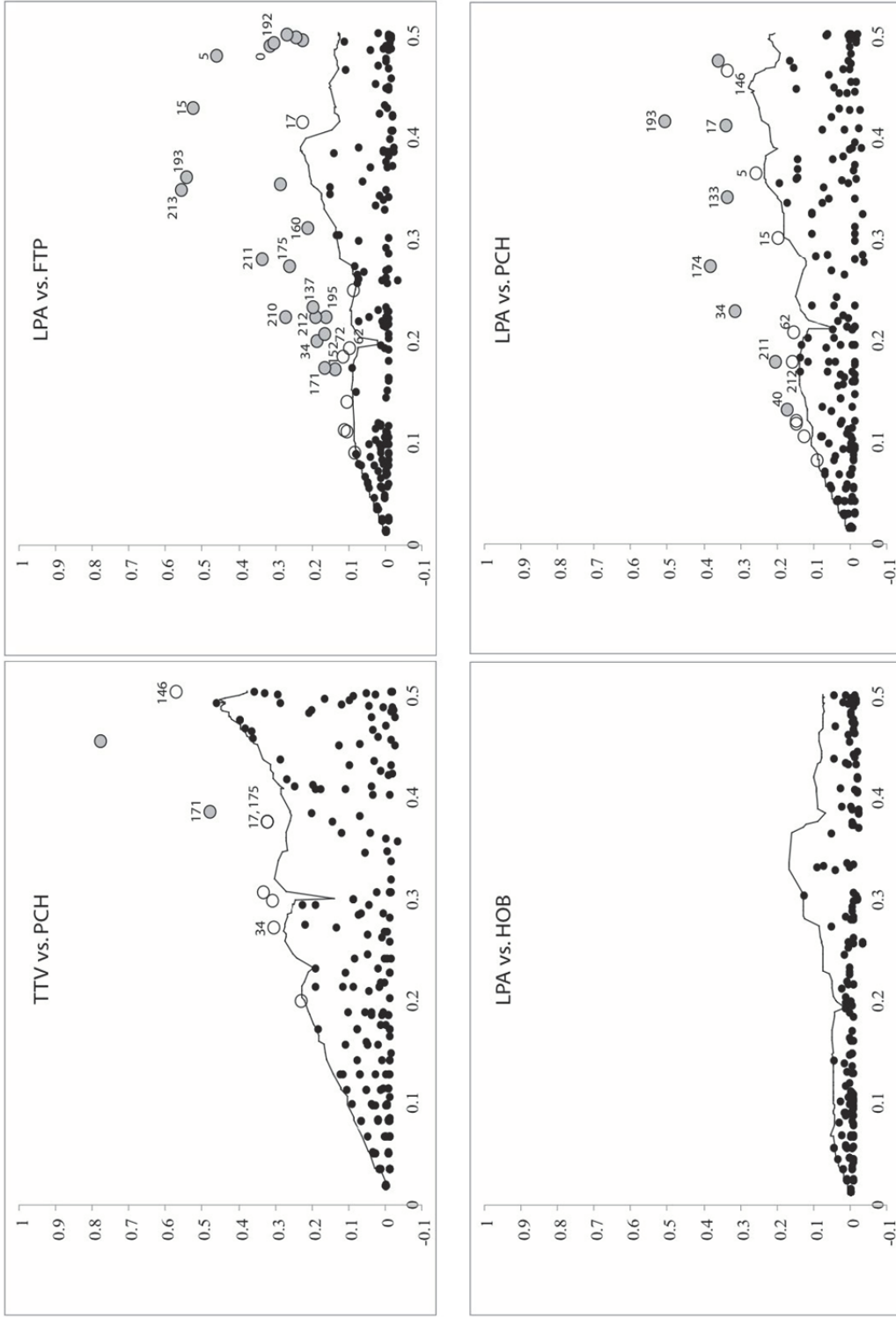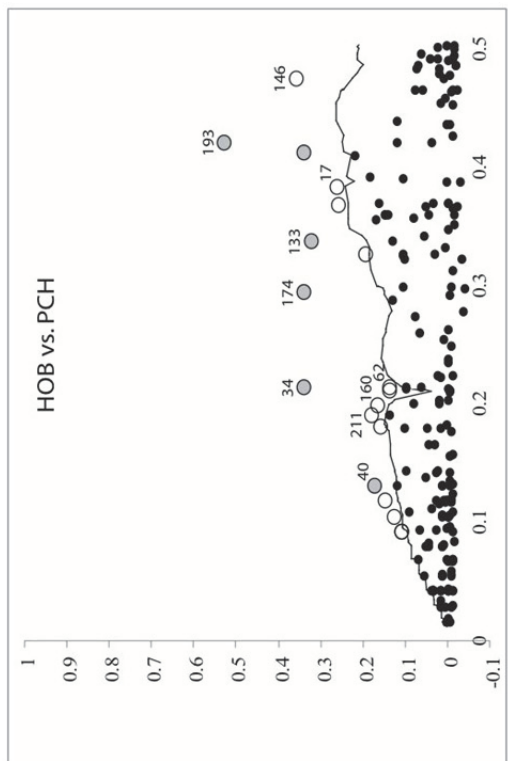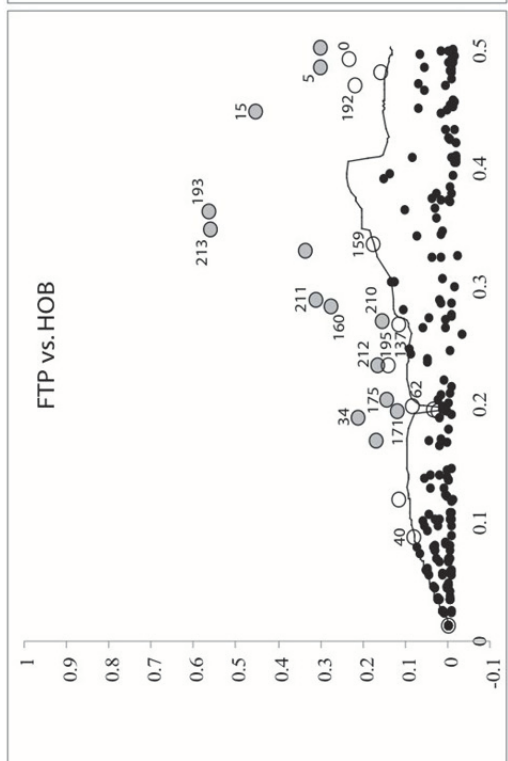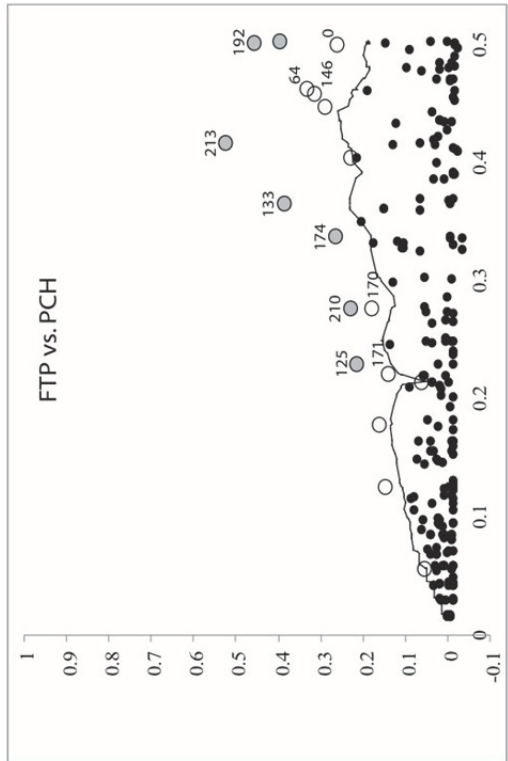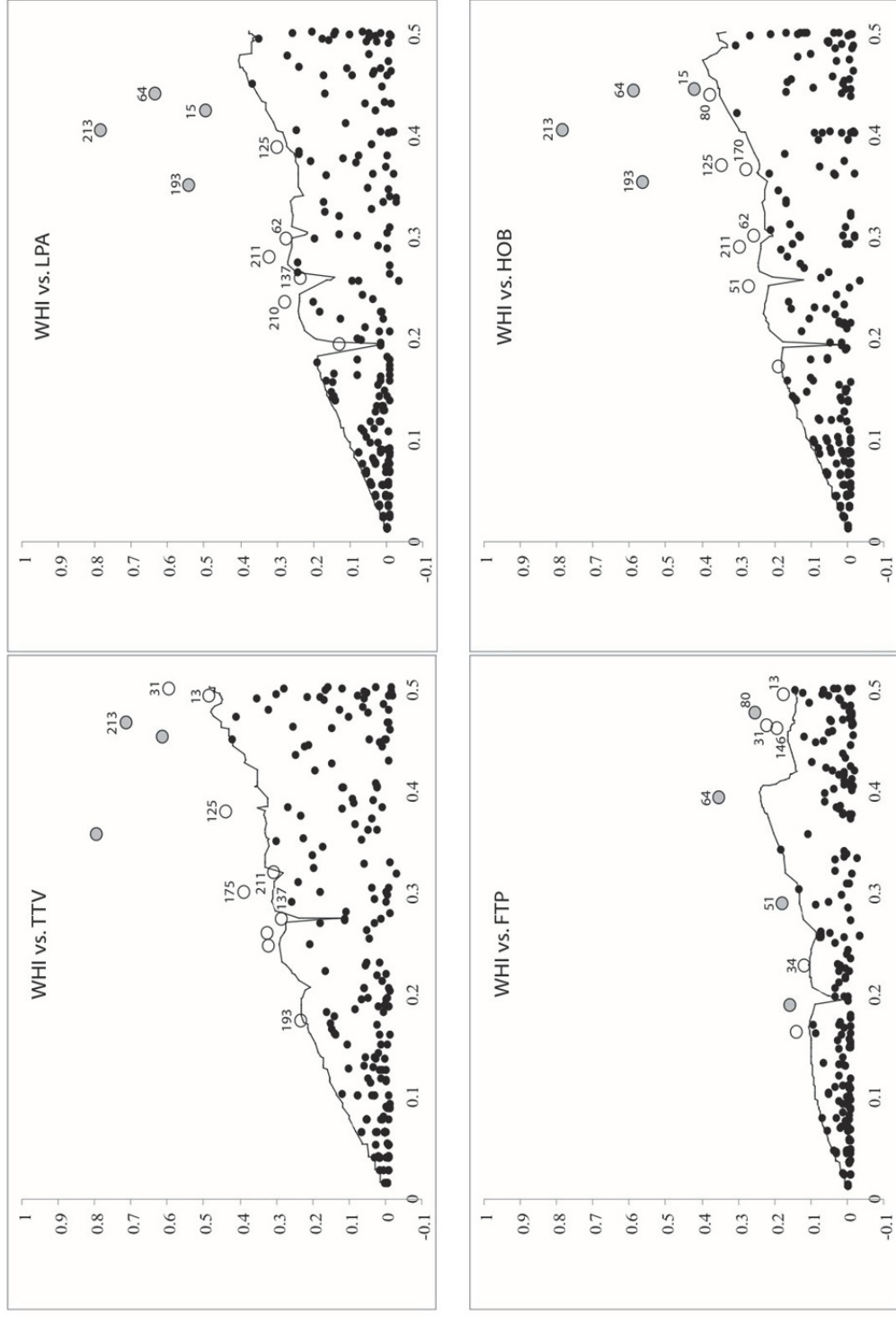
Fig. 2-2.B: The population comparisons comprised of populations on different sides of the cline. Comparisons are ordered by decreasing mean distance from WHI.

136

Fig. 2-2.B continued.

137

Fig. 2-2.B continued.

138
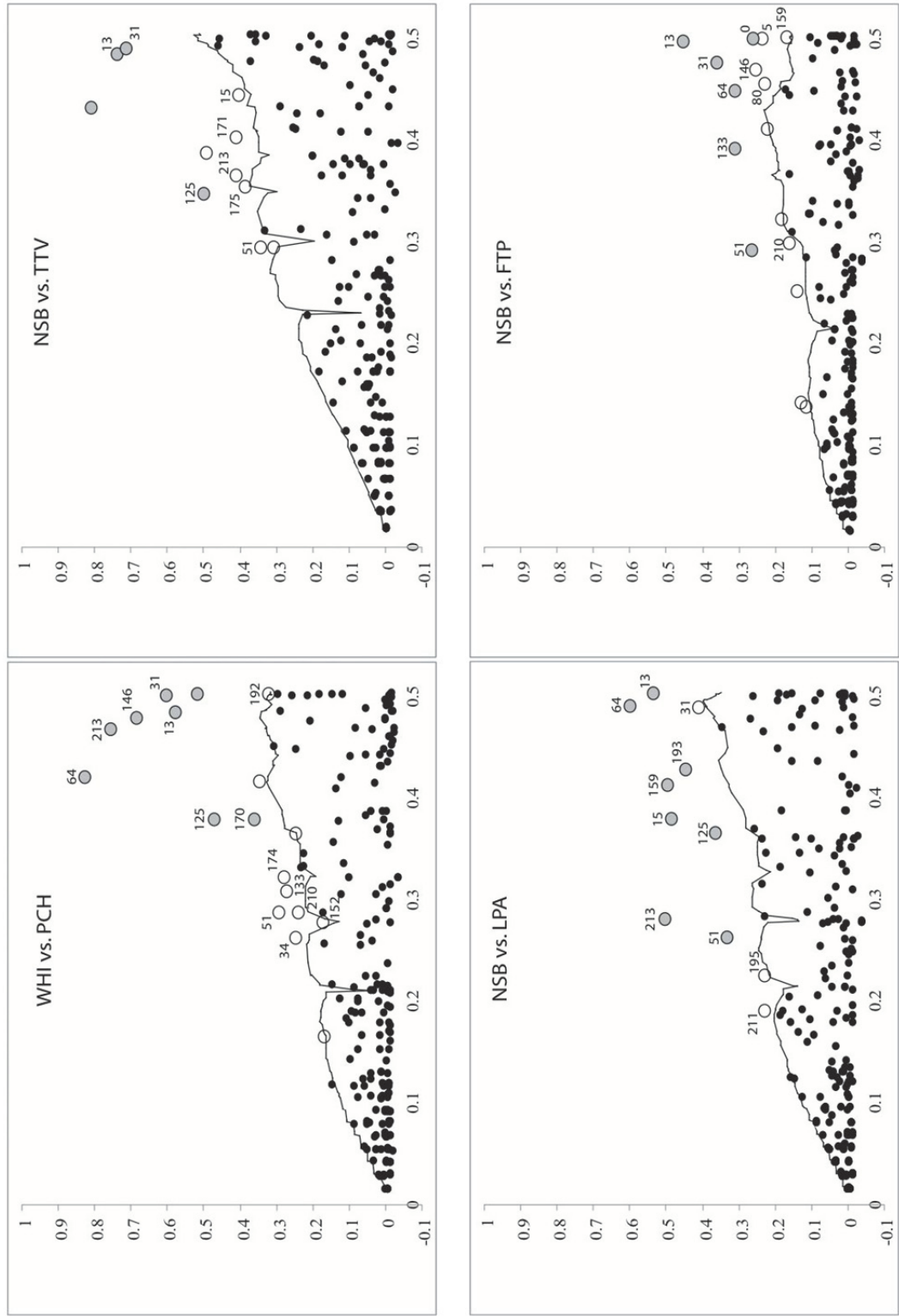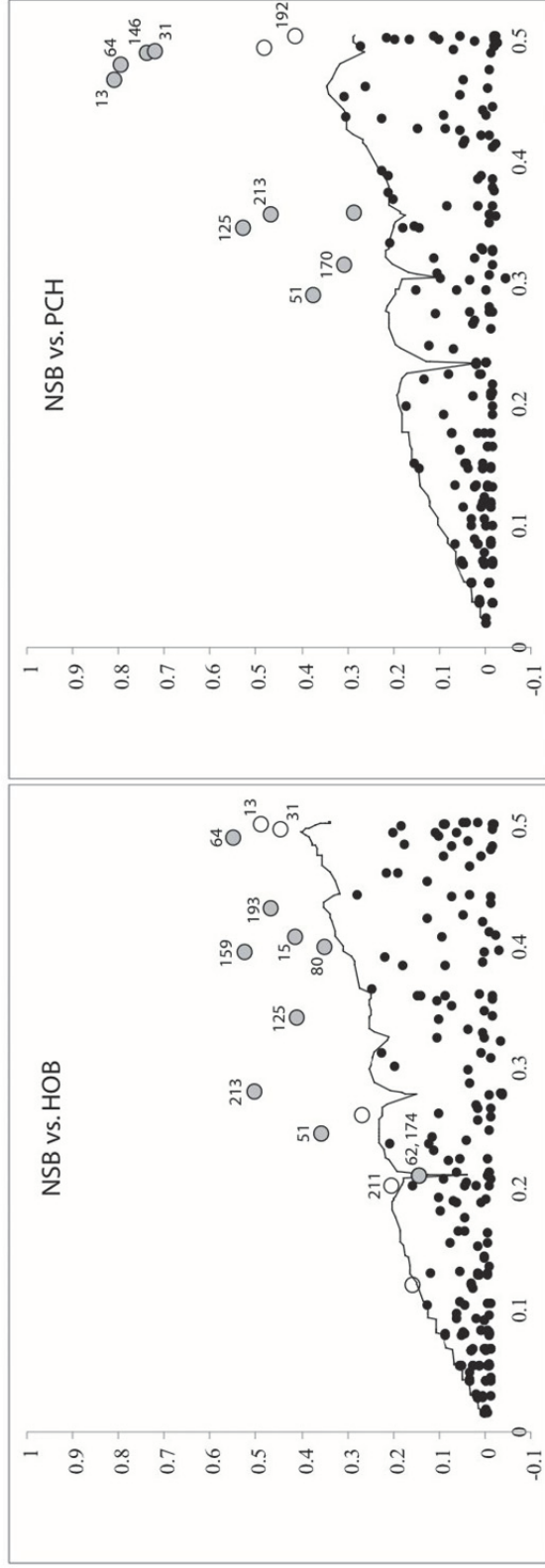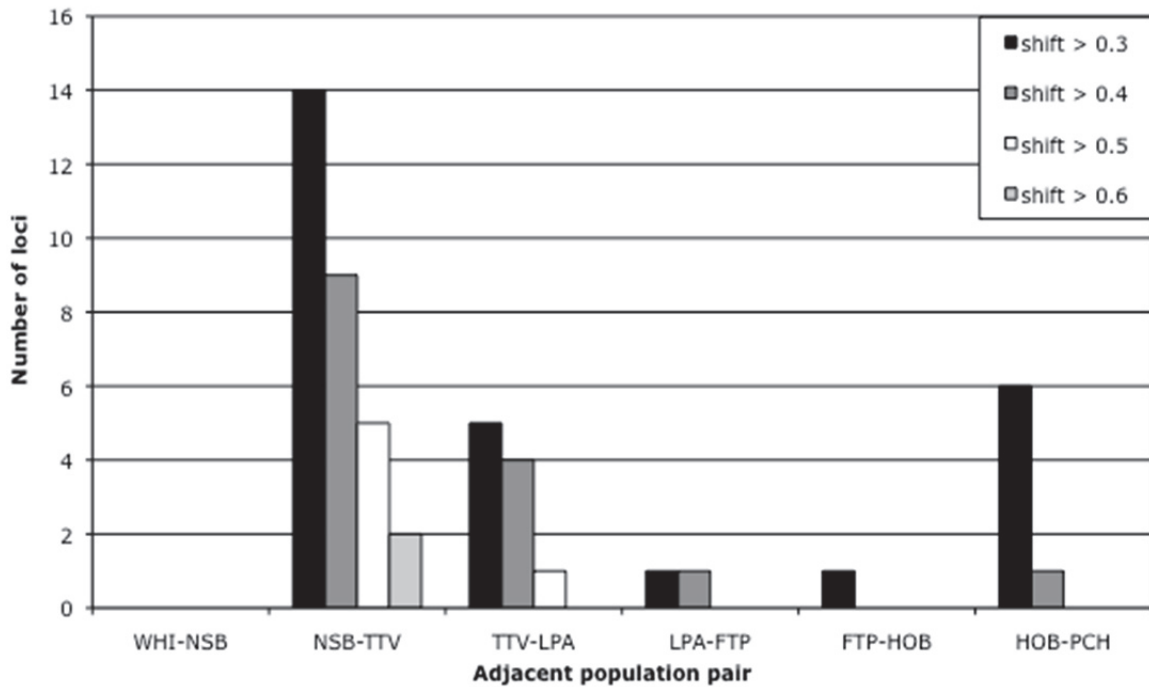
Fig. 2-3: Number of loci displaying a single allele frequency shift over adjacent population pairs. Loci counted when the allele frequency differences between adjacent population pairs were greater than 0.3, 0.4, 0.5, and 0.6, respectively. Note: the documented step cline occurs between NSB and TTV (Hare & Avise 1996).
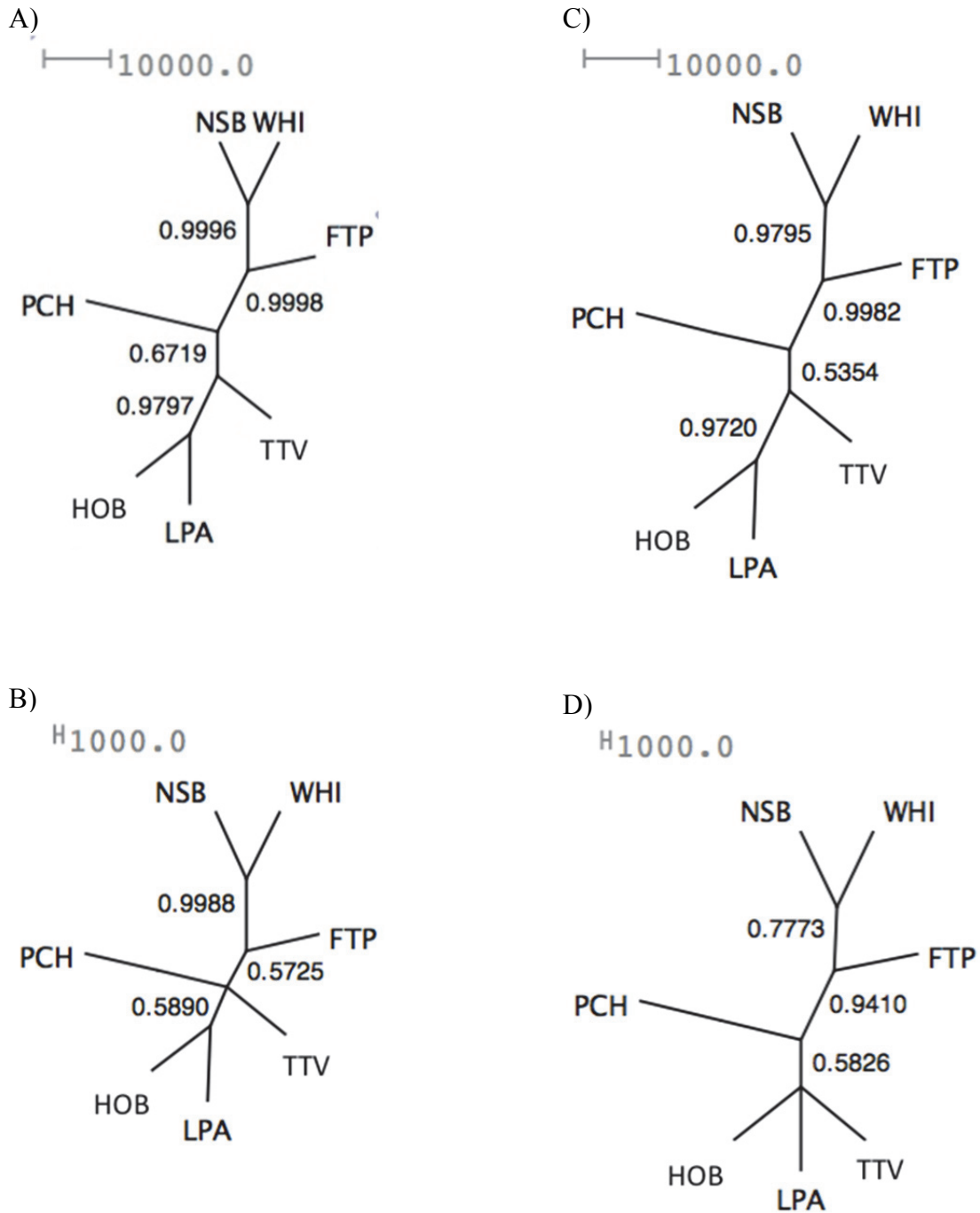
Fig. 2-4: Neighbor-joining consensus trees for each locus class using 50% majority rule. Bootstrap values above 0.5 are indicated at internodes. (A) All 217 loci. (B) Non-outlier loci at the 0.05 α threshold. (C) Outliers in multiple independent population comparisons (MI) at the 0.05 α threshold. (D) Outliers in single or multiple dependent population comparisons (SD) at the 0.05 α threshold.
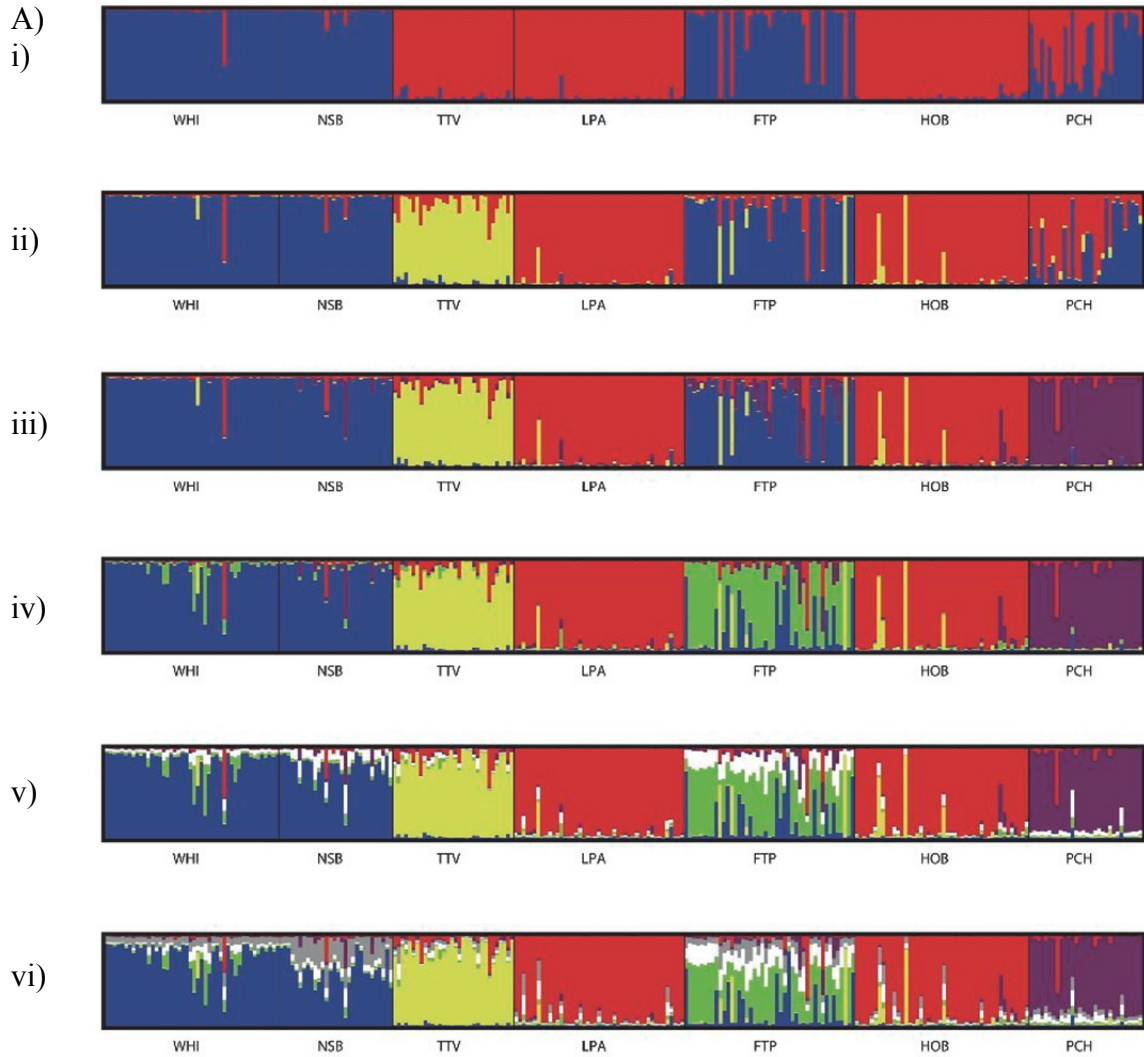
Fig. 2-5: Assignment tests using CLUMPP summary results for each class of loci at each K value greater than 1, where i – vi represent K=2, 3, 4, 5, 6, and 7, respectively. Individual bars represent individuals, which are grouped by collection site. Colors correspond to clusters, where multiple colors in a single individual represent proportions of mixed ancestry. (A) All 217 loci.
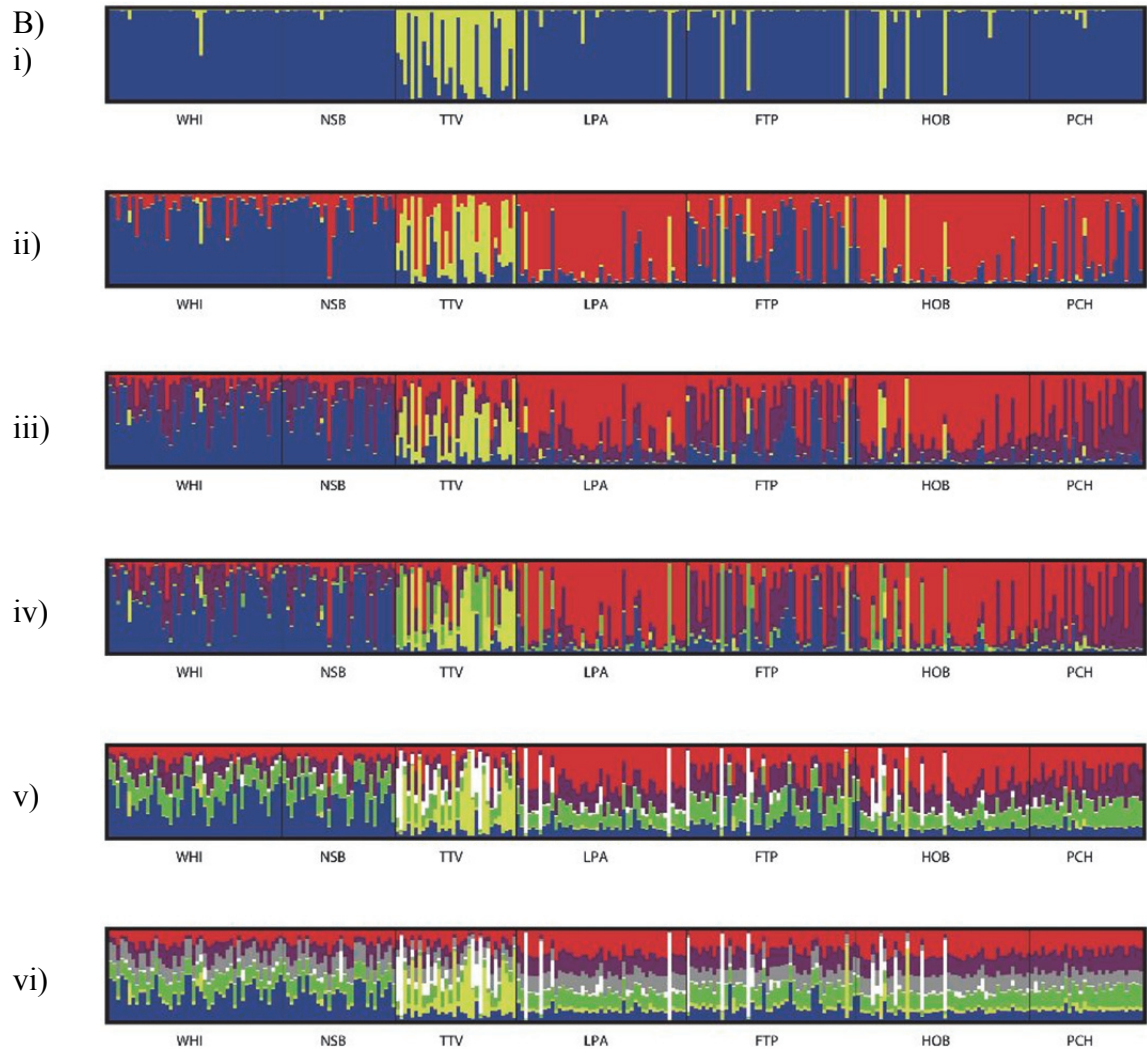
Fig. 2-5 continued.  (B) Non-outlier loci at the 0.05 α threshold.

142
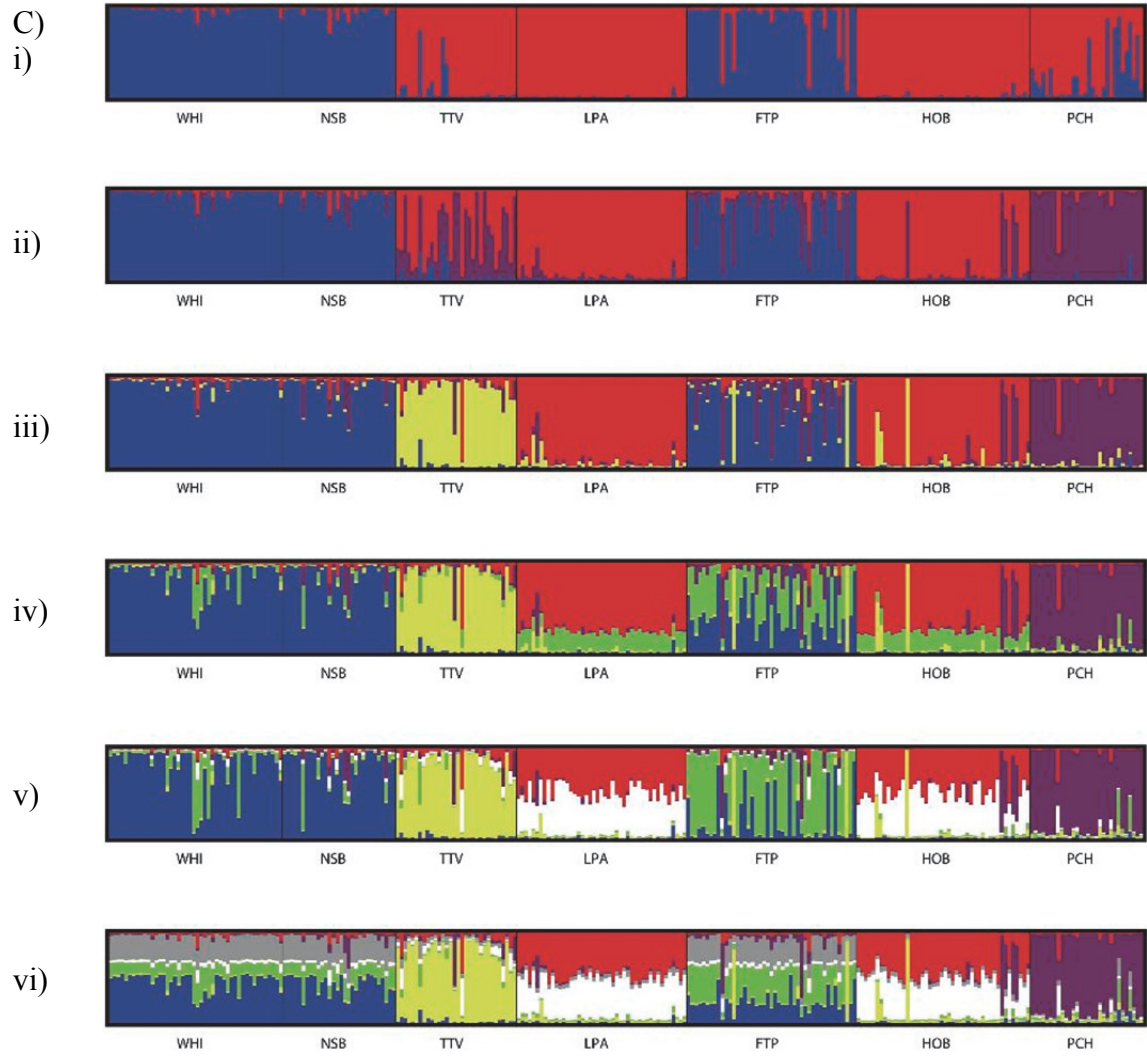
Fig. 2-5 continued. (C) Outliers in multiple independent population comparisons (MI) at the 0.05 α threshold.
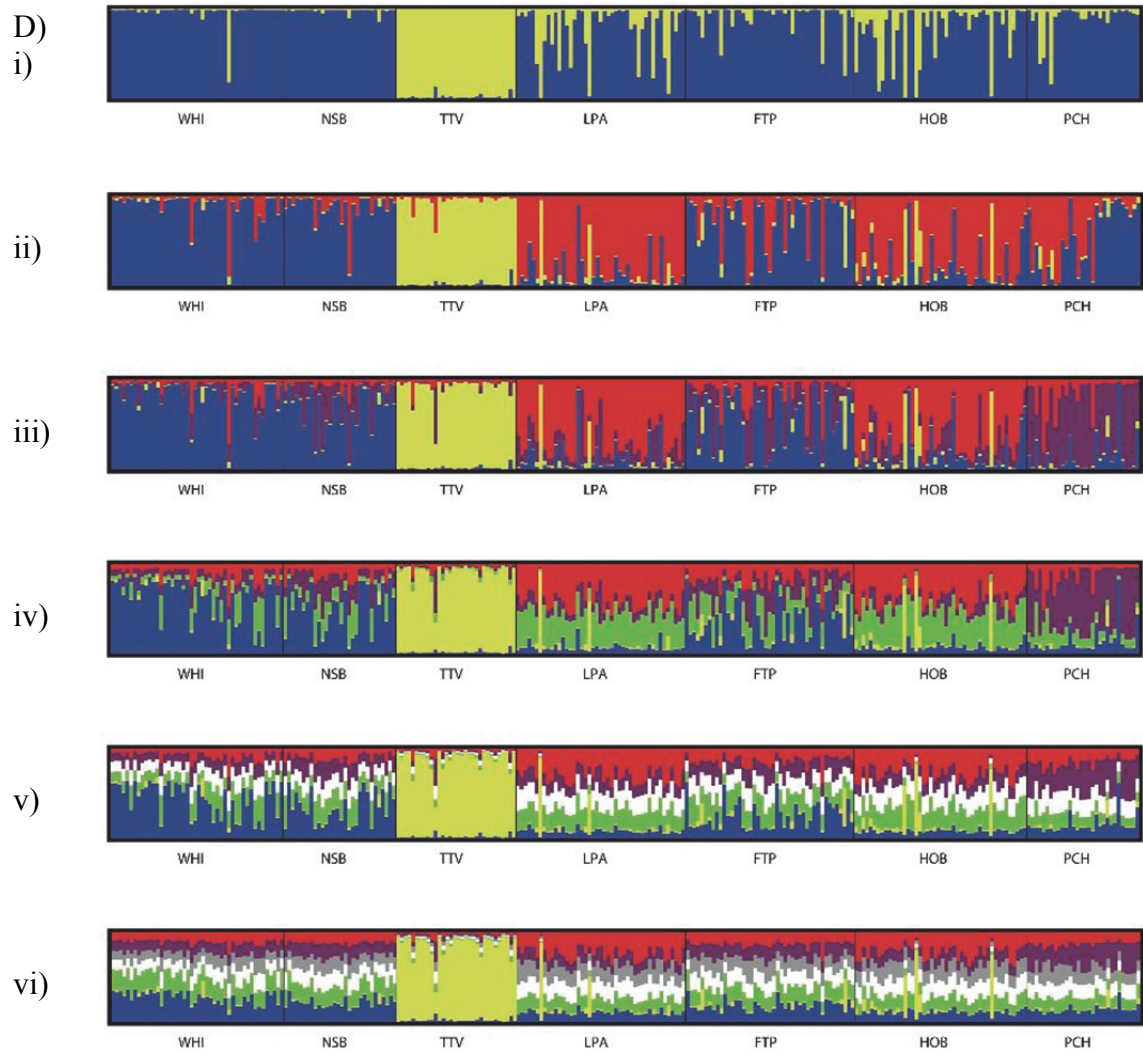
Fig. 2-5 continued. (D) Outliers in single and multiple dependent population comparisons (SD) at the 0.05 α threshold.
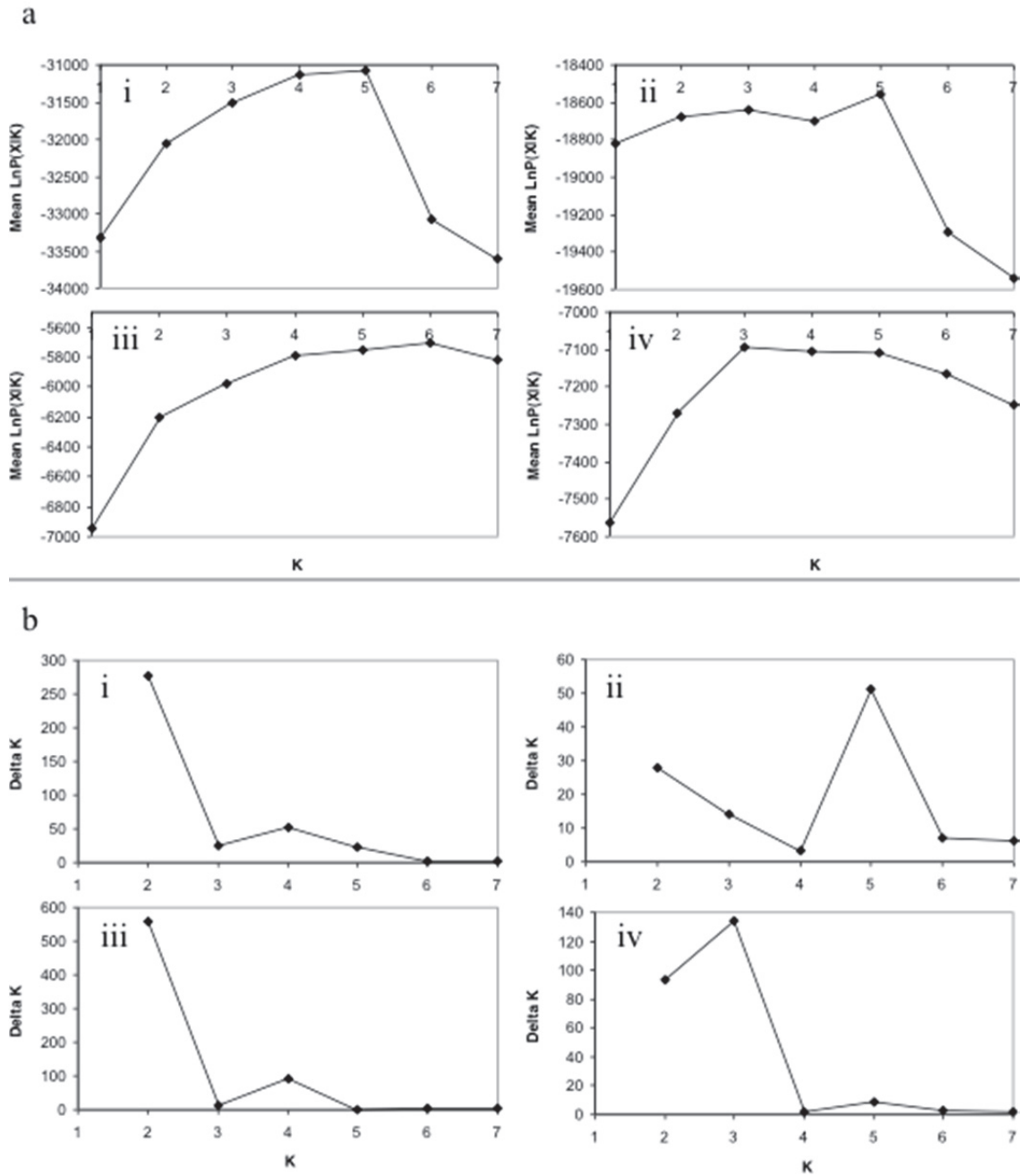
Fig. 2-6: Assessment of the optimal K values from STRUCTURE run data within each K. All 217 loci, non-outlier loci at the 97.5 threshold, outliers in multiple independent population comparisons (MI), and outliers in single and multiple dependent population comparisons (SD) are represented in i - iv, respectively. (A) Using the maximum log probability of the data method where the line begins to plateau (Pritchard et al. 2000). (B) Using the mode of the rate of change in log probability for consecutive K values (Evanno et al. 2005).
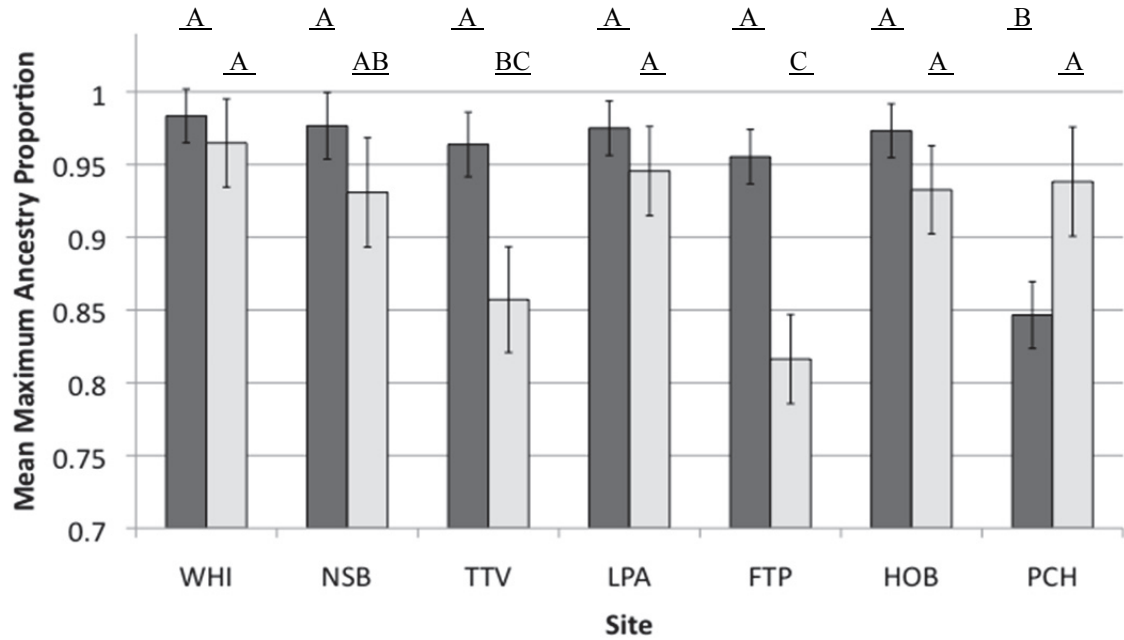
Fig. 2-7: Mean maximum ancestry proportion at each site for K=2 (dark gray) and K=4 (light gray) clusters. Error bars are for 1.96 times the standard error. Different letters over columns indicate significant difference after means comparisons for all pairs within a STRUCTURE analysis (K=2 or K=4) using a Tukey-Kramer HSD adjustment for multiple comparisons.
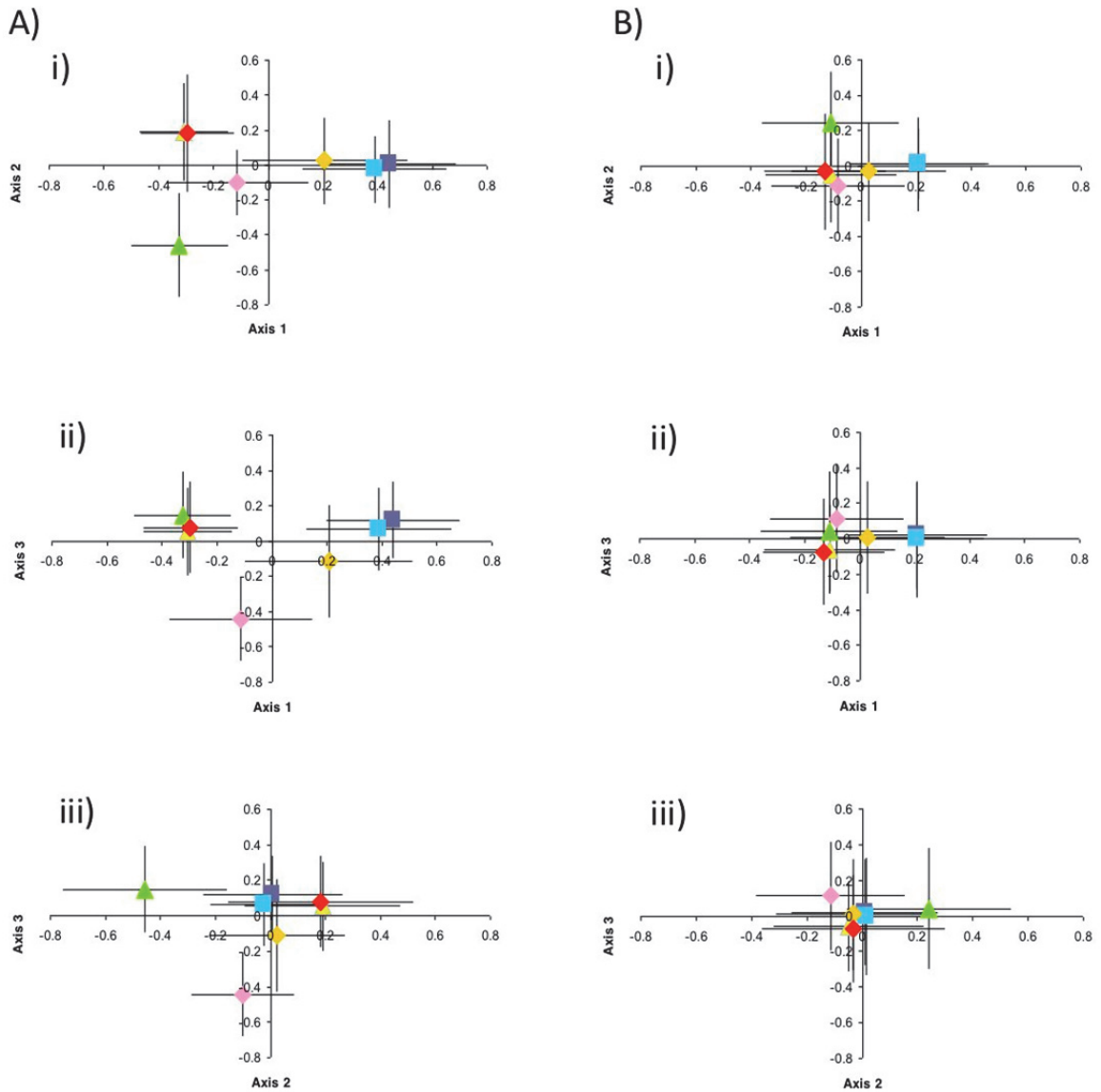
Fig. 2-8: Principal coordinate analysis results. For each site, the mean eigen value among individuals is shown by a dot, with error bars representing 1.96 times the standard error. Key: dark blue square = WHI; light blue square = NSB; green triangle = TTV; yellow triangle = LPA; orange circle = FTP; red circle = HOB; and pink diamond = PCH. (i) Axis 1 (x) versus Axis 2 (y). (ii) Axis 1 (x) versus Axis 3 (y). (iii) Axis 2 (x) versus Axis 3 (y). (A) All 217 loci. (B) Non-outlier loci at the 0.05 α threshold.
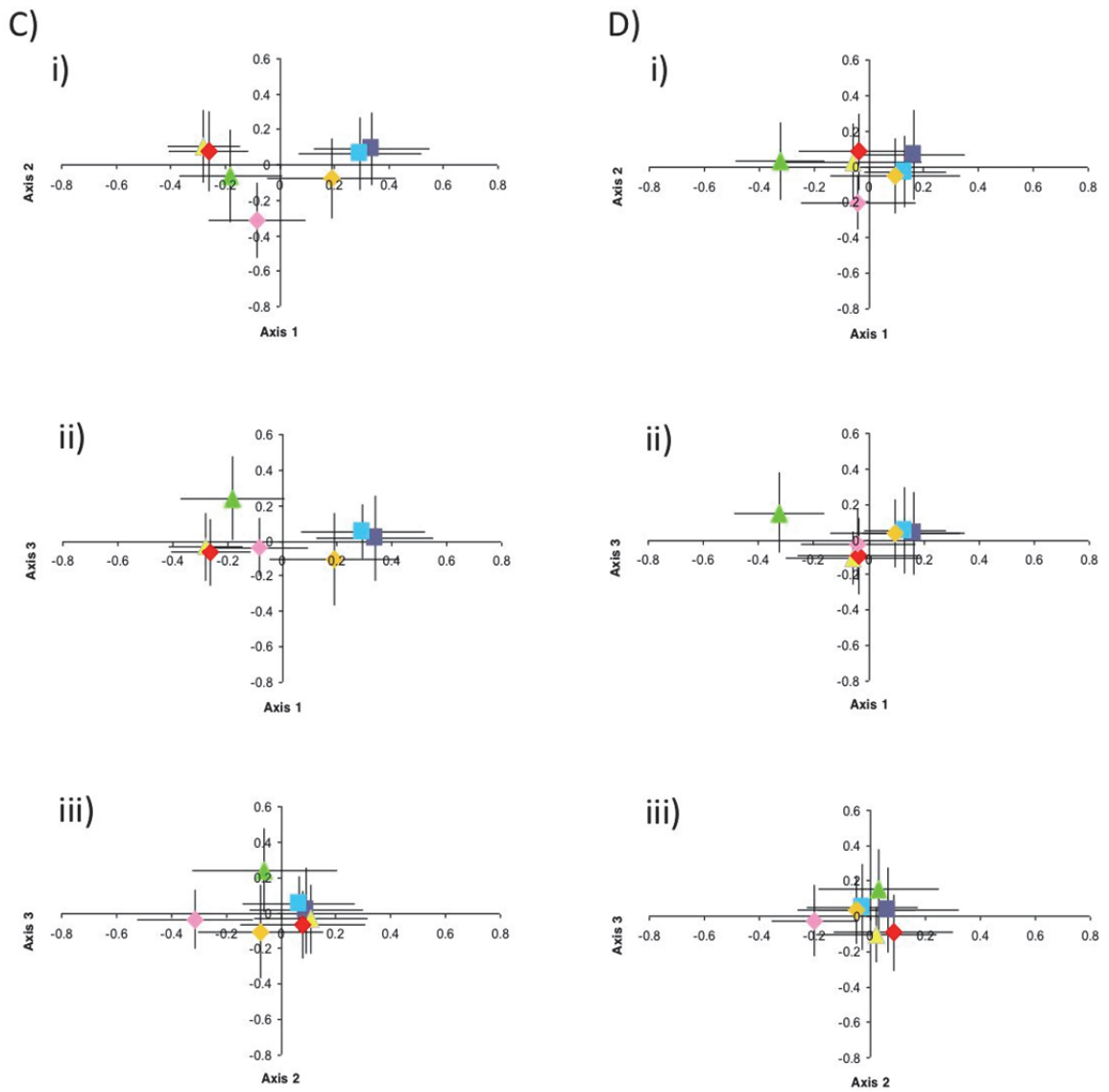
Fig. 2-8 continued. (C) Outliers in multiple independent population comparisons (MI). (D) Outliers in single and multiple dependent population comparisons (SD).
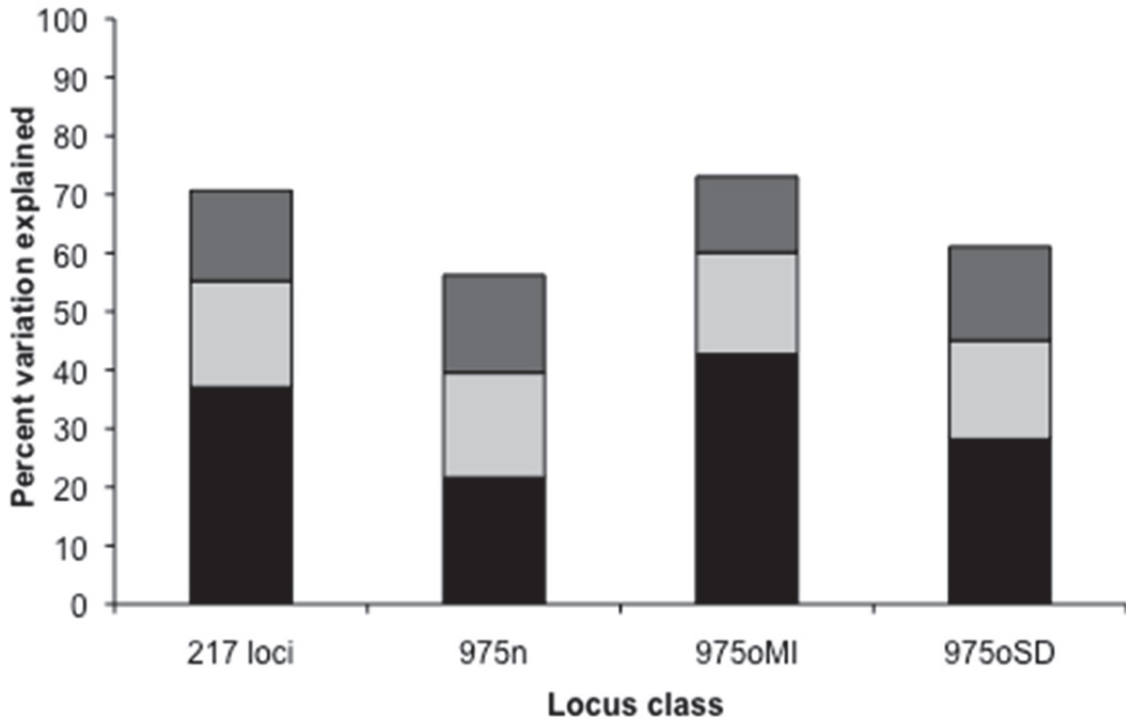
Fig. 2-9: Cumulative percentage of total variance explained by each axis of the principal coordinate analysis for each locus class. Black bars indicate the percent variance explained by axis 1. Light gray bars indicate the percent variance explained by axis 2. Dark gray bars indicate the percent explained by axis 3. Axes are not comparable among locus class data sets and do not necessarily describe the same aspects of the data signal. Rather, axes are labeled in order of the proportion of the variance within each locus class they each explain.

# CHAPTER 3: Exogenous Post-settlement Selection Against Non-native Genotypes in the Eastern Oyster Hybrid Zone

## Abstract

Hybrid zones exhibit variation in the mechanisms maintaining them. Examples of pre- and post-zygotic isolation barriers as well as of environment-independent and environment-dependent selection all can be found in nature and are not mutually exclusive. To assess the timing and mode of selection, it is necessary to determine whether selection acts exclusively against hybrids and to assess hybrid fitness relative to parental types across the hybrid zone and over the life cycle. The eastern Florida contact zone between Atlantic and Gulf populations of the eastern oyster (*Crassostrea virginica*) exhibits a sharp step cline at Cape Canaveral, despite the fact that a two to three week planktotropic larval phase gives this species a high potential for gene flow. Although the phylogeography and population genetics of this contact zone are well studied, it remains unknown whether the two populations hybridize and how selection acts to maintain this sharp differentiation.

In this study, I seek to determine the levels of hybridization in this contact zone, the timing of selection in the life cycle, and the nature of isolation barriers by conducting a static cohort analysis of wild juveniles (spat) and adults across the contact zone. Specifically, I first tested for the presence of a demographic barrier by measuring

recruitment across the reproductive season and along the contact zone. Using 217 AFLP loci, including seven candidate loci for differential selection between the two populations, I genotyped 1,011 spat over two seasons and 274 adults at sites along the contact zone. I tested for differences in the level of regional differentiation versus within-region differentiation among the life stages and examined how the genotypic distributions changed geographically and among life stages. I then tested for a difference in the proportions of hybrid-like and migrant-like genotypes across the two life stages and among three regions of the contact zone (North, Cape Canaveral, and South). Finally, I used genome scans to determine whether the same adult outlier loci were detectable as outliers in the spat stage. Results demonstrated: (1) lower recruitment and some mortality in the Cape Canaveral region; (2) significantly stronger regional genetic differentiation in adults than in spat; (3) distribution shifts from intermediate genotypes in spat to more parental genotypes in adults; (4) substantial intermediates present at the spat stage (~40%) dropped to 12% in adults, a similar reduction in migrant-like genotypes from spat (22-30%) to adult (3%) stages, and geographic differences in the pattern of these genotypic shifts over the life cycle; and (5) most regionally selected adult outlier loci were not repeated as outliers in spat. Overall, these results point to a major role for post-zygotic, environment-dependent selection in the maintenance of the hybrid zone between Atlantic and Gulf-type oyster populations.

# Introduction

Determining how selection operates to maintain differentiation in the face of persistent gene flow is important to deciphering the role adaptation plays in diversification. The broad suite of naturally occurring recombinants found in hybrid zones can be used to test how and when selection acts on different genotypes. A hybrid zone is defined as a region where genetically distinct groups meet, interact, and mate to produce at least some offspring of mixed ancestry (Harrison 1990; Arnold 1997). Such natural interbreeding between distinct groups occurs in a variety of taxa (Arnold 1997). In order for hybrid zones to be stable over time, they must involve some level of incomplete isolation. If selection is too weak or non-existent, persistent and free introgression will collapse the hybrid zone and result in the two groups becoming one unit. If isolation becomes very strong, separate species or subspecies may form. Understanding the nature of isolation barriers and balance of evolutionary forces in stable hybrid zones provides insight into the processes that generate and dissolve biological diversity when the balance is disrupted. In particular, hybrid zones are informative about certain types of speciation, especially for the selective divergence that occurs despite gene flow in ecological and parapatric speciation. Exploring how selection varies geographically, among genotypes, and over the life cycle in hybrid zones can illuminate mechanisms maintaining differentiation through a dynamic interplay between selection and gene flow.

**Evaluating Isolation Barriers in Hybrid Zones**

These isolation barriers in hybrid zones are the same classes of barriers that occur

between biological species not experiencing gene flow and can be endogenous or

exogenous (Dobzhansky 1951).  Endogenous barriers are based on reduced hybrid fitness

due to genetic incompatibilities between the hybridizing populations and act

independently of the environment (Barton & Hewitt 1985).  Barton and Hewitt termed

these hybrid zones under endogenous selection 'tension zones'.  Alternatively, the

environment fully drives exogenous barriers, whereby hybrids and immigrants experience

reduced fitness due to a mismatch between their phenotypes and their environment

(Endler 1977).  Implicit in exogenously maintained hybrid zones is that a geographic

ecological gradient produces corresponding geographic variation in fitness (Slatkin

1973).  Because hybrids, especially $F_1$ generation hybrids, often have intermediate

phenotypes, they potentially could succeed in an intermediate environment if one exists

(Moore 1977).  It is the exogenous barriers to genetic exchange that are uniquely relevant

to ecological speciation, where ecologically driven divergent selection causes isolation

barriers to evolve between populations (Coyne & Orr 2004; Rundle & Nosil 2005).

Distinguishing between endogenous and exogenous selection in contact zones is critical

to understand the role of adaptation in divergence.  When a hybrid zone is associated with

complex habitat structure or is geographically associated with climatic features, it is

common to invoke environmental gradient models (Rand & Harrison 1989; Moore &

Price 1993; Bert & Arnold 1995).  Alternatively, signs of developmental abnormalities,

reduced fertility or viability, and increased susceptibility to parasites or disease in hybrids

are often indicative of endogenous selection against hybrids (Barton & Hewitt 1989; Harrison 1990; Bert et al. 1993). In practice, these models are a challenge to distinguish because tension zone theory predicts that the hybrid zone will move until it reaches a barrier to gene flow (Barton & Hewitt 1985; Hewitt 1988), which can co-locate the zone with an ecotone (Wilhelm & Hilbish 1998). Therefore, it is not enough merely to show that a hybrid zone occurs in an area of environmental transition. One straightforward question that addresses a distinguishing feature of tension zone and ecological gradient models is whether selection varies spatially, because selection does not vary spatially within genotypes in tension zones but does in hybrid zones maintained by exogenous selection. However, endogenous and exogenous barriers to gene flow in hybrid zones are not mutually exclusive and have been observed to operate simultaneously within the same contact zone (Hewitt 1988; Bert & Arnold 1995; Nurnberger et al. 1995; Dorken & Pannell 2007).

Isolation barriers also can be classified depending on when they occur in the life cycle. Pre-mating barriers to gene flow include temporal, spatial, mechanical, and behavioral forms of isolation (Dobzhansky 1951). Post-mating barriers, on the other hand, involve gametic incompatibilities, local adaptation, and various types of reduced hybrid fitness, including inviability, sterility, and breakdown. Tracking the frequency of hybrids across the life cycle can clarify the relevance of pre- and post-zygotic barriers to the maintenance of a contact zone. If the frequency of hybrids in the populations is constant over the life cycle, then post-zygotic isolation is not important to the differentiation observed. Alternatively, if the frequency of hybrids reduces over the life cycle, then

selection likely acts against hybrids.  The presence of later generation hybrids points to at least some viability and fertility of $F_1$ hybrids.  If hybrids are less fit, mechanisms for reinforcement could evolve, whereby pre-mating reproductive isolation mechanisms evolve to reduce the frequency of hybridization and prevent genetic waste (Dobzhansky 1940).  However, as long as even some hybrids are viable and fertile, they present opportunities for introgression among populations and for subsequent novel genetic variation (Arnold 1997; Rieseberg et al. 2003).

Examining patterns and distributions of genetic variation among populations can confirm whether, and the level at which, hybridization occurs in a contact zone and can place the system on the continuum from unimodal to bimodal hybrid zones.  The existence of a contact zone between differentiated populations does not necessarily imply that these populations hybridize.  For example, pre-mating isolation barriers may be strong enough to completely eliminate the possibility of intermating.  If hybridization does occur, the prevalence may be so great as to indicate a lack of pre- and post-mating barriers or that hybrids have equal or higher fitness of compared to parent individuals.  In contrast, the frequency of hybrids may be so low as to indicate the rarity of hybridization or drastically lower hybrid fitness.  Jiggins & Mallet (2000) characterized the continuum of hybrid zones from unimodal to bimodal with the processes that maintain them.  Unimodal hybrid zones, or hybrid swarms, exhibit weak pre-zygotic isolation and weak endogenous and exogenous selection against hybrids.  Bimodal hybrid zones are further along in the speciation process and generally exhibit well developed, but incomplete, pre-zygotic isolation or selection against hybrids.  Therefore, documenting the distribution of parental

and hybrid genotypes in a contact zone can reveal the types of isolating mechanisms maintaining a hybrid zone.

## Understanding How Selection Acts on Hybrid Zones in the Wild

Despite the theory and growing study of the genetic structure of hybrid zones, there is still a need for a better understanding of how selection acts on hybrids in wild populations.  Hybrid zones offer a wide variety of recombinant genotypes, which can be analyzed to determine how selection maintains differentiation among populations in the face of genetic exchange (Barton & Hewitt 1985; Barton & Hewitt 1989; Harrison 1990; Moore & Price 1993; Bert & Arnold 1995).  Careful sampling schemes can exploit this spectrum of recombinants to reveal fitness heterogeneities in space and time among genotypes, enabling a dissection of post-mating, endogenous and exogenous isolation barriers (Grant & Grant 1996; Albert et al. 2006).  In a meta-analysis of the limited instances where studies measure fitness of hybrids in nature to date (1995), Arnold & Hodges discovered hybrids are not always unfit relative to that of the parental taxa, as predicted by the predominantly used tension zone model.  Rather, the authors found a mixture of increased, equivalent, and, in fewer cases, decreased fitness of hybrids relative to one or both parental taxa.  They advocate a need for more studies of hybrid zone genotype classes in the wild to distinguish between hybrid zone models.

Indeed, obtaining information on hybrid fitness is critical for distinguishing between the primary models that describe hybrid zone maintenance and position (Moore 1977; Endler 1977; Barton & Hewitt 1981; Barton & Hewitt 1985; Howard et al. 1993).  In particular

156

by tracking whether the relative frequency of hybrids and parental taxa changes across the life cycle using multilocus data, cohort analyses can be used to determine if selection acts against hybrids only to maintain hybrid zones.  In a cohort analysis, genotype frequencies or hybrid frequencies are compared across the life cycle.  Dynamic cohort analyses resample a single cohort through time (e.g., Dowling & Moore 1985; Kocher & Sage 1986; Howard et al. 1993).  An alternative is the static cohort analysis, which compares genotype frequencies across cohorts (e.g., Bert & Arnold 1995; Wilhelm & Hilbish 1998).  In either type of analysis, a significant reduction in the proportion of hybrids from younger to older individuals indicates selection against hybrids.

While field measurements of fitness are often challenging, cohort analyses are a powerful way to assess hybrid fitness because they estimate relative viability of hybrid and parental genotypes under natural conditions (Howard et al. 1993).  However, cohort analyses have their limitations, too.  Unless it is possible to sample embryonic and larval stages (but see Cruzan & Arnold 1994), the test is blind to the potential selection occurring in these early life stages.  It also cannot assess mating pattern and mating success (Howard et al. 1993).  Rather than simply sampling a single location for a cohort analysis, expanding the cohort analysis spatially along the hybrid zone can distinguish between endogenously and exogenously maintained hybrid zones (e.g., Bert & Arnold 1995; Albert et al. 2006).

Cohort analyses have not been employed frequently to examine hybrid zones.  However, in the studies that do exist, results indicate a mixture of genotypic distributions and timing and form of selection.  Early cohort studies in freshwater hybrid zones found

strong selection against hybrids and concluded that the hybrid inviability was due to intrinsic selection (cyprinid fish, Dowling & Moore 1985; leopard frogs, Kocher & Sage 1986). However, these studies lacked robust sampling designs, such as tracking relative proportions of each parental species over the life cycle or multiple sample sites over multiple years, which could distinguish between endogenous and exogenous selection. A more recent cohort analysis of a trout hybrid zone found genotype frequency distributions consistent with a hybrid swarm and no strong selection against hybrids, thereby eliminating the possibility of endogenous selection (Rubidge & Taylor 2004). Exogenous selection likely causes the hybrid reductions over the life cycle to maintain the hybrid zones in threespine sticklebacks (Gow et al. 2007).

In two terrestrial studies, hybrid zones of ground crickets (Howard et al. 1993) and irises (Cruzan and Arnold 1994) exhibited genotype-dependent viability where the various hybrid genotypes were not consistently less fit than parentals. Both of these studies infer that assortative mating maintains the hybrid zones and not hybrid inviability. A cohort analysis of an Atlantic eel hybrid zone showed geographically dependent increased hybrid survival relative to parentals and inferred a latitudinal temperature gradient as the exogenous selection agent (Albert et al. 2006). Similarly, intermediate genotypes did not exhibit lower fitness than both parents in two marine mussel hybrid zones (Wilhelm & Hilbish 1998; Toro et al. 2004). For both of these systems, it appears ecological selection on morphology gives a parental mussel species lowest survivorship. Toro and colleagues point out that, for species with a long planktonic larval stage, it cannot be assumed that local adults produced larvae and juveniles sampled at a particular site.

Having genetic markers that distinguish between parental types is critical for cohort analysis. In wild, non-model systems, amplified fragment length polymorphisms (AFLPs) are useful for providing a randomized genomic view of differentiation, especially in cases where that differentiation is morphologically cryptic (Brattstrom et al. 2010). They are ideal for intraspecific studies and for hybrid zones (Meudt & Clarke 2007). Moreover, they have been used to unravel the genetic architecture underlying differentiation and to isolate genomic regions important to adaptation (Hawthorne & Via 2001; Rogers et al. 2001; Rogers et al. 2007; Wood et al. 2008). AFLPs also can be used to identify hybrid individuals (Bensch et al. 2002; Bensch & Akesson 2005; Albert et al. 2006; Bonin et al. 2007). In addition, multilocus data are required to assess bimodality of genetic distributions of hybrid populations. Utilizing genomic regions under strong divergent selection is ideal for such a study because these loci should show clearer and greater bimodality (Rieseberg et al. 1999; Jiggins & Mallet 2000).

**The System**

The eastern oyster, *Crassostrea virginica* (Gmelin), serves as a test system for investigating isolation barriers. The species reproduces through external fertilization after broadcast spawning. The planktonic larval period of 14-21 days is the only opportunity for dispersal because larvae settle and affix themselves permanently to hard substrate. Over this larval period, dispersal can be tens to hundreds of kilometers (Kinlan & Gaines 2003) and was estimated to be 472 $km^2$ in the Chesapeake Bay (Rose et al. 2006). With such high dispersal potential, regional genetic differentiation is low across

much of the broad eastern oyster distribution from Maine to the Yucatan Peninsula. However, several loci exhibit a sharp genetic step cline located near Cape Canaveral, Florida (Chapter 2, Reeb & Avise 1990; Karl & Avise 1992; Hare & Avise 1996; Murray & Hare 2006). This regional differentiation has remained stable for at least 15 years (Murray & Hare 2006), but the mechanisms for its maintenance in a species with high gene flow potential are unknown.

The oyster step cline coincides with a biogeographic province boundary, an ecotonal transition along a temperature gradient, and a decrease in adult oyster density (Hare & Avise 1996; Avise 2004). The reason for this density trough is currently unknown. However, tension zone theory predicts that hybrid zones move toward isolation barriers and are, thus, attracted to density troughs and environmental transitions (Barton & Hewitt 1985). For the oyster populations, it is unclear whether the decrease in density indicates reduced recruitment due to a dispersal barrier or due to localized reductions in survivorship following strong larval settlement.

Avise and colleagues documented sharp differentiation between two groups they termed "Atlantic" and "Gulf" for populations north and south of northern Cape Canaveral, respectively (Reeb & Avise 1990; Karl & Avise 1992; Hare & Avise 1996). The extent of hybridization between these two groups has never been documented, leaving question to whether the two hybridize at all. Individual-based assignment tests of adult oysters in Florida did document some signal of individuals with mixed ancestry, but did not explicitly test for this (see Chapter 2). In addition, this study did not focus on regionally

differentiated loci and did not classify individuals as hybrids or parentals.  Therefore, examining the extent of inter-region hybridization in this system requires additional study of regionally differentiated loci.

While a sharp step cline has been documented in several loci, many other loci demonstrate geographic allele frequency patterns that are not consistent with this regional differentiation (see Chapter 2; Karl & Avise 1992; McDonald et al. 1996; Murray & Hare 2006).  AFLP studies revealed that allele frequency differences at most loci are consistent with neutral processes (see Chapter 2; Murray & Hare 2006).  Allele frequencies at other loci fluctuated greatly along the east coast of Florida and demonstrated a pattern of divergent selection over local scales within regions (see Chapter 2).  Still, some loci do show a geographic pattern in allele frequency that indicates divergent selection between regions.  Although the system exhibits genetic signals of adaptation at different spatial scales among loci, it is those loci involved in regional adaptation that can reveal how selection maintains clinal differentiation in this species with a high gene flow potential.  These loci are the best candidates to test for regional hybridization, for the life cycle timing of selection maintaining regional differentiation, and for spatial patterns of selection among genotypes.

Similar tests were conducted using a static cohort analysis in another coinciding bivalve hybrid zone (Bert & Arnold 1995).  While this hybrid zone was between two species of hard clams that live in different habitats (*Mercenaria mercenaria* in the estuarine lagoons and *M. campechiensis* in the open water), the hard clam life cycle is very similar to that

of the eastern oyster.  In an examination of the isolation barriers maintaining this hybrid zone, the authors found evidence for both endogenous and exogenous selection. Supporting endogenous selection, hybrid frequency decreased with age; linkage disequilibrium was highest at intermediate ages; and Hardy-Weinberg equilibrium decreased with age.  However, selection on hybrids and adults changed spatially and among hybrid genotypes, which the authors interpreted as evidence for exogenous selection.  Indeed, a separate study also documented increased hybrid susceptibility to disease, supporting an endogenous selection cause of the hybrid breakdown (Bert et al. 1993).  Although the results of this study cannot be transferred automatically to another species, they do provide insight into what types of selection to investigate in the eastern oyster contact zone.

This study will address the following questions:

1.  Is there successful recruitment in the Cape Canaveral area?

2.  To what extent does hybridization occur between Atlantic and Gulf types?

3.  Does selection occur pre-settlement or post-settlement?

4.  Is selection endogenous or exogenous?  That is, does selection vary geographically across the zone of contact?

## Materials and Methods

### Sample Collection

A total of 274 adult oysters and 2945 spat were collected from multiple locations in

Florida estuaries between 2002 and 2006 (Table 3-1). Adult oysters were collected as

detailed in Chapter 2. I chose to collect spat during the springs and summers of 2005 and

2006 because natural spat settlement peaks during these months (Wilson et al. 2005).

For spat collected in 2005, I posted two slate plates in the low intertidal to shallow

subtidal shoreline regions of sites between Jacksonville and West Palm Beach to collect

natural spat settlement. After minimally one-month intervals between March and

August, these plates were retrieved and inspected for spat settlement. If spat settled, they

were counted and collected. Then plates either were cleaned and reposted if no spat

settled or were replaced with new plates to collect new potential settlement over the next

period. Up to 60 individual spat were counted and collected per collection site and

period. Depending on the individual size, the entire spat tissue or just the gill and mantle

were dissected from shells and preserved in 95% ethanol.

To collect spat between April and August of 2006, I used a similar procedure to the one I

used in 2005, except: 1) net bags filled with 3-4 L of clean oyster shell were the

settlement substrate, 2) up to 100 individual spat were collected per site and settlement

period, and 3) all bags were removed and replaced with new ones at each site visit.

**Observational Data Collection and Analysis**

During each visit to a spat collection site, I recorded the water temperature, salinity,

number of spat collected, quantity of mortality if any, categorical quantities of types of

other settled organisms, and any qualitative observations about the site. Temperature and

salinity measurements were taken using a YSI 30 (YSI Incorporated).

To determine whether recruitment success varied temporally and spatially, I analyzed the

number of spat samples collected per site visit within each transect. I calculated the

proportion of the collection goal that actually settled for each year, where the collection

goal was the maximum number of spat that I would possibly collect (60 and 100 spat per

site visit in 2005 and 2006, respectively). To test for a temporal effect on settlement

success, I performed a Wilcoxon non-parametric one-way comparison of means on the

proportion of site settlement success using the spat year as the treatment. Unless

otherwise stated, all statistical analyses were conducted using JMP$^®$ 9.0 software (SAS

Institute, Inc.). For all statistical tests performed in JMP, the validity of any equal

variance assumption was tested. If treatments had unequal variance, I employed

nonparametric statistical methods.

To test whether recruitment differed among regions, I labeled sites north of Cape

Canaveral (North), sites within Cape Canaveral (Cape Canaveral), and sites to the south

(South) as regional treatments. I then specifically tested whether recruitment was lower

in the Cape Canaveral region than in the North and South regions with a one-way

analysis of means for each year to determine if Cape Canaveral is a barrier to gene flow

due to a lack of oyster supply.  If variances were unequal among regions, the one-way

analysis was nonparametric and used a Steel-Dwass correction for all possible

comparisons among regions.  If variances were equal, the one-way analysis compared all

regions using a Tukey-Kramer HSD correction for multiple comparisons.


## AFLP Data Collection

For all further analyses, I included all AFLP data on 274 adults collected from seven sites

found in Chapter 2.  For all samples, whole genomic DNA was extracted and diluted

following procedures in Chapter 2.  I performed a modified version of the AFLP assay

outlined by Vos and colleagues (1995) on spat DNA samples following the protocol set

forth by Murray & Hare (2006).  I scored the same loci as the final set of 217 loci in

Chapter 2 for utility of and comparison to the adult oyster AFLP data set.  Allele scoring

and use of duplicate samples followed the procedures outlined in Chapter 2.


Only certain spat samples were chosen for AFLP analysis (Table 3-1, Fig. 3-1).  These

samples contained the greatest number of individuals at the greatest number of sites for a

given collection interval within each year's transect.  For spat collected in 2005, I chose

to analyze 254 individuals collected in August from six populations.  For spat collected in

2006, I performed the AFLP analysis on 757 individuals collected in July from eight

populations.  Subsequent AFLP and RFLP (M. Hare pers. comm.) analysis of these

populations revealed that *Ostrea equestris* dominated one site collection (John U. Lloyd

Park, Fort Lauderdale in July 2006).  I removed this set of samples from further analyses,

resulting in seven final sites for 2006 spat.

**Allele Frequency Differences between Life Stages**

I used the program AFLP-SURV 1.0 (Vekemans 2002) to estimate allele frequencies, genetic diversity, and population differentiation ($F_{ST}$) of all AFLP data for all 217 loci among sites within transects following procedures outlined in Chapter 2. AFLP-SURV generated the global $F_{ST}$ estimates over all 217 loci between each site. For the adult transect, I used allele frequencies and global $F_{ST}$ estimates from Chapter 2. To test whether overall genomic differentiation increases between spat and adult life stages, I performed t-tests to determine if global $F_{ST}$ estimates across all sites differed among transects using a nonparametric one-way comparison with the Steel-Dwass method of correction for multiple comparisons. For each transect, I also asked whether genomic differentiation was higher among sites on opposite sides of the cline than for those on the same side of the cline using one-tailed t-tests assuming unequal variances.

To answer questions addressing whether post-settlement selection shapes the regional differentiation in allele frequencies between life stages, I compared summary genetic data between adults and spat. For these analyses, I restricted the genetic data to the best candidate loci determined to be influenced by regional divergent selection (outliers detected at the $\alpha = 0.01$ level only in adult population comparisons on different sides of the cline (DSO outliers, see Chapter 2 genome scan results). I took the following steps to calculate the regional allele frequency difference (RAFD) between sites on opposite sides of the cline for each transect. 1) I calculated the mean allele frequency among sites on the same side of the cline (or within a region). 2) I calculated the absolute value of the difference in these regional means; this value is the RAFD. I also calculated regional

genetic differentiation estimates ($F_{RT}$; Hartl & Clark 1997; Hedrick 2004) for each

transect weighted by the number of sample sites. I chose to use both RAFD and $F_{RT}$ as

estimates of regional differentiation because the $F_{RT}$ calculation assumes that the number

of samples in subpopulations is proportional to the proportion of the population

represented by a subdivision and that this proportion is reflected by the number of

populations, which may not be valid in this system. To address whether RAFD and $F_{RT}$

are higher in adults than spat, I performed one-tailed, paired t-tests, where estimates were

matched by locus. Specifically, I was interested in whether RAFD changed between life

stages in directions to produce a steeper cline between regions in adults than spat and

whether $F_{RT}$ estimates were higher in adults than spat. I performed similar tests for

differences in RAFD and $F_{RT}$ between spat transects using two-tailed paired t-tests. Over

all tests each for RAFD and $F_{RT}$ I used a Bonferroni adjustment to correct for multiple

comparisons (corrected $\alpha = 0.0167$).


## Outlier Repetition among Life Stages

In order to determine whether previously described genomic signals of divergent

selection (see Chapter 2) are established early in the life cycle or in the months and years

post-settlement, I performed genome scans on all pairwise spat population combinations

within a transect using genotype data for all 217 AFLP loci. I generally followed

procedures outlined in Chapter 2. However, I chose to use only 0.01 as the $\alpha$ value to

reduce the probability of type I error. For a one-tailed test, this equates to the 99.5

quantile of the null distribution simulated by Dfdist (Beaumont & Nichols 1996;

Beaumont & Balding 2004). A total of 15 and 21 pairwise population comparisons were

performed in triplicate for the 2005 and 2006 transects, respectively.  The mean statistics

from these triplicates were calculated and summarized as in Chapter 2.  For the results

from each transect, I then divided loci into two locus classes: NS (non-outliers and

outliers appearing in only a single comparison) and MO (outliers identified in multiple

comparisons).

Using the list of MO loci generated in adult genome scans at the $\alpha = 0.01$ level in

Chapter 2, I asked whether genome scans among spat identify the same loci as MO.

Specifically, I used a one-sided, two-sample test for proportions (Agresti & Caffo 2000)

with a Bonferroni $\alpha$ adjustment for two tests ($\alpha = 0.025$) to determine whether MO loci

from each spat transect are more likely to have been MO than NS loci in the adult

transect.  A lack of congruence between outlier sets would indicate that selection occurs

post-settlement and later in the spat to adult life.  On the other hand, if there is high

correspondence between the identity of adult MO loci and spat MO loci, selection either

occurs pre-settlement, soon after settlement, or in the presence of local recruitment and

low gene flow.

In order to have enough data in each contingency table, I had to group multiple outliers

that all shared a single population among the significant population comparisons with

those that were detected in at least two population comparisons without any populations

in common (MD and MI outliers, respectively, in Chapter 2).  Ideally, only MI outliers

would be used in this analysis because these are the best candidates for divergent

selection and are not likely to be anomalies, experimental or biological, of a single sample site.

**Genotype Index Distribution among Life Stages**

For each individual, I calculated a genotype index score to evaluate the multilocus genotype distribution shapes and trends of the transects. When constructing multilocus genotype distributions, studies often assign an arbitrary -1, 0, or +1 to parental type alleles and sum these over all loci for each individual, regardless of the allele frequencies in the parental populations (e.g., Howard et al. 1993; Cruzan & Arnold 1994; Wilhelm & Hilbish 1998; Rubidge & Taylor 2004). However, this approach may not be the most appropriate when opposing alleles are not differentially fixed at each locus.

Because no AFLP locus in Chapter 2 exhibited fixed differences between Atlantic and Gulf-types, I assigned each allele of every DSO outlier locus a representative allele diagnostic value (RADV) generally following Bert & Arnold (1995). First, I created representative parental pools for 'Atlantic' and 'Gulf' types using adults that had a 90% or greater inferred ancestry to one cluster based on the two-cluster STRUCTURE results from all 217 loci in Chapter 2. After assigning each of these mostly-pure individuals to Atlantic or Gulf populations, I ran the DSO genotypes for these parental representatives through AFLP-SURV to calculate the allele frequencies for these loci in the Atlantic and Gulf pools. I assigned each allele to Atlantic or Gulf, depending on the pool in which it was most common. The RADV for each allele was equal to the frequency of the allele in

the Gulf pool minus its frequency in the Atlantic pool. For each individual, I calculated its multilocus genotype index as the sum of its RADVs over all DSO loci.

To visualize the genetic index distributions for each transect, I graphed histograms summarizing genotype index scores over all individuals within each transect and for each site within each transect. To determine whether there was a geographic association along the transects with individual genotype indices, I conducted a linear regression of genetic index on transect position for each transect and tested whether the slope was not equal to zero and how much of the data variance the geographic model explained.

### Hybridity among Life Stages

To estimate the admixture proportions for each individual, I ran the AFLP data through STRUCTURE v. 2.2.3 (Pritchard et al. 2000), applying the extension for dominant data. This software utilizes Markov Chain Monte Carlo-based Bayesian algorithm to assign individuals to genetic clusters based on their multilocus genotypes. I chose to analyze the spat and adult data for two clusters, or populations, based on the Atlantic and Gulf parapatric clusters determined in Chapter 2. Using the admixture proportions of adult individuals in the K=2 analysis of all 217 loci in Chapter 2, I assigned individuals as Atlantic or Gulf if their estimated ancestry was greater than 90% of one of the two clusters. Individuals exhibiting greater admixture were not assigned to either cluster. I ran the K=2 model 20 times with the option to use this prior adult assignment information in parameter estimation on all individuals (spat and adults) for DSO genotype data only. Other options applied to this analysis included the admixture model and correlated allele

frequencies. Each run incorporated a burn-in period of $10^5$ MCMC iterations and a data

collection period of 5 x$10^5$ iterations. Results across runs were summarized and

graphically displayed using the software CLUMPP v. 1.1.2 (Jakobsson & Rosenberg

2007) and Distruct v. 1.1 (Rosenberg 2004), respectively.

I used the CLUMPP results to estimate the admixture proportion of each individual as the

proportion of ancestry belonging to the Atlantic-type cluster $\left(q_A^{(i)}\right)$, which STRUCTURE

calculated. To estimate the level of hybridity of individual multilocus genotypes, I

transformed these admixture proportion values to hybridity values ($h_i$) using the formula

$h_i = 0.5 - \left| 0.5 - q_A^{(i)} \right|$ (Carney et al. 2000; Duvernell et al. 2007). This hybridity value

ranges from 0 in pure parental individuals to 0.5 for $F_1$ hybrids. To test whether hybridity

differed among transects, I performed pairwise non-parametric comparisons of means

tests using the Steel-Dwass method, which corrected for multiple comparisons. To

distinguish this hybridity estimate from the allele frequency-based genotype index

estimate mentioned above, I hereafter refer to hybridity estimates as CLUMPP-based

hybridity.

I also categorically assigned individuals to an ancestry status (Atlantic, Gulf, Hybrid),

depending on the individual's STRUCTURE-generated 90% posterior probability

interval (90% PI) of its admixture proportion for DSO loci (Albert et al. 2006; Gow et al.

2007). If the 90% PI did not overlap 0 and did overlap 1, the individual was assigned to

the Atlantic (ATL) category. If the 90% PI did not overlap 1 and did overlap 0, the

individual was assigned to the Gulf (GLF) category. All other individuals were assigned

as Intermediates (INT).  I then assigned ATL and GLF individuals to source classes as

Native-like (NTV) or Migrant-like (MGR), depending on whether the individual's ATL

or GLF assignment agreed with the cline side of its sample site.  Using these

assignments, I calculated the frequency of Intermediates at each site for each transect.  It

is important to note that the combination of the extensive recombination that occurs in

oysters and the lack of loci with fixed differences prevents absolute categorization of

hybrids, migrants, and natives.  To represent this uncertainty, I have chosen to use the

following proxy source classes: Intermediates, Migrant-like, and Native-like,

respectively.


To determine whether the proportion of hybrid-like and migrant-like multilocus

genotypes differed between each transect and between spat and adults, I performed a

static cohort analysis using an $R$ X $C$ $G$-test for an effect of transect (Sokal & Rohlf

1995).  I also tested for such an effect within the North, Cape Canaveral, and South

regions to determine if the pattern varied geographically, using Bonferroni $\alpha$ adjustment

for three tests ($\alpha = 0.0167$).  For these analyses, I used a combination of $R$ X $C$ $G$-tests

and a Fisher's exact test, using the Fisher's exact test when there were small expected cell

values in the contingency table.  Regional differences in the life stage patterns of

hybridity would point to exogenous selection because endogenous selection against

Intermediate genotypes should not vary geographically.

## Results

### Environmental Trends

Water temperature and salinity measurements taken during spat site visits may have been influenced by time of day and/or recency of rain events (Table 3-1.B). I present regional means here to mitigate the time of day influence. During 2005, temperature was most strongly inversely proportional to latitude in late March over a range of $6.7^{O}C$ (regional means: $17.9^{O}C$ in the North, $23.1^{O}C$ in Cape Canaveral, $24.6^{O}C$ in the South). This relationship gradually diminished over time through May, spanning less than $3.5^{O}C$ each month (late April regional means: $22.0^{O}C$ in the North, $23.8^{O}C$ in Cape Canaveral, $25.4^{O}C$ in the South; late May regional means: $26.3^{O}C$ in the North, $29.3^{O}C$ in Cape Canaveral, $29.5^{O}C$ in the South). During 2006 temperature showed no relationship with latitude and spanned less than $1.5^{O}C$ each month, however these site visits occurred later in the year (late May regional means: $29.8^{O}C$ in the North, $30.3^{O}C$ in Cape Canaveral, $30.2^{O}C$ in the South; mid July regional means: $29.8^{O}C$ in the North, $30.3^{O}C$ in Cape Canaveral, $30.4^{O}C$ in the South; late August regional means: $31.0^{O}C$ in the North, $30.9^{O}C$ in Cape Canaveral, $29.7^{O}C$ in the South).

A comparison of field measurements of water temperature to climate data on air temperature over similar geographic sites (Southeast Regional Climate Center, http://www.sercc.com/climateinfo/historical/historical_fl.html) showed strong consistency between the two data sets for corresponding months. Because climate data are more comprehensive and less susceptible to the short-term effects of time of day and

recency of rain, I concentrated analyses on climate data. Climate data for monthly averages show large variance among months for: minimum temperature, maximum temperature, number of days reaching a minimum temperature below freezing, and number of days reaching a maximum temperature above $90^O$F (Fig. 3-2). I did not present statistical analysis of the relationship between temperature and latitude because this is well-established (e.g., Brocklesby 1868). Winter months exhibit strong negative correlation between latitude and temperature and strong positive correlation between latitude and number of days reaching freezing temperatures. These associations diminish during spring and fall months, while summer months show relatively equivalent temperatures across latitudes.

Field data on salinity demonstrated high variation among sites (Fig. 3-3; 2005 mean = 25.9 ppt, 2006 mean = 26.4 ppt). There was no relationship between salinity and latitude averaged over 2005 measurements. In 2006, there was a significant positive relationship between mean salinity and latitude, but much of the variation in the data was not explained by the relationship ($R^2$ = 0.2283, p = 0.0285). Differences between 2005 and 2006 salinity data demonstrate temporal variability that may be indicative of variation among years and/or between seasons (March-May 2005 versus May-August 2006).

**Mortality and Recruitment**

Mortality of settled spat was negligible for nearly all collections. However, in August 2005 approximately 25 gaped, empty spat shells were found at East Haulover Canal on Merritt Island in Cape Canaveral. In addition, at three other sites in the Cape Canaveral

region (Port St. John, Banana River, and Palm Shores) there were some half and some gaped adult oyster shells and no living oysters visible.  While not all spat transect sites had adult oysters, adults were alive at all other sites where adult shells were present.  All evidence of adult and spat mortality was isolated to the Cape Canaveral region.

Settlement success differed among regions (Fig. 3-4).  The Cape Canaveral region exhibited a significantly lower mean percentage of the sample size goal than the North and South regions during 2005 (p = 0.0064 and 0.0004, respectively).  While the percent of the sample size goal was lower in Cape Canaveral than in the North and South during 2006, the differences were not significant (p = 0.2173 and 0.1799, respectively, using Tukey-Kramer HSD).  The lack of significance for these tests among 2006 regions was possibly due to the small number of sites within each region.  For both spat transect years, there was no difference in settlement success between North and South regions (2005: p = 0.9994, 2006: p = 0.9563).  In general, settlement was significantly more successful in 2006 than in 2005 (mean percentage of sample size goal = 0.6252 and 0.1674, respectively; p < 0.0001), though these transects differed in the amount and quality of settlement substrate.  Overall, both mortality and recruitment observations indicated that the Cape Canaveral region does not have prolific oyster populations.

**Allele Frequency Differences between Life Stages**

As expected, allele frequency differences between Atlantic and Gulf populations and regional population differentiation ($F_{RT}$) gave similar results, even though the latter measure makes more assumptions (Table 3-2).  For DSO outliers, both regional allele

frequency differences (RAFD) and $F_{RT}$ were highly significantly greater in adults than in either spat transect (Table 3-2). Because adult RAFD and $F_{RT}$ estimates were highly significantly greater than those in each spat transect, there was no need to pool spat transect data. In contrast to adult-spat comparisons, the RAFD and $F_{RT}$ values were very similar between 2005 and 2006 spat. Taken together, these results indicate regional genetic differences are strengthened by selection as young spat mature to adulthood because DSO loci are the best candidates for regional selection (Fig. 3-5).

While calculated based on all 217 loci, global $F_{ST}$ results tell a similar regional story (Fig. 3-6). For the adult transect, the mean of the global $F_{ST}$ estimates for population pairs on opposite sides of the cline was significantly higher than the mean for population pairs on the same side of the cline ($p = 0.0224$). No such significant difference existed for either spat transect, though the 2006 transect exhibited marginal significance for a higher mean global $F_{ST}$ for populations on the opposite cline side ($p = 0.0652$). While there was a lack of significance for 2005 spat ($p = 0.3595$), mean global $F_{ST}$ was higher among population pairs on opposite cline sides. These quantitatively different, though nonsignificant, regional patterns in spat indicate that some overall regional genomic differentiation does exist at the spat stage.

Mean global $F_{ST}$ across all populations within a transect was significantly lower in the 2006 spat transect than both 2005 spat and adult transects ($p = 0.0011$ and $0.0009$, respectively; Fig. 3-6). There was no significant difference in mean global $F_{ST}$ between adult and 2005 spat transects ($p = 0.2435$). These patterns were not driven by genetic

diversity, as within-population gene diversity was highly similar among transects

(0.2232, 0.2127, and 0.2177 for adults, 2005 spat, and 2006 spat, respectively). The

difference in mean global $F_{ST}$ between the two spat transects indicate that migration may

be variable over time. The unexpected similarity between 2005 spat and adult transects

suggest recruitment was largely local in August of 2005.


## Outlier Repetition among Life Stages

Genome scan results were not congruent between life stages. Spat multiple outliers (MO)

were adult MO more than expected and were adult non-outliers or singleton outliers (NS)

less than expected (Spat 85 vs. Adult: $p = 0.0239$; Spat 76 vs. Adult: $p = 0.0003$).

However, a larger proportion of spat MO loci were adult NS loci than were adult MO.

For 2005 spat, only six of 16 (37.5%) MO loci were MO in adult genome scans.

Similarly, 14 of 36 (41.2%) of 2006 spat were MO in adult genome scans. These results

demonstrate that most of the adaptive population genetic differentiation in adults was not

established in these early spat transects, suggesting at least some selection occurs later in

the life cycle.


Examining the outlier status of the adult DSO loci revealed additional disparities. While

four of the seven DSO loci were MO loci in the 2006 spat transect, only one was detected

as MO in the 2005 spat transect. This latter locus was the only one identified as MO in

genome scans for both transects. Although a large portion of DSO loci were not repeat

MO loci in spat, three of the four that were repeated were detected only in spat

population pairs on different sides of the cline.  Therefore regional selection may occur before spat were sampled in the life cycle, or recruitment may be largely regional.

**Genotype Index Distribution among Life Stages**

Transects differed in their genotype index distributions (Fig. 3-7).  The adult distribution was more bimodal than the spat transects, indicating more Atlantic- and Gulf-like individuals and less intermediate genotypes in adult populations.  While the 2005 spat distribution exhibited a hint of bimodality, it was mainly platykurtotic.  Both spat transects were largely unimodal, with greater representation of genetic indices near zero than near the extremes.  Overall, there was greater representation of moderately Gulf-like indices, likely due to the larger number of Gulf sites for all transects.

Breaking these transect distributions down by site revealed geographic patterns (Fig. 3-8).  For all transects, the distribution shifts from lower genotype indices at the northern end of the transect to larger indices at the southern end.  Notably, adult sites, as compared to both spat transects, show a sharper transition in genotype indices between northern and Cape Canaveral regions.  In addition, adults have more native types and less non-natives at the extreme ends of the transect compared to spat.  Unexpectedly, non-native genotypes were not uncommon in northern spat sites, indicating non-negligible northward migration of southern genotypes in spat populations.  Furthermore, extreme native genotypes were rare at these northern spat sites.  The strongest shift in genetic index distribution across life stages occurs at these northern sites, where Gulf-like

genotypes present at the spat stage are eliminated by the adult stage and frequencies of Atlantic-like genotypes are boosted.

Linear regression analysis confirmed a geographic relationship with genetic index (Fig. 3-9). Site distance from St. Augustine (the northern-most site, BNB76) explained a highly significant amount of the variation in genetic index for all transects (p < 0.0001). The coefficient of determination ($R^2$) for adult transect data was more than twice that for spat and comparable among the two spat transects. Although the slopes of the linear regressions were small due to the difference in scale between the two variables, the increase in genetic index over southward distance was highly significant for all transects. Slopes were comparable among all transects. The smaller amount of unexplained variation in adult data and similarity in slope among transects indicate selection refines the genotypes across life stages. Note: a linear model may not be the most appropriate for data that has more of a sigmoid shape, as in the adult transect. However, no common data transformations, including logit, were appropriate for these genetic index data, nor did a single transformation fit all transects well.

**Hybridity among Life Stages**

Life stage had a strong effect on CLUMPP-based hybridity (Fig. 3-10). Mean CLUMPP-based hybridity for the adult transect was approximately half that measured in each spat transect. This difference was highly significant (p < 0.0001), demonstrating a higher proportion of hybrids in spat than adults. In contrast, mean CLUMPP-based hybridity scores of 2005 and 2006 spat were essentially equivalent (p = 0.9987). The consistency

179

in higher levels of CLUMPP-based hybridity for two different spat transect times

supports a reduction in hybrids between life stages.

Transects also differed strongly in their frequencies of the source classes (Fig. 3-11.A, p

< 0.0001).  In particular, the adult transect had the greatest differences between observed

and expected values.  The observed number of adult Intermediate (INT) and Migrant-like

(MGR) genotypes (33 and 8, respectively) were much lower than the expected values

(77.86 and 30.83, respectively).  Correspondingly, the number of adult Native-like

individuals was much higher than expected (233 observed versus 165.31 expected).  In

contrast, the observed numbers of INT and MGR were higher than expected, and number

of NTV lower than expected, for both spat transects.  These results suggest that a

significant portion of Intermediate and Migrant-like individuals perish as oysters age

from young spat to adults.

When I analyzed North, Cape Canaveral, and South regions separately for a transect

effect on source class composition, the pattern of transect effect differed among regions

(Fig. 3-11.B-D).  North of Cape Canaveral, transect had a highly significant effect on

source class proportions (p < 0.0001).  MGR were completely absent in the adult transect,

but represented 22% and 30% of 2005 and 2006 spat, respectively.  Furthermore, INT

were only 4% of adults, while they were 39% and 41% of 2005 and 2006 spat,

respectively.  In the North, there were strong conversions from INT- and MGR-

dominated spat populations to adult populations consisting of nearly all NTV (p <

0.0001).  Results were somewhat similar in the Cape Canaveral region, with a highly

significant transect effect (p < 0.0001). However, 2006 spat contained very few MGR (3%) and much more NTV (68%). It should be noted that there was only one 2005 spat site to analyze in this region, versus two sites in the other transects. In contrast to the North and Cape Canaveral regions, there were relatively similar source class compositions among transects for sites south of Cape Canaveral (p = 0.0376, non-significant after Bonferroni correction for multiple tests). In the South, there were few MGR and 21-29% INT individuals. The maintenance of source class proportions across life stages and presence of adult MGR suggests a higher tolerance of INT and MGR in the South region. The great reduction in INT and elimination of MGR from spat to adult stages at sites in the North and Cape Canaveral regions indicate stronger selection against these genotypes than in the South region.

## Discussion

This static cohort analysis, coupled with observational study of recruitment patterns, improved understanding of how selection is acting in these wild Florida oyster populations. Poor recruitment and increased mortality in the Cape Canaveral region demonstrated that this area has feeble populations, is less hospitable, and thus may serve as an incomplete barrier to gene flow between Atlantic and Gulf populations. This study serves as the first estimation of hybridization among these populations and demonstrates that the frequency of intermediates is non-negligible and is not geographically isolated to the immediate area spanning the Cape Canaveral step in the cline. Prezygotic and early endogenous selection against hybrids is weak if it exists at all, as indicated by large proportions of intermediates surviving to the spat stage. While non-significant regional differentiation was present in spat, adults exhibited strong regional differentiation. Furthermore, proportions of intermediate and non-native genotypes were drastically reduced in adults compared to spat, though the strength of this pattern varied among regions. Evidence presented here points to post-settlement, exogenous selection against intermediates and non-natives as the dominant, though perhaps not exclusive, player maintaining regional differentiation in this contact zone.

Static cohort analyses sometimes cannot distinguish between variation in initial abundance among cohorts and mortality within a cohort because they assume equal abundance among cohorts (Wilhelm & Hilbish 1998). In marine invertebrates, recruitment success can vary greatly among cohorts (Gaines & Bertness 1992; Levin

2006; Siegel et al. 2008). However, the adult cline has been stable for over 20 years (Reeb & Avise 1990; Karl & Avise 1992; Hare & Avise 1996), demonstrating that the allele frequencies are spatially stable and likely would be found in adult populations from cohort to cohort. In this study, the two years of spat cohorts generally exhibited equivalent patterns in most analyses. While these two cohorts differed greatly in abundance/recruitment and likely in environmental conditions, they had similar genetic composition. Though it is possible the spat cohort(s) that led to the adults sampled here had very different composition from the two spat cohorts sampled here, it seems unlikely given these lines of evidence.

### Feeble Oyster Populations in Cape Canaveral

Compared with sites to the north and south, recruitment was not successful in Cape Canaveral. Summed over both years of spat collections, more than 4.5 and seven times as many spat recruited to settlement traps at North and South sites, respectively, than to Cape Canaveral sites. The low numbers of spat collected in Cape Canaveral were unbiased by number of sites (summed over both years: 17 North sites, 22 Cape Canaveral sites, 24 South sites). The causes of reduced recruitment in this region are unknown. However, the region is home to a density trough of oyster populations (Hare & Avise 1996). Fewer adult oysters likely result in fewer progeny compared to more robust populations. In some spots, densities may be so low that populations are unsustainable, possibly due to an allee effect (Allee 1931; Courchamp et al. 1999; Stephens et al. 1999). On the other hand, the collection sites may not have been positioned where healthy populations exist in the Cape Canaveral region. Tidal amplitudes in the Indian River

lagoon range from 0-15 cm and less than 0.5 cm in the Banana River lagoon (Smith 1996). In addition, flushing time has been estimated to range from less than one week in the southern stretches of the lagoon to over a year in northern parts (Smith 1993). Because the region's hydrography is so weak, recruitment may be extremely localized to the immediate vicinity of these healthy populations. This scenario is unlikely to explain the broader regional patterns, as sites were chosen in the same fashion as in the other regions. Furthermore, a two to three week larval period is likely ample time to disperse larvae beyond the immediate area (Kinlan & Gaines 2003; Rose et al. 2006).

The higher mortality observed among Cape Canaveral spat and adults provides further insight into the region's density trough. The region likely does not provide a highly hospitable environment to support healthy oyster populations. For instance, the region is also a biogeographic province boundary, where the range limits of many species meet (northern limit for subtropical and southern limit for warm temperate species; Briggs 1974). Perhaps the same ecological conditions that prevent the spread of these species beyond the region also make it a challenging environment for *C. virginica.* Indeed, ranges are determined by ecological tolerances, with the highest densities found at optimum and intermediate positions leading to lower abundance as tolerance thresholds are reached (Endler 1977; Bridle & Vines 2007). The weak hydrography in the region could be suboptimal for feeding and fighting disease. Cape Canaveral is located along the frost line of central Florida, south of which the maximum depth of frost penetration is zero (Floyd 1978). This frost line may be one such threshold. Clines form where the direction of selection changes between genetic optima (Endler 1977). Migration load

from northern and southern regions of differential optima into the depauperate and geographically intermediate populations of Cape Canaveral may be limiting the ability of local populations to adapt and flourish in this zone of ecological transition (Garcia-Ramos & Kirkpatrick 1997). Therefore, populations in this region may be maladapted relative to the local phenotypic optimum, which may explain the higher mortality observed. Thus, there may be a habitat gap at Cape Canaveral. Due to feeble local populations, weak hydrography, and the linear nature of the Florida lagoons, Cape Canaveral indeed may be an incomplete barrier to gene flow between Atlantic and Gulf populations. Such features attract clines (Barton & Hewitt 1985). The confluence of numerous intraspecific, environmental, and ecological factors may determine the location of the cline and the robustness of local oyster populations.

**Frequent Hybridization and Lack of Pre-zygotic and Pre-settlement Selection**
Intermediates represented approximately a third of all spat genotyped in 2005 and 2006 and about one tenth of adults, demonstrating that intermediates are far from rare. To my knowledge, this is the first estimation of intermediate frequencies in this well studied contact zone. I was unable to resolve Intermediates into different hybrid classes (e.g. $F_1$, $F_2$, and backcross) due to the lack of fixed allele frequency differences in dominant DSO loci. Such resolution would have would have enabled direct estimation of hybridization and a study of differential selection pressure among recombinant types (e.g., Cruzan & Arnold 1994; Bert & Arnold 1995; Wilhelm & Hilbish 1998; Rubidge & Taylor 2004; Albert et al. 2006). This may be facilitated in the future by converting DSO or other differentially distinguishing loci into codominant markers. Nevertheless, the existence

of Intermediates at spat and adult stages suggests that strong pre-settlement selection against hybrids is unlikely.

Because oysters do not have well-established sexual behavior or other forms of mate recognition systems, pre-zygotic barriers are not major factors to consider. The same inference has been made for the coincident hard clam hybrid zone (Bert & Arnold 1995). As broadcast spawners, mating is likely to be random among neighboring individuals because spawning of one individual induces its neighbors to reciprocate, provided the neighbors have evolved similar responses to environmental cues. Genetically based differences in spawning cues have been documented for oysters from Long Island Sound and Delaware Bay in a common garden experiment (Barber et al. 1991). These differences likely evolved as countergradient adaptations to latitudinal gradients (Conover & Schultz 1995) and may operate in the Florida populations studied here. Such differentially adapted timing can contribute to assortative mating (Coyne & Orr 2004). Given the high prevalence of Intermediates, such differences, if pre-zygotic barriers do exist among Florida oysters, may not causing strong assortative mating.

Indeed, an investigation of gametic compatibility between Marineland and Vero Beach (the WHI and VER sites in this study, respectively) found no strong barriers to fertilization between these populations on opposite sides of the cline (Zhang et al. 2010). The authors did not detect any differences in fertilization success, paternity, or embryo survivorship and suggested that mechanisms other than strong reproductive barriers must be maintaining the oyster cline. Paternity did show subtle advantage for local sperm and

egg pairings over non-native crosses.  Slight advantages for sperm from northern males in competition with southern male sperm hint at a possible mechanism for the asymmetry of the cline, with northern alleles present in southern adult populations.  Similar results were found in an earlier study of single-pair crosses (Gaffney et al. 1993).  However, it is difficult to determine whether the results of Zhang and colleagues can be extrapolated to a regional story because only one site on each side of the cline was used for broodstock sampling.  Previous genetic evidence of localized adaptation within regions (see Chapter 2) and the increased variation in allele frequencies of southern populations (Hare & Avise 1996) open the possibility to the broodstock being unrepresentative of their source regions.  Nonetheless, the lack of strong reproductive barriers documented by Zhang and colleagues (2010) corroborate the large proportions of Intermediates sampled here.

As a third of all sampled spat, Intermediates are unlikely to suffer strong pre-settlement selection.  However, the possibility of some level or pre-settlement or early post-settlement (i.e., after settlement but before sampling) cannot be completely ruled out by this study because no samples were taken of planktonic larvae populations.  The frequency of Intermediates and hybrids in the plankton remains unknown.  If substantial selection occurs against hybrids pre-settlement or early post-settlement, positive linkage disequilibrium and Hardy-Weinberg disequilibrium would exist in spat and adult samples.  Using dominant markers, I was unable to test for such patterns.  In a previous study, Hardy-Weinberg proportions were not significantly different than equilibrium expectations among adults (Hare & Avise 1996).  Linkage disequilibrium patterns were not consistent across sites, but did reveal positive linkage disequilibrium (excess of

parental genotypes) for adults and genotypic independence for spat in Sebastian, Florida. However, these estimates have been shown to vary across life stages in other hybrid zones (Kocher & Sage 1986; Cruzan & Arnold 1994; Bert & Arnold 1995; Wilhelm & Hilbish 1998; Rubidge & Taylor 2004).  Based on the results of Hare & Avise (1996) and the large proportions of Intermediates spat sampled here, planktonic selection against hybrids is not strong.  If pre-zygotic barriers or pre-settlement selection do exist, it is unlikely that they play prominent roles in the maintenance of this hybrid zone.

### The Dominant Phase of Regionally Divergent Selection is Post-settlement

Focusing on the loci that are the best candidates for regional divergent selection (DSO outliers) permitted detection of differential selection between life stages.  Because adults exhibited much stronger regional differences at DSO loci than spat, it is clear that the majority of regionally divergent selection occurs post-settlement.  Indeed, weak clines in spat were sharper in the adult stage.  These results are consistent with previous work demonstrating that the anonymous, codominant nDNA locus CV-7.7, which was detected as having a signal of divergent selection in a genome scan (Murray & Hare 2006), exhibited a significantly shallower cline in spat when compared to adults and strong heterozygote deviations from Hardy-Weinberg expectations (Cammen honors thesis; Hare & Avise 1996).  It appears that oyster populations are more genetically homogenous at the early spat stage than as adults and are subject to subsequent regional selection, whereby non-Native-like alleles are removed from the population over the life cycle.

However, low levels of regional differentiation were present in the spat stage for DSO loci. It is possible that some regional selection takes place pre-settlement or in the early spat stage. Oysters and other highly fecund, low-parental-care marine organisms are known to suffer high mortality in early life stages (Kennedy 1996; Karleskint et al. 2010). Because the analyses of all 217 loci included DSO loci and other loci putatively subject to the effects of pre-settlement or pre-sampling regional and local divergent selection (see Chapter 2), non-neutral processes indeed may have influenced the observed levels of differentiation in spat. A non-exclusive, alternative reason for the low levels of regional differentiation in spat is that there may not be complete admixture in reproduction between the regions. Local or even regional recruitment patterns are often the primary explanation offered for population differentiation in marine organisms with high dispersal potential (Swearer et al. 1999; Barber et al. 2002; Cowen et al. 2006). These types of studies implicate oceanographic patterns as physical barriers to dispersal, which cannot be ruled out in this Florida system. Furthermore, if regionally differentiated adults have local mates due to limited gamete survival duration (1-5 hours; Galtsoff 1964), they would hybridize less frequently with more rare non-native genotypes in the vicinity. Alternatively, It is highly possible that recombination among non-fixed DSO alleles from native parents can generate the broad representation of Intermediate, Native-like, and Migrant-like spat genotypes. However, these broad genotypic distributions are narrowed by adulthood. Given the much greater adult regional differentiation in DSO loci, the majority of adaptive cline maintenance is due to regional post-settlement selection in these oysters.

In contrast, the high global differentiation observed in 2005 spat when all 217 loci were included was not expected. Although this differentiation was as high as in adults, the regional pattern of differentiation in the later life stage was largely absent in spat, suggesting that the differentiation observed in 2005 spat was geographically uniform or random. The fact that 2005 and 2006 spat regional differentiation was very low at DSO loci and no different from within region differentiation for all 217 loci does suggest that the high overall differentiation was due to neutral processes and not driven by regional selection. Random genetic drift may be the cause of this geographically independent differentiation observed in 2005 spat and may be in the form of the sweepstakes effect, a special form of genetic drift (Hedgecock 1994; Hedgecock & Pudovkin 2011). Under the sweepstakes hypothesis, chance events create a high variance in reproductive success among individuals of highly fecund marine populations subject to high early mortality. In fact, the development of this hypothesis was inspired by patterns observed in Pacific oysters, *Crassostrea gigas*. The sweepstakes hypothesis predicts: 1) significant, random allele frequency shifts among cohorts, and 2) reduced genetic diversity in recruits as compared to their source adult populations. The first prediction may be one possible explanation for the significant differences in global differentiation between the two spat transects. It is possible that the number of individuals effectively contributing to the spat populations was much lower in the 2005 samples than the 2006 spat, leading to random differentiation in 2005 and greater admixture in 2006 samples. However, the very similar levels of gene diversity among all transects, regardless of life stage, are not consistent with the second prediction of the sweepstakes hypothesis. Indeed, there has been limited empirical support of the sweepstakes hypothesis (Flowers et al. 2002; but see Hedgecock

& Pudovkin 2011).  Whether the oyster populations in Florida are subject to sweepstakes events, even sporadically, is a question that deserves further investigation.

These discrepancies between differentiation patterns among spat transects likely are due to neutral processes.  Local recruitment or lower recombination at DSO loci in 2005 spat could easily generate the higher global differentiation observed in this spat transect because recruits would be the non-recombinant progeny of the local populations, thereby carrying over any local differentiation that exists in adults to this 2005 set of progeny.  In contrast, geographically broader recruitment or recombination may have kept global $F_{ST}$ low in the 2006 spat sample by reducing local differentiation through migration or recombination.  Environmental variation between 2005 and 2006 or even within the recruitment seasons could have caused these temporal differences in dispersal and recruitment.  Smaller sample sizes in 2005 spat compared to 2006 spat could lead to error in population-level estimates and caused sites to appear more divergent.  However, regional differentiation was consistently low in both spat transects and was much greater in the adults.  The differences among types of loci and temporal and life stages indicate regional selection in this system is strong enough to override these sporadic neutral processes and re-sharpen the step cline at target loci after settlement but before adulthood.

Outlier repetition analyses offered mixed evidence for post-settlement selection.  Most adult multiple outliers were detected as outliers in zero or one spat population comparison out of a total of 36 spat site comparisons.  However, the detection of some

outliers suggests the possibility that not all selection is entirely post-settlement. The lack

of identity between spat and adult multiple outliers does suggest that some of the

statistically detected selection in spat may be temporally variable and not consistently

sustained through adulthood among all cohorts. Furthermore, if clinal selection was

established pre-settlement or in the early spat stage, adult DSO loci would be multiple

outliers and even DSO outliers in both spat transects. This was not the case for six out of

seven adult DSO loci, indicating that post-settlement selection establishes regional

differentiation. Some signal of post-settlement selection may be generated through local

recruitment by retaining selected loci near to the parent populations. Such local

recruitment and resulting reduction in gene flow have the potential to ease the way for

selection to generate or maintain regional adaptations (Bertness & Gaines 1993; Swearer

et al. 1999; Thorrold et al. 2001; Conover et al. 2006). While outlier patterns suggest that

some local recruitment and/or pre-settlement selection exists, they demonstrate that most

of the regional selection is largely post-settlement.


**Regionally Dependent Selection Against Intermediate and Migrant-like Genotypes**


Selection is against Intermediates

As compared to the general adult population, larger proportions of the spat populations

had Intermediate-like genotypes at DSO loci, both measured by genotype index and

CLUMPP-based source class. As Florida oysters transition from spat to adults,

individuals with intermediate genotypes at loci subject to regionally divergent selection

are removed from the general population. Bimodal genotypic distributions such as those

found in adults are indicative of strong reproductive barriers, either in the form of pre-zygotic isolation or disruptive selection against hybrids (Jiggins & Mallet 2000). It should be noted that the general adult population here exhibited a weakly bimodal distribution, which may be the result of less well-developed isolation barriers. While incomplete isolation likely plays a role (Zhang et al. 2010), the weak bimodality also may be a symptom of using non-diagnostic loci. The scarcity of diagnostic nuclear markers in these populations possibly is the result of: 1) incomplete lineage sorting due to insufficient time in allopatry between vicariant Atlantic and Gulf populations, 2) balancing selection for multiple alleles, or 3) secondary gene flow between Atlantic and Gulf populations (Hare & Avise 1998). This may have decreased the ability to resolve differences between pure Atlantic and pure Gulf genotypes, if they exist in natural Florida populations.

In contrast to the weak bimodality of adult genotypic distributions, broad unimodal distributions characterized both years of spat samples. Though there were slight differences in the two spat distributions, the consistency of broad distributions heavily concentrated in moderate genotype index values demonstrates the intermediacy of spat populations and their dearth of genotypically parental-like individuals. Unimodal distributions are associated with largely incomplete pre-reproductive isolation and/or an absence of selection, endogenous or exogenous, against hybrids (Jiggins & Mallet 2000). In the case of the oyster, intermediate multilocus genotypes of non-diagnostic loci also can be generated through recombination within a parental type. Higher CLUMPP-based hybridity, proportion of Intermediates, and frequency of intermediate genotype indices

present in spat demonstrate that intermediate-like individuals live to reach the spat stage. Furthermore, the presence of Intermediate CLUMPP-based source class individuals in adult populations demonstrates that they are viable through adulthood. Because further categorization of Intermediates could not be resolved to conclusively identify hybrids or distinguish $F_1$, $F_2$, and backcrosses, it is unclear whether certain hybrid classes vary in this viability. While much of the selection against intermediate genotypes appears to occur before adulthood, that selection is incomplete and occurs post-settlement. As a result, if these intermediates are genuine proxies for hybrids and if adult hybrids are even slightly fertile, they can serve as genetic bridges between the divergently selected Atlantic and Gulf populations. This gene flow would permit neutral and universally advantageous alleles to pass more freely across the cline than loci under the influence of divergent selection (Luikart et al. 2003; Emelianov et al. 2004; Teeter et al. 2008). Because low levels of intermediates survive to adulthood, it is possible that this regional differentiation is maintained in the face of at least some gene flow.

While information continues to grow about the genetics underlying post-zygotic isolation (Coyne & Orr 2004), knowledge is limited regarding outcomes of hybrids and post-zygotic isolation in the wild (Gow et al. 2007). Marrying the cohort analysis with genotypic distribution analysis enabled interpretation of both the timing of selection and the nature of it. At the spat stage, isolation barriers are weak to non-existent. However, isolation barriers are not absent from the system altogether, which would have been the interpretation if I had only examined spat. Instead, the bimodal and site-specific genotypic distribution of adults suggests isolation barriers present themselves later in the

life cycle.  Extreme or native genotypes, which are more rare in spat, survive while intermediates or hybrids perish, shifting genotypic distributions somewhat toward the extremes through the life cycle.  The elimination of most intermediates before the adult stage means that they are less likely to contribute reproductively to future generations (e.g., Kocher & Sage 1986; Cruzan & Arnold 1994; Gow et al. 2007).  This life stage-dependent presence of isolation barriers rules out strong pre-zygotic barriers and supports substantial disruptive selection against intermediates.  These distribution results are consistent with the lack of strong reproductive barriers observed by Zhang and colleagues (2010) and the significant reductions between spat and adult life stages in CLUMPP-based hybridity and in the proportion of Intermediate CLUMPP-based source class individuals observed here.  Therefore, disruptive selection against intermediates or hybrids appears to contribute to the maintenance of this hybrid zone.


<u>Selection is also against Migrant-like Genotypes:</u>

This study represents the first documentation that individuals with migrant-like genotypes do recruit to the alternate region from their ancestry and do so at non-trivial levels.  However, it remains unclear if these individuals with genotypes similar to migrants: (1) actually have migrated across the cline, (2) are simply the progeny of local individuals with non-native genotypes, or (3) are recombinants of local individuals with native-like genotypes.  With a 14 to 21 day planktonic larval stage, oysters have a high potential for long distance dispersal (Galtsoff 1964).  For these types of marine species, dispersal distance can be 20-120 kilometers (Kinlan & Gaines 2003; Shanks et al. 2003; Siegel et al. 2003; Shanks 2009).  This level of dispersal would permit migration across this cline's

sharp step that exhibits dramatic allele frequency shifts over only 20 km (Hare & Avise 1996). Prevailing near-shore currents converge in the Cape Canaveral region (Bumpus 1973), facilitating dispersal of Atlantic and Gulf genotypes toward each other if larvae exit and re-enter the lagoons through inlets to the ocean. Even if gene flow occurs in a stepping stone fashion and not through long distance migration, substantial selection would be required to maintain the hybrid zone at these potential levels of migrant recruitment.

Generally, the proportion of migrants-like individuals was higher in spat than adults. The reductions in migrant-like individuals observed between life stages demonstrate selection against non-native-like genotypes. Combining this with the observation that intermediates also suffer mortality implicates the role of exogenous viability selection in cline maintenance. This creates the subtly bimodal adult genotype distribution and a geographic pattern of site genotype index distributions that moves from one extreme to the other with more intermediate genotypes near the center of the cline. The most used model for describing hybrid zone dynamics, the tension zone model, states that clines are maintained by a balance of gene flow from parental populations and selection against hybrids due to genomic incompatibilities between cohesively evolved parental genomes (Barton & Hewitt 1985). Although I cannot rule out effects of endogenous selection here, it seems unlikely that genomic incompatibilities are the primary cause of reduced fitness in intermediates because migrant-like individuals with parental genotypes also suffer elimination, which would not be an effect of endogenous selection. Intermediates and migrant-like individuals exhibit signs of reduced fitness due to a mismatch between

their genotypes and the environment. The increase in native-like genotypic proportions from spat to adult life stages suggests that parental-type individuals experience increased fitness in their home environment. Outside of their native environment, these same parental genotypes are removed from the population over this span of the life cycle.

However, selection is not likely to completely eliminate non-native-like genotypes before spat become reproductively competent, especially in the South. Spat can reach sexual maturity/adulthood in their first season in lower latitudes (i.e., by the end of the summer; Thompson et al. 1996; Shumway 1996; personal observation), making gene flow possible. The presence of migrant-like and hybrid-like individuals may indicate that the cline is maintained minimally in the face of dispersal, if not gene flow. Dispersal from areas of higher fitness to those of lower fitness can cause maladaptation due to migration load of unfit immigrants (Lenormand 2002). If the selection gradient is broader than the characteristic length, the ratio of the standard deviation of parent-offspring distance ($\sigma$) to the square root of selection strength (s), then maladaptation can be avoided (Slatkin 1973). This may be true in the Florida system because the transition from salt marsh grass to mangroves occurs over a distance of 400 km between St. Johns and Palm Beach counties (USDA 2011), whereas dispersal distance estimates for Chesapeake Bay eastern oysters was 21.9 km (Rose et al. 2006, where this is the linear square root of the estimated 479 km$^2$ average squared dispersal distance). Florida oysters may experience longer distance dispersal more often by exiting the lagoons through inlets and re-entering after riding nearshore currents to more distant inlets or by riding occasionally well timed hurricanes (Bert & Arnold 1995). Viability selection due to environmental gradients

maintains divergence in the face of dispersal for other marine species with high fecundity and larval dispersal (Koehn et al. 1980; Diehl & Koehn 1985; Bertness & Gaines 1993; Wilhelm & Hilbish 1998). It seems exogenous divergent selection counteracts the potential for homogenizing dispersal in multiple coastal marine organisms and supports the environmental-dependent model of hybrid zone dynamics (Endler 1977).

Indeed, an unexpected pattern of what could be long-distance migration emerged here. Gulf-like spat were found in North sites, whereas the opposite occurred with comparative rarity. The moving hybrid zone hypothesis and the geological history of the area would predict northward movement of the hybrid zone, leaving Atlantic-like populations in southern sites (Hare & Avise 1996). While data in this study do not refute and some data support this prediction, the presence of Gulf genotypes in North spat does not fall under this hypothesis. Such northward migration may be stepwise within the lagoons or may use nearshore currents. The Florida Current also moves northward, but moves farther away from shore near Cape Canaveral and becomes the Gulf Stream. However, this unexpected pattern could be generated through recombination of local individuals without the need to invoke long-distance migration. Although these genotypes were eliminated in adult samples, the possibility of an intriguing dispersal patterns in spat warrant further study. The differences between each of these life stages have implications for what may be interpreted about migration, gene flow, recombination, and selection. Future studies should be careful to look at spat to examine migration questions and adults to examine gene flow.

<u>Selection is not Spatially Homogeneous:</u>

Further evidence supporting a dominant role for exogenous selection in the maintenance

of the oyster cline comes from geographic differences in cohort composition.  By

definition endogenous selection is independent of environmental heterogeneity (Barton &

Hewitt 1985).  This cohort analysis revealed differences among North, Cape Canaveral,

and South regions in the magnitude of reduction in hybrid- and migrant-like genotypes

between life stages.  This argues against spatially uniform selection and therefore against

an endogenous selection-only model.

One model of selection maintaining hybrid zones exogenously is the bounded hybrid

superiority model (Moore 1977; Moore & Price 1993).  This predicts superior fitness for

hybrids over parental types in intermediate environments.  In the case of the oyster hybrid

zone, the identity and nature of the selection agents remains unclear.  Therefore, it is

difficult to test for superior hybrid fitness in an intermediate environment.  However,

there is some evidence here that intermediates are at least less unfit in the South than in

the North and Cape Canaveral because they are present at higher proportions in the adult

stage in the South.  While this study focused mainly on eastern Florida sites, it is possible

that the environment in southern Florida is intermediate to those found toward the north

and in the Gulf of Mexico.  The limited study sites in the Gulf of Mexico preclude

determination of whether hybrids have increased fitness in southern Florida only or if

their fitness is similar to that of GLF types south of Cape Canaveral and in the Gulf of

Mexico.  Despite the unknown intermediacy of the environment in South sites, the

maintenance of migrant- and hybrid-like genotype proportions through the adult stage

suggests that selection pressures are weaker in this region than in the North and Cape

Canaveral and that they tolerate a variety of genotypes.

Spatially fluctuating allele frequencies can be indicative of patchy selection pressures.

Such genetic patterns are present in the oyster cline system and are prevalent particularly

in the region south of Cape Canaveral (Chapter 2; Karl & Avise 1992; Hare & Avise

1996). Although fluctuating allele frequencies in the South could be due to remnant

Atlantic-type populations after northward movement of the hybrid zone (Hare & Avise

1996), this neutral process does not explain the signal of selection observed in several

loci, including the DSO loci examined here. Indeed, many of the outliers detected in

multiple genome scans (MI outliers in Chapter 2) are candidates for indicators of smaller-

scale, localized selection. While these loci demonstrated signals of divergent selection,

they did not show strong regional differentiation. Rather, these loci likely are subject to

the influence of spatially heterogeneous selection, which shapes their fluctuating allele

frequencies. In particular, the South pattern of non-native-like genotypes surviving into

adulthood may be driven largely by unique conditions at one of the adult sample sites

(FTP), where most of the Atlantic-like adult genotype indices sampled in the South

existed. This site may exhibit a unique selective regime within this patchwork.

This type of patchy selection underlies the maintenance of mosaic hybrid zones, another

form of exogenously maintained hybrid zones whereby genotypes and environmental

factors covary over microgeographic scales (Arnold 1997). For example, ground crickets

*Allonemobius fasciatus* and *A. socius* exhibit patchy and mixed distributions in their

hybrid zone due to patchy distributions of different climatic regimes (Howard & Waring 1991). A classic, well-studied cricket hybrid zone features striking mosaic patterns associated with loam and sand soil habitats on small spatial scales (*Gryllus pennsylvanicus* and *G. firmus*, respectively; Rand & Harrison 1989; Ross & Harrison 2002). Habitat specialization also controls genotype distributions in a mussel hybrid zone featuring high gene flow potential, where *Mytilus edulis* dominates open sea and sheltered habitats and *M. galloprovincialis* dominates exposed and wave-beaten sites (Bierne et al. 2002). Therefore, regional selection may overlay smaller scale exogenous selection particularly south of Cape Canaveral in the oyster hybrid zone.

Finally, the geographic shift from low to high genotype indices (Atlantic- to Gulf-like genotypes) along the transects follows the inverse relationship between winter low temperatures and latitude. Indeed, gradients in mean temperature and number of days reaching freezing temperatures are steepest in the winter. Although I did not conduct any tests for selective agents, winter low temperature is a potential candidate for an agent of selection. Its geographic pattern is consistent with adult shifts from Atlantic alleles to Gulf alleles in putatively selected loci. While summer temperatures are similar along the transect, fall and winter temperatures have strong, inverse relationships with latitude. Even though the climate data demonstrated a broader latitudinal temperature gradient than the stepped oyster cline, clines along a shallow environmental gradient may be steepened by other factors including convergent or partial barriers to gene flow and density troughs (Barton & Hewitt 1985; Hare et al. 2005). The timing of fall and winter lows also is consistent with evidence here for post-settlement selection acting after the

summer collection dates of spat here.  Freezing conditions in the North could make it difficult for southern-like individuals, adapted to a warmer climate, to establish themselves.  Meanwhile non-native-like genotypes may have higher fitness in the South than in the North and Cape Canaveral because the Atlantic-type genotypes are already adapted to similar summer temperatures and are able to tolerate any occasional cold spells that they may encounter in the South.  This could explain the relatively greater reduction in, and therefore stronger selection against, hybrid- and migrant-like individuals over the life cycle at North and Cape Canaveral sites compared to South sites.  If fall and winter cold temperatures are a selection agent in this system, they would create a selection gradient that is strongest in the North and decreases southward.

**Future Prospects**

While this study made strides toward uncovering the mechanisms maintaining this oyster hybrid zone with the timing and general targets of selection, the agents of selection remain unknown.  One possibility discussed above is that the latitudinal temperature gradient, where temperature increases with decreasing latitude (Brocklesby 1868), may cause viability selection against genotypes not appropriately adapted to local conditions.  However, a latitudinal temperature gradient exists throughout the range of *Crassostrea virginica*, which spans from the Gulf of St. Lawrence, Canada through the Yucatan Peninsula, Mexico and exhibits the strongest genetic homogeneity at clinal loci in the portion north of Cape Canaveral (Hare & Avise 1996).  At these loci, there is no continuous change in allele frequency that follows the temperature gradient along its length.  It is possible, however, that oysters populations reach a temperature threshold in

eastern Florida.  The frost line cutting across Cape Canveral may be such a threshold

(Floyd 1978; Hull & the Federal Geodetic Control Committee 1989).  There is some

evidence of adaptation among oyster populations to different temperatures (Long Island

Sound versus Delaware Bay and Chesapeake Bay populations: Barber et al. 1991;

Dittman 1997; Dittman et al. 1998).  Temperature tolerance experiments should be

conducted on Atlantic, Gulf, and hybrid individuals to determine if temperature is a

selection agent contributing to the maintenance of the hybrid zone.  These should be

designed with careful consideration of documented short-term tolerances (Shumway

1996) and may require long-term temperature stress.

Beyond temperature gradients, the hybrid zone coincides with an ecotone, or transition

from warm temperate to subtropical ecologies (Engle & Summers 2000).  While this is

visible as a transition from salt marsh cordgrass to mangroves, the ecological transitions

go beyond these plants.  It is a biogeographic province boundary, where the ranges of

many species meet (Briggs 1974; Engle & Summers 1999; Avise 2004).  Many of these

transitions and range limits occur within the Canaveral National Seashore

(www.nps.gov/cana; personal observation).  Therefore, it will be difficult to test which of

the many changing elements may play a role in the selection gradient acting on this

system.  A multivariate landscape genetics study could illuminate which ecological

factors act as selection agents (Manel et al. 2003; Joost et al. 2007; Kozak et al. 2008;

McCairns & Bernatchez 2008; Schmidt et al. 2008; Manel et al. 2010).  To confirm the

role of selection agent, landscape genetics should be followed up with formal common

garden experiments using the potential agents as treatments (Kawecki & Ebert 2004).

Although this study suggests that the Cape Canaveral region suffers from poor recruitment success, the causal factors remain unknown.  The possibility that low flow experienced in the local lagoons creates a dispersal barrier for the area should be tested, at the minimum through modeling.  Such a model ideally would incorporate the hydrographic patterns of Florida lagoons, oyster densities, and what dispersal patterns might be for oyster larvae at different times in the reproductive season.  Common garden experiments located *in situ* of Cape Canaveral lagoons also should be conducted to determine whether the habitat is inhospitable and causes mortality of Atlantic, Gulf, and hybrid spat and adults.

Improvements can be made on this study to address certain questions about the nature and timing of selection more directly.  A dynamic cohort analysis that incorporates more continuous spatial sampling of spat over multiple years through adulthood and at different times within each year might reveal at which age(s) viability selection occurs. Distinguishing different spat cohorts from the same season will require great care. Ideally, such a study would sample the larval life stage as well to determine whether any selection takes place early in the life cycle (e.g., Bierne et al. 2003).  In addition, using codominant loci and following subclasses of hybrids in a cohort analysis would help distinguish differential selection among hybrid genotypes (Arnold & Hodges 1995). Tying this extensive of a dynamic cohort analysis with landscape genetics using ecological data over the course of the study could help identify selection agents (Gow et al. 2007).

The results of this study have management and conservation implications. Similar types of studies should be performed when designing restoration and conservation projects for other hybrid zones. Although most oyster restoration efforts in Florida focus of reef construction, there are some ongoing efforts to restore Florida oyster populations through oyster gardening and stock enhancement, or outplantings of young individuals bred and preliminarily raised in laboratories (e.g., Oyster Gardening and Restoration Project and Florida Oceanographic Oyster Restoration project through the Florida Oceanographic Society; Brumbaugh & Coen 2009). However, much effort should be made to ensure broodstock are from the region of the planned transplant location. The native conditions also should be taken into consideration when optimizing the hatchery conditions. Otherwise, resources will go to waste when maladapted, non-native genotypes suffer viability selection. Although these regionally adapted populations of *Crassostrea virginica* are considered a single species currently, their status (i.e., as subspecies, ecotypes, or races) may need to be reviewed to ensure appropriate conservation measures can be enacted.

This study supports growing evidence that the tension zone model does not explain universally how all hybrid zones are maintained. A meta-analysis of hybrid zone studies concluded that hybrids are not uniformly unfit across all hybrid zones, as the tension zone model would predict (Arnold & Hodges 1995). The variation in mechanisms maintaining hybrid zones calls for additional in-depth studies in nature. Indeed, one coincident hybrid zone is controlled by a complex suite of endogenous and exogenous selection and fits

both tension zone and environmental gradient models (Bert & Arnold 1995). Because the

hard clam and oyster hybrid zones share several characteristics and are geographically

co-located with other phylogeographic transitions (Avise 1992; Avise 2004), their further

study should help illuminate selection and dispersal mechanisms affecting the

distributions of other species in the area. Furthermore, these systems highlight the need

for researchers to consider the interplay between dispersal and endogenous and

exogenous selection as they design their studies and draw conclusions from their results.

# Conclusions

Interest in hybrid zones stems from their characteristic of differentiation despite gene flow. Understanding the balance between selection and gene flow in hybrid zones is informative about the process of divergence. However, documentation of the nature and timing of selection in nature has been relatively limited among hybrid zone studies. Cohort analyses have proved useful for distinguishing between endogenous and exogenous isolation barriers in hybrid zones. Such studies of hybrid zones in the wild can illuminate the mechanisms mediating differentiation in the face of gene flow and the role of selection in biological diversification in nature.

Exogenous, post-settlement selection appears to play a prominent role in the maintenance of the well-studied eastern oyster hybrid zone in Florida. In Cape Canaveral, the region spanning the sharpest differentiation in the cline, recruitment is substantially lower than in the North and South and likely assists in creating an incomplete barrier to gene flow. This study represents the first documentation of the levels of intermediate individuals in spat and adults for this hybrid zone. Substantial levels of intermediates surviving to the spat stage, and some even to adulthood, argue against strong endogenous viability selection. Further study of candidate loci for regionally selective differentiation in a cohort analysis detected strong viability selection against non-native-like genotypes between spat and adult life stages. This regional dependence on the direction of selection in intermediate and parental genotypes suggests environmental differences play a larger role than endogenous genomic incompatibilities to maintain differentiation in the face of

possibly broad dispersal and gene flow.  The documentation of natural intermediates and

viability selection against non-native-like genotypes will be useful for designing

appropriate and successful conservation and restoration projects for *Crassostrea*

*virginica*, a species that provides important economic and ecosystem services.

**Tables**

Table 3-1.A: Adult and spat transect sites in order along the coastline from North to South to the Gulf of Mexico. Cline side indicates the side of the previously identified cline (Hare & Avise 1996) on which the site exists. If a sample was used for AFLP genetic analysis, AFLP data collection states the collection abbreviation determined by the site abbreviation, the month, and last digit of the year the samples were collected. AFLP data collection is only the site abbreviation for adult samples. An 'X' indicates if the site was used for a particular transect and the year collected. Note that the Titusville sample collected in 2006, marked by an X-a, consisted of adult individuals.

| Cline Side | Region | AFLP data collection | Site | Latitude (oN) | Longitude (oW) | 2005 spat | 2006 spat | 2004 adults | 2002 adults |
|---|---|---|---|---|---|---|---|---|---|
| Atlantic | North | | FL Fish & Wildlife, Jacksonville Beach | 30.290 | 81.419 | X | | | |
| Atlantic | North | | Pine St., St. Augustine | 29.903 | 81.313 | X | | | |
| Atlantic | North | BNB76 | B&B Woodworking, St. Augustine | 29.881 | 81.323 | | X | | |
| Atlantic | North | | Devil's Elbow, Crescent Beach | 29.753 | 81.250 | X | | | |
| Atlantic | North | WHI, WHI85, WHI76 | Whitney Lab Dock, Marineland | 29.670 | 81.216 | X | X | X | |
| Atlantic | North | | Flagler by the Sea, Beverly Beach | 29.527 | 81.154 | X | | | |
| Atlantic | North | | High Bridge Park, Mound Grove | 29.408 | 81.100 | X | X | | |
| Atlantic | North | | Cassen Marina, Ormond Beach | 29.286 | 81.054 | X | | | |
| Atlantic | North | | Bethune Point Park, Daytona | 29.201 | 81.100 | X | | | |
| Atlantic | North | | Rose Bay, Port Orange | 29.103 | 80.972 | X | | | |
| Atlantic | North | NSB | New Smyrna Beach | 29.080 | 80.956 | | | X | X |
| Atlantic | North | | Critter Fleet, Ponce Inlet | 29.079 | 80.929 | X | | | |
| Atlantic | North | EDG85, EDG76 | George Kennedy Park, Edgewater | 28.994 | 80.904 | X | X | | |
| Atlantic | North | | River Breeze Park, Ariel | 28.898 | 80.852 | X | X | | |

Table 3-1.A continued.

| Cline Side | Region | AFLP data collection | Site | Latitude (oN) | Longitude (oW) | 2005 spat | 2006 spat | 2004 adults | 2002 adults |
|---|---|---|---|---|---|---|---|---|---|
| Gulf | Cape Canaveral | | LeFill's Fish Camp, Oak Hill | 28.866 | 80.833 | X | | | |
| Gulf | Cape Canaveral | | WSEG boat ramp, Mosquito Lagoon | 28.790 | 80.788 | X | | | |
| Gulf | Cape Canaveral | | Beacon boat ramp, Canaveral National Seashore | 28.756 | 80.765 | X | X | | |
| Gulf | Cape Canaveral | | Live Oak Rd, Canaveral National Seashore | 28.749 | 80.771 | X | | | |
| Gulf | Cape Canaveral | | Haulover Canal East end shells 2 | 28.740 | 80.751 | X | X | | |
| Gulf | Cape Canaveral | | Haulover Canal East end shells | 28.740 | 80.751 | X | | | |
| Gulf | Cape Canaveral | | Haulover Canal East end plates | 28.738 | 80.750 | X | | | |
| Gulf | Cape Canaveral | | Haulover Canal Boat Ramp | 28.733 | 80.757 | X | | | |
| Gulf | Cape Canaveral | | Haulover Canal Boat Ramp shells | 28.733 | 80.757 | X | X | | |
| Gulf | Cape Canaveral | | MINS Post 8, Canaveral National Seashore | 28.676 | 80.773 | X | | | |
| Gulf | Cape Canaveral | | MINS Post 4, Canaveral National Seashore | 28.656 | 80.778 | X | | | |
| Gulf | Cape Canaveral | TTV | Titusville | 28.618 | 80.803 | | X-a | | |
| Gulf | Cape Canaveral | PSJ76 | Port St. John Boat Ramp | 28.475 | 80.767 | X | X | | |
| Gulf | Cape Canaveral | LPA | Launch Pad A, Cape Canaveral | 28.437 | 80.582 | | | X | |
| Gulf | Cape Canaveral | | Bennett Bridge, Cocoa | 28.404 | 80.729 | X | | | |
| Gulf | Cape Canaveral | | Kelly Park, Banana River | 28.402 | 80.662 | X | X | | |
| Gulf | Cape Canaveral | PSH85, PSH76 | Palm Shores boat ramp, Palm Shores | 28.207 | 80.661 | X | X | | |
| Gulf | Cape Canaveral | | Eau Gallie Pier, Eau Gallie | 28.132 | 80.626 | X | | | |
| Gulf | South | | Melbourne Beach Pier, Melbourne Beach | 28.068 | 80.566 | X | | | |
| Gulf | South | | John Jorgensen's Landing, Grant | 27.938 | 80.530 | X | X | | |
| Gulf | South | | Sebastian Inlet, Sebastian | 27.847 | 80.442 | X | | | |
| Gulf | South | | Roseland at 83rd Ave. & 129th St., Sebastian River | 27.835 | 80.498 | | X | | |
| Gulf | South | | Riverview Park, Sebastian | 27.809 | 80.463 | | X | | |
| Gulf | South | | Sebastian | 27.754 | 80.427 | | | X | |
| Gulf | South | VER85, VER76 | McWilliams Park, Vero Beach | 27.654 | 80.369 | X | X | | |
| Gulf | South | | Jack Island Reserve, Fort Pierce | 27.502 | 80.309 | X | | | |
| Gulf | South | FTP | Fort Pierce | 27.491 | 80.315 | | | X | |
| Gulf | South | | A1A North park, Ft. Pierce | 27.474 | 80.322 | | X | | |
| Gulf | South | | Rtes 707 & 732, Jensen Beach | 27.251 | 80.227 | X | | | |
| Gulf | South | | FL Oceanographic Coastal Center, Jensen Beach | 27.214 | 80.184 | | X | | |
| Gulf | South | | Shepard Park, Stuart | 27.197 | 80.259 | | X | | |
| Gulf | South | HOB | Hobe Sound | 27.100 | 80.141 | | | X | |
| Gulf | South | HOB85 | Greenfield Park, Hobe Sound | 27.055 | 80.121 | X | | | |
| Gulf | South | JUP76 | Burt Reynolds Park, Jupiter | 26.941 | 80.085 | X | X | | |
| Gulf | South | | Juno Park, Juno | 26.853 | 80.066 | X | | | |
| Gulf | South | WPB85 | Currie Park, West Palm Beach | 26.736 | 80.050 | X | X | | |
| Gulf | South | | Memorial at bridge, West Palm Beach | 26.704 | 80.049 | | X | | |
| Gulf | South | | Boynton Beach Park, Boynton Beach | 26.546 | 80.053 | | X | | |
| Gulf | South | | Deerfield Park, Deerfield Beach | 26.315 | 80.081 | | X | | |
| Gulf | South | | A1A & Canal St., Pompano Beach | 26.253 | 80.087 | | X | | |
| Gulf | South | | John U. Lloyd Park ramp, Ft. Lauderdale area | 26.082 | 80.112 | | X | | |
| Gulf | South | | Oleta River Park, North Miami area | 25.926 | 80.136 | | X | | |
| Gulf | South | PCH | Port Charlotte | 26.429 | 81.867 | | | | X |

210

Table 3-1.B: Adult and spat transect sites in order along the coastline from North to South to the Gulf of Mexico. Cline side indicates the side of the previously identified cline (Hare & Avise 1996) on which the site exists. If a sample was used for AFLP genetic analysis, AFLP data collection states the collection abbreviation determined by the site abbreviation, the month, and last digit of the year the samples were collected. AFLP data collection is only the site abbreviation for adult samples. Mean salinity values in ppt were averaged over all site visits within a given year, where blank values indicate no salinity observations made. Recruitment proliferation is indicated by the percent of the collection goal (60 and 100 spat per site visit in 2005 and 2006, respectively) averaged over all site visits within a given year. Number of spat observed or collected up to the collection goal (n) is presented for each site-visit, where blank values indicate no observation made.

| Cline Side | Region | AFLP data collection | Site | 2005 Mean Salinity (ppt) | 2006 Mean Salinity (ppt) | 2005 Avg % of 60 spat goal | 2006 Avg % of 100 spat goal | 4/05 n | 5/05 n | 8/05 n | 7/06 n | 8/06 n |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Atlantic | North | | FL Fish & Wildlife, Jacksonville Beach | 16.9 | | 33% | | 0 | 0 | 60 | | |
| Atlantic | North | | Pine St., St. Augustine | 31.9 | | 0% | | 0 | 0 | | | |
| Atlantic | North | BNB76 | B&B Woodworking, St. Augustine | | 34.0 | | 100% | | | | 100 | |
| Atlantic | North | | Devil's Elbow, Crescent Beach | 31.7 | | 35% | | 0 | 3 | 60 | | |
| Atlantic | North | WHI, WHI85, WHI76 | Whitney Lab Dock, Marineland | 30.8 | 33.0 | 35% | 100% | 0 | 3 | | 100 | |
| Atlantic | North | | Flagler by the Sea, Beverly Beach | 28.3 | | 61% | | 0 | 55 | 55 | | |
| Atlantic | North | | High Bridge Park, Mound Grove | 25.0 | 32.7 | 35% | 71% | 0 | 3 | 60 | 42 | 100 |
| Atlantic | North | | Cassen Marina, Ormond Beach | 20.0 | | 0% | | 0 | 0 | | | |
| Atlantic | North | | Bethune Point Park, Daytona | 20.7 | | 0% | | 0 | 0 | | | |
| Atlantic | North | | Rose Bay, Port Orange | 22.4 | | 1% | | 0 | 1 | | | |
| Atlantic | North | NSB | New Smyrna Beach | | | | | | | | | |
| Atlantic | North | | Critter Fleet, Ponce Inlet | 31.5 | | 67% | | 0 | 60 | 60 | | |
| Atlantic | North | EDG85, EDG76 | George Kennedy Park, Edgewater | 31.1 | 34.8 | 33% | 85% | 0 | 0 | 60 | 85 | |
| Atlantic | North | | River Breeze Park, Ariel | 31.5 | 35.2 | 50% | 25% | | 0 | 60 | 13 | 36 |

Table 3-1.B continued.

| Cline Side | Region | AFLP data collection | Site | 2005 Mean Salinity (ppt) | 2006 Mean Salinity (ppt) | 2005 Avg % of 60 spat goal | 2006 Avg % of 100 spat goal | 4/05 n | 5/05 n | 8/05 n | 7/06 n | 8/06 n |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gulf | Cape Canaveral | | LeFill's Fish Camp, Oak Hill | 30.6 | | 0% | | 0 | 0 | | | |
| Gulf | Cape Canaveral | | WSEG boat ramp, Mosquito Lagoon | 29.4 | | 0% | | 0 | 0 | 0 | | |
| Gulf | Cape Canaveral | | Beacon boat ramp, Canaveral National Seashore | 30.7 | 32.6 | 0% | 1% | | 0 | 0 | 1 | 0 |
| Gulf | Cape Canaveral | | Live Oak Rd, Canaveral National Seashore | 29.7 | | 0% | | | 0 | 0 | | |
| Gulf | Cape Canaveral | | Haulover Canal East end shells 2 | 29.6 | 28.9 | 0% | 0% | | 0 | | | 0 |
| Gulf | Cape Canaveral | | Haulover Canal East end shells | 28.9 | | | | | | | | |
| Gulf | Cape Canaveral | | Haulover Canal East end plates | 28.7 | | 1% | | 0 | 1 | 0 | | |
| Gulf | Cape Canaveral | | Haulover Canal Boat Ramp | 29.1 | | 3% | | 0 | 6 | 0 | | |
| Gulf | Cape Canaveral | | Haulover Canal Boat Ramp shells | 29.9 | 31.7 | 0% | 2% | 0 | 0 | | 4 | 0 |
| Gulf | Cape Canaveral | | MINS Post 8, Canaveral National Seashore | 23.9 | | 0% | | 0 | 0 | 0 | | |
| Gulf | Cape Canaveral | | MINS Post 4, Canaveral National Seashore | 13.2 | | 0% | | 0 | 0 | 0 | | |
| Gulf | Cape Canaveral | TTV | Titusville | | | | | | | | | |
| Gulf | Cape Canaveral | PSJ76 | Port St. John Boat Ramp | 23.7 | 19.5 | 2% | 100% | | 2 | 0 | 100 | |
| Gulf | Cape Canaveral | LPA | Launch Pad A, Cape Canaveral | | | | | | | | | |
| Gulf | Cape Canaveral | | Bennett Bridge, Cocoa | 20.9 | | 0% | | 0 | 0 | | | |
| Gulf | Cape Canaveral | | Kelly Park, Banana River | 18.1 | 19.3 | 0% | 0% | 0 | 0 | | 0 | 0 |
| Gulf | Cape Canaveral | PSH85, PSH76 | Palm Shores boat ramp, Palm Shores | 18.5 | 23.2 | 13% | 99% | 0 | 1 | 23 | 99 | |
| Gulf | Cape Canaveral | | Eau Gallie Pier, Eau Gallie | 17.1 | | 0% | | 0 | 0 | | | |
| Gulf | South | | Melbourne Beach Pier, Melbourne Beach | 14.5 | | 7% | | 0 | 8 | | | |
| Gulf | South | | John Jorgensen's Landing, Grant | 23.4 | 27.7 | 2% | 18% | 0 | 2 | | 29 | 6 |
| Gulf | South | | Sebastian Inlet, Sebastian | 29.1 | | 9% | | 2 | 9 | | | |
| Gulf | South | | Roseland at 83rd Ave. & 129th St., Sebastian River | | 12.5 | | 0% | | | | 0 | 0 |
| Gulf | South | | Riverview Park, Sebastian | | 26.4 | | 13% | | | | | 13 |
| Gulf | South | | Sebastian | | | | | | | | | |
| Gulf | South | VER85, VER76 | McWilliams Park, Vero Beach | 20.8 | 27.5 | 67% | 100% | 0 | 60 | 60 | 100 | |
| Gulf | South | | Jack Island Reserve, Fort Pierce | 29.6 | | 1% | | 0 | | | | |
| Gulf | South | FTP | Fort Pierce | | | | | | | | | |
| Gulf | South | | A1A North park, Ft. Pierce | | 33.5 | | 76% | | 1 | | 51 | 100 |
| Gulf | South | | Rtes 707 & 732, Jensen Beach | 25.9 | | 9% | | 0 | 11 | | 100 | 100 |
| Gulf | South | | FL Oceanographic Coastal Center, Jensen Beach | | 32.3 | | 100% | | | | 100 | 100 |
| Gulf | South | | Shepard Park, Stuart | | 18.2 | | 100% | | | | 100 | |
| Gulf | South | HOB | Hobe Sound | | | | | | | | | |
| Gulf | South | HOB85 | Greenfield Park, Hobe Sound | 30.4 | | 68% | | 60 | 3 | 60 | | |
| Gulf | South | JUP76 | Burt Reynolds Park, Jupiter | 29.0 | 29.9 | 18% | 100% | 0 | 0 | 33 | 100 | |
| Gulf | South | | Juno Park, Juno | 28.9 | | 36% | | 2 | 3 | 60 | | |
| Gulf | South | WPB85 | Currie Park, West Palm Beach | 28.0 | | 33% | | 0 | 0 | 60 | | |
| Gulf | South | | Memorial at bridge, West Palm Beach | | 26.3 | | 100% | | | | | 100 |
| Gulf | South | | Boynton Beach Park, Boynton Beach | | 22.9 | | 100% | | | | | 100 |
| Gulf | South | | Deerfield Park, Deerfield Beach | | 11.6 | | 50% | | | | 99 | 0 |
| Gulf | South | | A1A & Canal St., Pompano Beach | | 25.0 | | 35% | | | | 19 | 50 |
| Gulf | South | | John U. Lloyd Park ramp, Ft. Lauderdale area | | 30.1 | | 91% | | | | 91 | |
| Gulf | South | | Oleta River Park, North Miami area | | 11.6 | | 100% | | | | 100 | |
| Gulf | South | PCH | Port Charlotte | | | | | | | | | |
| | | | Min | 13.2 | 11.6 | 0% | 0% | | | | | |
| | | | Max | 31.9 | 35.2 | 68% | 100% | | | | | |
| | | | Avg | 25.9 | 26.4 | 17% | 63% | | | | | |

212

Table 3-2: A) Regional differences for each transect using DSO loci for the absolute value of the difference in regional mean allele frequencies (regional allele frequency difference, RAFD) and regional differentiation ($F_{RT}$). $F_{RT}$ estimates are weighted by the number of sites within each region. Here regions are divided by the cline (Atlantic, north of the cline; Gulf, south of the cline). B) P-values of paired t-tests among transects, with significant comparisons in bold. Comparisons between adult and spat transects are one-tailed tests. Comparisons between 2005 and 2006 spat are two-tailed tests. P-values reflect the level of significance for the appropriate number of tails.

A)

| Locus | RAFD | | | $F_{RT}$ | | |
|---|---|---|---|---|---|---|
| | Adult | Spat 85 | Spat 76 | Adult | Spat 85 | Spat 76 |
| 13 | 0.4652 | 0.2185 | 0.2341 | 0.1835 | 0.0646 | 0.0627 |
| 31 | 0.4506 | 0.2569 | 0.2081 | 0.1754 | 0.0859 | 0.0543 |
| 51 | 0.3181 | 0.0066 | 0.0280 | 0.1792 | 0.0034 | -0.0024 |
| 64 | 0.5085 | 0.1109 | 0.1791 | 0.2571 | 0.0037 | 0.0481 |
| 80 | 0.3586 | 0.1592 | 0.1594 | 0.1239 | 0.0378 | 0.0258 |
| 146 | 0.2727 | 0.2554 | 0.2010 | 0.0954 | 0.0385 | 0.0593 |
| 170 | 0.2654 | 0.0143 | 0.0386 | 0.0932 | -0.0104 | -0.0010 |
| **Mean** | **0.3770** | **0.1460** | **0.1498** | **0.1582** | **0.0319** | **0.0353** |

B)

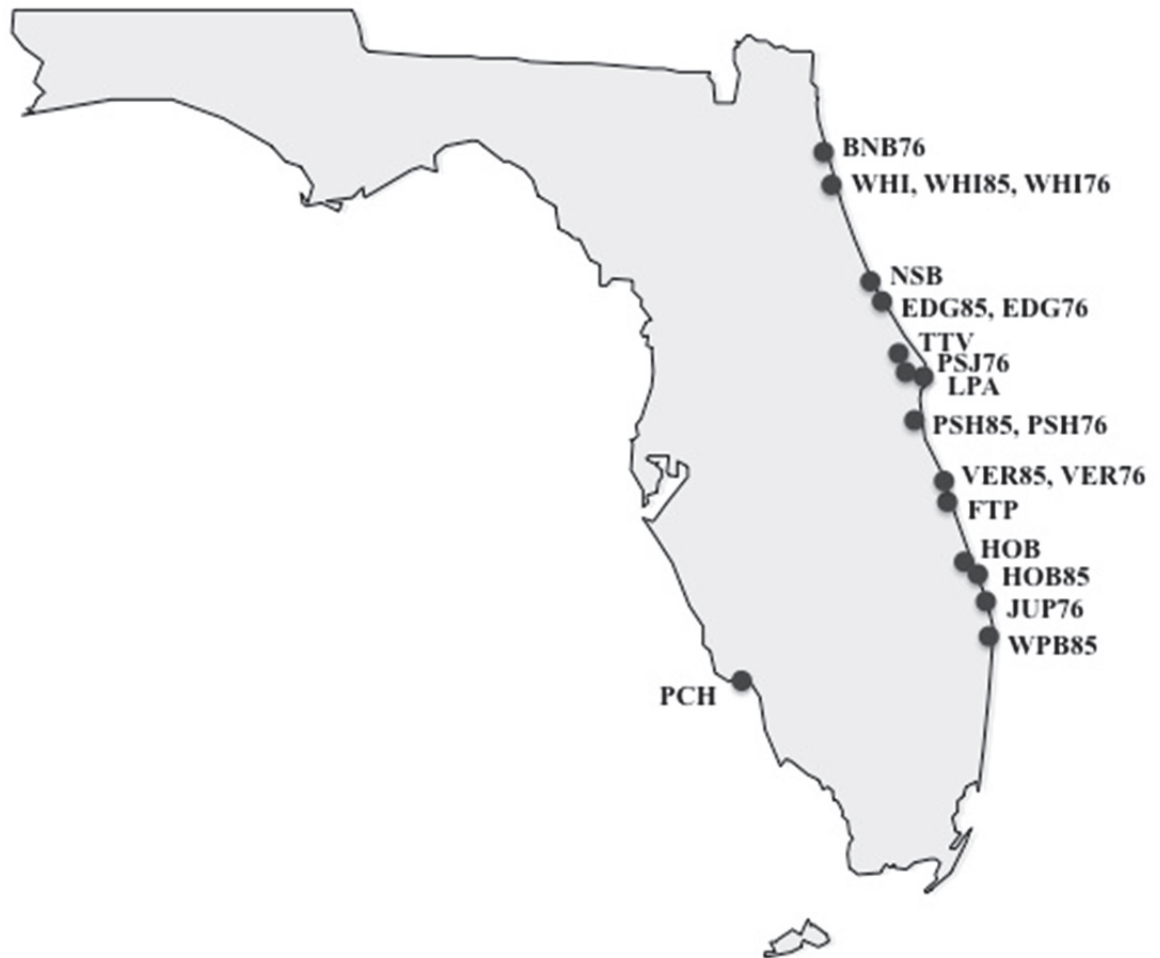| Transect A | Transect B | p-values | |
|---|---|---|---|
| | | RAFD | $F_{RT}$ |
| Adult | Spat 85 | **0.0001** | **0.0002** |
| Adult | Spat 76 | **< 0.0001** | **0.0002** |
| Spat 85 | Spat 76 | 0.4061 | 0.3300 |

**Figures**



Fig. 3-1: Sampling sites for genetic studies. Adult, 2005 spat, and 2006 spat are represented by three-letter codes, three-letter codes followed by 85, and three-letter codes followed by 76, respectively. Sites sampled for multiple transects have a list of those transects.
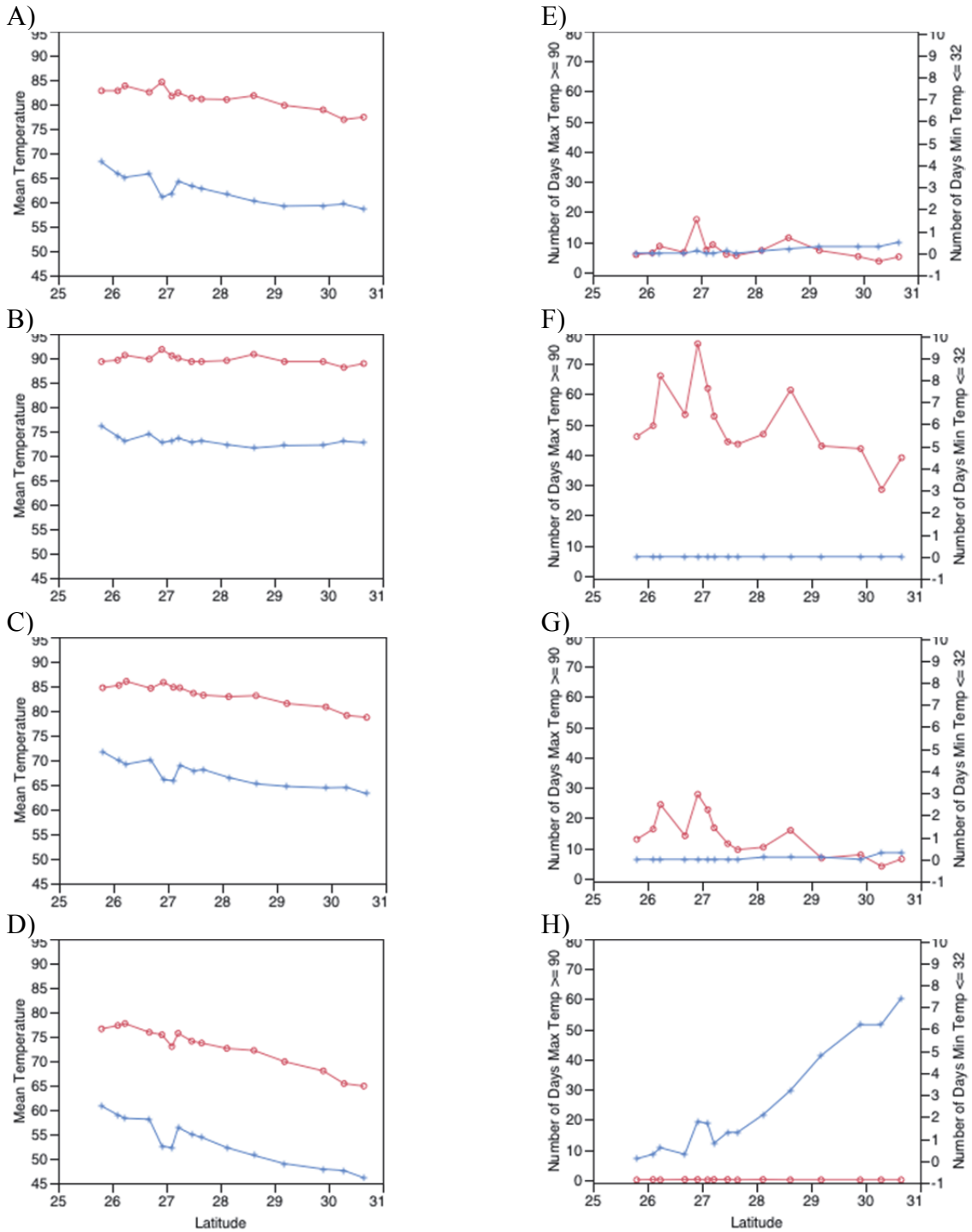
Fig. 3-2: Climate data for temperature statistics across latitudes for 1971-2000. A-D) Mean minimum (blue) and mean maximum (red) temperatures ($^o$F). E-H) Number of days when the maximum temperature reached or surpassed 90$^o$F along the left y-axis (red); number of days when the minimum temperature was equal or less than 32$^o$F along the right y-axis (blue). A,E) Spring: March – May. B,F) Summer: June – August. C,G) Fall: September – November. D,H) Winter: December – January. Data source: Southeast Regional Climate Center, www.sercc.com/climateinfo/historical/historical_fl.html).

215

A)
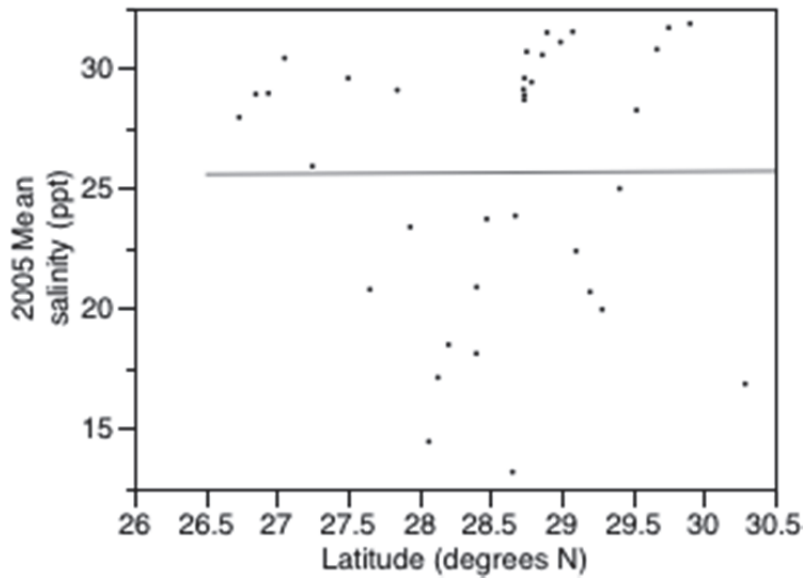


B)
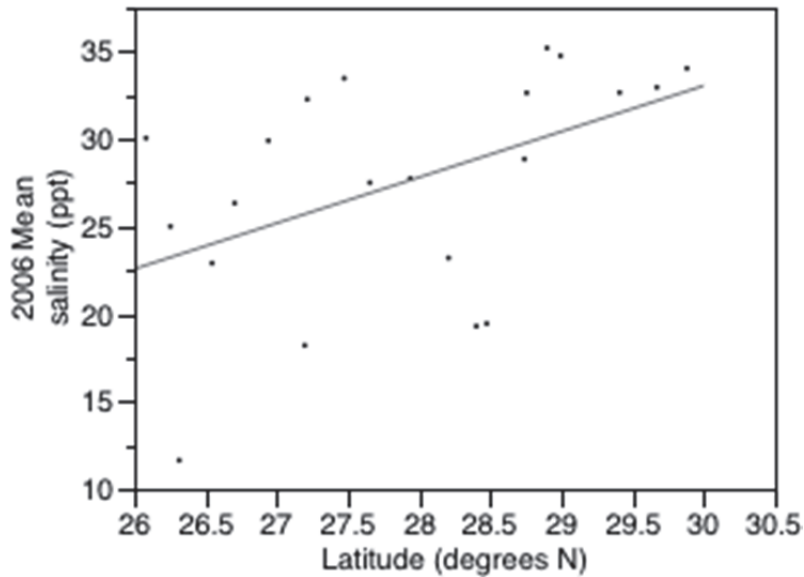


Fig. 3-3: Mean measured salinity at field sites for 2005 and 2006 by latitude (°N). Each dot indicates the mean salinity for a single field site. The solid gray line indicates the linear regression line. A) 2005 transect data: $R^2 = 0$, $p = 0.9742$; $y = 24.67 + 0.0347$(Latitude). B) 2006 transect data: $R^2 = 0.2283$, $p = 0.0285$; $y = -45.4769 + 2.6172$(Latitude).
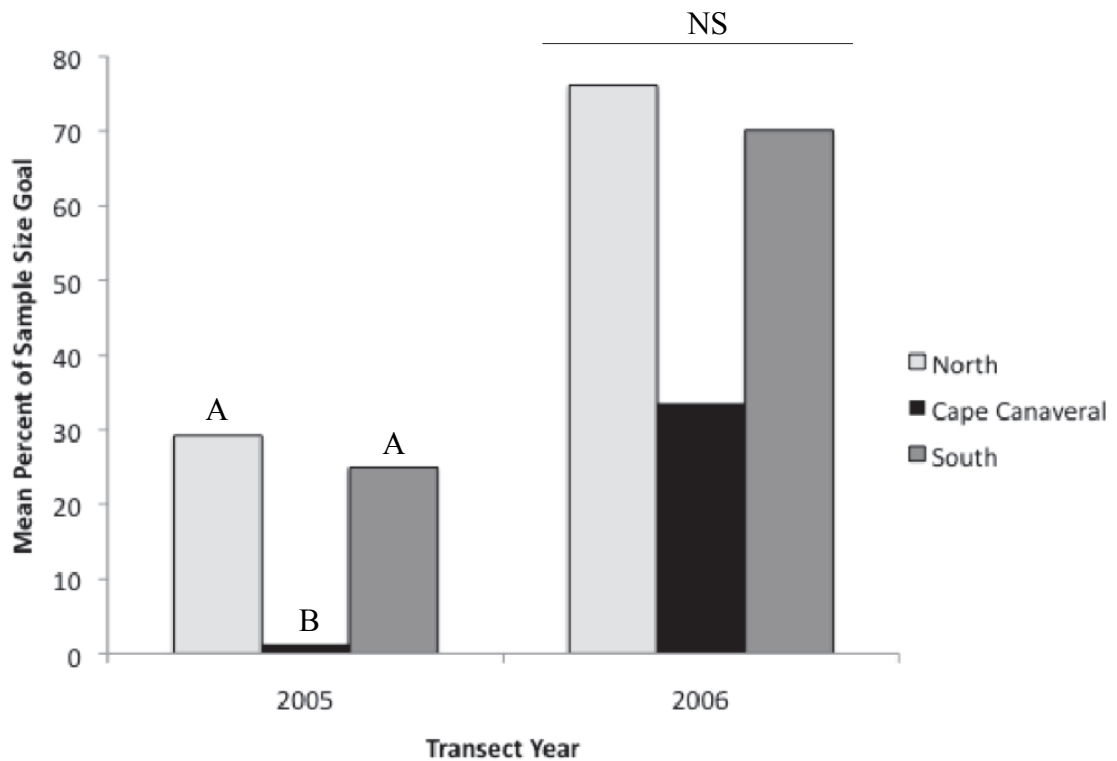
Fig. 3-4: Mean percent of the sample size goal met, averaged among sites within each of three regions, for settled spat collected from transects in 2005 and 2006. Letters over bars indicate significant differences among regions in 2005. Mean percents were not significantly different among regions in 2006.
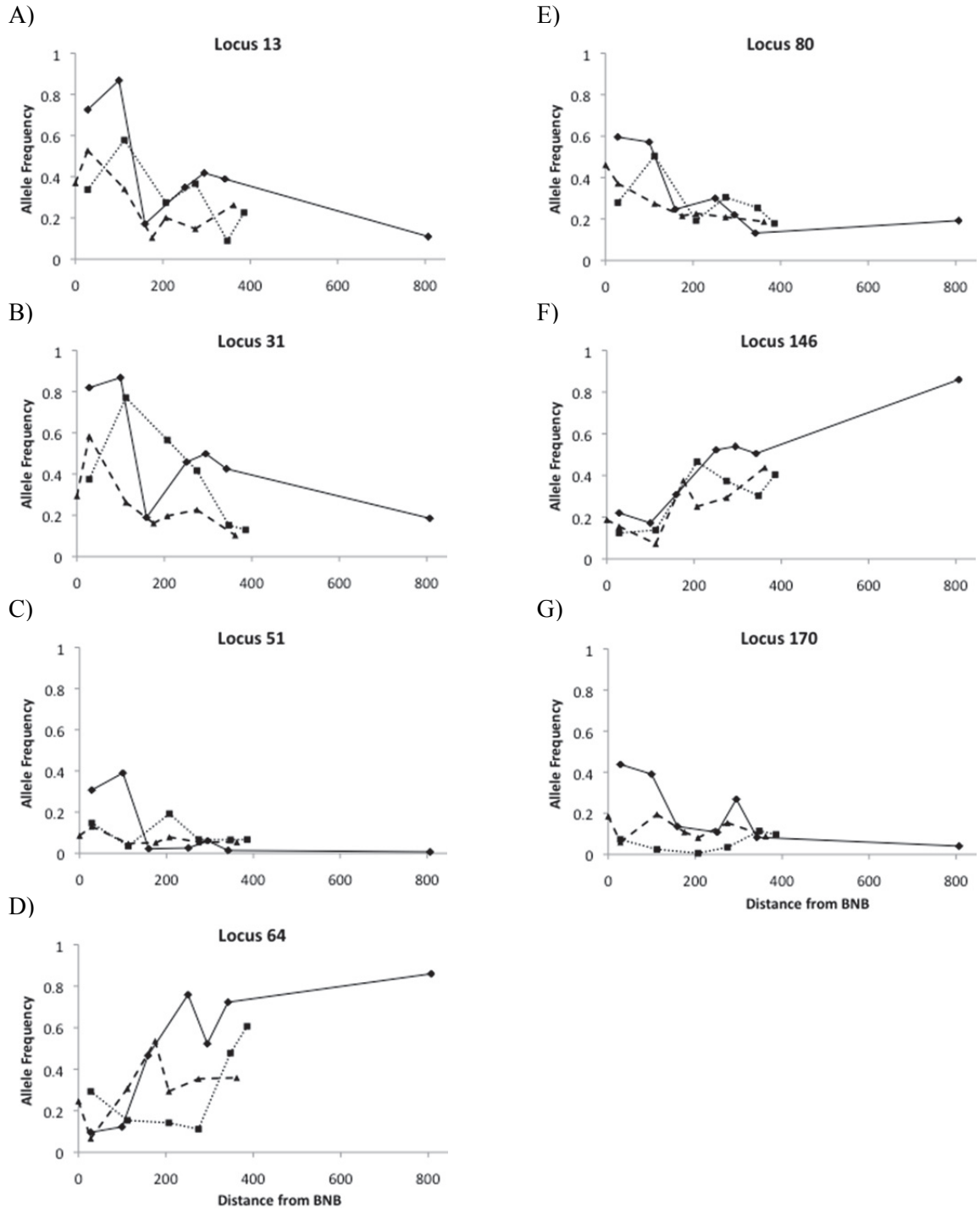
A)



E)



B)



F)



C)



G)



D)



Fig. 3-5: A-G) Allele frequencies of the presence allele for each DSO locus along each transect. Solid, dashed, and dotted lines represent adult, 2005 spat, and 2006 spat, respectively. Note: The center of the previously documented step cline is 142 km from BNB (Hare & Avise 1996).
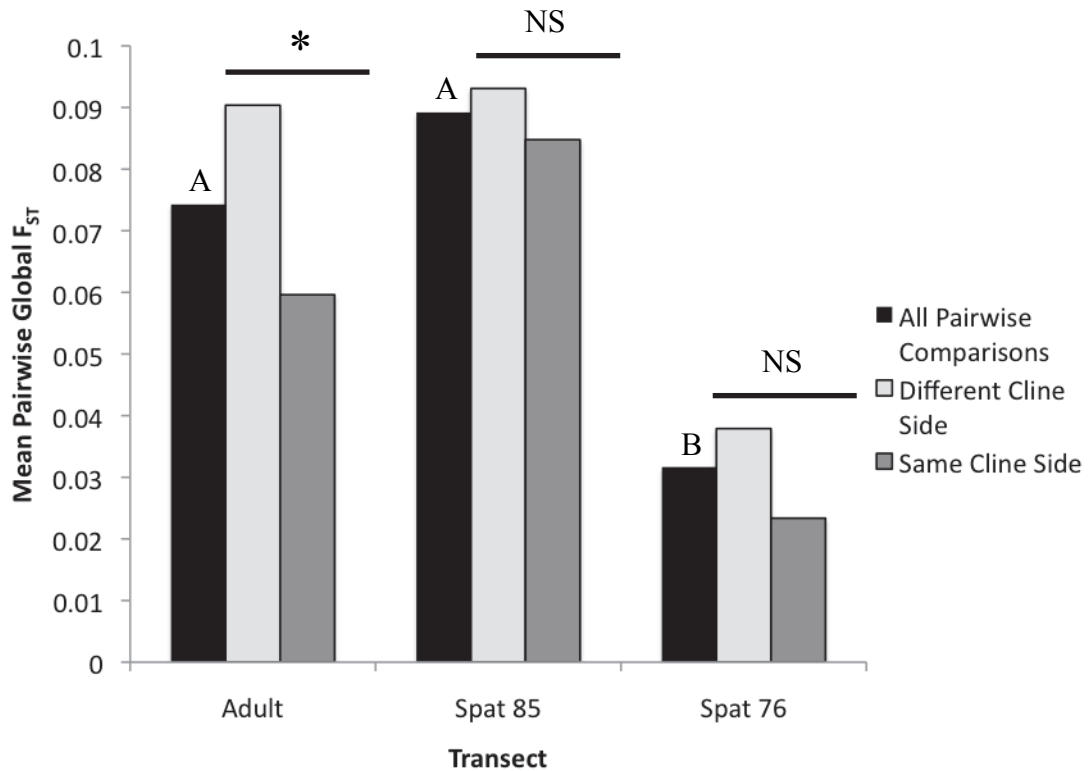
Fig. 3-6: Mean global $F_{ST}$. Global estimates averaged over all pairwise population comparisons, comparisons between populations on different sides of the cline, and those on the same side. Letters indicate significant differences among all pairwise comparison estimates among transects. Asterisk over bar indicates one-tailed significantly greater global $F_{ST}$ for adult pairwise comparisons on different sides of the cline than for those on the same side. NS over bars indicates nonsignificant difference in global $F_{ST}$ between population comparisons on the same and different sides of the cline.
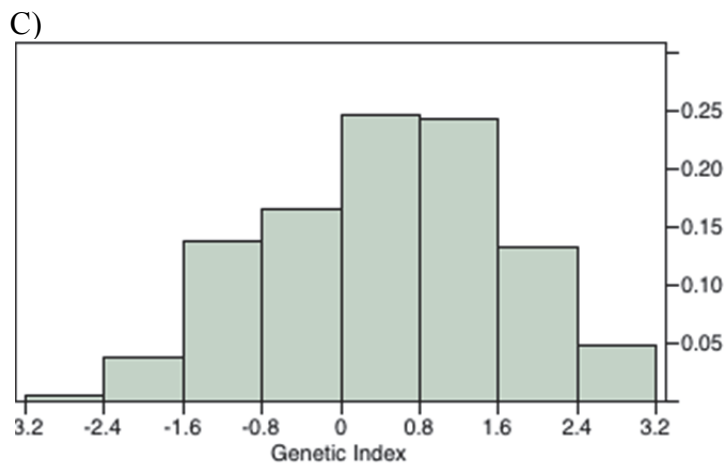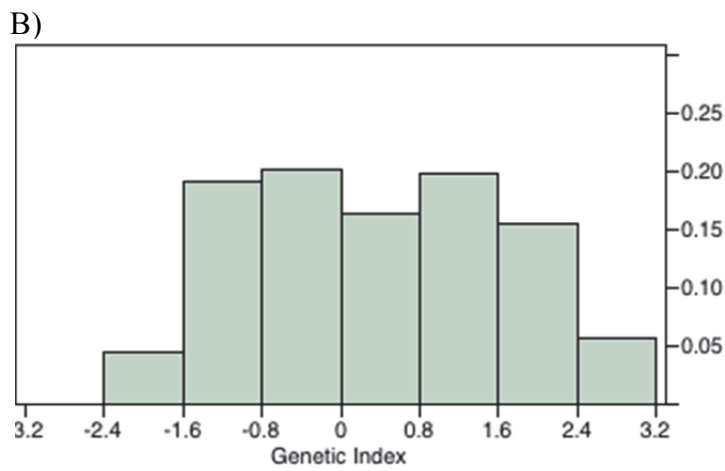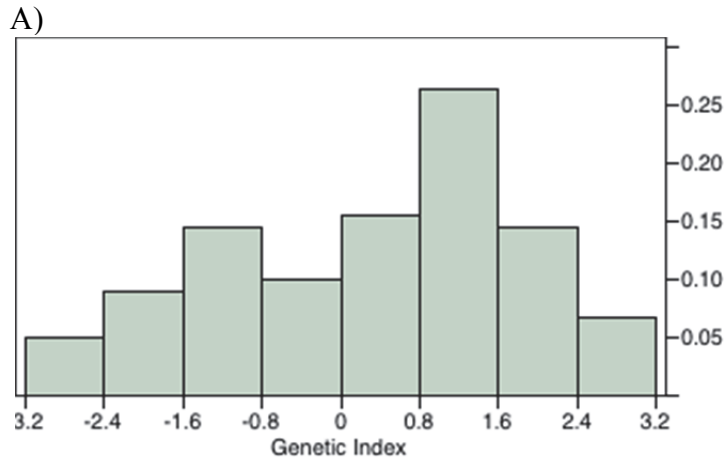
A)

B)

C)

Fig. 3-7: Distribution of genetic index values for each transect. Proportion of individuals with a genetic index value in a given binned range is shown along the y-axis. A) Adult transect; B) 2005 spat transect; C) 2006 spat transect.
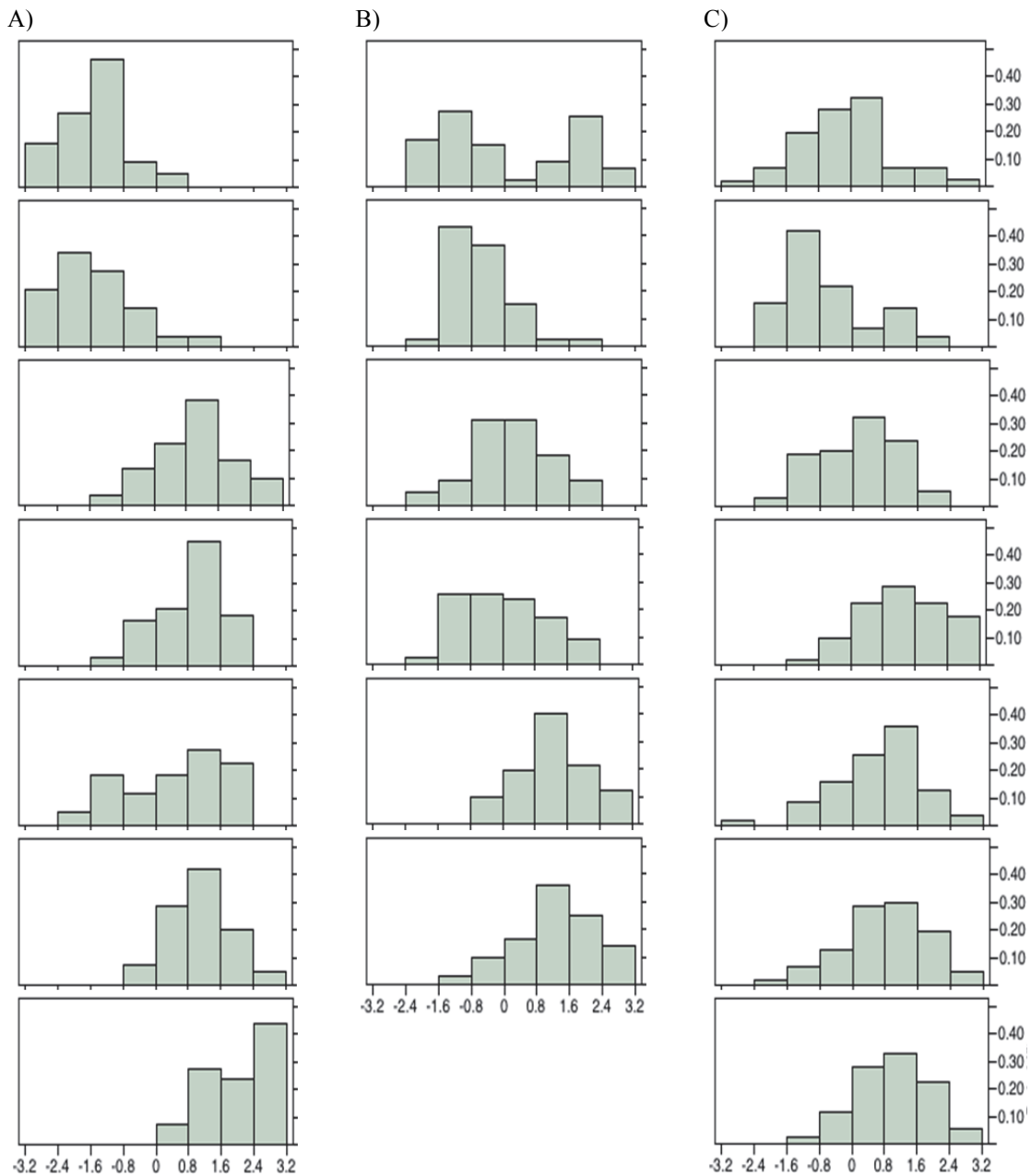
Fig. 3-8: Genotype index distributions for each site in each transect. The x-axis represents genotype index bins for DSO loci. The y-axis represents the proportion of individuals having genotype indices for each corresponding bin. Transects displayed from north to south: (A) Adult: WHI, NSB, TTV, LPA, FTP, HOB, PCH; B) Spat 85: WHI85, EDG85, PSH85, VER85, HOB85, WPB85; C) Spat 76: BNB76, WHI76, EDG76, PSJ76, PSH76, VER76, JUP76).
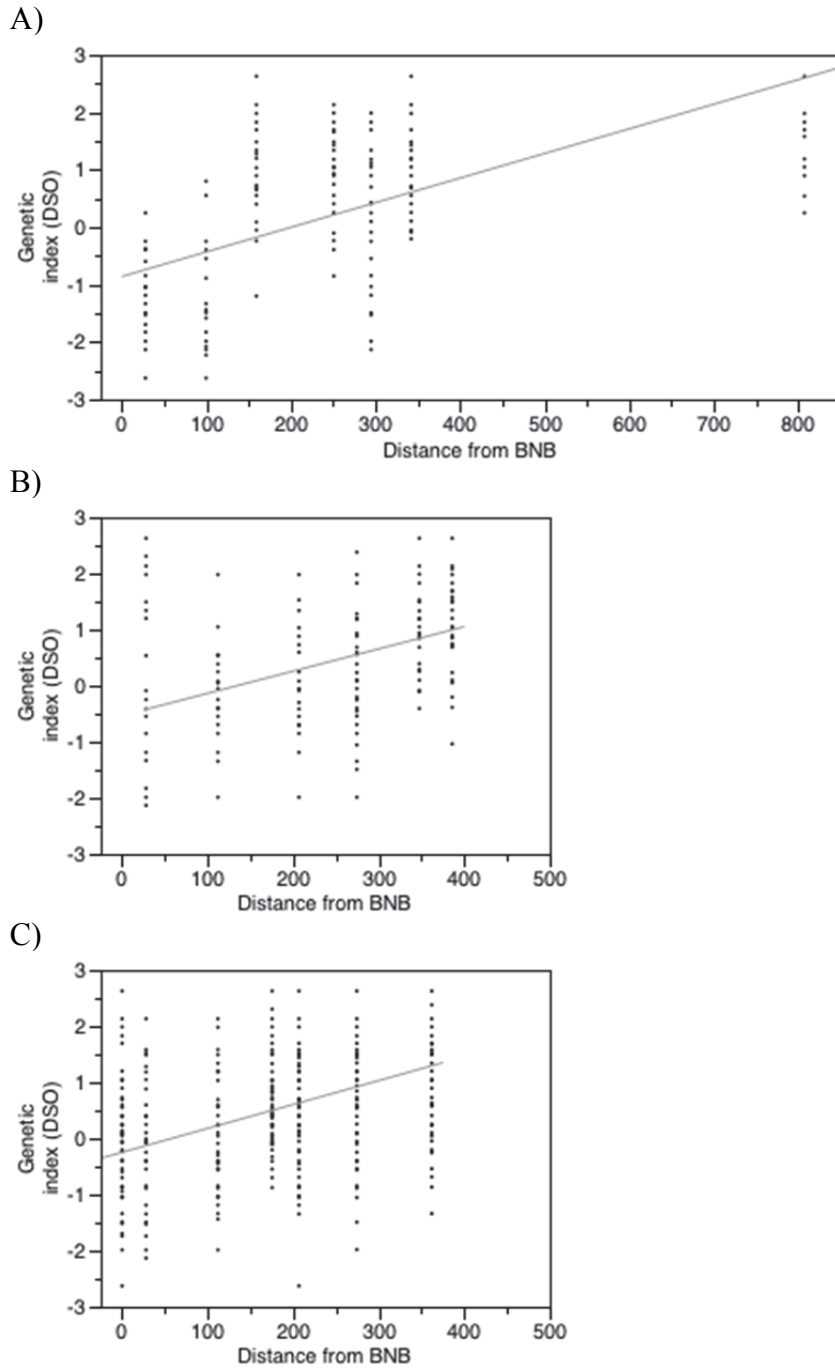
Fig. 3-9: Individual genetic index scores for DSO loci plotted on sample site distance (km) from St. Augustine (BNB). Best linear regression line in gray. A) Adult transect: $R^2 = 0.3975$, $p < 0.0001$; $y = -0.8595 + 0.0043(\text{Distance})$. B) Spat 85 transect: $R^2 = 0.1648$, $p < 0.0001$; $y = -0.5317 + 0.0040(\text{Distance})$. C) Spat 76 transect: $R^2 = 0.1887$, $p < 0.0001$; $y = -0.2468 + 0.0043(\text{Distance})$.
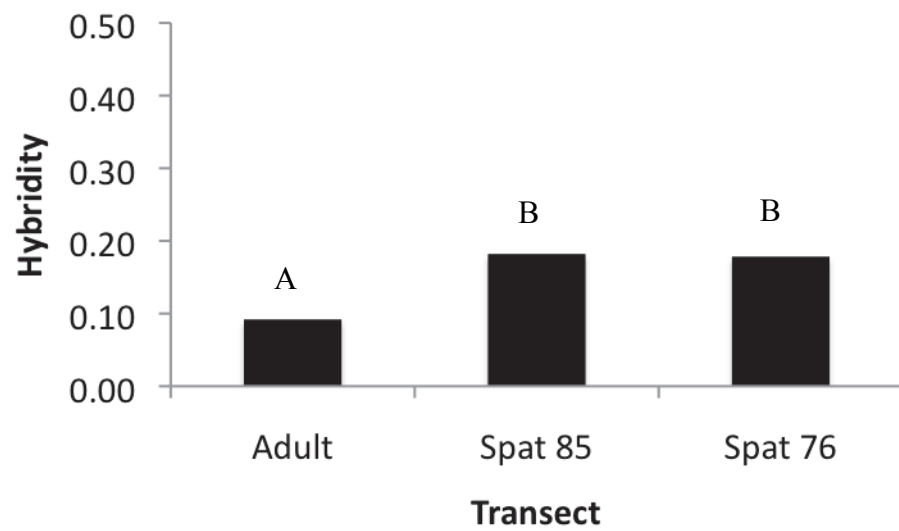
Fig. 3-10: Mean CLUMPP-based hybridity for each transect. Adult comparison to each spat transect: p < 0.0001. Spat 85 to Spat 76 comparison: p = 0.9610.

Fig. 3-11: Proportion of individuals in each transect belonging to Native-like (black), Intermediate (gray), and Migrant-like (white) source classes based on CLUMPP summary results. A) All regions grouped together. B-D) North, Cape Canaveral, and South regions, respectively.

# APPENDIX
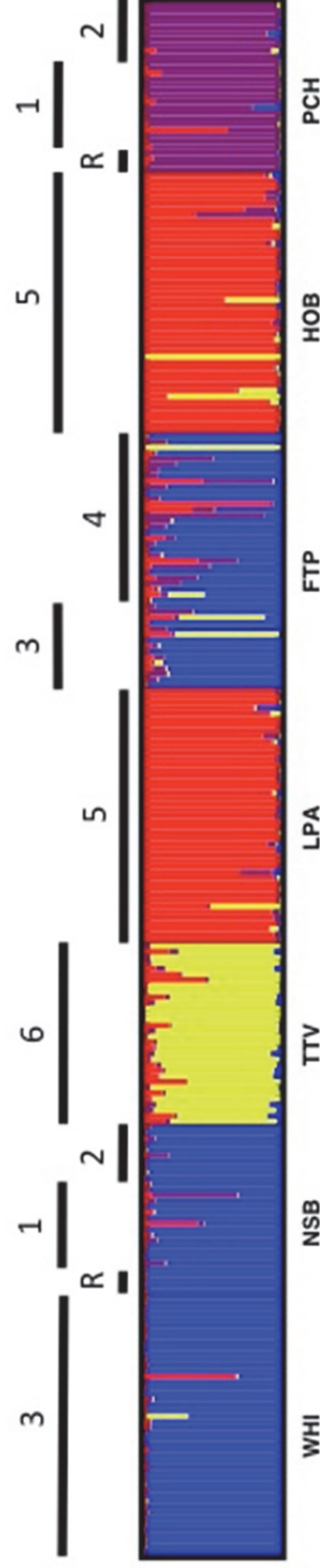


Fig. A-1: Bar plots of assignment test using CLUMPP summary results of a STRUCTURE model of K=4 for all 217 loci. Individual vertical ba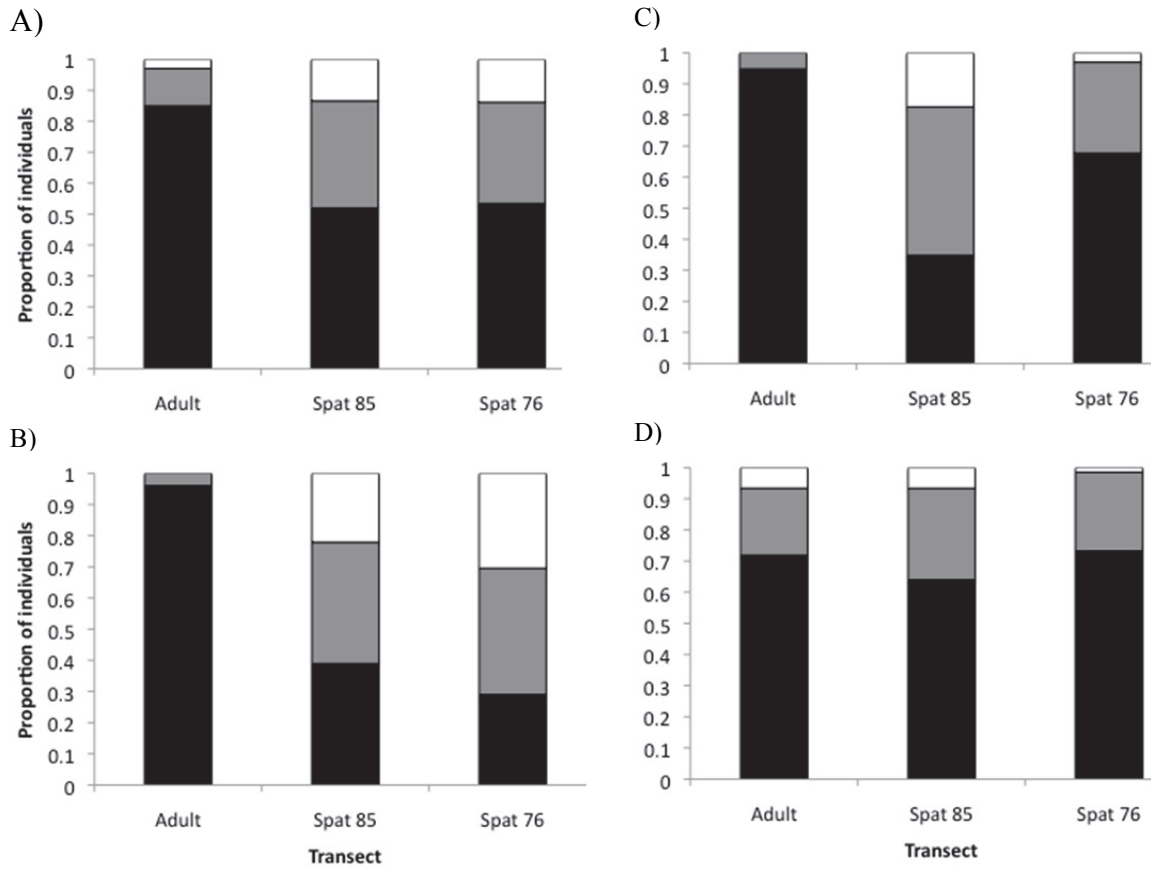rs represent individuals, which are grouped by collection site. Colors correspond to clusters, where multiple colors in a single individual represent proportions of mixed ancestry. Black horizontal bars overlay individuals from the same site whose AFLP profiles were analyzed on the same plate in the automated sequencer. The R denotes positive controls used as replicates on each plate. Numbers 1-6 correspond to the six distinct plates analyzed. Note that an ABI Prism 3100 Genetic Analyzer analyzed all plates except plate 6, which was analyzed by an ABI 3730xl DNA Analyzer (Applied Biosystems Inc.). On the 3100, each run was two lanes of a single 12-lane plate.
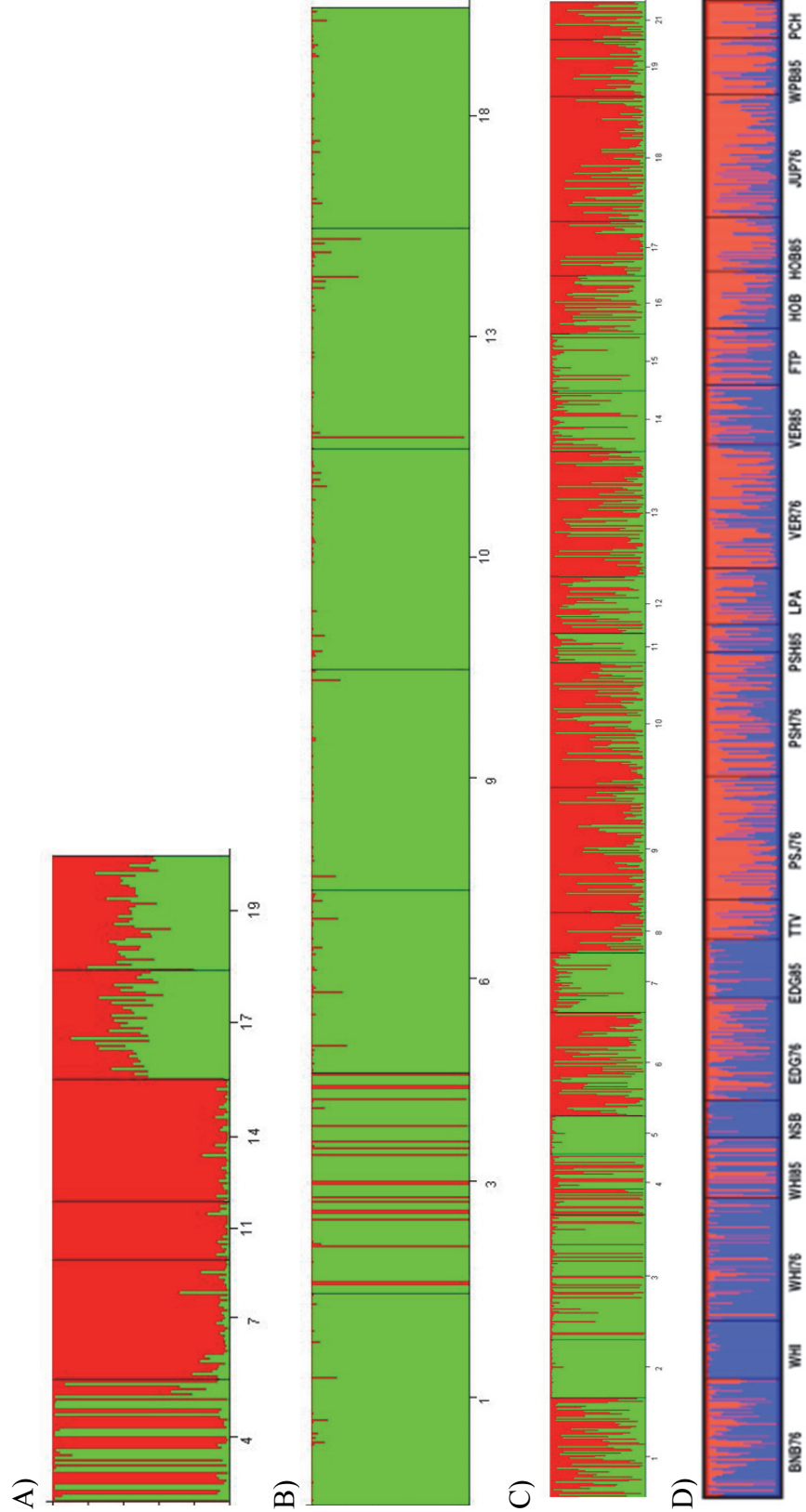
225

Fig. A-2: Bar plots of assignment tests using CLUMPP summary results of a STRUCTURE model of K=2. Analyses used no reference samples (A, B) or did use representative Atlantic- and Gulf-like adult individuals as reference samples for the two clusters (C, D). Individual vertical bars represent individuals, which are grouped by collection site. Sites are ordered from north to south. Colors correspond to clusters, where the amount of each color in a single individual represents the proportion of mixed ancestry belonging to that cluster. A) All 217 loci for Spat85. Site order: WHI85, EDG85, PSH85, VER85, FTP, HOB85, and WPB85. B) All 217 loci for Spat85. Site order: BNB76, WHI76, EDG76, PSJ76, PSH76, VER76, and JUP76. C) Outliers detected at the α = 0.01 level in at least two population comparisons that did not share a site (MI outliers, see Chapter 2). Site order: BNB76, WHI, WHI76, WHI85, NSB, EDG76, EDG85, TTV, PSJ76, PSH76, PSH85, LPA, VER76, VER85, FTP, HOB, HOB85, JUP76, WPB85, and PCH. D) DSO loci only for all adult and spat samples. Site order is the same as (C).

226

# LITERATURE CITED

Agresti, A., & Caffo, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *The American Statistician*, *54*(4), 280-288.

Akey, J. M., Eberle, M. A., Rieder, M. J., Carlson, C. S., Shriver, M. D., Nickerson, D. A. et al. (2004). Population history and natural selection shape patterns of genetic variation in 132 genes. *Plos BiologY*, *2*(10), 1591-1599-e286.

Albert, V., Jonsson, B., & Bernatchez, L. (2006). Natural hybrids in Atlantic eels (*Anguilla anguilla, A-rostrata*): evidence for successful reproduction and fluctuating abundance in space and time. *Molecular Ecology*, *15*(7), 1903-1916.

Albertson, R. C., Streelman, J. T., & Kocher, T. D. (2003). Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, *100*(9), 5252-5257.

Allee, W. C. (1931). *Animal aggregations, a study in general sociology.* Chicago, IL: University of Chicago Press.

Andolfatto, P. (2001). Adaptive hitchhiking effects on genome variability. *Current Opinion In Genetics & Development*, *11*(6), 635-641.

Arnold, M. L. (1992). Natural hybridization as an evolutionary process. *Annual Review Of Ecology And Systematics*, *23*, 237-261.

Arnold, M. L. (1997). *Natural Hybridization and Evolution.* Oxford: Oxford University Press.

Arnold, M. L., & Hodges, S. A. (1995). Are natural hybrids fit or unfit relative to their parents. *Trends In Ecology & Evolution*, *10*(2), 67-71.

Avise, J. C. (1992). Molecular population-structure and the biogeographic history of a regional fauna - a case-history with lessons for conservation biology. *Oikos*, *63*(1), 62-76.

Avise, J. C. (2000). *Phylogeography: The History and Formation of Species.* Cambridge, MA: Harvard University Press.

Avise, J. C. (2004). *Molecular Markers, Natural History, and Evolution* (2 ed.). Sunderland, MA: Sinauer Associates.

Balloux, F. (2001). EASYPOP (version 1.7): A computer program for the simulation of population genetics. *Journal Of Heredity*, *92*, 301-302.

Barber, B. J., Ford, S. E., & Wargo, R. N. (1991). Genetic-variation in the timing of gonadal maturation and spawning of the eastern oyster, *Crassostrea virginica* (Gmelin). *Biological Bulletin*, *181*(2), 216-221.

Barber, P. H., Palumbi, S. R., Erdmann, M. V., & Moosa, M. K. (2002). Sharp genetic breaks among populations of *Haptosquilla pulchella* (Stomatopoda) indicate limits to larval transport: patterns, causes, and consequences. *Molecular Ecology*, *11*(4), 659-674.

Barton, N. (2000). Genetic hitchhiking. *Philosophical Transactions Of The Royal Society B-Biological Sciences*, *355*(1403), 1553-1562.

Barton, N., & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridizing populations. *Heredity*, *57*, 357-376.

Barton, N. H., & Gale, K. S. (1993). Genetic analysis of hybrid zones. In R. G. Harrison (Ed.), *Hybrid Zones and the Evolutionary Process* (pp. 13-45). Oxford: Oxford University Press.

Barton, N. H., & Hewitt, G. M. (1981). The genetic-basis of hybrid inviability in the grasshopper *Podisma pedestris*. *Heredity*, *47*(Dec), 367-383.

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review Of Ecology And Systematics*, *16*, 113-148.

Barton, N. H., & Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature*, *341*(6242), 497-503.

Beaumont, M. A. (2005). Adaptation and speciation: what can F-st tell us? *Trends In Ecology & Evolution*, *20*(8), 435-440.

Beaumont, M. A., & Balding, D. J. (2004). Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, *13*(4), 969-980.

Beaumont, M. A., & Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings Of The Royal Society Of London Series B-Biological Sciences*, *263*(1377), 1619-1626.

Beckenbach, A. (1994). Mitochondrial haplotype frequencies in oysters: neutral alternatives to selection models. In G. Golding (Ed.), *Non-neutral Evolution: Theories and Molecular Data* (pp. 188-198). New York: Chapman & Hall.

Bensch, S., & Akesson, M. (2005). Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology*, *14*(10), 2899-2914.

Bensch, S., Helbig, A. J., Salomon, M., & Siebold, I. (2002). Amplified fragment length polymorphism analysis identifies hybrids between two subspecies of warblers. *Molecular Ecology*, *11*(3), 473-481.

Berec, L., Angulo, E., & Courchamp, F. (2007). Multiple Allee effects and population management. *Trends In Ecology & Evolution*, *22*(4), 185-191.

Berlocher, S. H. (2000). Radiation and divergence in the *Rhagoletis pomonella* species group: Inferences from allozymes. *Evolution*, *54*(2), 543-557.

Bert, T. M., & Arnold, W. S. (1995). An empirical-test of predictions of 2 competing models for the maintenance and fate of hybrid zones - both models are supported in a hard-clam hybrid zone. *Evolution*, *49*(2), 276-289.

Bert, T. M., Hesselman, D. M., Arnold, W. S., Moore, W. S., Cruzlopez, H., & Marelli, D. C. (1993). High-frequency of gonadal neoplasia in a hard clam (*Mercenaria* spp.) hybrid zone. *Marine Biology*, *117*(1), 97-104.

Bertness, M. D., & Gaines, S. D. (1993). Larval dispersal and local adaptation in acorn barnacles. *Evolution*, *47*(1), 316-320.

Bierne, N., Bonhomme, F., & David, P. (2003). Habitat preference and the marine-speciation paradox. *Proceedings Of The Royal Society Of London Series B-Biological Sciences*, *270*(1522), 1399-1406.

Bierne, N., Daguin, C., Bonhomme, F., David, P., & Borsa, P. (2003). Direct selection on allozymes is not required to explain heterogeneity among marker loci across a *Mytilus* hybrid zone. *Molecular Ecology*, *12*(9), 2505-2510.

Bierne, N., David, P., Langlade, A., & Bonhomme, F. (2002). Can habitat specialisation maintain a mosaic hybrid zone in marine bivalves? *Marine Ecology-Progress Series*, *245*, 157-170.

Black, W. C., IV, Baer, C. F., Antolin, M. F., & DuTeau, N. M. (2001). Population genomics: Genome-wide sampling of insect populations. *Annual Review of Entomology*, *46*, 441-469.

Bonin, A., Ehrich, D., & Manel, S. (2007). Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular Ecology*, *16*(18), 3737-3758.

Bonin, A., Taberlet, P., Miaud, C., & Pompanon, F. (2006). Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology And Evolution*, *23*(4), 773-783.

Brattstrom, O., Akesson, S., & Bensch, S. (2010). AFLP reveals cryptic population structure in migratory European red admirals (*Vanessa atalanta*). *Ecological Entomology*, *35*(2), 248-252.

Bridle, J. R., Vass-De-Zomba, J., & Butlin, R. K. (2002). Fine-scale ecological and genetic variation in a *Chorthippu*s grasshopper hybrid zone. *Ecological Entomology*, *27*(4), 499-504.

Bridle, J. R., & Vines, T. H. (2007). Limits to evolution at range margins: when and why does adaptation fail? *Trends In Ecology & Evolution*, *22*(3), 140-147.

Briggs, J. C. (1974). *Marine Zoogeography*. New York: McGraw-Hill.

Brocklesby, J. (1868). *Elements of Physical Geography, Together with a Treatise on the Physical Phenomena of the United States* (5). Philadelphia: E. H. Butler & Co.

Brumbaugh, R. D., & Coen, L. D. (2009). Contemporary approaches for small-scale oyster reef restoration to address substrate versus recruitment limitation: a review and comments relevant for the olympia oyster, *Ostrea lurida Carpenter 1864*. *Journal Of Shellfish Research*, *28*(1), 147-161.

Buerkle, C. A., & Rieseberg, L. H. (2001). Low intraspecific variation for genomic isolation between hybridizing sunflower species. *Evolution*, *55*(4), 684-691.

Bulmer, M. G. (1972). Multiple niche polymorphism. *American Naturalist*, *106*(948), 254-&.

Bumpus, D. F. (1973). A description of the circulation on the continental shelf of the east coast of the United States. *Progress in Oceanography*, *6*, 111-158.

Buroker, N. E. (1983). Population-genetics of the american oyster *Crassostrea virginica* along the Atlantic coast and the Gulf of Mexico. *Marine Biology*, *75*(1), 99-112.

Butlin, R. (2005). Recombination and speciation. *Molecular Ecology*, *14*(9), 2621-2635.

Butlin, R. K. (2010). Population genomics and speciation. *Genetica*, *138*(4), 409-418.

Caicedo, A. L., Stinchcombe, J. R., Olsen, K. M., Schmitt, J., & Purugganan, M. D. (2004). Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. *Proceedings Of The National Academy Of Sciences*, *101*(44), 15670-15675.

Cammen, K. (2007). *Temporal stability of the genetic cline observed in eastern oysters (*Crassostrea virginica*) along Florida's Atlantic coast.* Unpublished Senior Honors Thesis, University of Maryland--College Park, College Park, MD.

Campbell, D., & Bernatchez, L. (2004). Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology And EvolutioN*, *21*(5), 945-956.

Carney, S. E., Gardner, K. A., & Rieseberg, L. H. (2000). Evolutionary changes over the fifty-year history of a hybrid population of sunflowers (*Helianthus*). *Evolution*, *54*(2), 462-474.

Carriker, M. R., & Gaffney, P. M. (1996). Catalogue of selected species of living oysters (*Ostreacea*) of the world. In V. S. Kennedy, R. I. E. Newell, & A. F. Eble (Eds.), *The Eastern Oyster: Crassostrea virginica* (pp. 1-18). College Park, MD: Maryland Sea Grant College.

Case, R. A. J., Hutchinson, W. F., Hauser, L., Van Oosterhout, C., & Carvalho, G. R. (2005). Macro- and micro-geographic variation in pantophysin (PanI) allele frequencies in NE Atlantic cod *Gadus morhua*. *Marine Ecology-Progress Series*, *301*, 267-278.

Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, *15*(5), 538-543.

Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J. et al. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, *307*(5717), 1928-1933.

Conner, J. K., & Hartl, D. L. (2004). *A Primer of Ecological Genetics.* Sunderland: Sinauer Associates, Inc.

Conover, D. O., Clarke, L. M., Munch, S. B., & Wagner, G. N. (2006). Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. *Journal of Fish Biology*, *69*, 21-47.

Conover, D. O., & Schultz, E. T. (1995). Phenotypic similarity and the evolutionary significance of countergradient variation. *Trends in Ecology & Evolution*, *10*(6), 248-252.

Cordes, J. F., & Graves, J. E. (2003). Investigation of congeneric hybridization in and stock structure of weakfish (*Cynoscion regalis*) inferred from analyses of nuclear and mitochondrial DNA loci. *Fishery Bulletin*, *101*, 443-450.

Courchamp, F., Clutton-Brock, T., & Grenfell, B. (1999). Inverse density dependence and the Allee effect. *Trends in Ecology & Evolution*, *14*(10), 405-410.

Cowen, R. K., Paris, C. B., & Srinivasan, A. (2006). Scaling of connectivity in marine populations. *Science*, *311*(5760), 522-527.

Coyne, J. A., & Orr, H. A. (2004). *Speciation.* Sunderland, MA: Sinauer Associates, Inc.

Cruzan, M. B., & Arnold, M. L. (1994). Assortative mating and natural-selection in an iris hybrid zone. *Evolution*, *48*(6), 1946-1958.

Cunningham, C. W., & Collins, T. M. (1994). Developing model systems for molecular biogeography: vicariance and interchage in marine invertebrates. In B. Schierwater, B. Streit, G. P. Wagner, & R. DeSalle (Eds.), *Molecular Ecology and Evolution: Approaches and Applications* (pp. 405-433). Switzerland: Birkhauser Verlag Basel.

Diehl, W. J., & Koehn, R. K. (1985). Multiple-locus heterozygosity, mortality, and growth in a cohort of *Mytilus edulis*. *Marine Biology*, *88*(3), 265-271.

Dillon, R. T., & Manzi, J. J. (1989). Genetics and shell morphology in a hybrid zone between the hard clams *Mercenaria mercenaria* and *Mercenaria campechiensis*. *Marine Biology*, *100*(2), 217-222.

Dittman, D. E. (1997). Latitudinal compensation in oyster ciliary activity. *Functional Ecology*, *11*(5), 573-578.

Dittman, D. E., Ford, S. E., & Haskin, H. H. (1998). Growth patterns in oysters, *Crassostrea virginica*, from different estuaries. *Marine Biology*, *132*(3), 461-469.

Dobzhansky, T. (1940). Speciation as a stage in evolutionary divergence. *The American Naturalist*, *74*, 312-321.

Dobzhansky, T. (1951). *Genetics and the Origin of Species.* New York, NY: Columbia University Press.

Doebeli, M., & Dieckmann, U. (2003). Speciation along environmental gradients. *Nature*, *421*(6920), 259-264.

Doebeli, M., Dieckmann, U., Metz, J. A. J., & Tautz, D. (2005). What we have also learned: Adaptive speciation is theoretically plausible. *Evolution*, *59*(3), 691-695.

Dorken, M. E., & Pannell, J. R. (2007). The maintenance of hybrid zones across a disturbance gradient. *Heredity*, *99*(1), 89-101.

Dowling, T. E., & Moore, W. S. (1985). Evidence For selection against hybrids in the family *Cyprinidae* (Genus *Notropis*). *Evolution*, *39*(1), 152-158.

Drent, J. (2002). Temperature responses in larvae of *Macoma balthica* from a northerly and southerly population of the European distribution range. *Journal of Experimental Marine Biology and Ecology*, *275*, 117-129.

Duchesne, P., & Bernatchez, L. (2002). AFLPOP: a computer program for simulated and real population allocation, based on AFLP data. *Molecular Ecology Notes*, *2*(3), 380-383.

Duggins, C. F., Karlin, A. A., Mousseau, T. A., & Relyea, K. G. (1995). Analysis of a hybrid zone in *Fundulus majalis* in a northeastern Florida ecotone. *Heredity*, *74*, 117-128.

Duvernell, D. D., Schaefer, J. F., Hancks, D. C., Fonoti, J. A., & Ravanelli, A. M. (2007). Hybridization and reproductive isolation among syntopic populations of the topminnows *Fundulus notatus* and *F. olivaceus*. *Journal of Evolutionary Biology*, *20*, 152-164.

Egan, S. P., Nosil, P., & Funk, D. J. (2008). Selection and genomic differentiation during ecological speciation: Isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. *Evolution*, *62*(5), 1162-1181.

Eldon, B., & Wakeley, J. (2006). Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics*, *172*(4), 2621-2633.

Emelianov, I., Marec, F., & Mallet, J. (2004). Genomic evidence for divergence with gene flow in host races of the larch budmoth. *Proceedings of the Royal Society of London Series B-Biological Sciences*, *271*(1534), 97-105.

Endler, J. A. (1977). *Geographical Variation, Speciation and Clines.* Princeton: Princeton University Press.

Engle, V. D., & Summers, J. K. (1999). Latitudinal gradients in benthic community composition in Western Atlantic estuaries. *Journal of Biogeography*, *26*(5), 1007-1023.

Engle, V. D., & Summers, J. K. (2000). Biogeography of benthic macroinvertebrates in estuaries along the Gulf of Mexico and western Atlantic coasts. *Hydrobiologia*, *436*(1-3), 17-33.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecrting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), 2611-2620.

Eveno, E., Collada, C., Guevara, M. A., Leger, V., Soto, A., Diaz, L. et al. (2008). Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate

genes as revealed by genetic differentiation analyses. *Molecular Biology and Evolution*, *25*(2), 417-437.

Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity*, *103*(4), 285-298.

Falush, D., Stephens, M., & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Molecular Ecology Notes*, *7*(4), 574-578.

Feder, J. L. (1998). The apple maggot fly, *Rhagoletis pomonella*: Flies in the face of conventional wisdom about speciation? In D. A. Howard & S. Berlocher (Eds.), *Endless Forms* (pp. 130-144). Oxford: Oxford University Press.

Felsenstein, J. (2005). PHYLIP (Phylogeny Inference Package) version 3.68.

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.

Floyd, R. P., Lt. (1978). *Geodetic Bench Marks*. Washington, DC: U.S. Government Printing Office.

Ford, M. J. (2002). Applications of selective neutrality tests to molecular ecology. *Molecular Ecology*, *11*(8), 1245-1262.

Fry, J. D. (2003). Multilocus models of sympatric speciation: Bush versus Rice versus Felsenstein. *Evolution*, *57*(8), 1735-1746.

Gaffney, P. M., Bernat, C. M., & Allen, S. K. (1993). Gametic incompatibility in wild and cultured populations of the eastern oyster, *Crassostrea virginica* (Gmelin). *Aquaculture*, *115*(3-4), 273-284.

Gagnaire, P.-A., Minegishi, Y., Zenboudji, S., Valade, P., Aoyama, J., & Berrebi, P. (2011). Within-population structure highlighted by differential introgression across semipermeable barriers to gene flow in *Anguilla marmorata*. *Evolution*, *doi:10.1111/j.1558-5646.2011.01404.x*.

Gaines, S. D., & Bertness, M. D. (1992). Dispersal of juveniles and variable recruitment in sessile marine species. *Nature*, *360*(6404), 579-580.

Galindo, J., Moran, P., & Rolan-Alvarez, E. (2009). Comparing geographical genetic differentiation between candidate and noncandidate loci for adaptation strengthens support for parallel ecological divergence in the marine snail *Littorina saxatilis*. *Molecular Ecology*, *18*(5), 919-930.

Galtsoff, P. S. (1964). *The American oyster: Crassostrea virginica Gmelin* (64). Washington, DC: Government Printing Office.

Garcia-Ramos, G., & Kirkpatrick, M. (1997). Genetic models of adaptation and gene flow in peripheral populations. *Evolution*, *51*(1), 21-28.

Gardner, J. P. A. (1997). Hybridization in the sea. *Advances in Marine Biology, Vol 31*, *31*, 1-78.

Gaudeul, M., Taberlet, P., & Till-Bottraud, I. (2000). Genetic diversity in an endangered alpine plant, *Eryngium alpinum* L. (*Apiaceae*), inferred from amplified fragment length polymorphism markers. *Molecular Ecology*, *9*(10), 1625-1637.

Gavrilets, S., & Vose, A. (2005). Dynamic patterns of adaptive radiation. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(50), 18040-18045.

Gilg, M. R., & Hilbish, T. J. (2003). Saptio-temporal patterns in the genetic structure of recently settled blue mussels (*Mytilus* spp.) across a hybrid zone. *Marine Biology*, *143*, 679-690.

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

Gow, J. L., Peichel, C. L., & Taylor, E. B. (2007). Ecological selection against hybrids in natural populations of sympatric threespine sticklebacks. *Journal of Evolutionary Biology*, *20*(6), 2173-2180.

Grahame, J. W., Wilding, C. S., & Butlin, R. K. (2006). Adaptation to a steep environmental gradient and an associated barrier to gene exchange in *Littorina saxatilis*. *Evolution*, *60*(2), 268-278.

Grant, B. R., & Grant, P. R. (1989). Natural selection in a population of Darwins finches. *American Naturalist*, *133*(3), 377-393.

Grant, B. R., & Grant, P. R. (1996). High survival of Darwin's finch hybrids: Effects of beak morphology and diets. *Ecology*, *77*(2), 500-509.

Grosberg, R., & Cunningham, C. W. (2001). Genetic structure in the sea: From populations to communities. In M. D. Bertness, S. D. Gaines, & M. E. Hay (Eds.), *Marine Community Ecology* (pp. 61-84). Sunderland, MA: Sinauer Asoociates, Inc.

Haddrill, P. R., Thornton, K. R., Charlesworth, B., & Andolfatto, P. (2005). Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Research*, *15*(6), 790-799.

Haldane, J. B. S. (1930). A mathematical theory of natural and artificial selection. Part IV. Isolation. *Proceedings of the Cambridge Philosophical Society*, *26*, 220-230.

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

Hare, M. P., & Avise, J. C. (1998). Population structure in the American oyster as inferred by nuclear gene genealogies. *Molecular Biology and Evolution*, *15*(2), 119-128.

Hare, M. P., Guenther, C., & Fagan, W. F. (2005). Nonrandom larval dispersal can steepen marine clines. *Evolution*, *59*(12), 2509-2517.

Hare, M. P., Karl, S. A., & Avise, J. C. (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*(2), 334-345.

Harrison, R. G. (1990). Hybrid zones: windows on evolutionary process. *Oxford Surveys in Evolutionary Biology*, *7*, 69-128.

Harrison, R. G. (1998). Linking evolutionary pattern and process: The relevance of species concepts for the study of speciation. In D. J. Howard & S. H. Berlochers (Eds.), *Endless Forms: Species and Speciation* (pp. 19-31). Oxford: Oxford University Press.

Harrison, R. G., & Bogdanowicz, S. M. (1997). Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. *Evolution*, *51*(2), 493-505.

Hartl, D. L., & Clark, A. G. (1997). *Principles of Population Genetics* (Third ed.). Sunderland, MA: Sinauer Associates.

Hatfield, T., & Schluter, D. (1999). Ecological speciation in sticklebacks: Environment-dependent hybrid fitness. *Evolution*, *53*(3), 866-873.

Hawthorne, D. J., & Via, S. (2001). Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature*, *412*(6850), 904-907.

Hedgecock, D. (1994). Does variance in reproductive success limit effective poplation size of marine organisms? In A. Beaumont (Ed.), *Genetics and Evolution of Aquatic Organisms* (pp. 122-134). London: Chapman and Hall.

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. W. (2004). *Genetics of Populations* (Third ed.). Boston, MA: Jones & Bartlett Learning.

Hedrick, P. W. (2006). Genetic polymorphism in heterogeneous environments: The age of genomics. *Annual Review of Ecology Evolution and Systematics*, *37*, 67-93.

Hewitt, G. M. (1988). Hybrid zones - natural laboratories for evolutionary studies. *Trends in Ecology & Evolution*, *3*(7), 158-167.

Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, *58*(3), 247-276.

Hey, J., & Machado, C. A. (2003). The study of structured populations - New hope for a difficult and divided science. *Nature Reviews Genetics*, *4*(7), 535-543.

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.

Hodges, S. A., & Arnold, M. L. (1994). Floral and ecological isolation between aquilegia-formosa and aquilegia-pubescens. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(7), 2493-2496.

Hoover, C. A., & Gaffney, P. M. (2005). Geographic variation in nuclear genes of the eastern oyster, *Crassostrea virginica* Gmelin. *Journal of Shellfish Research*, *24*(1), 103-112.

Howard, D. J., Preszler, R. W., Williams, J., Fenchel, S., & Boecklen, W. J. (1997). How Discrete are Oak Species? Insights from a Hybrid Zone Between *Quercus grisea* and *Quercus gambelii*. *Evolution*, *51*(3), 747-755.

Howard, D. J., & Waring, G. L. (1991). Topographic diversity, zone width, and the strength of reproductive isolation in a zone of overlap and hybridization. *Evolution*, *45*(5), 1120-1135.

Howard, D. J., Waring, G. L., Tibbets, C. A., & Gregory, P. G. (1993). Survival of hybrids in a mosaic hybrid zone. *Evolution*, *47*, 789-800.

Hudson, R. R., & Turelli, M. (2003). Stochasticity overrules the "three-times rule": Genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution*, *57*(1), 182-190.

Huff, D. R., Peakall, R., & Smouse, P. E. (1993). RAPD variation within and among natural populations of outcrossing buffalograss *Buchloe dactyloides* (Nutt) Englem. *Theoretical and Applied Genetics*, *86*, 927-934.

Hull, S. L., Grahame, J., & Mill, P. J. (1996). Morphological di- vergence and evidence for reproductive isolation in *Littorina saxatilis* (Olivi) in northeast England. *Journal of Molluscan Studies*, *62*, 89-99.

Hull, W. V., Rear Adm., & Federal-Geodetic-Control-Committee. (1989). *Geometric geodetic accuracy standards and specifications for using GPS relative positioning techniques.* Rockville, MD: NOAA National Geodetic Survey.

Huson, D. H., Richter, D. C., Rausch, C., Dezulian, T., Franz, M., & Rupp, R. (2007). Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics*, *8*, 460.

Irwin, D. E., Bensch, S., Irwin, J. H., & Price, T. D. (2005). Speciation by distance in a ring species. *Science*, *307*(5708), 414-416.

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*(14), 1801-1806.

Jiggins, C. D., & Mallet, J. (2000). Bimodal hybrid zones and speciation. *Trends in Ecology & Evolution*, *15*(6), 250-255.

Jiggins, C. D., McMillan, W. O., & Mallet, J. (1997). The maintenance of species differences across a *Heliconius* hybrid zone. *Heredity*, *79*, 495-505.

JMP. (2010). (Version 9). Cary, NC: SAS Institute Inc.

Johannesson, K., Panova, M., Kemppainen, P., Andre, C., Rolan-Alvarez, E., & Butlin, R. (2009). Repeated evolution of reproductive isolation in a marine snail: Unveiling mechanisms of speciation. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *365*(1547), 1735-1747.

Joost, S., Bonin, A., Bruford, M. W., Despres, L., Conord, C., Erhardt, G. et al. (2007). A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology*, *16*(18), 3955-3969.

Karl, S. A., & Avise, J. C. (1992). Balancing selection at allozyme loci in oysters - implications from nuclear RFLPs. *Science*, *256*(5053), 100-102.

Karleskint, G. K., Jr., Turner, R., & Small, J. W., Jr. (2010). *Introduction to Marine Biology* (Third ed.). Belmont, CA: Brooks/Cole, Cengage Learning.

Kauer, M. O., Dieringer, D., & Schlotterer, C. (2003). A microsatellite variability screen for positive selection associated with the "Out of Africa" habitat expansion of *Drosophila melanogaster*. *Genetics*, *165*(3), 1137-1148.

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*(12), 1225-1241.

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

Key, K. H. L. (1968). Concept of stasipatric speciation. *Systematic Zoology*, *17*(1), 14-&.

Kimura, M. (1968). Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genetical Research*, *11*(3), 247-&.

Kinlan, B. P., & Gaines, S. D. (2003). Propagule dispersal in marine and terrestrial environments: A community perspective. *Ecology*, *84*(8), 2007-2020.

Kirkendale, L., Lee, T., Baker, P., & O Foighil, D. (2004). Oysters of the Conch Republic (Florida Keys): A molecular phylogenetic study of *Parahyotissa mcgintyi*, *Teskeyostrea weberi* and *Ostreola equestris*. *Malacologia*, *46*(2), 309-326.

Knowles, L. L., & Maddison, W. P. (2002). Statistical phylogeography. *Molecular Ecology*, *11*(12), 2623-2635.

Kocher, T. D., & Sage, R. D. (1986). Further genetic analyses of a hybrid zone between leopard frogs (*Rana pipiens* complex) in central Texas. *Evolution*, *40*(1), 21-33.

Koehn, R. K., Newell, R. I. E., & Immermann, F. (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*(9), 5385-5389.

Kozak, K. H., Graham, C. H., & Wiens, J. J. (2008). Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology & Evolution*, *23*(3), 141-148.

Latta, R. G. (2003). Gene flow, adaptive population divergence and comparative population structure across loci. *New Phytologist*, *161*, 51-58.

Latta, R. G., & Mitton, J. B. (1999). Historical separation and present gene flow through a zone of secondary contact in ponderosa pine. *Evolution*, *53*(3), 769-776.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, *17*(4), 183-189.

Levin, L. A. (2006). Recent progress in understanding larval dispersal: new directions and digressions. *Integrative And Comparative Biology*, *46*(3), 282-297.

Lewontin, R. C., & Krakauer, J. (1973). Distribution of gene frequency as a test of theory of selective neutrality of polymorphisms. *Genetics*, *74*(1), 175-195.

Liermann, M., & Hilborn, R. (2001). Depensation: evidence, models and implications. *Fish and Fisheries*, *2*, 33-58.

Loosanoff, V. L., & Nomejko, C. A. (1951). Existance of physiologically different races of oysters, *Crassostrea virginica*. *Biological Bulletin*, *101*, 151-156.

Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics*, *4*(12), 981-994.

Lynch, M., & Milligan, B. G. (1994). Analysis of population genetic-structure with RAPD markers. *Molecular Ecology*, *3*(2), 91-99.

Manel, S., Gaggiotti, O. E., & Waples, R. S. (2005). Assignment methods: matching biological questions techniques with appropriate. *Trends in Ecology & Evolution*, *20*(3), 136-142.

Manel, S., Joost, S., Epperson, B. K., Holderegger, R., Storfer, A., Rosenberg, M. S. et al. (2010). Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology*, *19*(17), 3760-3772.

Manel, S., Schwartz, M. K., Luikart, G., & Taberlet, P. (2003). Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, *18*(4), 189-197.

Mariac, C., Jehin, L., Saidou, A., Thuillet, A., Couderc, M., Sire, P. et al. (2011). Genetic basis of pearl millet adaptation along an environmental gradient investigated by a combination of genome scan and association mapping. *Molecular Ecology*, *20*(1), 80-91.

Marshall, J. C., & Sites, J. W. (2001). A comparison of nuclear and mitochondrial cline shapes in a hybrid zone in the *Sceloporus grammicus* complex (*Squamata : Phrynosomatidae*). *Molecular Ecology*, *10*(2), 435-449.

Maruyama, T. (1970). On fixation probability of mutant genes in a subdivided population. *Genetical Research*, *15*(2), 221-&.

Maynard Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research*, *23*, 23-35.

Mayr, E. (2001). Wu's genic view of speciation - Commentary. *Journal of Evolutionary Biology*, *14*(6), 866-867.

McCairns, R. J. S., & Bernatchez, L. (2008). Landscape genetic analyses reveal cryptic population structure and putative selection gradients in a large-scale estuarine environment. *Molecular Ecology*, *17*(17), 3901-3916.

McDonald, J. H., Verrelli, B. C., & Geyer, L. B. (1996). Lack of geographic variation in anonymous nuclear polymorphisms in the American oyster, *Crassostrea virginica*. *Molecular Biology and Evolution*, *13*(8), 1114-1118.

Meagher, S., & Dowling, T. E. (1991). Hybridization between the Cyprinid fishes *Luxilus albeolus, L. conutus,* and *L. cerasinus* with comments on the proposed hybrid origin of *L. albeolus*. *Copeia*, *4*, 979-991.

Meudt, H. M., & Clarke, A. C. (2007). Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends In Plant Science*, *12*(3), 106-117.

Millar, C. I., Strauss, S. H., Conkle, M. T., & Westfall, R. D. (1988). Allozyme differentiation and biosystematics of the Californian closed-cone pines (*Pinus subsect oocarpae*). *Systematic Botany*, *13*(3), 351-370.

Minder, A. M., & Widmer, A. (2008). A population genomic analysis of species boundaries: neutral processes, adaptive divergence and introgression between two hybridizing plant species. *Molecular Ecology*, *17*(6), 1552-1563.

Moore, W. S. (1977). Evaluation of narrow hybrid zones in vertebrates. *Quarterly Review of Biology*, *52*(3), 263-277.

Moore, W. S., & Price, J. T. (1993). Nature of selection in the northern flicker hybrid zone and its implications for speciation theory. In R. G. Harrison (Ed.), *Hybrid Zones and the Evolutionary Process* (pp. 196-225). New York, NY: Oxford University Press.

Moran, M. D. (2003). Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos*, *100*(2), 403-405.

Mueller, U. G., & Wolfenbarger, L. L. (1999). AFLP genotyping and fingerprinting. *Trends in Ecology & Evolution*, *14*(10), 389-394.

Murray, M. C., & Hare, M. P. (2006). A genomic scan for divergent selection in a secondary contact zone between Atlantic and Gulf of Mexico oysters, *Crassostrea virginica*. *Molecular Ecology*, *15*(13), 4229-4242.

Nachman, M. W., Hoekstra, H. E., & D'Agostino, S. L. (2003). The genetic basis of adaptive melanism in pocket mice. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(9), 5268-5273.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, *70*, 3321-3323.

Nei, M., & Chesser, R. K. (1983). Estimation of fixation indexes and gene diversities. *Annals of Human Genetics*, *47*(Jul), 253-259.

Nei, M., & Maruyama, T. (1975). Lewontin-Krakauer test for neutral genes - comment. *Genetics*, *80*(2), 395-395.

Newell, R. I. E., & Langdon, C. J. (1996). Mechanisms and physiology of larval and adult feeding. In V. S. Kennedy, R. I. E. Newell, & A. F. Eble (Eds.), *The Eastern Oyster: Crassostrea virginica* (pp. 185-230). College Park, MD: Maryland Sea Grant College.

Newkirk, G. F., Haley, L. E., Waugh, D. L., & Doyle, R. (1977). Genetics of larvae and spat growth rate in the oyster *Crassostrea virginica*. *Marine Biology*, *41*(1), 49-52.

Nikula, R., Strelkov, P., & Vainola, R. (2008). A broad transition zone between an inner Baltic hybrid swarm and a pure North Sea subspecies of *Macoma balthica* (Mollusca, Bivalvia). *Molecular Ecology*, *17*(6), 1505-1522.

Nosil, P., Egan, S. P., & Funk, D. J. (2008). Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. *Evolution*, *62*(2), 316-336.

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, *18*(3), 375-402.

Nurnberger, B., Barton, N., MacCallum, C., Gilchrist, J., & Appleby, M. (1995). Natural selection on quantitative traits in the *Bombina* hybrid zone. *Evolution*, *49*(6), 1224-1238.

Oetjen, K., & Reusch, T. B. H. (2007). Genome scans detect consistent divergent selection among subtidal vs. intertidal populations of the marine angiosperm *Zostera marina*. *Molecular Ecology*, *16*(24), 5156-5167.

Ogden, R., & Thorpe, R. S. (2002). Molecular evidence for ecological speciation in tropical habitats. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(21), 13612-13615.

Orr, H. A. (2005). The genetic theory of adaptation: A brief history. *Nature Reviews Genetics*, *6*, 119-127.

Ortiz-Barrientos, D., Reiland, J., Hey, J., & Noor, M. A. (2002). Recombination and the divergence of hybridizing species. *Genetica*, *116*, 167-178.

Palumbi, S. R., Cipriano, F., & Hare, M. P. (2001). Predicting nuclear gene coalescence from mitochondrial data: The three-times rule. *Evolution*, *55*(5), 859-868.

Payseur, B. A., Krenz, J. G., & Nachman, M. W. (2004). Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. *Evolution*, *58*(9), 2064-2078.

Peakall, R., & Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, *6*(1), 288-295.

Pennings, S. C., & Bertness, M. D. (2001). Salt marsh communities. In M. D. Bertness, S. D. Gaines, & M. E. Hay (Eds.), *Marine Community Ecology* (pp. 289-316). Sunderland, MA: Sinauer Associates, Inc.

Pigeon, D., Chouinard, A., & Bernatchez, L. (1997). Multiple modes of speciation involved in the parallel evolution of sympatric morphotypes of lake whitefish (*Coregonus clupeaformis,* Salmonidae). *Evolution*, *51*(1), 196-205.

Pogson, G. H. (2001). Nucleotide polymorphism and natural selection at the pantophysin (Pan I) locus in the Atlantic cod, *Gadus morhua* (L.). *Genetics*, *157*(1), 317-330.

Pogson, G. H., & Mesa, K. A. (2004). Positive Darwinian selection at the pantophysin (Pan I) locus in marine Gadid fishes. *Molecular Biology and Evolution*, *21*(1), 65-75.

Pogson, G. H., Mesa, K. A., & Boutilier, R. G. (1995). Genetic population-structure and gene flow in the Atlantic cod *Gadus morhua* - a comparison of allozyme and nuclear RFLP loci. *Genetics*, *139*(1), 375-385.

Pompanon, F., Bonin, A., Bellemain, E., & Taberlet, P. (2005). Genotyping errors: Causes, consequences and solutions. *Nature Reviews Genetics*, *6*(11), 847-859.

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

Raffaelli, D. G. (1978). Relationship between shell injuries, shell thickness and habitat characteristics of inter-tidal snail *Littorina rudis maton*. *Journal of Molluscan Studies*, *44*(Aug), 166-170.

Rand, D. M., & Harrison, R. G. (1989). Ecological genetics of a mosaic hybrid zone - mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. *Evolution*, *43*(2), 432-449.

Reeb, C. A., & Avise, J. C. (1990). A genetic discontinuity in a continuously distributed species - mitochondrial DNA in the American oyster, *Crassostrea virginica*. *Genetics*, *124*(2), 397-406.

Reeves, P. A., & Richards, C. M. (2009). Accurate inference of subtle population structure (and other genetic discontinuities) using principal coordinates. *PLOS One*, *4*(1), e4269.

Remington, C. L. (1968). Suture-zones of hybrid interaction between recently joined biotas. *Evolutionary Biology*, *2*, 321-428.

Rieseberg, L. H., & Buerkle, C. A. (2002). Genetic mapping in hybrid zones. *American Naturalist*, *159*, S36-S50.

Rieseberg, L. H., Raymond, O., Rosenthal, D. M., Lai, Z., Livingstone, K., Nakazato, T. et al. (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, *301*(5637), 1211-1216.

Rieseberg, L. H., Whitton, J., & Gardner, K. (1999). Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, *152*(2), 713-727.

Riginos, C., & Cunningham, C. W. (2005). Local adaptation and species segregation in two mussel (*Mytilus edulis* x *Mytilus trossulus*) hybrid zones. *Molecular Ecology*, *14*(2), 381-400.

Riginos, C., Sukhdeo, K., & Cunningham, C. W. (2002). Evidence for selection at multiple allozyme loci across a mussel hybrid zone. *Molecular Biology and Evolution*, *19*(3), 347-351.

Robertson, A. (1975). Lewontin-Krakauer test for neutral genes - comment. *Genetics*, *80*(2), 396-396.

Rogers, S. M., Campbell, D., Baird, S. J. E., Danzmann, R. G., & Bernatchez, L. (2001). Combining the analyses of introgressive hybridisation and linkage mapping to investigate the genetic architecture of population divergence in the lake whitefish (*Coregonus clupeaformis*, Mitchill). *Genetica*, *111*(1-3), 25-41.

Rogers, S. M., Isabel, N., & Bernatchez, L. (2007). Linkage maps of the dwarf and normal lake whitefish (*Coregonus clupeaformis*) species complex and their hybrids reveal the genetic architecture of population divergence. *Genetics*, *175*(1), 375-398.

Rose, C. G., Paynter, K. T., & Hare, M. P. (2006). Isolation by distance in the eastern oyster, *Crassostrea virginica,* in Chesapeake Bay. *Journal of Heredity*, *97*(2), 158-170.

Rosenberg, N. A. (2003). The shapes of neutral gene genealogies in two species: Probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. *Evolution*, *57*(7), 1465-1477.

Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, *4*(1), 137-138.

Rosenberg, N. A., & Nordborg, M. (2002). Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics*, *3*(5), 380-390.

Rosenblum, E. B. (2006). Convergent evolution and divergent selection: Lizards at the White Sands ecotone. *American Naturalist*, *167*(1), 1-15.

Ross, C. L., & Harrison, R. G. (2002). A fine-scale spatial analysis of the mosaic hybrid zone between *Gryllus firmus* and *Gryllus pennsylvanicus*. *Evolution*, *56*(11), 2296-2312.

Rubidge, E. M., & Taylor, E. B. (2004). Hybrid zone structure and the potential role of selection in hybridizing populations of native westslope cutthroat trout (*Oncorhynchus clarki lewisi*) and introduced rainbow trout (*O. mykiss*). *Molecular Ecology*, *13*(12), 3735-3749.

Rundle, H. D., Nagel, L., Boughman, J. W., & Schluter, D. (2000). Natural selection and parallel speciation in sympatric sticklebacks. *Science*, *287*(5451), 306-308.

Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, *8*(3), 336-352.

Sage, R. D., & Selander, R. K. (1979). Hybridization between species of the *Rana pipiens* complex in central Texas. *Evolution*, *33*, 1069-1088.

Sarver, S. K., Landrum, M. C., & Foltz, D. W. (1992). Genetics and taxonomy of ribbed mussels (*Geukensia* spp). *Marine Biology*, *113*(3), 385-390.

Savolainen, V., Anstett, M. C., Lexer, C., Hutton, I., Clarkson, J. J., Norup, M. V. et al. (2006). Sympatric speciation in palms on an oceanic island. *Nature*, *441*(7090), 210-213.

Schilthuizen, M. (2000). Ecotone: speciation-prone. *Trends in Ecology & Evolution*, *15*(4), 130-131.

Schmidt, P. S., & Rand, D. M. (1999). Intertidal microhabitat and selection at MPI: Interlocus contrasts in the northern acorn barnacle, *Semibalanus balanoides*. *Evolution*, *53*, 135-146.

Schmidt, P. S., Serrao, E. A., Pearson, G. A., Riginos, C., Rawson, P. D., Hilbish, T. J. et al. (2008). Ecological genetics in the North Atlantic: environmental gradients and adaptation at specific loci. *Ecology*, *89*(11), S91-S107.

Schneider, C. J., Smith, T. B., Larison, B., & Moritz, C. (1999). A test of alternative models of diversification in tropical rainforests: Ecological gradients vs. rainforest refugia. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(24), 13869-13873.

Shanks, A. L. (2009). Pelagic larval duration and dispersal distance revisited. *Biological Bulletin*, *216*(3), 373-385.

Shanks, A. L., Grantham, B. A., & Carr, M. H. (2003). Propagule dispersal distance and the size and spacing of marine reserves. *Ecological Applications*, *13*(1), S159-S169.

Shumway, S. E. (1996). Natural environmental factors. In V. S. Kennedy, R. I. E. Newell, & A. F. Eble (Eds.), *The Eastern Oyster: Crassostrea virginica* (pp. 467-513). College Park, MD: Maryland Sea Grant College.

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

Siegel, D. A., Mitarai, S., Costello, C. J., Gaines, S. D., Kendall, B. E., Warner, R. R. et al. (2008). The stochastic nature of larval connectivity among nearshore marine populations. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(26), 8974-8979.

Slatkin, M. (1973). Gene flow and selection in a cline. *Genetics*, *75*(4), 733-756.

Smith, N. P. (1987). An introduction to the tides of Florida's Indian River Lagoon: I. Water levels. *Florida Scientist*, *50*, 49-61.

Smith, N. P. (1993). Tidal and nontidal flushing of Florida Indian River Lagoon. *Estuaries*, *16*(4), 739-746.

Smith, N. P. (1996). Tidal and low frequency flushing of a coastal lagoon using a flexible grid model. In C. Pattiaratchi (Ed.), *Mixing in Estuaries and Coastal Seas* (pp. 171-183). Washington, DC: American Geophysical Union.

Sokal, R. R., & Rohlf, F. J. (1995). *Biometry* (Third ed.). New York, NY: W. H. Freeman and Company.

Sotka, E. E., Wares, J. P., Barth, J. A., Grosberg, R. K., & Palumbi, S. R. (2004). Strong genetic clines and geographical variation in gene flow in the rocky intertidal barnacle *Balanus glandula*. *Molecular Ecology*, *13*(8), 2143-2156.

Stephens, P. A., & Sutherland, W. J. (1999). Consequences of the Allee effect for behaviour, ecology and conservation. *Trends in Ecology & Evolution*, *14*(10), 401-405.

Stephens, P. A., Sutherland, W. J., & Freckleton, R. P. (1999). What is the Allee effect? *Oikos*, *87*(1), 185-190.

Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(16), 9440-9445.

Storfer, A., Murphy, M. A., Evans, J. S., Goldberg, C. S., Robinson, S., Spear, S. F. et al. (2007). Putting the 'landscape' in landscape genetics. *Heredity*, *98*(3), 128-142.

Storz, J. F., Payseur, B. A., & Nachman, M. W. (2004). Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. *Molecular Biology and Evolution*, *21*(9), 1800-1811.

Stuckas, H., Stoof, K., Quesada, H., & Tiedemann, R. (2009). Evolutionary implications of discordant clines across the Baltic *Mytilus* hybrid zone *(Mytilus edulis* and *Mytilus trossulus*). *Heredity*, *103*(2), 146-156.

Swearer, S. E., Caselle, J. E., Lea, D. W., & Warner, R. R. (1999). Larval retention and recruitment in an island population of a coral-reef fish. *Nature*, *402*(6763), 799-802.

Swenson, N. G., & Howard, D. J. (2005). Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *American Naturalist*, *166*(5), 581-591.

Szabo, V. M., & Burr, B. (1996). Simple inheritance of key traits distinguishing maize and teosinte. *Molecular and General Genetics*, *252*(1-2), 33-41.

Szymura, J. M., & Barton, N. H. (1991). The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: Comparisons between transects and between loci. *Evolution*, *45*(2), 237-261.

Teeter, K. C., Payseur, B. A., Harris, L. W., Bakewell, M. A., Thibodeau, L. M., O'Brien, J. E. et al. (2008). Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Research*, *18*(1), 67-76.

Thompson, R. J., Newell, R. I. E., Kennedy, V. S., & Mann, R. (1996). Reproductive processes and early development. In V. S. Kennedy, R. I. E. Newell, & A. F. Eble

(Eds.), *The Eastern Oyster: Crassostrea virginica* (pp. 335-370). College Park, MD: Maryland Sea Grant College.

Thorrold, S. R., Latkoczy, C., Swart, P. K., & Jones, C. M. (2001). Natal homing in a marine fish metapopulation. *Science*, *291*(5502), 297-299.

Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S. et al. (2007). Convergent adaptation of human lactase persistance in Africa and Europe. *Nature Genetics*, *39*(1), 31-40.

Toro, J., Innes, D. J., & Thompson, R. J. (2004). Genetic variation among life-history stages of mussels in a *Mytilus edulis-M. trossulus* hybrid zone. *Marine Biology*, *145*(4), 713-725.

Tsakas, S., & Krimbas, C. B. (1976). Testing heterogeneity of F-values - suggestion and a correction. *Genetics*, *84*(2), 399-401.

USDA, N. R. C. S. (2011). *The PLANTS Database.* National Plant Data Team, Greensboro, NC.

Vasemagi, A., Nilsson, J., & Primmer, C. R. (2005). Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Molecular Biology And Evolution*, *22*(4), 1067-1076.

Vekemans, X. (2002). AFLP-SURV version 1.0.

Via, S., & West, J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, *17*(19), 4334-4345.

Vitalis, R., Dawson, K., & Boursot, P. (2001). Interpretation of variation across marker loci as evidence of selection. *Genetics*, *158*(4), 1811-1823.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Vandelee, T., Hornes, M. Et Al. (1995). AFLP - a new technique for DNA-fingerprinting. *Nucleic Acids Research*, *23*(21), 4407-4414.

Wares, J. P., Gaines, S. D., & Cunningham, C. W. (2001). A comparative study of assymmetric migration events across a marine biogeographic boundary. *Evolution*, *55*(2), 295-306.

Warner, R. R. (1997). Evolutionary ecology: how to reconcile pelagic dispersal with local adaptation. *Coral Reefs*, *16*, S115-S120.

White, M. E., & Wilson, E. A. (1996). Predators, pests, and competitors. In V. S. Kennedy, R. I. E. Newell, & A. F. Eble (Eds.), *The Eastern Oyster: Crassostrea virginica* (pp. 559-579). College Park, MD: Maryland Sea Grant College.

Whitehead, A., & Crawford, D. L. (2006). Neutral and adaptive variation in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(14), 5425-5430.

Whitlock, R., Hipperson, H., Mannarelli, M., Butlin, R. K., & Burke, T. (2008). An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Molecular Ecology Resources*, *8*(4), 725-735.

Wilding, C. S., Butlin, R. K., & Grahame, J. (2001). Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology*, *14*(4), 611-619.

Wilhelm, R., & Hilbish, T. J. (1998). Assessment of natural selection in a hybrid population of mussels: Evaluation of exogenous vs. endogenous selection models. *Marine Biology*, *131*(3), 505-514.

Williams, G. C. (1975a). *Sex and Evolution* (8). Princeton, NJ: Princeton University Press.

Williams, G. C. (1975b). The elm-oyster model. In *Sex and Evolution* (pp. 35-43). Princeton, NJ: Princeton University Press.

Wilson, C., Scotto, L., Scarpa, J., Volety, A., Laramore, S., & Haunert, D. (2005). Survey of water quality, oyster reproduction and oyster health status in the St. Lucie Estuary. *Journal of Shellfish Research*, *24*(1), 157-165.

Wood, H. M., Grahame, J. W., Humphray, S., Rogers, J., & Butlin, R. K. (2008). Sequence differentiation in regions identified by a genome scan for local adaptation. *Molecular Ecology*, *17*(13), 3123-3135.

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

Wu, C. I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, *14*(6), 851-865.

Yatabe, Y., Kane, N. C., Scotti-Saintagne, C., & Rieseberg, L. H. (2007). Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics*, *175*(4), 1883-1893.

Young, W. P., Ostberg, C. O., Keim, P., & Thorgaard, G. H. (2001). Genetic characterization of hybridization and introgression between anadromous rainbow

trout (*Oncorhynchus mykiss irideus*) and coastal cutthroat trout (*O. clarki clarki*). *Molecular Ecology*, *10*(4), 921-930.

Zhang, H. B., Scarpa, J., & Hare, M. P. (2010). Differential Fertilization Success Between Two Populations of Eastern Oyster, *Crassostrea virginica*. *Biological Bulletin*, *219*(2), 142-150.

Zhivotovsky, L. A. (1999). Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology*, *8*(6), 907-913.