GENETIC CHARACTERIZATION OF A HYBRID ZONE IN KILLIFISH, (*FUNDULUS HETEROCLITUS*): EVIDENCE FOR ASSORTATIVE MATING OR SELECTION AGAINST HYBRIDS

by

JESSICA LAUREN MCKENZIE

B.Sc., The University of Victoria, 2002
M.Sc., Texas A&M University, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Zoology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

February 2013

Abstract

Hybrid zones act as natural experiments that can provide insights into the factors governing species formation and maintenance. In order to investigate these factors, I examined a hybrid zone between two subspecies of Atlantic killifish, *Fundulus heteroclitus*. Previous research has shown that these subspecies differ both genetically and phenotypically, but very little work had examined the hybrid zone between them. I used a suite of genetic markers to describe the genetic pattern within this hybrid zone and laboratory breeding experiments to investigate the forces responsible for its maintenance.

Based on hybrid indices calculated using microsatellite and SNP (single nucleotide polymorphism) loci, a trimodal hybrid zone located in Beaverdam Creek (in the Metedeconk river system in New Jersey) separates the two subspecies of killifish, suggesting that while some F1 hybrids are produced, backcross types are rare. This pattern persisted across several sampling sites and across two years, suggesting that this pattern was not a sampling artifact. By investigating the geographical patterns of genetic variation in 30 SNPs along the Atlantic coast, I found that clines in mitochondrial DNA markers and in SNPs in several nuclear genes with mitochondrially-associated functions were coincident, concordant and exceptionally steep compared to those of other loci. I used tension zone analyses to conclude that these clines are likely being maintained either by selection or by assortative mating. The observed cytonuclear disequilibria also suggested a role for cytonuclear epistasis in maintaining this hybrid zone.

Within Beaverdam Creek, there was no genetic differentiation between samples taken at locations differing in temperature and salinity, suggesting that habitat specialization on these abiotic variables is not involved in the maintenance of the hybrid zone. However, my results

from a "choice" breeding experiment among individuals originating from the extremes of the species' distributions suggested a possible role for positive assortative mating.

Taken together, my research provides evidence that differentiation in mitochondrial properties resulting in selection or assortative mating could be involved in the maintenance of distinct subspecies of *F. heteroclitus*, and points to a potential general role for divergence in energy metabolism as a mechanism in promoting or maintaining species differences.

Preface

Chapter Two is based on a study conducted by Jessica L. McKenzie under the supervision of Dr. Patricia M. Schulte. I participated in all field collections with Dr. Rashpal S. Dhillon, who also assisted in the design of the sampling strategy. I was responsible for all laboratory work including extracting DNA from all samples, designing the PCR-RFLP assay used for the diagnosis of mtDNA type, performing all PCRs of mitochondrial and microsatellite loci, and preparing the microsatellite loci for genotyping. I also scored all of the microsatellite genotypes. I performed all data analysis and generated all of the figures and tables in this chapter. I was also responsible for writing this chapter with editing by Dr. Schulte.

Chapter Three is based on a study conducted by Jessica L. McKenzie under the supervision of Dr. Patricia M. Schulte. I participated in all field collections with Dr. Rashpal S. Dhillon, who also assisted in the design of the sampling strategy. Dr. Schulte selected the loci for inclusion in the SNP panel. I prepared all of the samples for SNP genotyping (with the exception of fish from the locations as labeled in chapter three, information for which was provided by Dr. Marjorie F. Oleksiak). I also performed all data analysis and produced all figures and tables that occur in this chapter. I wrote this chapter with the editing help of Dr. Schulte.

A version of chapters two and three has been submitted to Molecular Ecology and is currently in revision.

Chapter Four is based on a study conducted by Jessica L. McKenzie under the supervision of Dr. Patricia M. Schulte. I was responsible for the collection and processing of all samples originating from Beaverdam Creek (both Summer 2008 and Fall 2009 sampling seasons). I also performed all DNA extraction, PCRs, sample preparation and genotype scoring regarding these samples. Dr. Carol Bucking and I conducted all fish care and performed all *in*

vitro fertilization experiments. Dr. Bucking collected and cared for all eggs and/or hatchlings resulting from the "choice" breeding experiment while I was on maternity leave. DNA extraction and microsatellite genotyping of these offspring was performed by Amanda L. Moreira under my supervision. I was responsible for all data analysis in this chapter as well as the production of all figures and tables in it. I was also responsible for writing the chapter with editing by Dr. Schulte.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to extend my most sincere and utmost thanks to my supervisor, Dr. Patricia Schulte, for her tireless guidance, support, patience and understanding. I would also like to thank the members of my committee, Drs. Darren Irwin, Dolph Schluter, and Eric Taylor for their invaluable advice at all points during my time as a student at the University of British Columbia. Thanks also goes to lab mates past and present for the camaraderie that can only come of the shared experience of having, at one point or another, been a graduate student.

I could not have completed all of these years of schooling without the continued support and encouragement of my family and friends. I would like to thank my parents, Koop and Brenda, and my brother Damon, for their continuous love and encouragement. I am also one of the few fortunate people who can honestly say that they have the best in-laws imaginable. Betty and Ernest, I am grateful for all of your support over the years. Lastly, I thank Amy, Diana, Em, Jess, and Sarah, for always being there. Time to go to Vegas....again...

Lastly, I sincerely would not have been able to complete this degree without the never-ending support of my husband Kurt. I cannot put into words how much it has meant to have someone dedicate so much of their time and effort to the pursuit of a goal not their own. And Owen, my bubba, I thank you for inspiring me and providing me with the motivation to complete this thesis.

Chapter One: Introduction

When two recently diverged taxa encounter one another they have the potential to interbreed if reproductive isolation is not complete. If hybridization occurs, it can then result in a transition between the morphological, physiological and ultimately genetic traits of the two taxa as one crosses the hybrid zone, moving from the territory of one of the parent taxa into that of the other (Dobzhansky 1940; Hewitt 1988). These phenotypic and genetic patterns are commonly referred to as clines (Dobzhansky 1940; Barton & Hewitt 1989; Sotka & Palumbi 2006). Hybrid zones, and the resulting phenotypic and genetic clines that they create, provide natural laboratories for the study of speciation and natural selection (Hewitt 1988) and can provide insight into the genes underlying morphological, physiological, and reproductive differences among taxa (e.g. Teeter *et al.* 2008). In my dissertation, I have performed a detailed genetic characterization of a hybrid zone between two subspecies of a common estuarine fish, *Fundulus heteroclitus*, and have examined the resulting genetic clines to infer which types of forces that might be involved in maintaining the distinction between these subspecies. In this introductory chapter, I provide the background necessary to situate my work in the context of the field, and to justify my hypotheses and approach.

Types of Hybrid Zones

There are two main ways in which a hybrid zone may originate, which can be classified as either primary or secondary intergradation. A primary hybrid zone results when a species diverges in response to an environmental gradient in a continuous habitat (Barton & Hewitt 1985; Hewitt 1988). Alternatively, a species' distribution may have, at one time, been physically

divided somewhere in its range, allowing the separated populations to diverge in allopatry for some time before coming back into contact after the barrier is removed (Endler 1977). This leads to the formation of a secondary hybrid zone. It is, however, challenging to distinguish which of these two alternatives played the major role in the formation of a given hybrid zone, because the geographic patterns generated in the formation of primary and secondary zones are often indistinguishable (Endler 1982). Endler (1977) concludes that unless the zone is observed within several hundred generations of its inception, it will be impossible to determine whether the species' divergence was initiated in parapatry or allopatry (Endler 1977; Endler 1982).

Some of the processes occurring within a hybrid zone can be inferred by examining patterns of multi-locus genetic variation in the zone of contact (e.g. Jiggins & Mallet 2000), using the "hybrid index", which is an estimate of the similarity of each hybrid individual to the parental types (where a hybrid index of "0" represents individual of one pure parental type and a hybrid index of "1" indicates an individual of the other pure parental type). By examining the frequency distribution of hybrid indices in a variety of hybrid zones, Jiggins & Mallet 2000 suggested that hybrid zones can be divided into three main types: unimodal, bimodal, and "flat" (Jiggins & Mallet 2000). A flat hybrid zone (which is intermediate between the other two types) consists of populations containing approximately even mixtures of parental and hybrid individuals. In a unimodal hybrid zone, hybrid types predominate, forming a hybrid swarm. In a bimodal hybrid zone, most individuals are similar to the two parental types, and hybrids are present only at a low frequency if at all. However, these classes of hybrid zone are, in fact, part of a continuum ranging from fully bimodal to fully unimodal.

When two previously allopatric taxa first come into contact, the pattern within the contact zone will be bimodal in shape. As the taxa begin to interbreed, and hybrids are produced, the

bimodal character of the zone begins to break down. If there is no reproductive isolation of any form, the zone will gradually approach Hardy-Weinberg equilibrium at all loci, forming a fully unimodal zone. Thus, the presence of a bimodal pattern is indicative of near-complete speciation between the two parental taxa in a stable hybrid zone, while a fully unimodal pattern is indicative of weak isolation between the parental taxa. However, it is important to note that within a single hybrid zone different loci may show differing degrees of bimodality, depending on whether selection is acting on a particular locus (Jiggins & Mallet 2000), or if assortative mating based on phenotypes encoded by that locus is occurring. Strong disruptive selection or assortative mating is likely to produce bimodal patterns within a hybrid zone, whereas high levels of introgression (and thus unimodal patterns) may occur in the absence of selection or strong assortative mating. In contrast, flat patterns can be indicative of various combinations of forces, or may be the result of a relatively recent hybrid zone that is in transition between the initial bimodal state and the final state of Hardy-Weinberg equilibrium. Thus determining the pattern of genetic variation within a hybrid zone is an important step in identifying the factors that may maintain the zone.

Scale of sampling is critically important when determining whether a hybrid zone is unimodal or bimodal (Jiggins & Mallet, 2000). Samples must be taken from locally panmictic populations at the center of the zone, and populations from the edges of the zone (where parental types predominate) cannot be mixed during analysis. Mixing samples containing two distinct parental types would result in the apparent detection of high levels of linkage disequilibrium and heterozygote deficit, creating the appearance of a bimodal zone where none exists (Jiggins & Mallet, 2000). Therefore, careful characterization of a hybrid zone is required prior to the application of such analyses.

The Maintenance of Hybrid Zones

Hybrid zones can typically be maintained by any combination of three main factors:

1. Endogenous selection
2. Exogenous selection
3. Strong habitat selection or positive assortative mating.

Endogenous selection results from environment-independent factors that result in decreased fitness of hybrid offspring (Jiggins & Mallet 2000). These post-zygotic barriers include such factors as hybrid inviability, hybrid sterility, and hybrid breakdown (Dobzhansky 1940; Barton & Hewitt 1985; Barton 2001). The resulting tension zones are maintained by a balance of dispersal and selection: dispersal of pure parental types into the zone and selection against the hybrid offspring that these parental forms produce (Barton & Hewitt 1985; Moore & Price 1993).

Exogenous selection can be the result of different parental genotypes being favoured in different environments, resulting in hybrid genotypes being less fit in the context of each parental environment (Endler 1977). The hybrid zones that result in this case are often referred to as geographical-selection gradients, and they occur when the preferred habitat type of one parental taxon differs from that of the other (Moore & Price 1993). Alternatively, hybrid genotypes could be maintained if hybrid individuals have higher fitness than either parental type in the local environment of the contact zone, but lower fitness outside of it. The bounded hybrid superiority model is used to describe such hybrid zones (Moore 1977).

Positive assortative mating between parental types can also result in the maintenance of a hybrid zone (Brodin & Haas, 2009). Similarly, if hybrids have intermediate appearance or

behaviour, these traits can make them less attractive as mates, and this form of assortative mating can also act to maintain a hybrid zone (Bridle *et al*. 2006). These forms of assortative mating may result in whole-genome signatures, or may result in effects only at specific loci that shape the phenotypes involved in mate choice. The genetic patterns resulting from assortative mating within a hybrid zone can thus be difficult to distinguish from the patterns resulting from endogenous or exogenous selection.

These processes can also be coupled. Post-zygotic barriers, whether in the form of endogenous or exogenous selection against the hybrid offspring, can result in selection driving the evolution of pre-zygotic barriers (Servedio & Noor 2003). Pre-zygotic barriers include increased habitat specialization among parental types and/or positive assortative mating. These barriers evolve because they relieve the costs that would be incurred by the parental types due to the production of unfit hybrid offspring. The interplay between these two general types of reproductive barriers, wherein declines in post-zygotic hybrid fitness encourage the strengthening of pre-zygotic barriers is called reinforcement (Servedio & Noor 2003).

Just as pre- and post-zygotic barriers can become tightly linked, so can endogenous and exogenous selection (Figure 1-1). If two divergent parental types encounter one another and produce hybrids, pre-existing post-zygotic barriers built-up during their time in allopatry can immediately result in the production of less fit hybrid offspring due to endogenous genomic incompatibilities, resulting in a tension zone. The locations of tension zones are not fixed, however, and are free to move until impeded by either a physical or environmental barrier. If the parental types respond differentially to the conditions presented on either side of the barrier, the resulting exogenous (environment-dependent) selection can even further encourage the

accumulation of differences among parental types, resulting in the production of even less fit hybrid offspring in a cycle first described by Bierne *et al.* (2011).

Mitochondria and Hybrid Zones

In addition to the typical linkage disequilibrium among nuclear genes characteristic of a bimodal hybrid zone, these zones can also be associated with a form of linkage disequilibrium termed cytonuclear disequilibrium. Cytonuclear disequilibrium involves a nonrandom association between the nuclear genome and the genome of cytoplasmic organelles such as mitochondria or chloroplasts (Asmussen *et al.* 1987). Mitochondria are the primary site of energy production in most animal cells. The mitochondrial genome, however, encodes very few of the proteins involved in its function. For example, only 13 of the polypeptides involved in forming subunits of the electron transport chain are mitochondrially encoded. The remaining 77 polypeptides of the subunits of the electron transport chain, as well as the approximately 1,500 other proteins localized in the mitochondria (as estimated using the human genome), are all encoded by the nuclear genome (Wallace 2005). Thus, the functional compatibility of the nuclear and mitochondrial genomes is critical for the functioning of the cell, and ultimately of the organism that contains it.

When the distributions of parental taxa possessing both divergent nuclear and mitochondrial genomes meet, assortative mating may encourage the persistence of cytonuclear disequilibria. In addition, possessing mismatched cytonuclear types may have a detrimental effect on the fitness of the resulting hybrids and lead to cytonuclear disequilibria being observed among individuals in hybrid zones (Asmussen *et al.* 1987). Such cytonuclear incompatibilities have been implicated in hybrid breakdown in a variety of species (Edmands & Burton 1999; Burton *et al.* 2006; Ellison & Burton 2008; Niehuis *et al.* 2008).

Calculations of cytonuclear disequilibrium depend on the interaction of two mitochondrial types (for example M and m) and three nuclear genotypes (for example AA, Aa, and aa). Thus, six cytonuclear combinations are possible from pure and reciprocal crosses among the parental types. It is the distribution of the observed cytonuclear combinations of these possible six outcomes among the individuals sampled from a hybrid zone that is of interest. The patterns, with respect to observed cytonuclear combinations when compared to the expected values if mating was random within a hybrid zone, can be used to gain insight into the mating system operating in the hybrid zone (Avise 2000).

Patterns of Genetic Variation in Hybrid Zones

The patterns of genetic variation within a hybrid zone can be used to provide insight into the forces maintaining the zone. Observations of Hardy-Weinberg Equilibrium and linkage equilibrium at multiple loci can be taken as evidence of the presence of a hybrid swarm, characteristic of unimodal hybrid zones. Alternately, the presence of heterozygote deficit and linkage disequilibrium (and cytonuclear disequilibrium) can be used as evidence that one or some combination of both pre- and/or post-zygotic reproductive barriers are acting to prevent the generation of hybrid offspring. Ultimately, the calculation of individual hybrid index values using multilocus genotype data can be used to determine the shape of the hybrid zone in question (Jiggins & Mallet 2000).

Patterns of genetic variation can also be measured across hybrid zones. Clinal analyses examining variations in allele frequencies among sampled populations can be used to calculate the centre and widths of the clines across the area sampled. Width estimates can in turn be compared with width estimates expected under a neutral model as a way of indirectly inferring the strength of selection operating to maintain the hybrid zone (Szymura & Barton 1986;

Cheviron & Brumfield 2009). If selection against the resulting hybrids is weak, then the steep clines present upon its initial formation will gradually erode, becoming wider over time until the once separate species' have become homogenized with respect to the particular trait (or traits) in question. Such wide, gradual clines, indicate that selection for the maintenance of the differences in allele frequencies between populations is weak (Barton & Hewitt 1985). Steep clines with associated cline widths narrower than the width predicted by a neutral model would be indicative of clines being maintained by selection (Haldane 1948; Slatkin 1973). The narrower the width, the stronger the selection required to maintain it. In addition, although assortative mating is seldom considered in the context of clinal analysis (and indeed most tension zone models assume random mating), the steep clinal patterns suggested to be the result of natural selection could, in principle, also be produced by assortative mating (Soularue & Kremer 2012).

Fundulus heteroclitus: A Model System

For my dissertation research I used the Atlantic killifish, *Fundulus heteroclitus*, as a model system to determine which of the possibilities outlined above is responsible for maintaining the hybrid zone between two forms of this species. *F. heteroclitus* is an abundant topminnow that resides in the brackish waters of estuaries along the eastern coast of North America from Newfoundland to northern Florida (Hardy 1978). Killifish, renowned for their sedentary lifestyle, usually maintain a home range of 36-38 m (Abraham 1985), although some individuals may move substantially longer distances (e.g. up to 375 m (Lotrich 1975); up to 1km (Able *et al.* 2012)). Adults range in length from 50 to 100 mm and typically have a life span of three to four years (Abraham 1985). Killifish are sexually dimorphic, with the males exhibiting exaggerated colouration during the breeding season (Hardy 1978). Due to their abundance in the wild and the ease with which large populations can be maintained in the laboratory owing to the

extreme hardiness of the species, killifish have been used in a variety of studies including those of physiology, toxicology, and ecological and evolutionary genetics (see Burnett *et al.* 2007 for a review).

Character clines in a wide variety of phenotypic traits and genetic markers have been observed as one moves from more northern to more southern populations of this species (Powers & Place 1978; Cashon *et al.* 1981; Ropson *et al.* 1990; Morin & Able 1983; Able & Felley 1986; Gonzalez-Villasenor & Powers 1990; Adams *et al.* 2006; Duvernell *et al.* 2008; Strand *et al.* 2012). Many of these clines show abrupt breaks from one predominant type to another between 40°N and 41°N indicating the presence of a hybrid zone. Results of earlier studies distinguishing morphological, embryological, and genetic differences between subspecies support a model of secondary intergradation wherein two previously isolated forms diverged in allopatry only to come back into contact with one another in this zone off of the coast of New Jersey (Morin & Able 1983; Able & Felley 1986; Powers *et al*. 1986; Gonzalez-Villasenor & Powers 1990). Thus, two subspecies are recognized (*F. h. macrolepidotus* represents the northern subspecies, and *F. h. heteroclitus* represents the southern subspecies) on the basis of a multitude of differences in physiological, endocrinological, biochemical, histological, meristic, morphological, genetic, habitat preference, reproductive, behavioural, larval, eggs, and embryonic traits occurring between individuals residing at the extremes of the species' distribution (Appendix A Table A.1).

The coastal clines in a variety of traits distinguishing the subspecies of *Fundulus heteroclitus* have been found to be replicated up the arms of the larger rivers of the Chesapeake Bay. Smith *et al.* (1998) found a significant excess of individuals possessing a northern mitochondrial type at the more freshwater ends of rivers in the Chesapeake Bay system and an

excess of southern mitotypes at the more seawater mouth. Consistent with these observations, laboratory experiments have shown that the northern subspecies has a more efficient mechanism for survival in freshwater than its southern counterpart (Scott *et al.* 2004). More recent work by Whitehead *et al.* (2011) found similar differences among individuals collected from more freshwater ends of large rivers (Potomac River > 600 km in length) of the Chesapeake Bay system when compared to those collected from the more seawater mouths of the same rivers. Individuals collected from the more freshwater end of the Potomac River (possessing mitochondrial and nuclear genotypes characteristic of the northern subspecies, *F. h. macrolepidotus*) showed higher tolerance to hypoosmotic conditions than individuals from the seawater populations (possessing mitochondrial and nuclear genotypes characteristic of the southern subspecies, *F. h. heteroclitus*; Whitehead *et al.* 2011). These differences in salinity tolerance are not restricted to adults. Able & Palmer (1988) found that while crosses among individuals from the northern end of the species' range were successful in freshwater, such crosses were never successful among individuals originating from the southern extreme of the distribution. Furthermore, larvae resulting from fertilizations among pure crosses of *F. h. heteroclitus* suffered significantly higher mortality in freshwater when compared to those larvae generated from pure crosses of the northern subspecies.

Significant differences between the subspecies are also tightly correlated with the different temperature regimes most commonly experienced by individuals occupying opposite ends of the species' range with the northern subspecies performing suboptimally at warmer temperatures more commonly experienced in the southern end of the range and vice versa. Specifically, embryos resulting from pure northern crosses were unable to reach the hatch stage when reared at 30°C. Similarly, pure southern embryos failed to hatch when reared at 15°C

(DiMichele & Westerman 1997). These subspecies also differ significantly in their critical thermal maximum and minimum (the maximum and minimum temperatures that can be tolerated immediately before loss of equilibrium; Fangue *et al.* 2006), with fish from southern populations exhibiting CTMax and CTMin values that are on average 1.5°C higher than those of fish from northern populations (Fangue *et al.* 2006). Such relationships between temperature and fitness at all stages in these subspecies' life histories could result in strong differences in habitat preference between the subspecies, but this has never been investigated in areas where these two subspecies co-occur along the coast.

There are also morphological differences between these subspecies which could allow them to distinguish among one another in a hybrid zone and facilitate positive assortative mating. Sexually mature males of the northern subspecies have a black ocellus on their dorsal fin while this marking is usually absent among members of the southern subspecies (Able & Felley 1986). Sexually mature southern females typically have a vertical banding pattern of melanophores on their lateral surface; in mature northern females, this banding is broken and more closely resembles a series of spots (Able & Felley 1986). Fish of the northern subspecies are typically significantly smaller than fish of the southern subspecies (mean body mass = 7.29g and 16.2g, respectively; Podrabsky *et al.* 2000). This difference persists in the center of the hybrid zone, with individuals possessing the northern mitochondrial type being significantly smaller than those possessing the southern mitotype (Haakons 2010). More subtle differences in anal sheath length of females could also provide a cue for positive assortative mating (Able & Felley 1986).

These subspecific differences in adult morphology are correlated with differences in egg morphology. Northern eggs are covered by long, thin hair-like chorionic fibres and a thin, foam-like jelly (Brummett & Dumont 1981). Among southern eggs however, these chorionic filaments

are shorter and less numerous (Able & Castagna 1975) and may even be absent (Brummett 1966), and the egg is covered by a thicker jelly coat (Brummett & Dumont 1981). Such differences may themselves present a significant obstacle to sperm differentially adapted to fertilizing eggs of their respective females or may be indicators of less obvious differences such as those at the level of gamete recognition proteins. While previous studies have reported no barriers to fertilization success at the gametic level (Powers & Place 1978; DiMichele & Westerman 1997) high fertilization success can often be achieved among highly divergent taxa if eggs are saturated with excess sperm (Palumbi & Metz 1991; Levitan 2002).

Of the numerous divergent characters known to distinguish the northern from the southern subspecies, many have been found to be correlated with the genotype at the heart-type isozyme of lactate dehydrogenase (LDH-B; Appendix A Table A.1, which is expressed in the heart, liver and red blood cells in this species). Individuals from populations spanning the shoreline of the east coast of North America show a moderately steep cline in LDH-B allele frequency, with the northern allozyme being rapidly replaced by the southern allozyme across a zone roughly corresponding to that previously observed for morphological, embryological, and mitochondrial traits (Place & Powers 1978). Extensive research involving this glycolytic enzyme has demonstrated that variation in frequencies of the southern and northern alleles ($LDH\text{-}B^{a}$ and $LDH\text{-}B^{b}$, respectively) is consistent with the operation of natural selection (Schulte *et al*. 1997). Among adult fish, functional differences at this locus exist, with individuals possessing the southern form of the enzyme exhibiting a superior catalytic efficiency in warmer waters, with the northern form of the enzyme being more efficient at cooler temperatures (Place & Powers 1984a). Furthermore, individuals possessing the northern genotype exhibit significantly faster critical swimming speeds at low temperatures when compared to individuals harboring the

southern genotype (DiMichele & Powers 1982b). These differences in swimming performance have been attributed to divergence in hemoglobin-oxygen affinity between individuals of differing LDH-B types. The hemoglobin of fish having the northern LDH-B type shows a decreased affinity for oxygen following exhaustive exercise compared to the hemoglobin of fish having the southern LDH-B type and a correspondingly high P_{50} (partial pressure of oxygen required to achieve 50% saturation of hemoglobin) suggesting that this genotype has a more efficient mechanism for transporting oxygen to its tissues during strenuous exercise at low temperatures when compared to individuals owning the southern genotype (DiMichele & Powers 1982a). It is hypothesized that this difference in hemoglobin function is due to the effect of LDH-B on ATP levels within the red blood cell (Powers *et al*. 1979; DiMichele and Powers 1982a). ATP is a potent allosteric modifier of hemoglobin in fishes and acts to change the P_{50} of hemoglobin, thus providing a functional link between LDH-B and hemoglobin (Powers *et al*. 1979). There is also a strong relationship between LDH-B genotype and embryonic metabolic rate with northern-type embryos exhibiting higher metabolic rates than individuals possessing the southern LDH-B type (Paynter *et al*. 1991). Such differences in metabolic rate translate into differences in hatch times among LDH-B genotypes with northern embryos emerging significantly earlier than those from the southern end of the range. This pattern has been observed *in situ* as well as in the laboratory across a range of incubation temperatures (DiMichele & Westerman 1997). Northern embryos also experienced higher heat stress mortalities than their southern counterparts (DiMichele & Westerman 1997). Differences at the level of *Ldh-B* gene expression show that northern fish exhibit transcription rates twice that of their southern counterparts, as is expected among individuals adapted to life in cooler environments (Crawford & Powers 1992). Such differences in transcription rates are thought to

be driven by various nucleotide differences in the promoter sequence of the *Ldh-B* gene (Segal *et al*. 1996, Schulte *et al*. 2000). All of these divergent character traits could have a profound negative impact on the fitness of an individual that finds itself in an environment better suited to the phenotype of the other parental type. Such an impact could ultimately limit the range of the subspecies, preventing the northern subspecies from establishing themselves in a southern environment and vice versa. Thus, genes involved in or related to metabolic processes may experience differential selective pressures based on different environmental temperatures such as those experienced by the northern and southern subspecies of *Fundulus heteroclitus* in a manner similar to that predicted under a geographical-selection gradient model. Alternatively (or in addition), co-adapted gene complexes may not function properly in alternate nuclear or cytoplasmic genetic backgrounds, leading to an increase in pre- and/or post-zygotic incompatibilities as predicted under a tension zone model.

Thesis Objectives

Over the past three decades, studies have focused on providing further genetic evidence for the separation of *F. heteroclitus* into two subspecies and also on the continued debate over whether this hybrid zone formed via primary or secondary intergradation. However, despite the extensive previous work on this species, the contact zone of these two groups has yet to be fully characterized. In particular, no studies have simultaneously examined multiple nuclear and mitochondrial markers in the contact zone, making analyses of potential cytonuclear disequilibrium impossible. In addition, most previous work has focused on exogenous selection as a potential mechanism for maintaining the hybrid zone, and other potential mechanisms such as endogenous selection and assortative mating have been neglected. Thus, for my Ph.D. dissertation I used a multi-locus genetic analysis, including both nuclear and mitochondrial loci,

as well as experimental studies of mating preference to provide additional insight into the various factors that may act to maintain the hybrid zone between subspecies of killifish, *Fundulus heteroclitus*. Together, this work provides insights into the relative relationships of the various factors that either promote or impede the process of speciation.

The primary objectives of my thesis were as follows:

1) Describe the patterns of genetic variation within the *F. heteroclitus* hybrid zone,
2) To determine if endogenous and/or exogenous selection might be involved in the maintenance of the hybrid zone,
3) To determine whether pre-zygotic factors such as habitat specialization, gametic incompatibility or assortative mating might play a role in the maintenance of the hybrid zone.

Objective 1:

The patterns of genetic variation within a hybrid zone, along the continuum from a bimodal zone to a unimodal zone, can provide insight into the factors that maintain the zone. In addition, individual loci may differ in their patterns, thus it is critical to utilize a multilocus approach, all the while being mindful of these per locus differences, to characterize the zone. In addition, because cytonuclear incompatibilities may be important in hybrid zones, both nuclear and mitochondrial loci should be examined. To address this objective, in Chapter Two I performed a genetic characterization of the *F. heteroclitus* hybrid zone using a combination of mitochondrial markers and putatively neutral nuclear microsatellite markers.

Objective 2:

The shapes of the clines formed as a result of hybridization can provide critical insight as to whether selection is acting on these loci. To address this objective, in Chapter Three I utilized

data from mitochondrial and nuclear SNPs in coding genes from *F. heteroclitus* populations along the Atlantic coast of North America to assess the potential for selection at these loci. I selected a subset of SNPs with high F_{ST} between the geographically extreme populations in Maine and Georgia that I predicted to be diagnostic or semi-diagnostic of contribution of the northern and southern *F. heteroclitus* genome in hybrid individuals, from a larger panel of randomly selected SNPs (Williams *et al*. 2010; Strand *et al*. 2012). Given that previous studies have shown mitochondrial DNA and mitochondrial Malate dehydrogenase clines are particularly steep in *F. heteroclitus* compared to the clines in most nuclear loci, (Strand *et al*. 2012), suggesting that they are acting as outlier loci, I predicted that if the clines of some or all of my diagnostic nuclear SNP loci were coincident and concordant with the mitochondrial DNA cline, then the type and strength of selection (either exogenous or endogenous) maintaining these clines, or the effect of these loci on assortative mating, is likely to be the same for these nuclear SNP loci as for the mitochondrial DNA.

Objective 3:

Pre-zygotic factors such as habitat specialization, positive assortative mating, or gametic incompatibilities can act alone or in concert in the maintenance of a hybrid zone. To address this objective, in Chapter Four I used a combination of mitochondrial, nuclear microsatellite, and nuclear SNPs to examine fine-scale population structure to test whether habitat specialization could be a pre-zygotic barrier operating within the hybrid zone. In order to address the question of gametic incompatibilities operating in the maintenance of the zone I looked for differences in fertilization success among pure and reciprocal crosses of *F. heteroclitus* in the laboratory. Lastly, I performed a breeding experiment in which fish were allowed free choice between

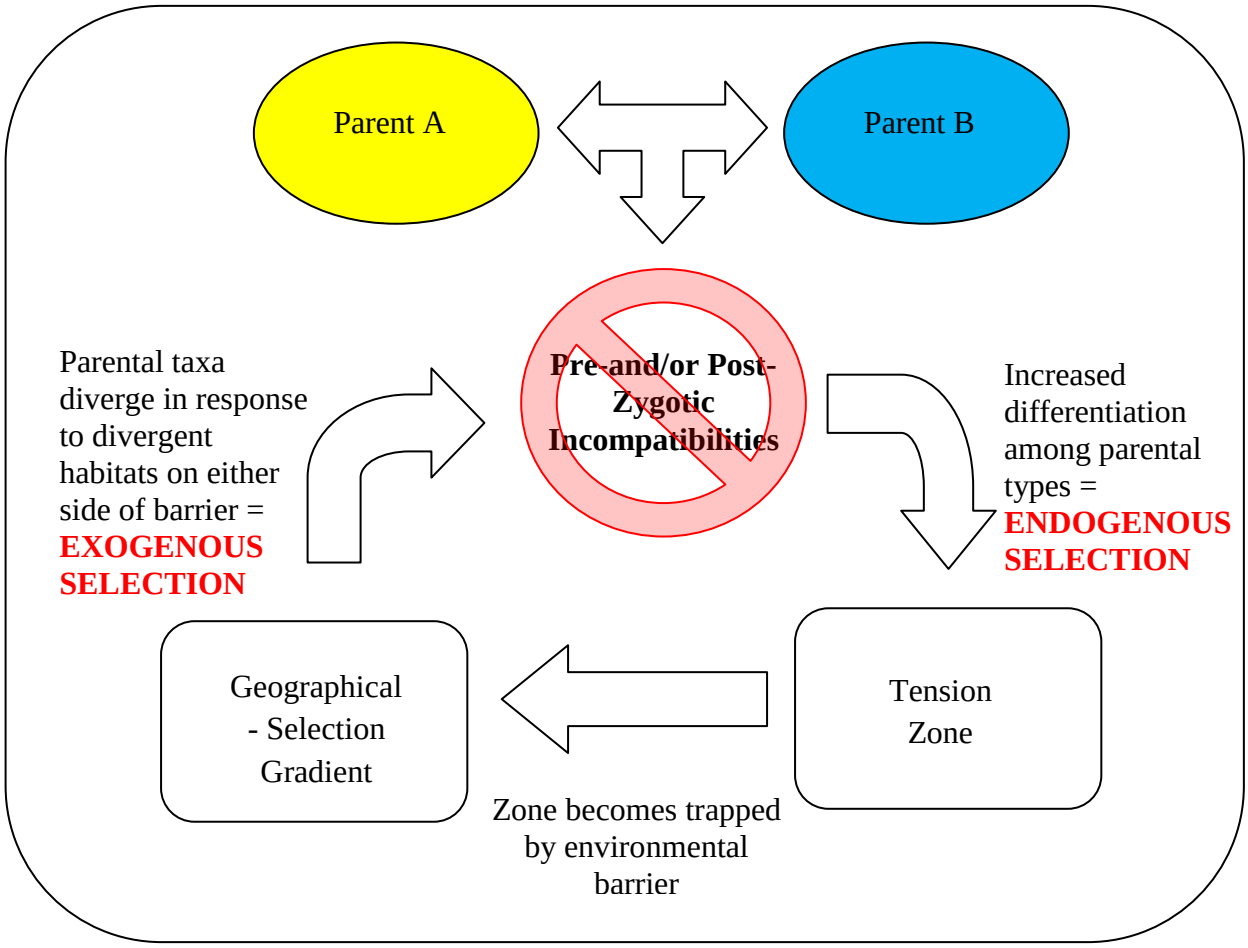
individuals of both subspecies to determine the possible contribution of assortative mating as a pre-zygotic barrier establishing the genetic patterns observed within the hybrid zone.

Thesis Organization

Following this introductory chapter, I have organized my thesis into three data chapters, each addressing different aspects of my thesis objectives. In Chapter Two, I provide a characterization of patterns of genetic variation through the species range of *F. heteroclitus* using microsatellite markers, with sampling focused in the putative range of the hybrid zone in New Jersey to determine patterns of variation in this hybrid zone along the continuum from bimodal to unimodal. In Chapter Three, I use 30 SNPs in the nuclear genome and two in the mitochondrial genome to characterize patterns of variation in the hybrid zone at this different class of markers, and then in combination with published data from Williams *et al*. (2010), kindly provided by Margorie Oleksiak, I analyze the shapes of the clines along the coast in order to identify potential loci under selection. In Chapter Four, I present data from intensive sampling of a single marsh located at the mitochondrial cline centre, sampled at multiple locations across two seasons, to determine whether the patterns observed in Chapters Two and Three are stable across sampling sites and years, and to assess the possibility of microhabitat specialization as an isolating barrier. I also present data from a laboratory mating-choice experiment to assess the possibility of gametic incompatibility or assortative mating.

Overall, my thesis research was designed to provide a case study of the genetic structure and potential factors maintaining a specific hybrid zone, which could then be used to shed light on some of the processes that may be operating to maintain species identity in general.

Figure 1-1. Coupling Exogenous and Endogenous Selection in a hybrid zone (as described in Bierne *et al.* 2011).



Chapter Two: Genetic evidence for reproductive isolation in a killifish (*Fundulus heteroclitus*) hybrid zone

Synopsis

The genetic structure of a hybrid zone can provide insights into the relative roles of the various factors that maintain the zone. Here, we use a multilocus approach to characterize a hybrid zone between two subspecies of killifish (*Fundulus heteroclitus*) found along the Atlantic coast of North America. We first analyzed clinal variation along the Atlantic coast using a single nucleotide polymorphism (SNP) in the mitochondrial DNA (mtDNA) displacement loop (D-loop) and a panel of nine nuclear microsatellite markers to locate the centre of this hybrid zone. A marsh located near the centre of the mtDNA cline demonstrated a clear trimodal distribution of nuclear hybrid index values for microsatellite genotypes. These data suggest that selection against backcross genotypes may be important in regions were more than one mitochondrial type are segregating. We also found some evidence for deviations from Hardy-Weinberg expectations and linkage disequilibrium across the putative hybrid zone. These data provide potential evidence of one or more reproductive isolating barriers operating between the two subspecies of *Fundulus heteroclitus*.

Introduction

When two recently diverged taxa encounter one another at their range limits and hybridization occurs, a transition between the morphological, physiological and ultimately genetic traits of the two taxa is observed as one passes from the territory of one of the parent taxa into that of the other, resulting in clines in a variety of traits (Dobzhansky 1940; Hewitt 1988; Barton & Hewitt 1989; Sotka & Palumbi 2006). Thus, by studying the evolutionary fates of hybrids (e.g., relative frequencies through time and space) hybrid zones provide "natural laboratories" for studies of speciation and adaptation (Barton & Hewitt 1985, 1989; Hewitt 1988). At the centre of a hybrid zone, genetic variation may be classified along a continuum from "unimodal", in which hybrid types predominate, to "bimodal" in which parental types predominate. Bimodality in a hybrid zone is an indication that speciation between the two taxa is almost complete, whereas unimodal hybrid zones result from random mating of parental types and subsequent survival of the hybrid offspring in the area of sympatry (Brown & Wilson 1956; Barton & Hewitt 1985). Patterns of genetic variation that accompany a bimodal distribution of hybrid indices include a high occurrence of heterozygote deficit and linkage and cytonuclear disequilibrium while such patterns are not observed in unimodal zones where reproductive barriers are not in place. It has been argued that bimodal hybrid zones are always characterized by strong assortative mating or assortative fertilization (pre-zygotic barriers) among parental types, while both types of zone may involve post-zygotic barriers to gene flow (Jiggins & Mallet 2000). Thus, identifying the genetic structure within a hybrid zone is a critical step in assessing the relative impacts of pre-zygotic and post-zygotic barriers to gene flow, which are the fundamental processes via which hybrid zones are maintained.

The Atlantic killifish, *Fundulus heteroclitus*, is an abundant topminnow (Pisces: Cyprinodontidae) that resides in the brackish waters of estuaries along the eastern coast of North America from Newfoundland, Canada, to northern Florida, USA (Hardy 1978). Previous biochemical genetic studies in killifish identified clines at various enzyme-coding loci along the Atlantic coast, with northern alleles being replaced by southern alleles along the New Jersey shoreline (Powers *et al.* 1986; Powers & Place 1978). Subsequent work at the molecular level has detected clines in mitochondrial DNA (Gonzáles-Villasenor & Powers 1990; Smith *et al.* 1998), putatively neutral nuclear microsatellite loci (Adams *et al.* 2006; Duvernell *et al.* 2008), and a variety of nuclear SNPs (Strand *et al*. 2012). In addition, there are gametic (Morin & Able 1983), embryonic (DiMichele & Westerman 1997), morphological (Able & Felley 1986) and physiological differences between those individuals occupying the northern part of the species range and those residing farther south (Crawford & Powers 1989, 1992; DiMichele & Powers 1982, 1991; Powers & Schulte 1998; Scott *et al.* 2004; Fangue *et al.* 2006; 2009; Whitehead *et al.* 2010; Dhillon and Schulte 2011). As a result of this variation, controversy has surrounded the taxonomic status of the killifish for almost a century (Morin & Able 1983).

The most compelling evidence for the separation of *F. heteroclitus* into two subspecies is that most of the aforementioned character clines show abrupt breaks from one predominant type to the other at various locations along the coast of New Jersey. As a result of these concordant findings, it is now generally accepted that *Fundulus heteroclitus* consists of two subspecies: *Fundulus heteroclitus heteroclitus* Linnaeus 1766 (hereafter referred to as "southerns") ranging from New Jersey south to northern Florida and *Fundulus heteroclitus macrolepidotus* Walbaum 1792 (hereafter referred to as "northerns") ranging from Connecticut north to Newfoundland (Able & Felley 1986). A hybrid zone, located along the coast of New Jersey, connects the two

subspecies. Most previous research on the contemporary population structure of *F. heteroclitus* has used only one of allozyme, microsatellite or mitochondrial markers and no studies have characterized the hybrid zone in detail, or used multilocus data to calculate a hybrid index, which is a critical step in distinguishing unimodal from bimodal hybrid zones (Jiggins & Mallet, 2000). Here, we used a combination of mitochondrial DNA and nine microsatellite markers (i) to examine populations of killifish from marshes located within and immediately surrounding the hybrid zone for evidence of heterozygote deficit, linkage and cytonuclear disequilibrium, indicating selection against hybrids, and (ii) to determine the shape of the genetic variation within the hybrid zone of the *Fundulus heteroclitus* complex with respect to the unimodal-bimodal continuum through the calculation of hybrid indices.

Materials and Methods

Fish Collection

Fundulus heteroclitus (520 individuals) were collected from various locations along the Atlantic coast of North America, with sample collection concentrated in and immediately adjacent to the presumed hybrid zone between the two subspecies in New Jersey (Table 2-1; Figure 2-1). Samples from the contact zone were collected from May-June of 2008 from 11 locations along the coast (at the mouth of each marsh). Minnow traps (G-type) were used and trap set time ranged from two to six hours. Fish length was recorded and a fin clip was taken from each individual and preserved in 95% ethanol. Additional samples representative of the pure populations of the two subspecies (Table 2-1; Fangue *et al.* 2006) were collected near Brunswick, Georgia, and Hampton, New Hampshire.

Mitochondrial D-loop PCR and Genotyping

DNA was extracted from all samples using Qiagen's DNeasy® Blood and Tissue Kit and all polymerase chain reactions were performed using a MJ Research PTC-200 Peltier Thermal Cycler. A restriction digestion assay was developed to allow the rapid diagnosis of mitochondrial type based on a single fixed difference found during a preliminary screening of northern (n = 20) and southern (n = 20) mitochondrial DNA D-loop sequences. Preliminary screening was performed using primers K (forward) and G (reverse) (Lee *et al*. 1995) to amplify approximately 1100 base pairs (bp) of the mtDNA D-loop and reactions were prepared in 12.5 µl volumes containing 10x *Taq* Buffer with KCl (Fermentas), 2.0 mM $MgCl_2$, 0.2 mM dNTPs, 0.4 µM each primer, 1 unit *Taq* DNA polymerase (Fermentas), the appropriate volume of ddH_2O, and 1.0 µl genomic DNA of varying concentrations. The amplification profile consisted of an initial denaturing step of 2 minutes at 95°C followed by 5 cycles of denaturation at 94.0°C for 30 seconds, primer annealing at 60°C for 45 seconds, extension at 72°C for 1 minute 30 seconds, and 34 cycles of denaturation at 94°C for 30 seconds, primer annealing at 65°C for 45 seconds, extension at 72°C for 1 minute 30 seconds. The final extension was at 72°C for 5 minutes. The resulting PCR products were then electrophoresed on a 1% agarose gel pre-stained with ethidium bromide and visualized using the GeneGenius Bio Imaging System (Syngene). Successful amplifications were purified using Qiagen's QIAquick® PCR Purification Kit and direct sequenced using the BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer Corporation, Foster City, California). This reaction contained 3.0 µl BigDye® dilution mix, 2.0 µl purified PCR product, 1.0 µM of K primer, and ddH_2O for a final volume of 10 µl. The cycle sequencing profile consisted of an initial denaturing step of 96°C for 1 minute followed by 26

cycles of 96°C for 10 seconds, 50°C for 5 seconds, and 60°C for 4 minutes. The sequences were edited and aligned using the program MEGA 4.0.2 (Tamura *et al.* 2007). The online program NEBcutter V2.0 (Vincze *et al.* 2003) was then used to identify a restriction enzyme that would cut at one of the several diagnostic sites that exhibited fixed differences between northern and southern fish.

The restriction digestion assay was performed by PCR amplifying the mtDNA D-loop as above. 8.0 µl reactions containing 5 µl of PCR product, 10x Buffer ScaI (Fermentas), 0.25 units of the restriction enzyme *Sca*I (Fermentas), and the appropriate volume of ddH_2O were incubated for 3 hours at 37°C. The undigested 1100 bp fragment represented individuals with a northern mitochondrial haplotype while the digested 900 bp fragment was diagnostic of the southern mitotype. This PCR-RFLP assay was then applied to all 520 fish collected from 13 locations (Table 2-1). The fragment polymorphisms generated by the restriction digest were easily discernible on a 1.5% agarose gel pre-stained with ethidium bromide. Ten northern, nine southern, and 28 hybrid zone fish were then sequenced at random to confirm the results of the RFLP assay. The resulting sequences have been submitted to GenBank (Accession Numbers JQ518219-JQ518265)**.**

Microsatellite Genotyping

Individuals from locations 1 - 13 inclusive were also genotyped at nine microsatellite loci (Adams *et al.* 2005): FhCA-1, FhATG-B101, FhATG-B128, FhATG-2, FhATG-4, FhATG-6, FhATG-17, FhATG-18, and FhATG-20. PCR conditions for each of these microsatellite markers required modification from those described in Adams *et al.* (2005) for successful amplification (Appendix B Table B.1). Reactions involving a single round of denaturation, annealing, and

extension were performed for 35 cycles. For any reaction involving two rounds of denaturation, annealing, and extension, the first round proceeded for 5 cycles, followed by a second round of 30 cycles. All final extensions were performed at 72° for 5 min. FhATG-17, FhATG-18, and FhATG-20 were combined in a single multiplex reaction, as were FhATG-2 and FhATG-4. Although FhATG-B101 and FhATGB-128 used identical reaction conditions as FhATG-2 and FhATG-4, they were combined in a separate multiplex reaction to avoid complications as a result of size-fragment overlap. Final reaction volumes were 10 µl with reagent concentrations as follows: 10x *Taq* Buffer with KCl (Fermentas), 2.0 mM $MgCl_2$, 0.2 mM dNTPs, 0.4 µM LDH-F1, 0.4 µM LDH-R2, 0.5 units *Taq* DNA polymerase (Fermentas), 1.0 µl genomic DNA, and ddH_2O for a final reaction volume of 12.5 µl. All reactions were performed with these concentrations except for the multiplex reaction involving FhATG-2 and FhATG-4 which required 1 unit of *Taq* polymerase. PCR products were then dried down in 96-well plates and sent to the Boston Children's Hospital Molecular Genetics Core Facility for genotyping on an Applied Biosystems 3730 DNA Analyzer. Alleles were scored using Peak Scanner™ Software v1.0 (Applied Biosystems).

Analyses

Microchecker V2.2.3 (Van Oosterhout *et al.* 2004) was used to screen the microsatellite data for the presence of null alleles, large allele dropout, and scoring errors. The program Genepop 4.0.10 (Raymond & Rousset 1995) was used to conduct the exact test for Hardy-Weinberg equilibrium. Tests for heterozygote excess and heterozygote deficiency were performed for each locus in each population. Markov chain parameters were set to the defaults (dememorization number = 1000; 100 batches; 1000 iterations per batch). Genetix 4.05.2

(Belkhir *et al.*1999) was used to test for linkage disequilibrium among the different microsatellite loci. Each locus pair in each population was tested for evidence of linkage disequilibrium using the Black and Krafsur procedure (Cockerham & Weir 1977).

Clinal Analyses

Cfit6 (Gay *et al.* 2005) was used to estimate the slope and centre of the mitochondrial cline using data collected from populations 1 – 13 (Table 2-1). The cline was fit using the scaled-logit function and with the Hardy-Weinberg genotypic structure selected. The program was run ten times and the cline with the highest log-likelihood was then plotted using the logistic 4p model in the program JMP® 9.0.2 (2010 SAS Institute Inc.). Briefly, this model fits a sigmoid curve on the presented data using four parameters: theta1 (upper asymptote), theta2 (lower asymptote), theta3 (slope) and theta4 (centre). Theta1 and theta2 were fixed based on the maximum and minimum observed allele frequencies, respectively (Brumfield *et al.* 2001) and theta3 and theta4 were entered as calculated using Cfit6. Clines for the microsatellite data were fit by fixing theta1 and theta2 at the maximum and minimum observed allele frequencies, and the resulting estimates for slope (theta3) and cline centre (theta4) were read directly from the JMP output.

Cytonuclear Disequilibrium

The program CNDm (available at http://statgen.ncsu.edu/cnd/CNDd.php; Asmussen & Basten 1996; Basten & Asmussen 1997) was used to test for cytonuclear disequilibrium between mitochondrial type and each microsatellite locus within marshes 2 – 8 (Table 2-1).

Hybrid Index

INTROGRESS (Gompert & Buerkle 2009, 2010) was used to calculate a hybrid index for each individual using the multilocus microsatellite and SNP genotypes separately. For both analyses individuals from marshes 2 and 8 (Table 2-1) were used as parental types in order to train the program. Hybrid index values for each individual in each marsh were then binned in 0.10 increments and compiled in a histogram representing the SNP and microsatellite hybrid indices present in each marsh. The number of clusters present in the frequency distributions of hybrid index values was estimated using the program MCLUST (Fraley *et al.* 2012). Deviations from normality were resolved by the addition of random noise to the data as recommended by the program authors. The number of clusters (either 1, 2, or 3) resulting in the highest Bayesian Information Criterion (BIC) was interpreted as best-fitting the shape of the hybrid index frequency distributions.

Results

Identification of the Hybrid Zone

Five hundred and twenty individuals from 13 locations were genotyped at one mtDNA SNP and nine microsatellite markers to confirm the previously observed locations of the cline centres for these nuclear and mitochondrial markers (Bernardi *et al.* 1993; Gonzáles-Villasenor & Powers 1990; Adams *et al.* 2006; Duvernell *et al.* 2008).

All microsatellite loci and the mtDNA SNP marker exhibited clinal variation along the coast (Figure 2-2, Figure 2-3; Table 2-2), as has been previously observed (Bernardi *et al.* 1993; Gonzáles-Villasenor & Powers 1990; Adams *et al.* 2006; Duvernell *et al.* 2008). The mtDNA

cline centre was extremely close to our sampled location #4 (Metedeconk creek). Two of the microsatellite cline centres similarly fell within the borders of the New Jersey coastline (between ~1100 km and 1260 km from Brunswick, GA; Figure 2-3G and 2-3H), six had cline centres located north of this border (Figure 2-3A-F), and one had its centre located to the south of this border (Figure 2-3I).

HWE and Linkage Disequilbrium

Among the microsatellites, there was no evidence of null alleles or large allele dropout, but there was some evidence of significant departures from HWE (Table 2-3). Significant heterozygote deficit was observed for at least one locus in each of marshes 2, 3, 6, 11, 12, and 13 (Table 2-3). Significant heterozygote excess was observed at three loci (FhATG18, FhATG17, and FhATG-CA1) in marsh 12 and one locus (FhCA-1) among individuals collected from marsh 4. However, only one of these values (heterozygote deficit at locus FhATGB101 in marsh 6) remained significant following an adjustment for false discovery rate (FDR) using the classical one-stage method (Pike 2010).

Among the 13 locations sampled, a total of 61 pairs of microsatellite loci were found to be in significant linkage disequilibrium prior to correction for multiple testing (Table 2-3). Of the eighteen linkage disequilibria that remained significant after the FDR-adjustment for multiple comparisons, ten were detected within marshes located within the border of New Jersey, and two pairs of loci in significant linkage disequilibrium were in individuals collected from a marsh located just outside of this border (marsh 9). No microsatellite loci were in significant disequilbrium with the mitochondrial genome at any location.

Hybrid Index

Microsatellite genotypes of 311 fish from seven locations within and immediately surrounding the hybrid zone (locations 2 – 8) were used to estimate hybrid index using the program INTROGRESS (Gompert & Buerkle 2009). Genotypes of individuals from marshes 2 (RUMFS) and 8 (Cheesequake) were used as southern and northern parents, respectively, in order to train the program. Individuals from these locations at the edges of the hybrid zone (as opposed to the extremes of the species' distribution), were used because they are more likely to more closely resemble the true parental types of individuals located within the centre of the hybrid zone. There was a gradual transition from a more southern hybrid index in the south of the hybrid zone to a more northern hybrid index towards the northern edge of the zone (Figure 2-4). However, in the central marsh (#4, Metedeconk), the microsatellite data show a lower frequency of intermediate genotypes and a higher frequency of pure northern and southern types upon visual inspection. Formal analysis using the program MCLUST indicated that three modes fit the distribution of hybrid indices the best (Table 2-4). Thus, the pattern of hybrid indices observed in Metedeconk is best explained by a trimodal distribution. In contrast, a nearby marsh (#5, Navesink) had an essentially even distribution of all hybrid indices. However, there was a relatively small sample size from this marsh (n = 13), so it is difficult to make firm conclusions about the shape of the distribution of hybrid indices based on this location.

Discussion

In this study, we used a combination of mitochondrial DNA and microsatellite markers and a multi-locus analytical approach to characterize genetic variation in killifish, *Fundulus heteroclitus*. We thus extended the previous analyses of Duvernell *et al*. (2008) and Adams *et al*. (2006), who examined variation at these microsatellites, but did not calculate a hybrid index, and those of Gonzalez-Villasenor (1990), who examined mtDNA variation, but did not examine variation in the nuclear genome. We confirmed that the location of the centre of the mtDNA cline is in north central New Jersey and that the majority of the microsatellite markers had cline centres north of New Jersey. However, here we emphasize that not all the microsatellites share the same cline centre, suggesting additional complexity to the genetic patterns in this species that have not been fully considered in all previous analyses. In addition, we found evidence of linkage disequilibrium concentrated within the borders of New Jersey as well as the presence of a trimodal distribution of nuclear hybrid indices at a marsh at the mtDNA cline centre. Together, these data suggest the possibility of reproductive isolating barriers that act against backcrossed types at the mtDNA cline centre to maintain this genetic structure. Thus, these data suggest that contemporary forces, and not just historical factors, are important in structuring patterns of genetic variation in killifish.

The nature of the hybrid zone between the northern and southern subspecies of Atlantic killifish, and even the taxonomic status of the subspecies, has been a matter of debate for several decades, despite the fact that this species has been extensively studied and is cited as a "textbook example" of intraspecific latitudinal genetic variation (e.g. Avise 2004; Freeman & Herron 2004).

Perhaps the most plausible scenario put forth to explain the mode of formation of the hybrid zone invokes the near extirpation of northern populations by the advancing ice sheet during the last glaciation, with the persistence of isolated populations in pockets of suitable habitats at the glaciers edge. Selection or drift could then operate on the standing genetic variation in the polymorphic ancestral population (similar to contemporary southern populations), resulting in differentiation between the current northern and southern types. The physiological differences between the contemporary northern and southern subspecies suggest that selection for individuals adapted to the conditions at the retreating glacial front, which is thought to have been characterized by cooler fresher water habitats (Powers *et al.* 1986), may have played a role (Whitehead *et al.* 2011). Such a scenario, or some variation of this version of secondary contact (Smith *et al*. 1998, Adams *et al*. 2006, Duvernell *et al.* 2008), is now widely accepted as the most likely mode of formation of the hybrid zone connecting *F. h. heteroclitus* and *F. h. macrolepidotus*. However, whether or not the hybrid zone is being maintained by contemporary processes or is gradually eroding had not been rigorously examined.

An important genetic signature of strong reproductive isolation is the presence of a bimodal hybrid zone, in which parental types dominate a particular location, with a deficit of hybrid types. At the other extreme of this continuum, unimodal hybrid zones are characterized by a predominance of hybrid genotypes. However, we detected a trimodal distribution of hybrid indices in Metedeconk Creek (sampling location #4), which is located at the centre of the cline in mtDNA genotype along the coast. Such a signature indicates that F1 hybrids are being produced, albeit at a low frequency, but backcross types, particularly those in the bins 0.6-0.699 to 0.8-0.899, appear to be lacking, suggesting the possibility of selection against advanced generation hybrids. Similarly, in the Navesink River (sampling location #5) we detected an essentially flat

profile of hybrid indices. Such a pattern is also indicative of selection against hybridization, but our sample size was low in this location, so this conclusion requires additional data.

The results of the hybrid index analyses coupled with the observation of elevated occurrences of linkage disequilibrium in some of the marshes near the centre of the mtDNA cline (marshes 2-8) reinforce the conclusion that there is some form of reproductive isolation between the northern and southern mitochondrial types of *F. heteroclitus* influencing population structure in this region. The main factors that could promote the formation of the patterns we observed are incomplete premating and/or pre-zygotic isolation, selection against backcross types, and the inability of recombination to breakdown coadapted gene complexes (as evidenced by the presence of linkage disequilibrium among genetic markers within the hybrid zone; Jiggins & Mallet 2000). Our data suggest that several of these factors are operating in Metedeconk marsh.

The maintenance of nonrandom associations of particular genotypic combinations in an area where two subspecies are known to be sympatric can be the result of several phenomena. For example, selection against hybrid offspring produced by matings between "more pure" northern and southern individuals adapted to alternative environments that are constantly migrating in from neighbouring populations can create a balance between dispersal and selection, resulting in the emergence of a tension zone (Barton & Hewitt 1985). Alternatively, phenomena such as assortative mating can also produce these genetic associations in a hybrid zone (Jiggins & Mallet 2000). Indeed, these phenomena are not mutually exclusive, and both may be operating. For example, strong selection against hybrids harbouring maladapted combinations of genotypes as the result of matings between genetically distinct parental types is likely to result in reinforcement through the strengthening of pre-zygotic barriers (i.e., assortative mating, ecological character displacement; Servedio & Noor 2003). Such pre-zygotic barriers tend to be

favored in response to existing post-zygotic barriers (i.e., hybrid sterility, hybrid inviability) because they operate early in an organisms life cycle and as such will reduce an individual's incurred cost of wasting gametes in production of less fit hybrid offspring (Coyne & Orr 2004). At this point it is reasonable to conclude that killifish residing in Metedeconk Creek are experiencing some combination of pre-zygotic and/or post-zygotic isolation that results in our observation of reduced numbers of backcross genotypes in the population.

Conclusions

In summary, we have found evidence of a trimodal hybrid zone connecting the northern and southern subspecies of the killifish, *Fundulus heteroclitus,* for multilocus genotypes assessed with highly variable microsatellite markers. This trimodal pattern in the centrally located marsh in the Metedeconk River estuary where the mtDNA D-loop cline centre is located, coupled with the observed linkage disequilibrium among loci in individuals from other centrally located marshes, suggests some combination of pre- and/or post-zygotic reproductive isolation is operating to prevent the merging of these two subspecies (Jiggins & Mallet 2000). The steepness of the mitochondrial cline and the location of its cline centre proximate to a marsh that also presents a trimodal pattern of nuclear hybrid indices suggests that mitochondrial processes may be playing a role in keeping the two subspecies of killifish separate. Together, these results emphasize the importance of collecting data on both the nuclear and cytoplasmic genomes and considering multilocus genotypes when trying to uncover the underlying genetic structure of closely related hybridizing taxa.

Figure 2-1. Map of collection locations. Numbers correspond to those listed in Table 2-1. Inset of North America indicates location of sampling area.

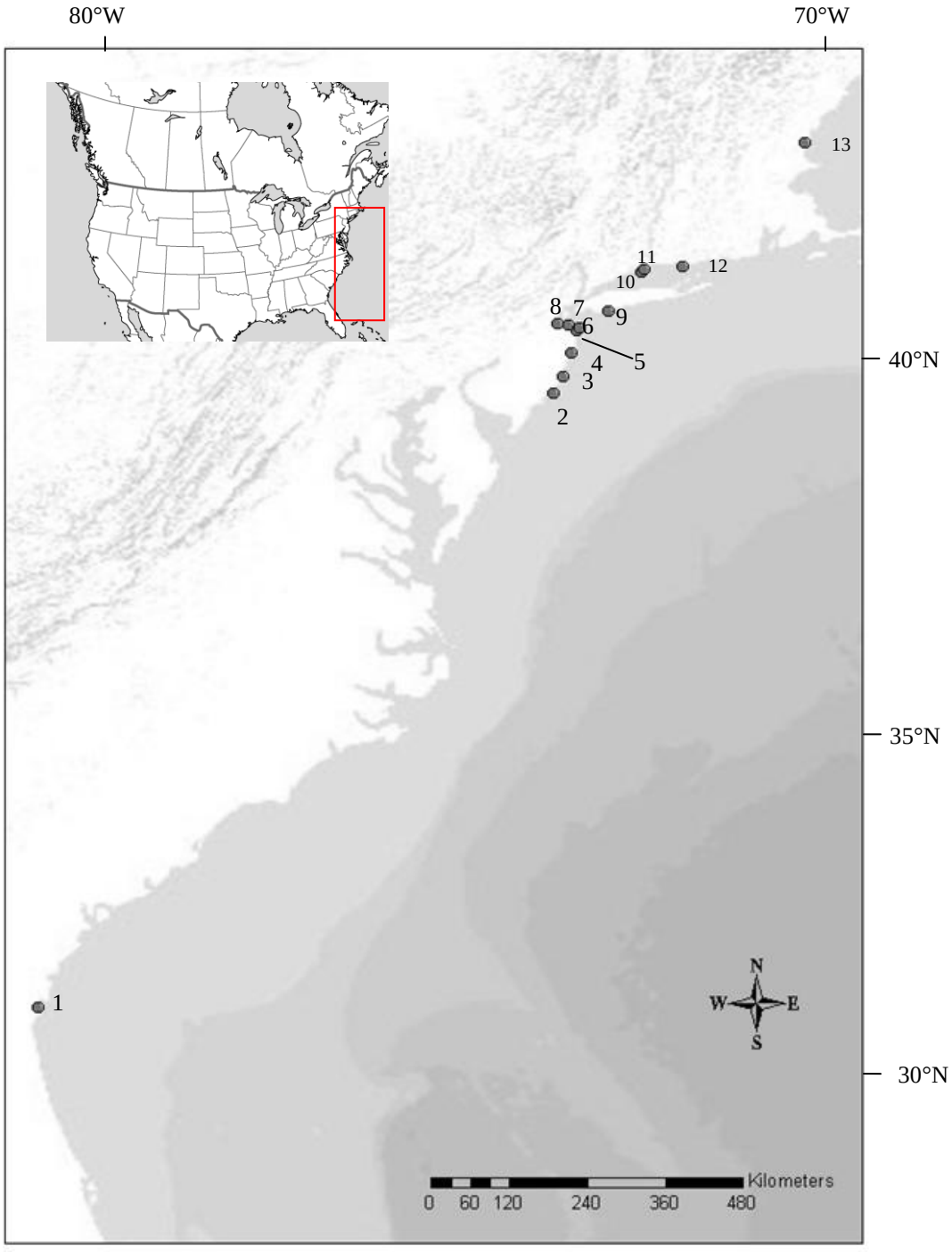


Figure 2-2. Shape of the *F. heteroclitus* mtDNA^{N} (northern mitotype) cline along the coast of North America. Distance refers to distance from Georgia as in Table 2-2.

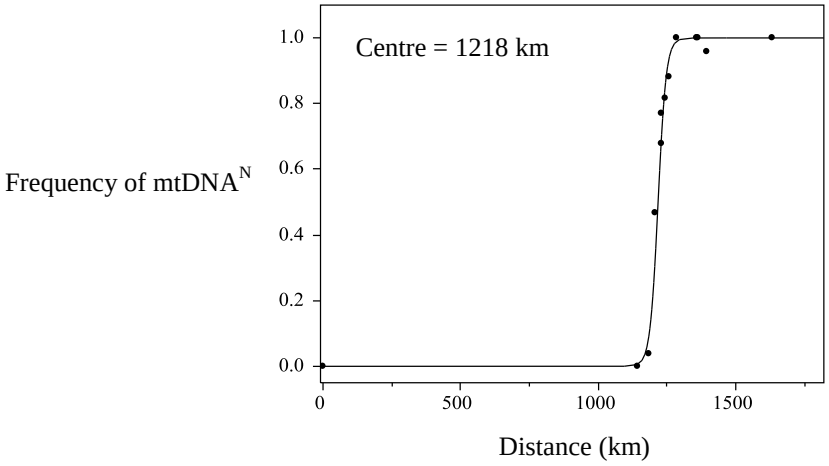


Figure 2-3. Clines in the frequency of the most common northern allele (y-axis) for 9 microsatellite loci: A) FhCA-1, B) FhATG18, C) FhATG20, D) FhATGB128, E) FhATG17, F) FhATG4, G) FhATGB101, H) FhATG2, I) FhATG6. The x-axis depicts distance from Georgia as in Table 2-2.

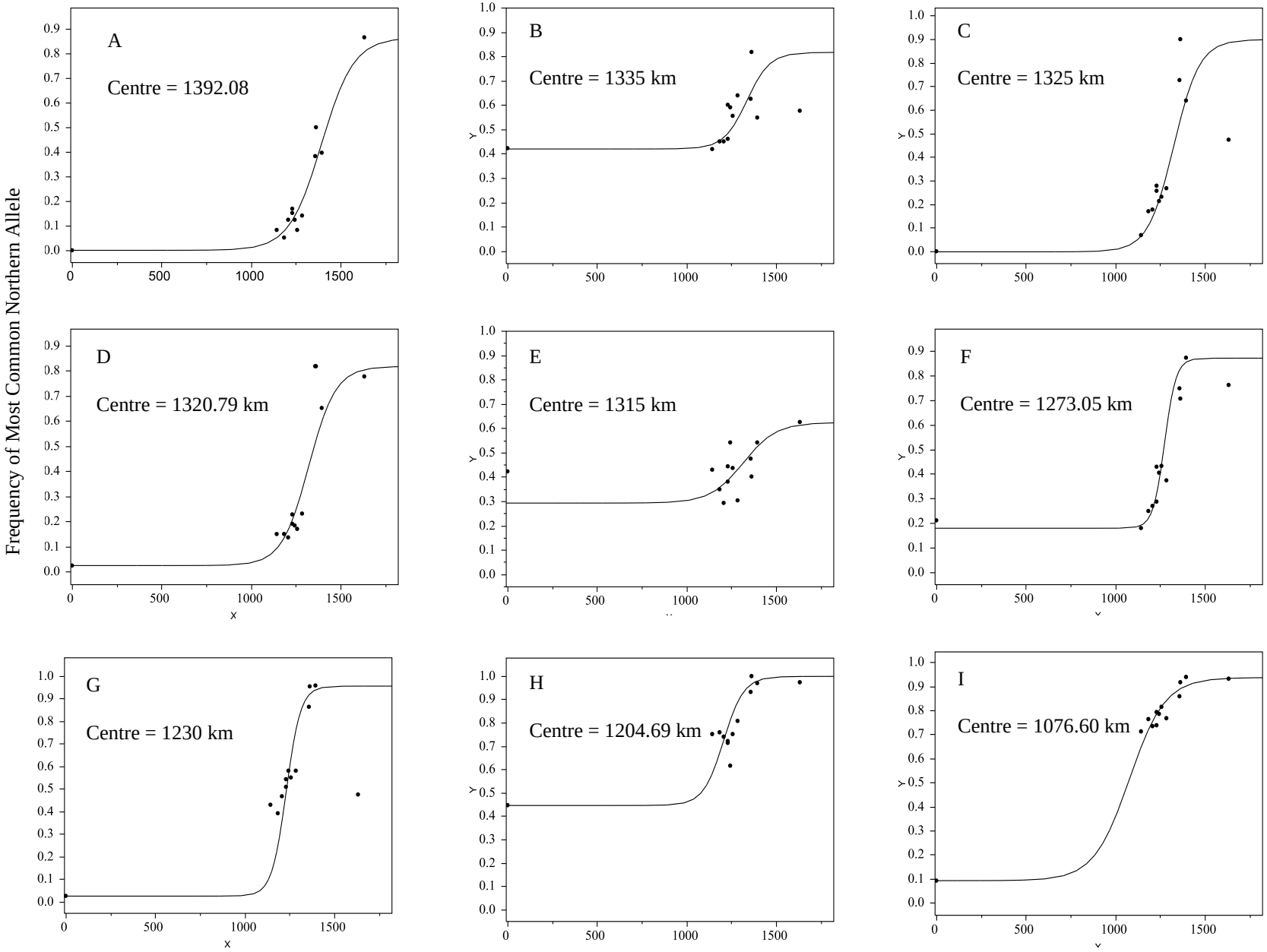


Figure 2-4. Frequencies of hybrid indices calculated using nine microsatellite loci for five sampling locations within the killifish hybrid zone. A hybrid index of 0 represents a pure southern genotype (RUMFS (location 2)-type) while a value of 1 indicates a pure northern (Cheesequake (location 8)-type) genotype.

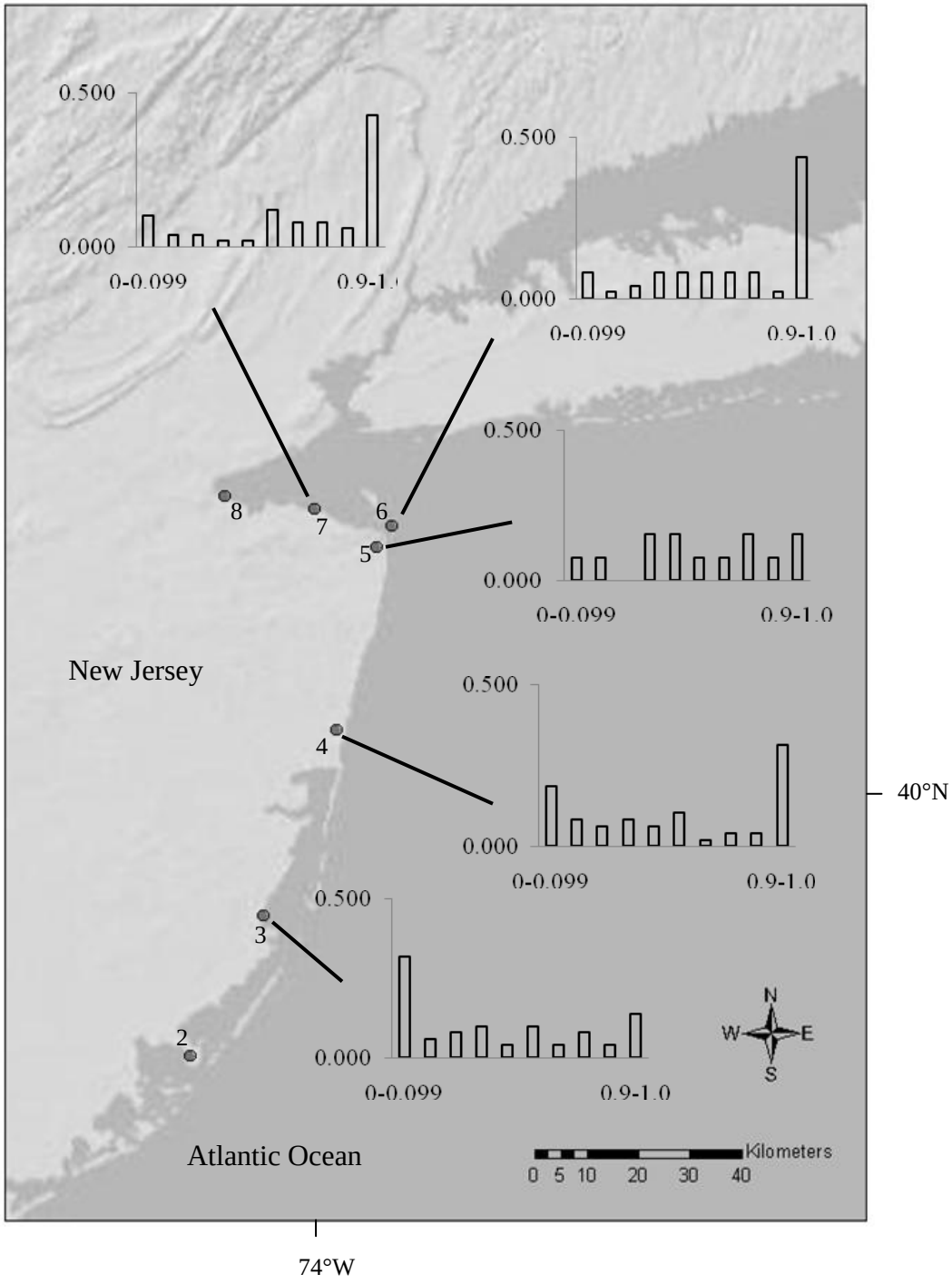


Table 2-1. Sample location and sample size along the Atlantic coast of North America.

Location Name	**Latitude (°N)**	**Longitude (°W)**	**n**
1. Brunswick, Georgia (GA)	31.116130	-81.462250	19
2. Rutgers University Marine Field Station (RUMFS), New Jersey (NJ)	39.512905	-74.320450	50
3. Laurel Harbour, NJ	39.750349	-74.192734	50
4. Metedeconk Creek, NJ	40.065616	-74.065320	49
5. Navesink River, NJ	40.376297	-73.993993	13
6. Sandy Hook Bay, NJ	40.412904	-73.969487	50
7. Belford Creek, NJ	40.440968	-74.103963	49
8. Cheesequake Creek, NJ	40.463420	-74.258888	50
9. Wantagh, New York (NY)	40.638689	-73.563080	39
10. Housatonic River, Connecticut (CT)	41.173856	-73.103081	22
11. Roger's Creek, CT	41.213379	-73.057564	12
12. Hammonassett Creek, CT	41.251971	-72.537634	48
13. Hampton, New Hampshire (NH)	42.937055	-70.839844	20

Table 2-2. Frequency of the northern mtDNA genotype and the most common northern allele for each microsatellite in *F. heteroclitus*. Location numbers correspond to those given in Table 2-1. Distance from Georgia is calculated as the straight line distance along the coastline. The total number of alleles at each locus and at each location are given in parentheses.

Location	Distance from Georgia (km)	$mtDNA^N$	FhATG18 (10)	FhATG20 (13)	FhATG17 (7)	FhATG B101 (16)	FhATG B128 (17)	FhATG2 (8)	FhATG4 (22)	FhATG6 (7)	FhCA-1 (34)
1.	0	0.000	0.421 (6)	0 (8)	0.421 (4)	0.026 (8)	0.026 (7)	0.447 (6)	0.211 (7)	0.094 (6)	0 (15)
2.	1145.77	0.000	0.420 (6)	0.070 (7)	0.430 (2)	0.420 (10)	0.150 (8)	0.750 (5)	0.180 (10)	0.714 (4)	0.082 (14)
3.	1186.03	0.040	0.450 (5)	0.170 (7)	0.350 (2)	0.390 (12)	0.150 (9)	0.760 (5)	0.250 (14)	0.765 (5)	0.052 (18)
4.	1208.94	0.469	0.449 (8)	0.176 (9)	0.294 (6)	0.427 (13)	0.136 (9)	0.741 (4)	0.268 (12)	0.735 (4)	0.122 (13)
5.	1228.41	0.769	0.602 (5)	0.258 (8)	0.444 (3)	0.563 (11)	0.227 (11)	0.713 (3)	0.285 (8)	0.792 (3)	0.167 (10)
6.	1231.88	0.680	0.460 (7)	0.280 (7)	0.380 (4)	0.510 (12)	0.190 (9)	0.720 (4)	0.430 (8)	0.740 (5)	0.150 (18)
7.	1242.06	0.816	0.592 (5)	0.214 (9)	0.543 (3)	0.582 (10)	0.184 (9)	0.615 (5)	0.406 (9)	0.786 (4)	0.122 (13)
8.	1257.54	0.880	0.555 (6)	0.233 (8)	0.438 (3)	0.556 (11)	0.169 (9)	0.753 (5)	0.432 (9)	0.816 (4)	0.082 (14)
9.	1285.31	1.000	0.641 (5)	0.269 (9)	0.303 (3)	0.579 (9)	0.231 (8)	0.808 (2)	0.372 (11)	0.769 (4)	0.141 (14)
10.	1357.23	1.000	0.625 (5)	0.725 (5)	0.475 (3)	0.864 (3)	0.818 (4)	0.932 (2)	0.750 (4)	0.857 (4)	0.381 (8)
11.	1363.41	1.000	0.818 (3)	0.900 (3)	0.400 (3)	0.955 (2)	0.818 (2)	1.000 (1)	0.708 (5)	0.917 (3)	0.500 (5)
12.	1393.09	0.958	0.548 (7)	0.639 (5)	0.541 (5)	0.962 (3)	0.652 (4)	0.970 (3)	0.873 (5)	0.938 (6)	0.396 (7)
13.	1631	1.000	0.575 (6)	0.475 (4)	0.625 (3)	0.475 (7)	0.775 (4)	0.974 (2)	0.763 (3)	0.933 (3)	0.867 (3)

Table 2-3. Deviations from HWE (F_{IS}) for microsatellite loci among coastal populations of *F. heteroclitus*. Positive F_{IS} indicate heterozygote deficit while negative F_{IS} indicate heterozygote excess. F_{IS} values significant at $\alpha = 0.05$ in bold. F_{IS} values significant after FDR-adjustment are marked with an asterisk (*). The final column indicates the number of pairs of loci showing significant linkage disequilibrium. The number of loci remaining significant after FDR- adjustment are shown in parentheses. Grey shading indicates sites located within New Jersey.

		Locus								# pairs of loci in linkage disequilibrium
Location	**FhATG18**	**FhATG20**	**FhATG17**	**FhATGB101**	**FhATGB128**	**FhATG2**	**FhATG4**	**FhATG6**	**FhATG-CA1**	
1	0.081	0.029	0.161	-0.057	-0.186	-0.049	-0.067	-0.041	-0.000	6 (0)
2	**0.154**	0.030	0.196	0.018	0.000	-0.015	-0.055	-0.018	0.009	7 (2)
3	0.091	-0.068	0.123	-0.025	-0.086	-0.055	**0.061**	-0.025	**-0.083**	7 (2)
4	0.041	-0.057	0.123	0.068	0.134	0.038	-0.099	0.044	**-0.112**	5 (2)
5	-0.206	-0.039	0.189	0.028	0.052	-0.375	0.219	-0.146	-0.143	2 (0)
6	-0.055	0.019	0.103	**0.127***	0.045	-0.005	0.050	-0.069	-0.050	11 (8)
7	0.018	0.042	0.201	0.044	-0.019	0.217	-0.047	-0.053	0.007	2 (0)
8	**-0.168**	0.064	-0.061	-0.055	0.066	0.100	-0.006	-0.079	-0.067	4 (2)
9	-0.110	0.083	0.070	**0.254**	-0.0041	-0.226	-0.021	0.002	-0.055	7 (2)
10	-0.176	-0.097	-0.301	0.091	-0.147	-0.050	0.216	-0.086	-0.314	3 (0)
11	-0.127	-0.029	0.018	NA	-0.177	NA	0.160	**0.500**	-0.065	2 (0)
12	**-0.211**	**-0.097**	**-0.235**	-0.025	0.029	**0.490**	-0.090	-0.029	**-0.147**	2 (0)
13	0.101	-0.162	-0.106	-0.062	**0.360**	NA	0.060	-0.018	-0.087	3 (0)

Table 2-4. Summary of MCLUST analysis fitting one, two, and three clusters to hybrid index frequency distribution of individuals of *Fundulus heteroclitus* from Metedeconk, NJ. (BIC = the Bayesian Information Criterion)

Number of Clusters	BIC
One	-53.72
Two	-9.51
Three	4.25

Chapter Three: Steep clines in mitochondrial and nuclear-encoded genes suggest that selection acts to maintain a hybrid zone in Atlantic killifish, *Fundulus heteroclitus*

Synopsis

Steep clines in genetic traits can result from recent secondary contact between previously isolated taxa. If selection maintaining such clines is weak or absent, neutral processes will gradually erode and widen clines over time. Alternatively, clines resulting from historical secondary contact can be maintained by endogenous selection against hybrids and/or by exogenous ecological selection or by assortative mating. We used 30 nuclear and two mitochondrial SNPs to determine if selection or assortative mating is involved in maintaining clines in a small topminnow, *Fundulus heteroclitus*. Fourteen of the 32 loci surveyed had cline widths inconsistent with neutral expectations, using a dispersal distance derived from mark-recapture data, which is likely to underestimate true dispersal and provide a conservative estimate of the neutral cline width. Increased heterozygote deficit and cytonuclear disequilibrium were observed among individuals from more centrally located marshes, suggesting that some form of reproductive isolation, such as assortative mating or selection against hybrids, may be acting in this hybrid zone. The narrowest clines were observed for the mitochondrial SNPs. The nuclear encoded mitochondrial genes phosphate carrier protein and HD domain containing isoform 2 (HDDC2), and several protein synthesis and oxygen-related genes formed a subset of steep coincident and concordant nuclear clines similar to the mitochondrial clines, suggesting cytonuclear epistasis as a potential selective factor. However, the lack of coincidence and concordance among other clines argues for additional factors operating in the zone. Together these data suggest that some form of selection is acting to maintain this hybrid zone.

Introduction

Broadly defined, hybrid zones are areas where divergent taxa meet, mate, and produce at least some hybrid offspring (Barton & Hewitt 1989). Hybrid zones may be formed transiently when previously isolated taxa come into contact, but two main classes of stable hybrid zone are recognized. A tension zone is maintained by a balance between dispersal of parental taxa into the zone and selection against hybrids produced by the mating of two genetically distinct parental types (Barton & Hewitt 1985). In this case selection against hybrids is endogenous, with these "disharmonious recombinants" suffering from hybrid inviability, sterility, or generally decreased fitness, irrespective of the environment in which they find themselves (Dobzhansky 1940; Moore & Price 1993). The second class of hybrid zone is located along an environmental gradient where exogenous selection operates on parents and offspring differentially suited to divergent habitats (Moore & Price 1993). However, endogenous and exogenous selection can interact, and even become coupled in a positive feedback loop (Bierne *et al*. 2011). In addition, assortative mating can act (either independently or in concert with endogenous or exogenous selection) to maintain a hybrid zone (Brodin & Haas 2009).

Fundulus heteroclitus, or the Atlantic killifish, is an abundant topminnow found primarily in estuarine marshes along the east coast of North America, ranging from Newfoundland to northern Florida (Hardy 1978). Currently, two subspecies are recognized: *Fundulus heteroclitus heteroclitus* spans the southern range of the distribution while *Fundulus heteroclitus macrolepidotus* is found in the northern end of the range. Morphological, embryological, and genetic differences between the subspecies support a model of secondary intergradation wherein two previously isolated forms diverged in allopatry only to come back into contact with one another along the coast of New Jersey (Morin & Able 1983; Able & Felley

1986; Gonzalez-Villasenor & Powers 1990; Ropson *et al*.1990; Strand *et al*. 2012). This history of secondary contact has resulted in the formation of geographical clines at many loci in this species (Strand *et al.* 2012). Although neutral demographic processes due to secondary contact (likely following the last glaciation) have been important in forming observed allele frequency clines, the role of selection in maintaining the hybrid zone remains unclear.

Extensive previous data, particularly focusing on clines in the heart-type isozyme of lactate dehydrogenase (LDH-B; Appendix A Table A.1), present the possibility that exogenous selection may be acting in this hybrid zone. Individuals from populations spanning the shoreline of the east coast of North America show a moderately steep cline in *Ldh-B* allele frequency, with the northern genotype being rapidly replaced by the southern genotype across a zone roughly corresponding to that previously observed for morphological, embryological, and mitochondrial traits (Powers & Place 1978). Genetic variation in this glycolytic enzyme is functionally significant (Place & Powers 1984a; DiMichele & Powers 1982a; Powers *et al*. 1979; DiMichele & Powers 1982b; Paynter *et al*. 1991; DiMichele & Westerman 1997), and may be under selection in response to temperature differences along the coast. Similarly, a variety of analyses suggest that variation in gene expression may also be under environmental selection in this species (Schulte *et al.* 1997; Schulte *et al*. 2000, Whitehead & Crawford 2006). However, the steepness of the LDH-B cline is not significantly different from neutral expectations (Strand *et al*. 2012), suggesting that other factors may be important in maintaining this hybrid zone.

The data presented in Chapter Two provide evidence for the existence of a trimodal hybrid zone separating the two subspecies of Atlantic killifish, suggesting that selection is operating against advanced generation backcross types. While the microsatellites provide a reliable snap-shot of the contemporary genetic structure of these populations, the locations and

associations of these markers in the nuclear genome are not known. The use of SNPs known to lie within or close to functional coding regions of the genome can provide further insight into the underlying evolutionary processes responsible for the contemporary patterns elucidated by the microsatellite data. Thus, to provide insight into possible factors maintaining the hybrid zone in *F. heteroclitus* we used genotypic data from 32 diagnostic or semi-diagnostic SNPs (30 located in the nuclear genome and two in the mitochondrial genome) to assess levels of heterozygote deficit, linkage disequilibrium and cytonuclear disequilibrium among individuals collected from marshes spanning the distribution of *F. heteroclitus*. We performed clinal analyses on these loci to determine which, if any, appeared to depart from neutral expectations. Loci exhibiting clines that are coincident (having the same cline centre) and concordant (having the same width and shape) could be the result of historical patterns of secondary contact, or could be responding to the same set (and magnitude) of evolutionary forces (e.g., endogenous selection for particular genotypic combinations or exogenous selection acting in the same way across multiple loci). In contrast, clines that are non-concordant might be the result of differing selective pressure across loci or stochastic neutral variation as separation between the two parent taxa gradually erodes. In addition to these clinal analyses, we also performed a comparison of our cline in lactate dehydrogenase B allele frequencies with *Ldh-B* allele frequency information from Powers & Place (1978) to see if the shape of this cline remained the same over the course of the last three and a half decades, implying either that the cline is stable or that not enough time has passed to observe a shift in the cline.

Materials and Methods

Fish Collection

Fundulus heteroclitus were collected from 35 locations along the Atlantic coast of North America and included samples from previous studies by Powers & Place (1978) and Williams *et al.* (2010) (Table 3-1; Figure 3-1). Samples collected by McKenzie were collected in May-June of 2008 using minnow traps and trap set time ranged from two to six hours. Fish length was recorded and a fin clip was taken from each individual and preserved in 95% ethanol. DNA was extracted from these samples using Qiagen's DNeasy® Blood and Tissue Kit. Multi-locus SNP genotypes for individuals collected from locations 2, 6, 8, 9, 21, 28, 30, and 35 (Table 3-1; Figure 3-1) have been previously published (Williams *et al*. 2010). Information for fish from all remaining locations was obtained from Powers & Place (1978) and was only used for the comparison of past and more recent cline parameters of the locus LDHB1033.

Single Nucleotide Polymorphism (SNP) Genotyping

Four hundred and eighty five individuals from 15 locations (Table 3-1, Figure 3-1) were genotyped at 30 nuclear SNPs and two mitochondrial SNPs selected from an existing panel of 458 genome-wide SNPs from *F. heteroclitus* (Williams *et al*. 2010). To develop this SNP panel, we used JMP Genomics 3.2 for SAS 9.1.3 to conduct SNP case-control trait association tests to identify SNPs that differed significantly in genotype frequency between the extreme northern and southern populations surveyed by Williams *et al.* (2010) (Wiscassett, ME and Sapelo Island, GA; 20 individuals at each site). We further narrowed this set by selecting only those SNPs that were fixed (or were close to fixation) in the more genetically diverse southern population. We made this selection to increase the fraction of SNPs in the SNP panel that were diagnostic or

semi-diagnostic for "northern" or "southern" genomic contributions in hybrid individuals. A total of 41 SNPs met our selection criteria and these sequences were used by the McGill University and Génome Québec Innovation Centre to design a custom genotyping assay using Sequenom® iPLEX®Gold Genotyping Technology. Nine of these SNPs failed to meet the quality control standards of the assay, and exhibited poor amplification or unreliable identification of hybrid individuals, and thus were dropped from the final SNP panel, which consisted of 30 SNPs in the nuclear genome and 2 SNPs in the mitochondrial genome that could be scored with high accuracy (Appendix C Table C.1). Genomic DNA was quantified and diluted to 20 ng/µl and at least 30µl of each sample were sent on dry ice to the McGill University and Génome Québec Innovation Centre for genotyping using this SNP panel.

Genepop 4.0.10 (Raymond & Rousset 1995) was used to conduct an exact test for Hardy-Weinberg equilibrium, testing for heterozygote excess and deficiency for each SNP locus in each population. In addition, this program was used to test for cytonuclear disequilibrium among the SNPs. Markov chain parameters were set to the defaults (dememorization number = 1000; 100 batches; 1000 iterations per batch). The Black and Krafsur procedure, as implemented in Genetix 4.05.2, was used to test each pair of SNP loci among individuals from each population for evidence of linkage disequilibrium (Cockerham & Weir 1977; Belkhir *et al.*1999).

Cline Analyses

ClineFit (Porter *et al*. 1997) was used to estimate the width, centre, and minimum and maximum frequencies of asymptotic polymorphisms (p_{min} and p_{max}, respectively) of all 32 SNPs. For these analyses, our most southern sampling location (Sapelo Island, Georgia) was set to zero and the distance of each more northerly population from this one was measured as the straight line distance along the coast using ArcGIS ArcMap 10 (ESRI 2010). Clines were fit using new

cline shape parameters (as opposed to testing the fit of the data to a predefined set of parameters), and each run was initialized with rough settings (as opposed to the user providing starting values for the center, width and asymptotes of each SNP). All clines were then plotted using the logistic 4p model in the program JMP® 9.0.2 (2010 SAS Institute Inc.). This model fits a sigmoid curve on the presented data using the four parameters, theta1 (p_{max}), theta2 (p_{min}), theta3 (slope) and theta4 (centre) as determined using ClineFit. ClineFit also calculates 2-unit support limits, analogous to 95% confidence intervals, for each parameter. The resulting cline width estimations can then in turn be used to calculate the number of generations that have elapsed under the assumption of a neutral cline using the equation:

$$T = \left(\frac{w}{\left(\sqrt{2\pi}\right)\sigma}\right)^2 \qquad \text{Equation (1)}$$

where T represents the number of generations since contact under the assumption of a neutral cline, and σ represents dispersal per generation (Barton & Gale 1993).

We also used the results from the SNP data to calculate an estimate of dispersal as described by Barton & Gale (1993). This model assumes the presence of a tension zone characterized by random mating and selection against hybrids. First, we estimated average pairwise linkage disequilibrium ($\bar{D}$) using:

$$var\ (z) = \frac{1}{2n}\left(\bar{z}(1-\bar{z}) - var(p)\right) + \frac{1}{2\left(1-\frac{1}{n}\right)\bar{D}} \qquad \text{Equation (2)}$$

where z = hybrid index as calculated using the program INTROGRESS (Gompert & Buerkle 2009, 2010) for individuals from the marsh located at the centre of the mitochondrial cline (Metedeconk) using individuals from Wiscasset, ME and Sapelo Island, GA as northern and southern parental types, respectively. Thus, $\bar{z}$ and var (z) are the mean and variance of the

hybrid index for these individuals, respectively, and n = the number of loci. The variance of allele frequency across the n loci was calculated as $var\ (p) = {}^{1}/_{n}\sum[(pi] - \bar{p})^{2}$, where p_i = the frequency of the northern allele at the i^{th} locus and $\bar{p}$ = the mean frequency of northern alleles across all loci.

The resulting value for $\bar{D}$ was then substituted into:

$$\bar{D} = \frac{\sigma^2}{r\bar{w}^2} \qquad \text{Equation (3)}$$

where r = recombination rate (0.5) and $\bar{w}$ = mean cline width. Equation (3) was subsequently solved for σ.

Ldh-B Cline in 1978 vs. Current Ldh-B Cline

Both 1978 and contemporary clines were plotted using the logistic 4p model in JMP® 9.0.2 (2010 SAS Institute Inc.) as previously described (Chapter Two). Theta1 and theta2 were fixed based on the maximum and minimum observed allele frequencies, respectively (Brumfield *et al.* 2001) and the resulting estimates for slope (theta3) and cline centre (theta4) were read directly from the JMP output.

Results

Heterozygote Deficit, Linkage Disequilibrium, and Cytonuclear Disequilibrium

Of the 30 nuclear SNPs genotyped, 20 deviated significantly from Hardy-Weinberg Equilibrium (HWE) expectations among individuals in at least one of the 15 locations sampled, and these deviations were dominated by heterozygote deficits (92% had positive values of F_{IS}), with the majority of the deviations (78%) occurring in populations within the borders of New

Jersey (Table 3-2). Ten of these loci still showed significant deviations from HWE following adjustment for false discovery rate (FDR; Pike 2010). Of the 24 significant deviations that withstood FDR adjustment, 23 were heterozygote deficits and all occurred among individuals from collection locations within the borders of New Jersey.

Following FDR-adjustment, there was little evidence of significant linkage disequilibrium. However, this approach to multiple test correction might be conservative. In any given marsh, a total of 435 possible pairwise tests for linkage disequilibrium were performed. If 5% of these tests result in false rejections of the null hypothesis that two loci are in linkage equilibrium with one another, we would expect 21 significant associations in any given marsh by chance. The result of these analyses on individuals from all locations except location 13 (Cheesequake, NJ) all exceed this level (Table 3-2; Appendix C Table C.2) with positive values of Rij (indicating an association among alleles from like parental types) being more common than negative values of Rij.

There was also a greater than expected occurrence of cytonuclear disequilibrium within the more centrally located marshes (Table 3-2). Marshes peripheral to the hybrid zone showed no evidence of loci in disequilibrium with the two mitochondrial SNPs genotyped. However, within the hybrid zone, individuals from four marshes showed significant associations between mitochondrial and nuclear genotypes. In marsh 13 (Cheesequake, NJ) the SNP in NACA was found to be significantly associated with cytochrome B. In marshes 15 (Belford Creek, NJ) and 16 (Sandy Hook, NJ), titin cap was found to be in significant association with both of the mitochondrial SNPs, cytochrome B and cytochrome oxidase subunit I (COXI). Lastly, in marsh 19 (Metedeconk) the nuclear SNP in activated protein kinase was found to be in significant

disequilibrium with COXI, and phosphate carrier protein, a nuclear-encoded mitochondrial gene, was found to be in cytonuclear disequilibrium with both COXI and cytochrome B.

Width of Clines in SNP Loci

It is possible to predict the expected width of a neutral cline resulting from secondary contact making some assumptions about the time since contact and dispersal distance per generation. *F. heteroclitus* has a generation time of approximately 1 year, as individuals are capable of breeding at one year of age, and continue breeding each year thereafter up to their typical maximum lifespan in nature of approximately 2-3 years (Abraham 1985). We assumed that the commencement of the last glacial retreat marks the beginning of contact between the two subspecies, and thus it has been approximately 15,000 years (T), or 15,000 generations, since contact occurred. Within subtidal creeks, mark recapture data suggest that individual *F. heteroclitus* are likely to move as much as 2 km within a season in a tidal creek (Fritz *et al*. 1975), and as much as 3.7% of a population may travel greater than 1 km within a complex marsh system consisting of marsh pools and intertidal and subtidal creeks within a single year (Able *et al*. 2012). However, little is known about dispersal between marshes, which would require dispersal across potentially unfavorable habitat along the coast. We used 2 km per generation as a conservative estimate of dispersal distance (σ) because we found no evidence of genetic isolation by distance between marshes at this scale (data not shown), similar to the results of Brown & Chapman (1991), suggesting that dispersal distances along the coast must be at least of this magnitude. Using these assumptions, rearrangement of Equation 1 allows us to predict a cline width (w) of 614 km if diffusion is neutral. The widths of the clines produced by 14 of the SNP loci are significantly less than this predicted value (i.e., 614 km falls outside of the upper 2-unit support limit; Table 3-3 and Figure 3-2). The 2-unit support limits for 13 of the SNP loci

contain the value 614 km; hence, these cline widths are not significantly different from that predicted under a neutral model. The 2-unit support limits of the last five loci, on the other hand, do not contain this value predicted under a neutral model (614 km is less than the lower 2-unit support limit calculated for these loci) thus, these clines are significantly wider than would be predicted under a neutral model, given a 2 km dispersal distance (Table 3-3).

As an alternative approach to estimating dispersal distance, we utilized tension zone theory to calculate a genetic estimate of dispersal distance using average pairwise linkage disequilibrium. This approach yielded a predicted per generation dispersal distance of 14.5 km emphasizing the conservative nature of our dispersal estimate of 2 km per generation used above. Thus, our conclusions relating to the expected neutral cline width and ultimately levels of selection on the narrow clines are, if anything, similarly a conservative underestimate because the greater dispersal distance calculated using linkage disequilibrium suggests that the width of the neutral cline is wider than 614km.

Location of Cline Centres for SNP Loci

As the majority of previously reported clines have their centres located within the borders of the New Jersey coastline, this seemed a reasonable landmark to use when classifying the clines examined in this study. As such, cline centres can be roughly grouped into three categories: 1) those located to the south of New Jersey, and thus south of the mitochondrial cline centres, 2) those that have centres similar to the centre of mitochondrial cline (within the borders of the New Jersey coastline), and 3) those located north of New Jersey and thus to the north of the mitochondrial cline centre (Figure 3-3). Figure 3-4 provides examples of the shapes of clines falling into each of these categories. A figure presenting the shapes of all the clines can be found in Appendix C (Figures C1- C3). We found that five loci exhibited cline centres to the south of

the mitochondrial cline centre, 17 loci yielded cline centre estimates within the bounds of the New Jersey shoreline, and 10 loci had predicted cline centres located to the north of the mitochondrial cline centre (Figure 3-3 and Appendix C Figures C.1 - C.3).

Some general patterns emerge upon closer inspection of the associated cline widths of the loci that fall into these three cline centre categories (compare Figures 3-2 and 3-3). Of the clines having their centres located to the south of the mitochondrial cline centre (loci 1-5 in Figure 3-3), all of their associated widths were either no different or were significantly wider than predicted under a neutral model using our conservative dispersal assumption (Figure 3-4). In contrast, in the ten clines with their centres located to the north of the mitochondrial cline centre (loci 23-32 in Figure 3-3), three showed narrow cline widths (e.g. locus 24: Elastase; locus 27: Hemoglobin a2; locus 30 Hemoglobin a1; Figure 3-4C), while the remaining seven loci had cline widths that were either not significantly different from (four loci) or were significantly wider (three loci) than that predicted in a neutral scenario using conservative dispersal assumptions.

On the other hand, of the 17 loci (15 nuclear and 2 mtDNA) that had their cline centres located within the borders of New Jersey (loci 6-22; Figure 3-3), 11 (9 nuclear and 2 mtDNA) showed cline widths significantly less than predicted under our neutral model (loci 9, 12, 13, 15, 18, 19, 20, 21, 22; Figure 3-2). All of these putatively non-neutral clines in nuclear loci had centres very similar to those of the mtDNA clines. The remaining six loci with cline centres in New Jersey (loci 6, 7, 8, 10, 11, 14; Figure 3-2) produced clines whose widths were not significantly different from the neutral prediction, and most of these had centres south of the centre of the mtDNA cline. Table 3-4 presents a summary of the nuclear SNP loci exhibiting steep clines, clines with centres similar to the centres of the mtDNA clines, and/or evidence of deviations from HWE or cytonuclear disequilibrium.

The comparison of the cline in northern allele frequencies for LDH-B allozymes generated using data from Powers & Place (1978) and those generated from our more recent sampling efforts show that the cline center has remained the same over the last three and a half decades (Figure 3-5). A comparison of the cline widths, however, suggests that the contemporary cline may be narrower than the cline generated from data collected in 1978. An examination of the 2-unit support limits for the width of our LDHB1033 cline (Figure 3-5A) as calculated in ClineFit (696.3 - 1,201.3 km) shows that the width of the Powers & Place cline (Figure 3-5B) is slightly outside of these limits (1,298.7 km). We could not provide 2-unit support limits for parameter estimates because the Powers & Place data could not be analyzed using ClineFit.

Discussion

In this study we performed an analysis of clines in allele frequencies for 32 SNPs in the nuclear (30 SNPs) and mitochondrial genomes (two SNPs) using 15 populations of killifish distributed along the east coast of North America. In addition, we compared the cline centre and width calculated from *Ldh-B* allele frequency data generated over 30 years ago (Powers & Place 1978) to our more recent findings. We also examined these data for deviations from Hardy-Weinberg Equilibrium and for evidence of linkage disequilibrium and cytonuclear disequilibrium to compare to the results for microsatellites presented in Chapter Two. The main conclusions of this study were that i) individuals collected from marshes located in the centre of the hybrid zone (i.e., along the coast of New Jersey) exhibited high levels of heterozygote deficit and cytonuclear disequilibrium when compared to marshes outside of the zone, ii) a subset of very narrow clines in nuclear SNPs were coincident and concordant with the mtDNA cline, and had their cline centres in New Jersey, coincident with previous data from morphological and meristic clines, iii) the majority of remaining nuclear clines had significantly different centres, which were located

either north or south of New Jersey. These data, which are summarized in Table 3-4, indicate that the *F. heteroclitus* hybrid zone is being maintained by contemporary forces that may include exogenous selection, endogenous selection and/or assortative mating.

Cline Steepness and Position

The widths of 14 clines were significantly less than that predicted by a model of neutral diffusion, using a conservative estimate of dispersal based on mark-recapture data (Table 3-3; Figure 3-2; summarized in Table 3-4). The widths of these clines ranged from 75 km (cytochrome b) to 457 km (hemoglobin a1). Clines this steep would have to have been formed between approximately 225 to 1,300 years ago in the absence of selection, assuming a dispersal distance (σ) of 2 km per generation. These data suggest that at least some of the clines in our 32 diagnostic and semi-diagnostic SNP loci are being maintained by some force such as selection or assortative mating. These results are broadly consistent with other work on *Fundulus heteroclitus* by Strand *et al*. (2012) who found a small subset of loci with extremely steep clines among the 310 loci that they surveyed, with other loci having much broader clines. Note, however, that we detected a much higher proportion of such steep cline loci (14 of 32 versus ~ eight of 310). This difference reflects the fact that in this study we targeted loci with high levels of differentiation between the geographically extreme populations, which will bias our sample towards loci with steeper clines. Because we used a subset of the same set of SNPs utilized by Strand *et al*. (2012), we can compare our set of SNPs that violate neutral expectations with those identified by Strand *et al.* (2012) using a different analytical approach. Of the loci detected as significant outliers by Strand *et al*. (2012) that were also assayed here, we also detected these as deviating from neutral expectations.

The majority of clines in morphological, embryological, and genetic traits distinguishing the northern from the southern subspecies of *F. heteroclitus* have been previously described as having their cline centres located at various points along the New Jersey shoreline (Morin & Able 1983; Able & Felley 1986; Gonzalez-Villasenor & Powers 1990; Ropson *et al*.1990). Similarly, we found that 11 of our 14 steepest clines, including the clines in mtDNA, had their centres located within the borders of the state of New Jersey. These results are consistent with those of Strand *et al*. (2012) who also identified a small subset of coincident and concordant clines that had centers similar to that of the mtDNA cline.

Information about the functional roles of the genes containing SNPs with steep clines and coincident centres has the potential to help generate hypotheses regarding the forces maintaining these clines, although it is important to note that the SNPs themselves may not be functionally important, but may simply be markers of functionally important genomic regions. The steepest clines were detected for mtDNA SNPs (Figure 3-2). We detected nine nuclear SNPs that had steep clines that were coincident and concordant with the mtDNA clines. These SNPs are located within genes that are involved in a variety of functions. Two of these genes are nuclear encoded mitochondrial proteins: phosphate carrier protein and SNP 280. Phosphate carrier protein is involved in carrying inorganic phosphate into the mitochondria. It is thus essential for energy metabolism. The function of the gene containing SNP 280 had not previously been annotated. However, a draft of the *Fundulus heteroclitus* genome is nearing completion and we used this draft information to search for the locations of this unknown SNP. We found SNP 280 to be located in the intron of the gene HDDC2, which is a nuclear-encoded mitochondrial gene of unknown function. One of the SNPs whose cline was coincident and concordant with the mtDNA cline was located in the gene encoding myoglobin, which is an oxygen storage

molecule, suggesting a function in the oxygen transport pathway from the environment to the mitochondria. Another of these SNPs was located in the gene coding for Warm acclimation related protein, which is known to increase in expression in response to a variety of stressors (including temperature), and to be related to handling of oxidative stress. Two of these coincident and concordant clines were for SNPs involved in processes related to muscle contraction: tropomyosin, and SNP 65 (which we identified as being located in the intron of an actin binding LIM family gene). Finally, SNPs in several genes involved in protein synthesis (40S ribosomal protein S17), and metabolism (glyceraldehyde 3 phosphate dehydrogenase) were also detected as coincident and concordant with the mitochondrial clines. The final remaining SNP (1176) is in an un-annotated region of the genome, and thus its relationships with known genes remain unclear. Interestingly, these coincident and concordant clines are in SNPs found in loci spanning a range of physiological functions in which these two subspecies of killifish are known to differ significantly (See Appendix A Table A.1). For example, tropomyosin is a gene involved in muscle contraction, and is plausibly associated with differences in swimming performance between *F. heteroclitus* subspecies (Fangue *et al*. 2008). Similarly, variation in the expression of glyceraldhyde-3-phosphate dehydrogenase is associated with variation in cardiac metabolism between *F. heteroclitus* subspecies (Podrabsky *et al*. 2000) and has been suggested to be under selection in this genus (Pierce & Crawford 1997). Therefore, it is possible that these processes are under selection in *F. heteroclitus*, or are responsible for phenotypes that result in assortative mating or other factors that could result in the maintenance of steep clines. However, it is important to note that these SNPs may simply be markers of genomic locations under selection, rather than functionally important sites, per se.

We were also able to perform a comparison of our data for the locus *Ldh-B* with data first published by Place & Powers in 1978. The cline centre for the SNP LDHB 1033, which is responsible for the allozyme variation detected by Place & Powers (1978) (Powers *et al*. 1991) has remained the same over 34 years, suggesting that this cline has been stable over this time period.

Evidence of Selection Against Hybrids or Assortative Mating

We found evidence of selection against hybrids or assortative mating in marshes located in New Jersey, near the mitochondrial cline centres. Specifically, we observed significant departures from Hardy-Weinberg equilibrium, with clear evidence of elevated levels of heterozygote deficit, concentrated within these New Jersey marshes. Although departures from Hardy-Weinberg equilibrium can be caused by a variety of factors, including population subdivision, natural selection, avoidance of hetero-specific mating, or assortative mating by phenotype, many of these processes would be expected to have effects on a whole-genome scale. Only selection or assortative mating by phenotype would be predicted to have locus-specific effects. Although heterozygote deficit was clearly more common in New Jersey than elsewhere in the species range, we also observed a single instance of heterozygote excess, and many loci were in Hardy-Weinberg equilibrium, clearly indicating that the heterozygote deficits were caused by a locus-specific phenomenon such as selection or assortative mating by phenotype.

Following FDR-adjustment, nine SNP loci exhibited significant heterozygote deficit, and all of these departures from HWE were detected in New Jersey populations (Table 3-2). Three of these loci (atrial natriuretic peptide, TC19110, and ribosomal protein S2) had relatively wide clines (with widths significantly larger than the predicted neutral width). We identified SNP TC19110 as being located in 3'UTR (untranslated region) of the signaling protein 14-3-3 epsilon,

known to be important in osmosensing signal transduction in the gill. This is significant because the northern and southern subspecies of *F. heteroclitus* are known to differ in salinity tolerance, particularly during development (Whitehead *et al*. 2011; Scott *et al*. 2004), and it is plausible that adults might prefer habitats of differing salinity for laying their eggs, which opens the possibility that assortative mating by phenotype might be influencing genetic patterns at this locus.

Of the remaining six loci exhibiting significant departures from HWE following FDR correction, two had intermediate cline widths consistent with neutral predictions (60S ribosomal protein L35, 60 ribosomal protein L6), while the last four (65, warm acclimation related protein, 280, 1176) displayed extremely narrow cline widths that deviated from neutral expectations. The observed heterozygote deficits coupled with the observed narrow cline widths suggest that these four clines might, at least in part, being maintained by selection against hybrids, or very strong assortative mating by phenotype.

An additional line of evidence supporting selection against hybrids or assortative mating by phenotype is the possible existence of linkage disequilibrium (specifically those having positive values of Rij) in marshes within the borders of New Jersey (Appendix C Table C.2). Linkage disequilibrium could be due to excessive immigration of pure parental types into the hybrid zone, assortative mating by phenotype (Sites *et al*. 1996) or post-zygotic barriers causing hybrid inviability, sterility, or hybrid breakdown (Szymura & Barton 1986; Mallet & Barton 1989). However, the relative roles of these processes in producing the observed weak pattern of linkage disequilibria are difficult to assess.

Evidence for Cytonuclear Epistasis

A special case of linkage disequilibrium, called cytonuclear disequilibrium, occurs when there is an excess of mitochondrial types of one parental form associated with the nuclear type of

the same parental form. While some of the subunits of the complexes of the mitochondrially-housed electron transport chain are encoded by the mitochondrial genome, the majority are transcribed from the nuclear genome and subsequently transported to the mitochondrion to form the multi-subunit complexes of the electron transport chain. Thus, there is a functional interaction between nuclear and mitochondrially-encoded genes (epistasis; Nei 1993). Such epistatic interactions are expected to promote selection favouring the occurrence of matching and thus more functionally compatible cytonuclear types within an individual. *Fundulus heteroclitus* subspecies are known to differ in mitochondrial properties, particularly when acclimated to cold temperatures (Fangue *et al*. 2009; Dhillon & Schulte 2011), suggesting the possibility of functional mismatches in hybrid individuals. Four SNP loci (activated protein kinase, NACA, titin cap, and phosphate carrier protein) were in significant cytonuclear disequilibrium with either one or both of the mitochondrial SNPs (cytochrome b and/or cytochrome oxidase subunit I) in at least one marsh within the hybrid zone. However, the signal of cytonuclear disequilibrium was not particularly clear or unequivocal. For example, activated protein kinase, NACA, and titin cap all had cline centres located significantly south of the mitochondrial cline centres, and their clines were relatively wide, and did not deviate significantly from the neutral prediction derived from conservative estimates of dispersal. Such variability among loci suggests that these three loci are not experiencing the same amount or type of selection as the mitochondrial SNPs. These data imply that the observed cytonuclear disequilibria at these three loci may be the result of Type I error (false rejection of a true null hypothesis). On the other hand, the cline centre of phosphate carrier protein very similar to the cline centres for both mitochondrial SNPs, its lower 2-unit support limit being only 14 km from encompassing the mitochondrial cline centres, and the concordance of cline widths suggests that in this instance the

observed cytonuclear disequilibrium is not spurious. Furthermore, a functional linkage exists between these two markers with phosphate carrier protein being responsible for the transport of phosphate groups, one of the substrates of the mitochondrial F_1F_0ATPase, from the cytosol into the mitochondrial matrix (Becker *et al.* 2003). Thus, at least in this one case (phosphate carrier protein), the evidence supporting cytonuclear epistasis is quite strong, suggesting that endogenous selection due to incompatibilities between nuclear encoded and mitochondrially encoded mitochondrial proteins could be playing a role in maintaining this hybrid zone. However, we did not detect statistically significant cytonuclear disequilibrium for SNP 280, which is located within the nuclear encoded mitochondrial protein HDDC2, which suggests that cytonuclear incompatibilities involve specific aspects of mitochondrial function, most likely those for processes with interacting protein subunits that are encoded in the nuclear or cytoplasmic genomes, such as the electron transport chain and the proton ATPase.

Cytonuclear incompatibilities have been implicated in hybrid breakdown as a result of the co-occurrence of two differentially adapted parental genomes in a variety of species (Edmands & Burton 1999; Burton *et al*. 2006; Ellison & Burton 2008; Niehuis *et al*. 2008). The problem becomes even more complicated when one considers the complexes of the electron transport chain (ETC) which are composed of protein subunits, some of which are encoded in the nuclear genome while others are encoded in the mitochondrial genome (i.e., COXI and CytB). Negative consequences of cytonuclear incompatibilities have been demonstrated in a variety of taxa. For example, in hybrid frog embryos, the activity of Complex IV of the ETC was found to be lower among embryos that possessed the nuclear genome of one parental type but the mitochondrial genome of the other. Normal cytochrome oxidase activity was restored in embryos that contained at least a haploid set of chromosomes originating from the same parent as the mitochondrion

(Liepins & Hennen 1977). Subsequent work by Burton and colleagues has shown similar breakdown among hybrids resulting from crosses between divergent populations of the intertidal copepod *Tigriopus californicus*. These studies have implicated mismatches between nuclear- and mitochondrial-encoded proteins of complex IV of the ETC (Edmands & Burton 1999) as well as between the nuclear-encoded mitochondrial RNA polymerase and its transcription of the mitochondrial genome (Ellison & Burton 2008). Cytonuclear incompatibilities might be common in cases where the distribution of an organism spans a range of thermal habitats, where selection for a particular enzyme variant may be strong as a result of its increased efficiency in a given temperature, as observed with the different *Ldh-B* alleles (Blier *et al*. 2001; Place & Powers 1978). In our case, the intimate association of the phosphate carrier protein with the membrane of the mitochondrion provides a mechanistic explanation for selection of the maintenance of these associations.

Several loci that did not exhibit significant associations with mitochondrial type had clines coincident and concordant with the mitochondrial clines (Table 3-4). The similarity of the cline centres for these loci with those in the mitochondrial markers suggests that the clines may be being maintained by the same force as is the mitochondrial cline (Singhal & Moritz 2012). However, since we detected no signal of cytonuclear epistasis at these loci, it seems unlikely that endogenous selection via this mechanism is acting. Instead, some other mechanism of endogenous or exogenous selection, or assortative mating based on some phenotype associated with all these markers could be producing these patterns.

In summary, of the 12 nuclear clines that were observed to be narrower than neutral expectations derived from conservative assumptions of dispersal, nine were coincident and concordant with the mtDNA clines, and several of these showed evidence of heterozygote deficit

or cytonuclear disequilibrium, strongly suggesting that they are responding to similar selective forces or are associated with a phenotypic complex that results in assortative mating by phenotype. The remaining three SNPs that had very narrow clines were hemoglobin a1, hemoglobin a2, and elastase, but these clines showed no evidence of heterozygote deficit, and their cline centres were significantly different from those of the mtDNA clines suggesting that neither endogenous selection or assortative mating by phenotype would be likely to account for these patterns, and opening the possibility that exogenous environmental selection might be operating. While the potential role of exogenous selection at elastase is not obvious, it is possible that the narrow cline width observed at this locus is the result of physical linkage of this gene to loci under selection. In contrast, significant differences in hemoglobin-oxygen affinity are known to exist between *F. heteroclitus* subspecies and have been found to be significantly associated with differences in swimming performance (DiMichele & Powers 1982a). Swimming performance is a good fitness proxy in fishes (e.g., Dalziel *et al*. 2012), suggesting that exogenous selection may play a role in maintaining these clines. Again, however, it is important to bear in mind that these SNPs may in fact simply be markers for particular genomic regions, and that selection may not be acting on these specific loci.

Interactions Between Endogenous and Exogenous Selection

Environment-independent tension zones originally maintained by endogenous selection against the production of disharmonious hybrid types can become trapped by an extrinsic environmental barrier(s), leading to further divergence among parental types in response to these distinct habitats (Bierne *et al*. 2011). Thus, endogenous and exogenous selection need not be mutually exclusive. The patterns of genetic variation we observed in *F. heteroclitus* suggest that the observed clines are the result of secondary intergradation between previously isolated taxa

and that the resulting tension zone is largely maintained by either by endogenous selection operating against their less fit hybrid offspring primarily due to interactions between genes associated with mitochondrial function that form coincident and concordant clines along the shoreline of New Jersey, or by assortative mating by phenotypes that are influenced by these loci. However, there is also the potential for these genes to be responding to exogenous selection. The northern and southern subspecies of *F. heteroclitus* differ in metabolic rate as adults (Healy & Schulte 2012) and during embryonic development (DiMichele & Powers 1991), and these phenotypes are associated with differences in developmental rate, which are thought to be under selection by the thermal regime along the coast (DiMichele & Westerman 1997; Williamson & DiMichele 1997). The northern subspecies has a higher metabolic and development rate than does the southern subspecies which is thought to be important in adaptation to lower temperatures and reduced growing season length – a pattern that is known as countergradient variation (Conover & Schultz, 1995). Variation in mtDNA sequence and mitochondrial function has been associated with metabolic rate variation in a variety of taxa (Boratynski *et al*. 2011; Tranah *et al*. 2011; Arnqvist *et al*. 2010), which opens the possibility for exogenous selection to be acting on mitochondrially associated clines in *F. heteroclitus*. Interestingly, genotype at the LDH-B allozymes encoded by SNP LDHB1033 has the opposite association with metabolic rate: the southern genotype is associated with higher metabolic and development rate than is the northern genotype. Selection for increased introgression is potentially consistent with the wide cline associated with this locus.

Implications for Estimating Clines Centres

The results of this study highlight the importance of considering cline centres and widths on a locus by locus basis since loci experiencing (or linked to loci experiencing) different levels

of selection (as demonstrated by varying widths) could also have dramatically different cline centres. Determining a cline centre based on admixture proportions calculated from multilocus microsatellite data and calculating a single value for the cline centre (as in Adams *et al*. 2006, Duvernell *et al*. 2008, Flight *et al*. 2011) can result in misinterpretation of patterns of genetic variation because this analysis incorporates loci that may be subjected to varying levels of selection. As a result, it has been repeatedly reported that the mitochondrial cline centre and nuclear cline centre of *Fundulus heteroclitus* are discordant, with the nuclear cline centre being located roughly 70 km north of the mitochondrial cline centre, near the mouth of the Hudson River (Adams *et al*. 2006, Duvernell *et al*. 2008, Flight *et al*. 2011). From our data, it is clear that this is not the case. More than half of our nuclear cline centres were located within the borders of New Jersey (either coincident or nearly coincident with the mitochondrial cline centres), south of the mouth of the Hudson. Furthermore, our narrowest clines were found to be both coincident and concordant with the mitochondrial cline centres which were located very close to the marsh Metedeconk, located in the centre of the New Jersey coastline. Adams *et al*. (2006) assumed that the nuclear cline centre is located near the mouth of the Hudson River and concluded that the Hudson River presented a "formidable barrier" to the movement of northern fish southward and vice versa. We find evidence to the contrary, because we observe loci with cline centres north of this point.

Using an average of multiple cline centres can also lead to potentially incorrect conclusions. For example, in their recent study, Flight *et al*. (2011) observed a differential effect of oxygen levels on the expression of a mitochondrial elongation factor ($EF\text{-}Ts_{mt}$) in killifish from New Jersey possessing northern or southern mitochondrial types. Citing the mouth of the Hudson River as the nuclear cline centre, Flight *et al*. (2011) concluded that these differences in

EF-Ts_{mt} expression were the result of fish harbouring the northern mitochondrial type in a southern nuclear background and thus possessing mismatched cytonuclear types, which caused a "poor response" to hypoxic stress. However, we see here that the cline centres for loci most likely to have a significant effect on fitness have their cline centres located close to Metedeconk, a marsh located only 3 km due north west of Mantoloking, NJ where Flight *et al*. collected their fish. In addition, on average, our fish from Metedeconk were homozygous for northern and southern genotypes at 30% and 36%, respectively, of the 30 nuclear SNPs we genotyped (McKenzie unpublished data). Thus, the fish collected by Flight *et al*. (2011) are likely to have been characterized by a high proportion of pure homozygous northern or southern nuclear genotypes at a variety of loci. Thus, it is possible that the divergent responses in EF-Ts_{mt} expression, exhibited by fish possessing alternate mitochondrial types observed by Flight *et al*. (2011) are a reflection of differences between pure types at functionally important nuclear loci.

Conclusions

The high levels of heterozygote deficit, linkage disequilibrium, and cytonuclear disequilibrium observed here, particularly for loci with their cline centers located within the borders of the New Jersey, suggests that endogenous or exogenous selection against heterozygotes, or assortative mating by phenotype is acting in this hybrid zone. In addition, the observation of several clines in nuclear allele frequencies with significantly narrow widths, indicative of strong selection operating in their maintenance, being coincident and concordant with clines observed at the mitochondrial loci open the possibility that cytonuclear epistasis may also play an important role in the preservation of these two distinct subspecies. In addition SNPs in at least two genes show a pattern consistent with exogenous selection as evidenced by significantly narrow cline widths but having cline centres deflected to the north of the

mitochondrial cline centres. Lastly, the results of the *Ldh-B* analyses indicate that these are likely stable clines, at least over the relatively short timeframe of 34 years; thus, if these latter results are any indication of processes occurring among the other clinally varying traits (genetic, morphological, physiological, embryological, biochemical, etc.) we can conclude that northern and southern subspecies of the killifish may represent evolutionarily stable units.

Figure 3-1. Collection locations for *Fundulus heteroclitus*. Location numbers correspond to those given in Table 3-1. Inset represents locations along the coast of New Jersey.

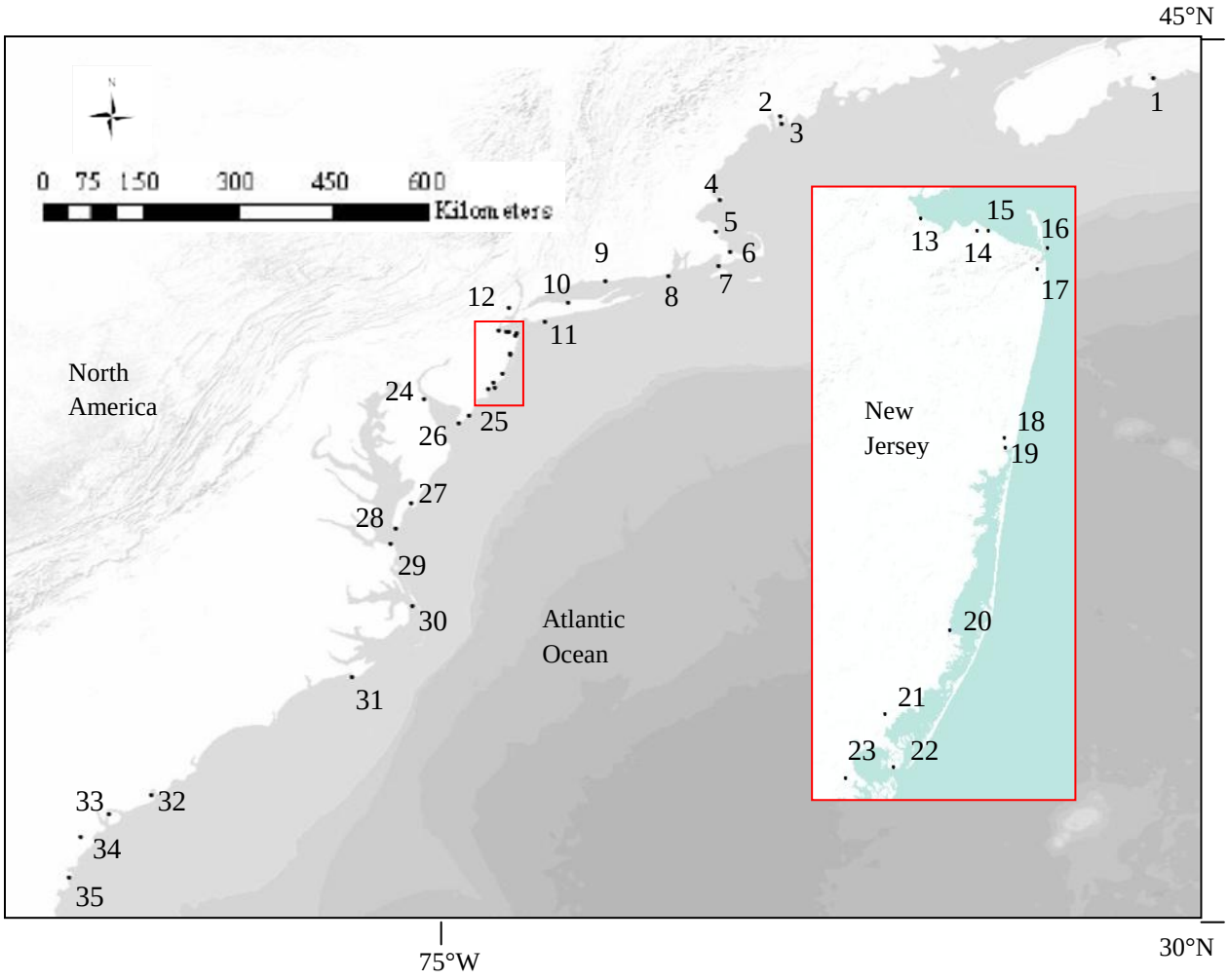


Figure 3-2. Cline widths and 2-unit support limits from ClineFit for 32 SNP loci in *Fundulus heteroclitus*. Mitochondrial loci highlighted in red. Loci are arranged in order of increasing distance of the cline centre from Georgia (see Figure 3-2).Numbering on the x-axis represents the following SNP loci: 1) Atrial Natriuretic Peptide, 2) Chymotrypsinogen, 3) Ribosomal protein S2, 4) Translationally controlled tumor protein, 5) Ribosomal protein 6) Activated protein kinase, 7) NACA, 8) Titin cap, 9) 1176, 10) 60S ribosomal protein L6, 11) Parvalbumin, 12) 65, 13) Myoglobin, 14) 60S ribosomal protein L35, 15) 40S ribosomal protein S17, 16) COXI, 17) CytB, 18) Phosphate carrier protein, 19) 280, 20) Warm acclimation related protein, 21) Glyceraldehyde 3 phosphate dehydrogenase, 22) Tropomyosin, 23) LDHB 654, 24) Elastase, 25) Nucleotide diphosphate kinase I, 26) LDHB 1033, 27) Hemoglobin a2, 28) Cytochrome p450, 29) 19110, 30) Hemoglobin a1, 31) Nucleotide diphosphate kinase 2, 32) 1173. Dashed line represents width of neutral cline calculated from mark-recapture estimate of dispersal.

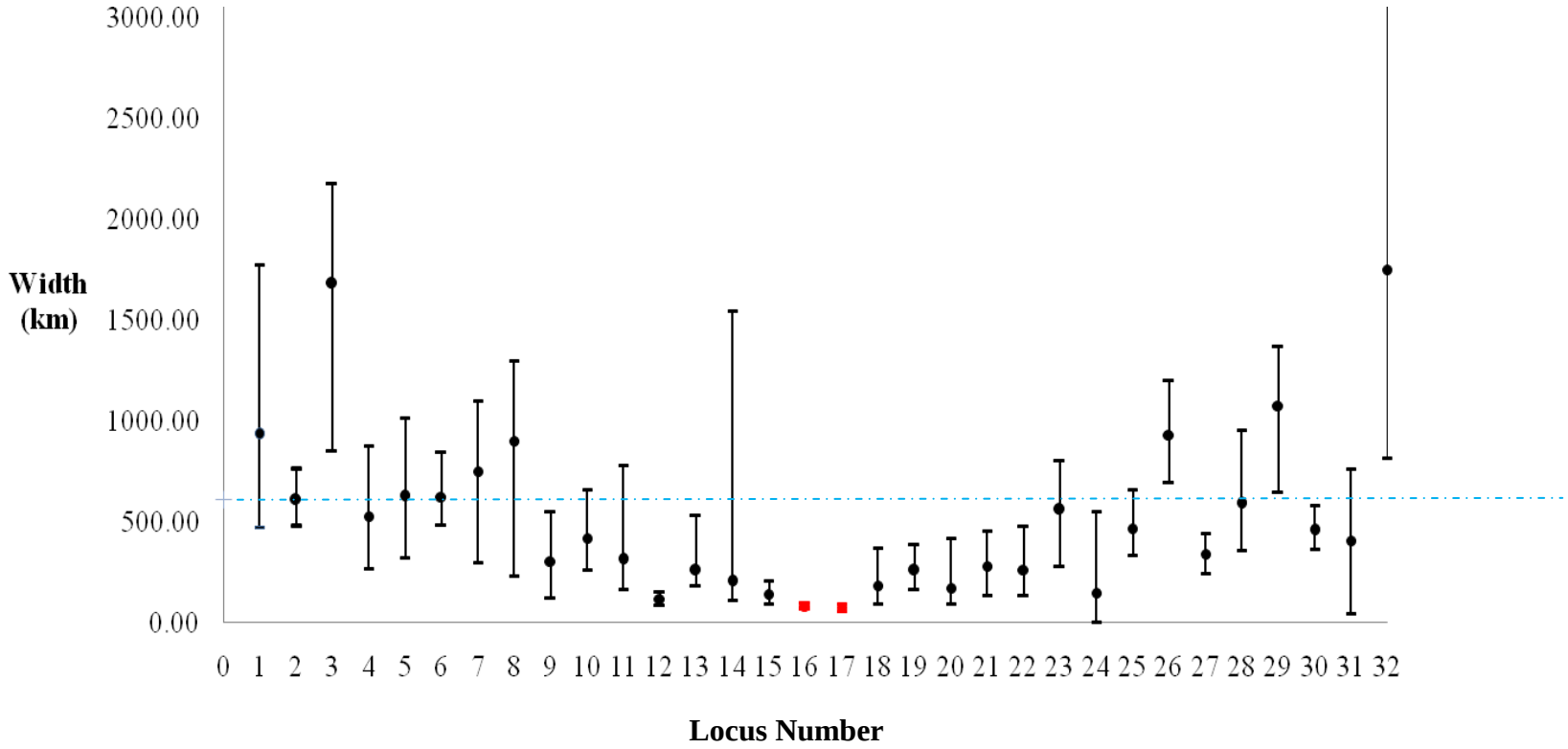

3000.00
2500.00
2000.00
1500.00
1000.00
500.00
0.00
Width (km)
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
Locus Number

Figure 3-3. Cline centres and 2-unit support limits from ClineFit for samples of *Fundulus heteroclitus*. Mitochondrial loci in red; shaded area illustrates New Jersey border. x-axis represents locus names for each single nucleotide polymorphism as in Figure 3-2.

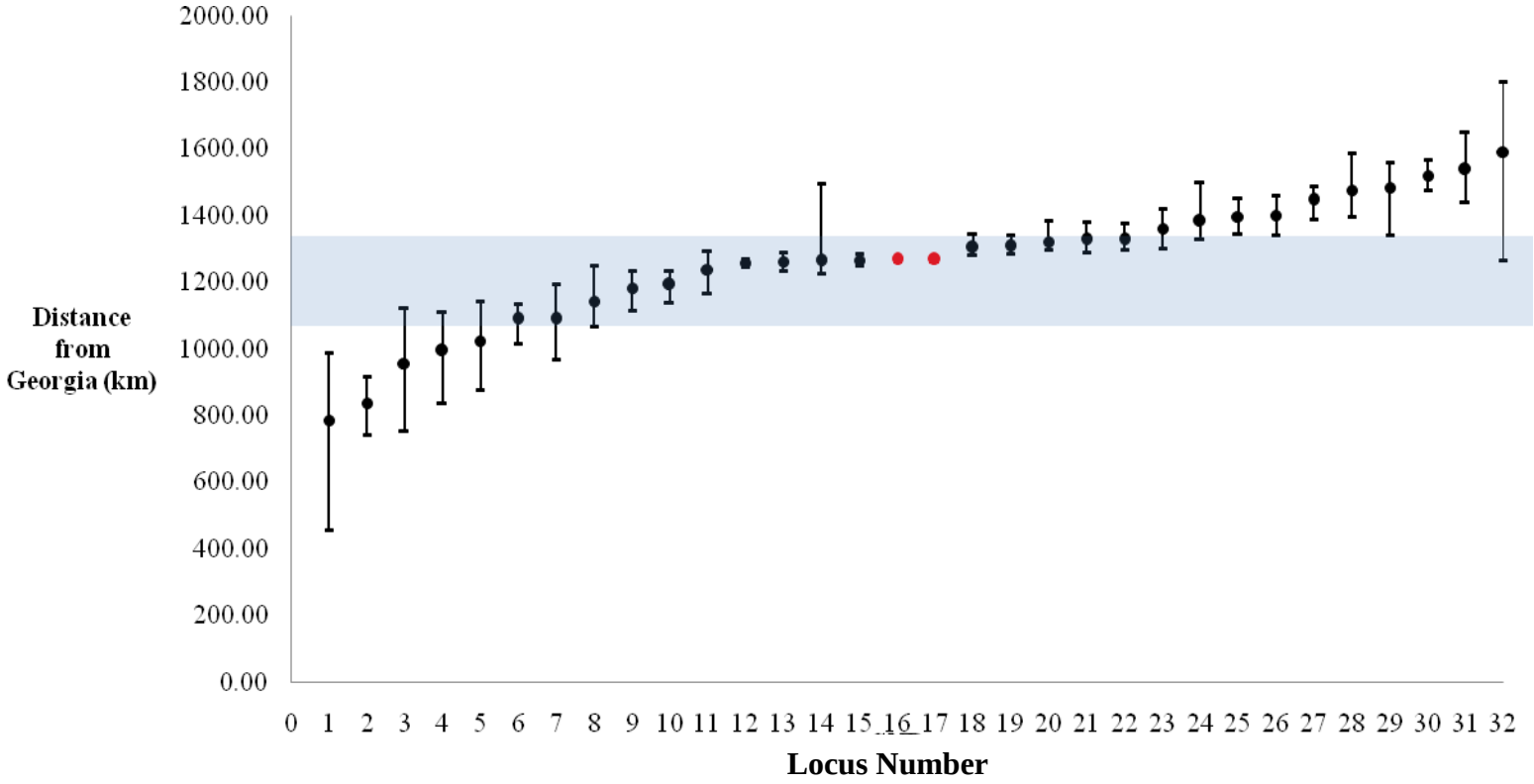


Figure 3-4. Examples of clines, resulting from parameter estimates from ClineFit for samples of *Fundulus heteroclitus*, with centre falling A) south, B) within, and C) north of New Jersey borders.

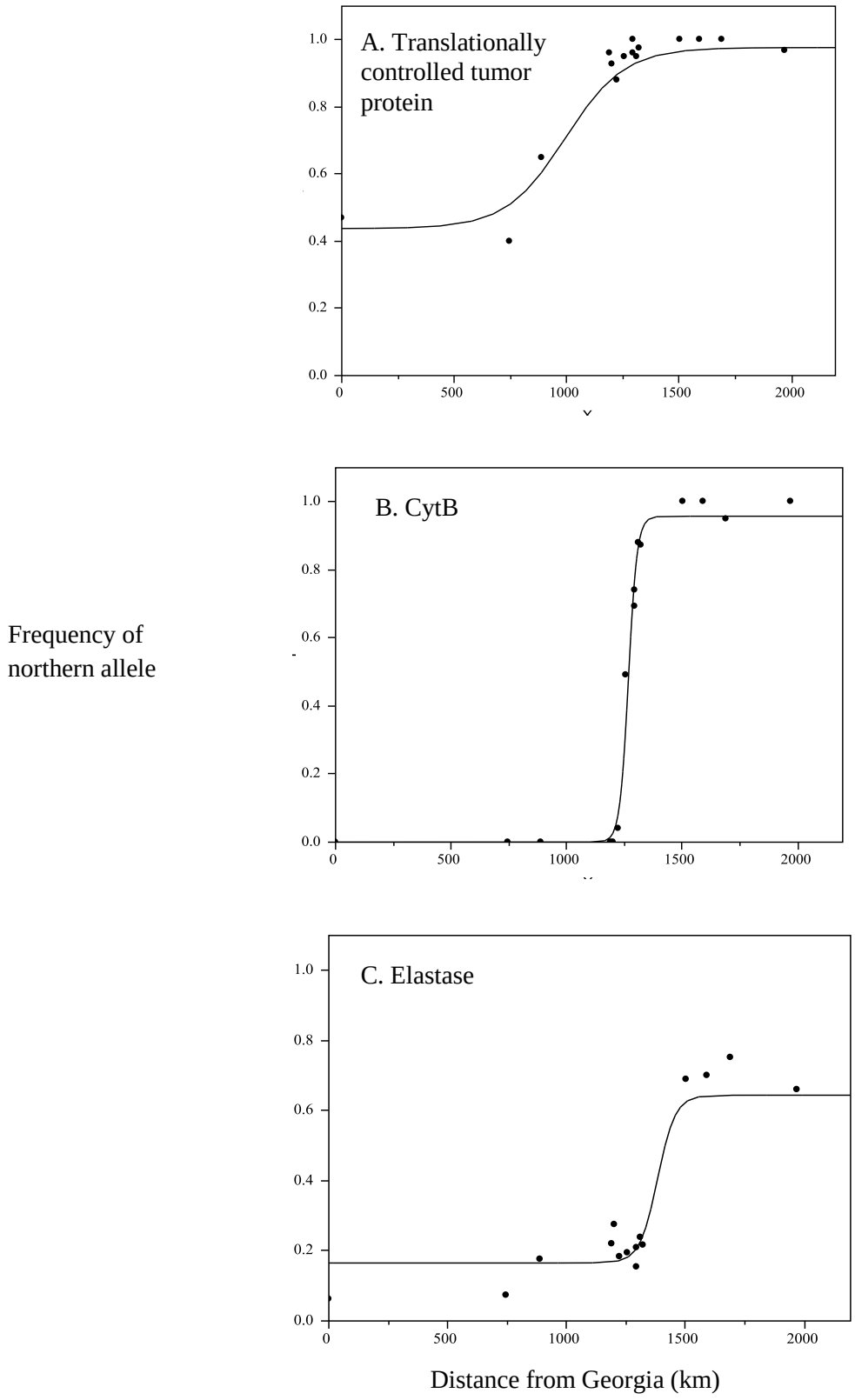


Figure 3-5. Cline in *Ldh-B* allele frequency from data from A) the current data set and B) Powers and Place (1978) as calculated using Cfit for samples of *Fundulus heteroclitus.*

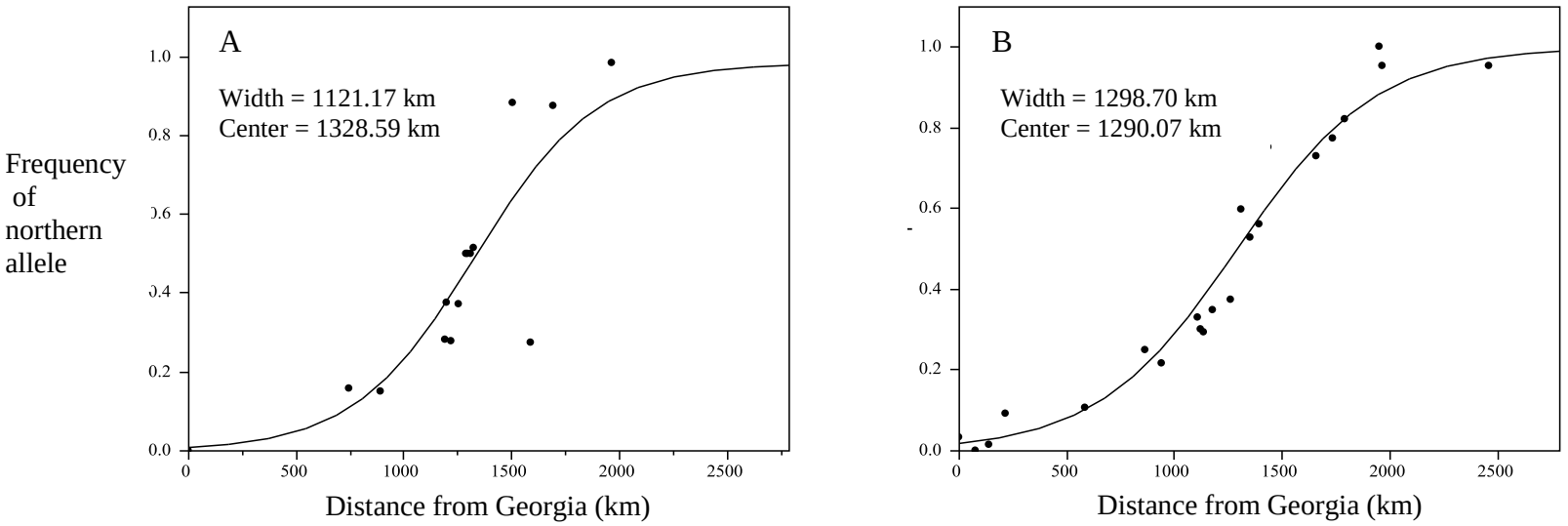


Table 3-1. Summary of collection location information and sample size of *Fundulus heteroclitus*. Distance refers to distance from most southerly location, Sapelo Island, Georgia. Locations marked with an asterix (*) were genotyped at 30 nuclear and two mitochondrial SNPs.

Location	Distance (km)	Collection	Latitude (°N)	Longitude (°W)	n
1. Halifax, Nova Scotia	2459.39	Powers & Place (1978)	44.64888	-63.57531	98
2. Wiscasset, Maine (ME)*	1968.44	Powers & Place (1978), Williams *et al.* (2010)	44.00289	-69.66558	91, 30
3. Boothbay, ME	1954.15	Powers & Place (1978)	43.87639	-69.63361	58
4. Gloucester, Massachusetts (MA)	1791.10	Powers & Place (1978)	42.61593	-70.66199	65
5. Marshfield, MA	1732.77	Powers & Place (1978)	42.09175	-70.70559	82
6. Sandwich, MA*	1691.84	Williams *et al.* (2010)	41.75896	-70.49393	20
7. Woods Hole, MA	1662.02	Powers & Place (1978)	41.5265	-70.67309	245
8. Point Judith, Rhode Island (RI)*	1591.72	Williams *et al.* (2010)	41.36538	-71.48672	21
9. Clinton, Connecticut (CT)*	1504.04	Williams *et al.* (2010)	41.27872	-72.52761	21
10. Stony Brook, New York (NY)	1439.29	Powers & Place (1978)	40.92565	-73.14094	60
11. Jones Beach, NY	1392.12	Powers & Place (1978)	40.6016	-73.50209	48
12. Vince Lombardi, NY	1349.41	Powers & Place (1978)	40.84038	-74.09070	36
13. Cheesequake, New Jersey (NJ)*	1322.31	Current study	40.463420	-74.258888	39
14. Keansburg, NJ	1311.10	Powers & Place (1978)	40.44177	-74.12986	46
15. Belford, NJ*	1308.90	Current study	40.440968	-74.103963	50
16. Sandy Hook, NJ*	1297.07	Current study	40.412904	-73.969487	50
17. Navesink, NJ*	1292.51	Current study	40.376297	-73.993993	13
18. Point Pleasant, NJ	1259.35	Powers & Place (1978)	40.08317	-74.06819	75
19. Metedeconk, NJ*	1257.39	Current study	40.065616	-74.065320	49
20. Laurel, NJ*	1220.73	Current study	39.750349	-74.192734	49
21. Tuckerton, NJ*	1200.06	Williams *et al.* (2010)	39.60317	-74.34015	20
22. RUMFS, NJ*	1189.90	Current study	39.512905	-74.320450	50
23. Leeds Point, NJ	1180.28	Powers & Place (1978)	39.49206	-74.42904	53
24. Woodland Beach, Delaware (DE)	1135.47	Powers & Place (1978)	39.33345	-75.47464	48
25. Stone Harbour, NJ	1123.67	Powers & Place (1978)	39.05095	-74.75794	73
26. Cape May, NJ	1105.51	Powers & Place (1978)	38.93511	-74.90601	47
27. Wachapreague, Virginia (VA)	942.65	Powers & Place (1978)	37.6043	-75.68965	174
28. Magotha, VA*	890.28	Williams *et al.* (2010)	37.18014	-75.94882	20
29. Cape Henry, VA	861.97	Powers & Place (1978)	36.93154	-76.01993	20
30. Manteo, North Carolina (NC)*	744.30	Williams *et al.* (2010)	35.90823	-75.67573	20
31. Beaufort, NC	584.60	Powers & Place (1978)	34.71822	-76.66382	131
32. Charleston, South Carolina (SC)	213.12	Powers & Place (1978)	32.77657	-79.93092	87
33. Lady's Island, SC	137.88	Powers & Place (1978)	32.4532	-80.63571	31
34. Savannah, Georgia (GA)	77.94	Powers & Place (1978)	32.08354	-81.09983	48
35. Sapelo Island, GA*	0	Powers & Place (1978), Williams *et al.* (2010)	31.39745	-81.27871	110, 33

Table 3-2. F_{IS} values for loci that deviated significantly from HWE and number of pairs of loci in significant linkage disequilibrium (LD) and cytonuclear disequilibrium (CND) at each location sampled for *Fundulus heteroclitus*. Positive values of F_{IS} indicate heterozygote deficit while negative values indicate heterozygote excess. A $^{\delta}$ next to the locus number indicates a SNP showing significant heterozygote deficit at two or more locations. Shaded cells indicate locations within the borders of New Jersey. F_{IS} values in bold are significant at $\alpha = 0.05$; values significant after FDR-adjustment are marked with an asterisk (*). Location numbers refer to locations as listed in Table 3-1. Shaded area indicates locations within New Jersey. Locus names are represented by numbers as defined in Figure 3-2. NA = calculation not applicable because only a single allele was present among individuals from a particular sampling site. The column LD indicates the number of pairs of loci exhibiting significant linkage disequilibrium at $\alpha = 0.05$. The number remaining significant following FDR-adjustment is in parentheses. +Rij = number of pairs of loci in significant linkage disequilibrium such that alleles from like parental type are more often together: -Rij = number of pairs of loci in significant linkage disequilibrium such that alleles from unlike parental types are more often found together. The last column represents the number of occurrences of significant cytonuclear disequilibrium (CND) within a given marsh.

Location	Locus Number										
	1 $^{\delta}$	2 $^{\delta}$	3 $^{\delta}$	4 $^{\delta}$	7	8	9 $^{\delta}$	10 $^{\delta}$	11	12 $^{\delta}$	13
2	NA	NA	NA	1.000	NA	-0.080	0.353	-0.019	-0.191	NA	NA
6	NA	NA	-0.027	NA	NA	NA	-1.166	0.465	-0.000	NA	-0.166
8	NA	NA	-0.132	NA	NA	-0.054	-0.379	NA	0.052	-0.029	NA
9	-0.053	NA	-0.047	NA	-0.056	0.022	0.403	0.098	-0.081	-0.053	-0.026
13	-0.088	NA	-0.118	-0.014	-0.086	**-0.438***	0.139	0.205	-0.219	**0.659***	0.063
15	**0.510***	**0.662**	**0.466***	-0.044	-0.153	0.091	**0.469***	0.142	**0.406***	**0.597***	0.226
16	0.010	**0.662**	0.101	-0.032	-0.181	-0.121	0.117	0.043	-0.037	**0.404**	0.165
17	0.268	NA	0.442	NA	0.294	-0.043	0.111	-0.043	-0.200	**0.662**	0.268
19	0.085	-0.032	0.227	**0.793***	-0.063	-0.145	**0.466***	**0.311**	0.020	**0.547***	**0.299**
20	**0.343**	0.035	**0.453***	0.082	-0.190	0.260	**0.473***	**0.358**	0.177	**0.613***	0.154
21	0.151	-0.056	0.254	-0.056	0.163	0.146	**0.673**	**0.518**	0.073	-0.059	0.207
22	**0.466***	-0.077	0.280	0.487	**0.365**	-0.073	**0.280**	0.081	0.043	**0.814***	-0.001
28	0.224	-0.280	**0.508**	0.146	0.247	0.384	NA	**1.000**	-0.188	-0.086	-0.059
30	-0.100	-0.280	-0.261	-0.226	-0.307	-0.166	-0.097	**1.000***	0.159	-0.027	NA
35	**0.571**	NA	-0.127	**0.465**	0.068	-0.016	**1.000**	0.480	0.283	NA	-0.053

	Locus Number									LD			CND
Location	14 $^{\delta}$	15	19 $^{\delta}$	20 $^{\delta}$	22	23	25	26	29 $^{\delta}$	Total	+Rij	-Rij	0
2	-0.018	NA	-0.018	-0.018	NA	NA	NA	NA	-0.057	24 (7)	22	2	0
6	-0.378	NA	NA	-0.056	-0.188	-0.086	NA	-0.118	-0.188	24 (2)	21	3	0
8	-0.240	NA	-0.053	0.216	-0.188	0.405	0.162	**0.639**	0.397	24 (0)	17	8	0
9	NA	NA	-0.081	0.098	-0.026	-0.081	-0.026	0.341	0.034	23 (0)	20	3	**1**
13	0.089	-0.104	0.242	**0.480***	0.078	0.118	0.209	0.040	**0.689***	18 (0)	11	7	**2**
15	**-0.290**	0.078	0.226	0.167	0.087	0.147	0.207	0.117	**0.655***	32 (0)	25	7	**2**
16	**0.365**	0.242	**0.530***	0.267	0.030	-0.137	-0.032	-0.190	**0.878***	23 (0)	18	5	**0**
17	0.111	-0.043	0.100	0.385	0.111	**-0.714**	-0.333	-0.294	0.446	27 (0)	17	10	**3**
19	0.077	0.147	**0.481***	**0.309**	**-0.313**	0.031	0.052	-0.219	**0.668***	29 (0)	20	9	**0**
20	0.203	0.201	**0.873***	**0.418**	0.177	-0.052	0.191	-0.041	0.291	24 (0)	12	12	**0**
21	-0.357	0.122	0.227	**0.624**	0.227	-0.000	-0.180	0.066	-0.152	34 (0)	18	16	0
22	-0.087	0.030	**0.351**	0.222	-0.095	**0.349**	**0.299**	-0.081	**0.511**	22 (0)	14	8	0
28	-0.267	0.105	NA	**0.832**	0.500	0.307	-0.310	0.240	-0.241	21 (0)	12	9	0
30	-0.086	**-0.491**	NA	**0.587**	0.151	-0.188	-0.216	-0.161	0.654	41 (0)	30	11	0
35	-0.049	-0.049	**0.532**	**0.532**	-0.103	-0.067	0.200	NA	NA	29 (0)	21	8	0

Table 3-3. Cline shape parameters for 32 SNP loci in populations of *Fundulus heteroclitus*, arranged in order of increasing cline width. Values in brackets are the 2-unit support limits on the cline shape parameters. General functional categories are also given where known. The asterisk (*) indicates loci with width less than conservative neutral prediction (614km); § indicates SNPs with 2 unit support limits on width estimate that contain the neutral prediction.

Locus Name	Width (km)	Centre (km)	pmin	pmax	Function
Cytochrome B*	75.16 (56.31, 93.84)	1267.09 (1258.56, 1274.24)	0.00 (0.00, 0.01)	0.96 (0.90, 0.99)	Mitochondrial
Cytochrome c oxidase subunit I*	83.78 (70.41, 98.93)	1266.82 (1260.31, 1273.55)	0.00 (0.00, 0.02)	1.00 (0.98, 1.00)	Mitochondrial
65*	117.61 (88.33, 153.91)	1256.71 (1245.24, 1267.36)	0.00 (0.00, 0.03)	0.87 (0.80, 0.92)	Unknown
40S ribosomal protein S17*	140.71 (93.61, 205.54)	1265.91 (1247.06, 1282.56)	0.24 (0.16, 0.33)	0.97 (0.93, 0.99)	Protein synthesis
Elastase*	150.05 (0.10, 551.90)	1382.38 (1326.52, 1499.86)	0.17 (0.07, 0.22)	0.64 (0.55, 0.75)	Tissue breakdown
Warm acclimation related protein*	173.65 (94.07, 419.00)	1319.63 (1296.15, 1382.30)	0.15 (0.10, 0.22)	0.82 (0.74, 0.91)	Temperature
Phosphate carrier protein*	184.11 (93.27, 367.30)	1303.14 (1281.50, 1344.20)	0.08 (0.03, 0.15)	0.78 (0.69, 0.86)	Mitochondrial
60Sribosomal protein L35§	204.85 (113.87, 1547.02)	1262.23 (1223.82, 1494.24)	0.12 (0.01, 0.20)	0.66 (0.57, 1.00)	Protein synthesis
Myoglobin*	258.80 (181.19, 532.76)	1261.74 (1231.34, 1287.52)	0.06 (0.00, 0.13)	0.85 (0.77, 0.91)	O_2 transport
280 *	260.32 (165.00, 390.74)	1306.84 (1284.68, 1337.72)	0.06 (0.02, 0.13)	0.96 (0.90, 0.99)	Unknown
Tropomyosin*	264.42 (134.79, 476.09)	1326.79 (1295.54, 1375.65)	0.14 (0.07, 0.23)	0.85 (0.76, 0.92)	Muscle
Glyceraldehyde 3phosphate dehydrogenase*	279.99 (136.68, 451.49)	1326.38 (1287.44, 1378.44)	0.39 (0.30, 0.49)	0.99 (0.95, 1.00)	Metabolism
1176*	305.53 (122.42, 550.46)	1182.75 (1114.94, 1232.99)	0.01 (0.00, 0.05)	0.59 (0.48, 0.70)	Unknown
Parvalbumin§	313.67 (165.73, 781.46)	1238.19 (1165.99, 1293.13)	0.18 (0.10, 0.26)	0.73 (0.64, 0.83)	Muscle
Hemoglobin a2*	341.43 (241.08, 441.31)	1445.74 (1386.30, 1487.74)	0.10 (0.06, 0.15)	1.00 (0.92, 1.00)	O_2 transport
Nucleotide diphosphate kinase 2§	406.89 (43.61, 764.39)	1536.48 (1437.40, 1647.93)	0.17 (0.09, 0.24)	0.90 (0.72, 1.00)	Growth
60S ribosomal protein L6§	418.27 (259.25, 658.27)	1191.22 (1137.27, 1231.96)	0.05 (0.01, 0.12)	0.86 (0.77, 0.94)	Protein synthesis
Hemoglobin a1*	457.40, (363.90, 580.64)	1518.09 (1473.77, 1565.98)	0.03 (0.01, 0.07)	1.00 (0.93, 1.00)	O_2 transport
Nucleotide diphosphate kinase 1§	467.51 (331.31, 659.73)	1395.28 (1345.25, 1449.92)	0.18 (0.11, 0.28)	1.00 (0.95, 1.00)	Growth
Translationally Controlled Tumor Protein§	526.69 (265.31, 876.49)	993.59 (836.94, 1108.65)	0.44 (0.29, 0.57)	0.98 (0.92, 1.00)	Stress
Lactate dehydrogenase B654§	561.49 (282.65, 802.19)	1360.23 (1298.24, 1417.55)	0.12 (0.05, 0.21)	0.99 (0.88, 1.00)	Metabolism
Cytochromep450§	591.49 (355.06, 954.05)	1474.93 (1395.25, 1584.35)	0.08 (0.02, 0.17)	0.91 (0.77, 1.00)	Metabolism
Chymotrypsinogen§	608.09 (482.21, 763.74)	837.38 (740.31, 913.86)	0.01 (0.00, 0.10)	1.00 (0.98, 1.00)	Digestion
Activated protein kinase§	622.71 (486.43, 847.31)	1087.80 (1015.94, 1131.71)	0.00 (0.00, 0.00)	0.93 (0.87, 0.98)	Protein synthesis
Ribosomal protein§	627.07 (322.39, 1015.45)	1024.05 (877.03, 1142.12)	0.16 (0.06, 0.32)	0.93 (0.87, 0.98)	Protein synthesis
NACA§	752.47 (295.31, 1096.48)	1087.95 (966.98, 1192.90)	0.15 (0.05, 0.30)	1.00 (0.94, 1.00)	Protein synthesis
Titin Cap§	900.33 (234.08, 1298.01)	1141.10 (1064.77, 1248.39)	0.00 (0.00, 0.16)	0.90 (0.80, 0.99)	Muscle
Lactate dehydrogenase B1033	931.92 (696.26, 1201.27)	1399.80 (1339.45, 1458.73)	0.00 (0.00, 0.07)	1.00 (0.90, 1.00)	Metabolism
Atrial Natriuretic Peptide	934.20 (473.95, 1772.92)	783.56 (453.75, 986.05)	0.14 (0.00, 0.31)	0.89 (0.82, 0.97)	Stress
TC19110	1076.98 (645.76, 1367.98)	1482.38 (1337.63, 1556.16)	0.00 (0.00, 0.06)	1.00 (0.77, 1.00)	Unknown
Ribosomal Protein S2	1681.58 (854.98, 2179.24)	954.38 (752.26, 1121.91)	0.00 (0.00, 0.17)	1.00 (0.83, 1.00)	Protein Synthesis
1173	1749.25 (815.18, 3100.53)	1590.03 (1265.50, 1798.30)	0.08 (0.00, 0.25)	1.00 (0.71, 1.00)	Unknown

Table 3-4. Summary of loci exhibiting steep clines, cline centers coincident with the mtDNA cline centre, heterozygote deficit and/or cytonuclear disequilibrium.

Locus name	**Steeper than neutral prediction**	**Cline Centre Coincident with mtDNA cline Centres**	**Heterozygote deficit (significant after FDR)**	**Cytonuclear disequilibrium**
65 (actin binding LIM family)	Y	Y	Y	
40S ribsomal S17	Y	Y		
Elastase	Y			
Warm acclimation related protein	Y	Y	Y	
Phosphate carrier protein	Y	Y		Y
Myoglobin	Y	Y		
280 (HDDC2)	Y	Y	Y	
Tropomyosin	Y	Y		
G3PDH	Y	Y		
1176 (unknown)	Y	Y	Y	
Hemoglobin A2	Y			
Hemoglobin A1	Y			
60S ribosomal protein L6		Y		
Titin cap		Y	Heterozygote excess	Y
Parvalbumin		Y	Y	
Ribosomal Protein S2			Y	
TC19110_377 (14-3-3 Protein)			Y	
Translationally controlled tumor protein			Y	
Atrial Naturietic peptide			Y	
NACA				Y
Activated protein kinase				Y
60S ribosomal protein L35		Y		

Chapter Four: Evidence for assortative mating among geographically isolated populations of *Fundulus heteroclitus* suggests a possible mechanism for maintaining a trimodal hybrid zone

Synopsis

Populations of Atlantic killifish, *Fundulus heteroclitus* exhibit latitudinal clinal variation at a variety of loci that is correlated with environmental temperature along the Atlantic coast, and similar clines correlated with salinity are present within the Chesapeake Bay. We used nine microsatellite loci and 31 SNPs (30 nuclear loci and 1 mitochondrial locus) to determine if such clines were also present at a smaller scale in a single creek in New Jersey, near the centre of the coastal mitochondrial cline, which might be indicative of a role for microhabitat partitioning as a reproductive isolating mechanism. While clines were not detected at this scale, the previously diagnosed trimodal hybrid zone, elevated levels of heterozygote deficit, linkage disequilibrium, and cytonuclear disequilibrium were evident across sampling sites and between seasons. In addition, an unusual form of cytonuclear disequilibrium, wherein females and males were significantly more likely to possess southern and northern mitochondrial types, respectively, occurred in this marsh, suggesting that sex and mitochondrial genotype are playing an important yet elusive role in the maintenance of this hybrid zone. Results of fertilization experiments among individuals from the extremes of the species' distribution indicated that gametic incompatibilities are not likely operating in the persistence of the hybrid zone. However, results from a "choice" breeding experiment among pure northern and southern parents suggest a role for either positive assortative mating or early hybrid inviability in the maintenance of the

trimodal hybrid zone observed between the two subspecies. Specifically, northern males and southern females showed a significant preference for mating with consubspecific types.

Introduction

The Atlantic killifish, *Fundulus heteroclitus*, is an abundant topminnow found in the brackish waters of estuaries along the eastern coast of North America from Newfoundland, Canada to northern Florida, USA (Hardy 1978). Killifish are sexually dimorphic, with the males exhibiting exaggerated colouration during the breeding season (Hardy 1978). These fish are external fertilizers, and pairs of males and females come together in close association during the release of gametes, in a virtually monogamous mating dance (Newmann 1907). Two subspecies are recognized on the basis of divergent characteristics in many aspects of their biology: *Fundulus heteroclitus heteroclitus* forms the "southern" subspecies ranging from southern New Jersey, USA to northern Florida, USA, while *Fundulus heteroclitus macrolepidotus* forms the "northern" subspecies, occupying coastal habitat from northern New Jersey to Newfoundland, Canada. A trimodal hybrid zone, located in the central New Jersey coast, separates the two subspecies (Chapter Two). However, the factors important in the maintenance of this trimodal hybrid zone are unknown.

Trimodal hybrid zones can be maintained by a suite of incomplete pre- and post-zygotic reproductive barriers between the participating pairs of taxa. Premating pre-zygotic barriers arise when taxa display differences in habitat specialization (ecological divergence), mate preference (positive assortative mating), and reproductive synchrony (temporal divergence), while postmating pre-zygotic barriers occur as a result of differences in fertilization success (gametic incompatibilities or sperm competition between taxa) between divergent populations. Post-zygotic mechanisms include hybrid inviability or sterility and general hybrid breakdown (Mayr

1970; Palumbi 1994). Pre- and post-zygotic barriers can be closely associated, with high levels of reduced hybrid fitness having the potential to reinforce the development of pre-zygotic barriers limiting the production of unfit hybrid offspring (Brown & Wilson 1956; Barton & Hewitt 1985; Servedio & Noor 2003).

In *F. heteroclitus*, evidence exists suggesting that habitat specialization could act as a pre-zygotic barrier to maintain the trimodal hybrid zone. Specifically, adult northern fish are more tolerant of freshwater environments and cooler temperatures while southern fish are less tolerant of freshwater and can withstand warmer temperatures (Scott *et al*. 2004; Fangue *et al*. 2006; Whitehead *et al*. 2011). Furthermore, differences in salinity tolerance have been implicated in the maintenance of population structure along the large rivers (> 600 km long) of the Chesapeake Bay system (Smith *et al*. 1998; Whitehead *et al*. 2011). Thus, differences in habitat specialization due to microhabitat differences in temperature and/or salinity could limit the extent to which these two subspecies interbreed in areas of sympatry (Palumbi 1994). Such ecological separation would result in the production of occasional hybrid offspring but would have the potential to present an effective pre-mating pre-zygotic barrier to gene flow if the two subspecies rarely encountered one another at the time of reproduction.

Morphological features also differ among the subspecies of *F. heteroclitus* and are particularly prominent in sexually mature adults providing potential cues for the promotion of positive assortative mating (Able & Felley 1986). Thus, the failure of divergent parental forms to produce hybrid offspring could also be the result of divergence in morphological traits that allow the different taxa to recognize one another and to avoid dissimilar types (Nagel & Schluter 1998).

In situations where individuals show no difference in preference between conspecific and heterospecific mates, divergent traits at the gametic level may be responsible for reductions in numbers of hybrid offspring (Levitan 2000; Palumbi & Metz 1991). Differences in fertilization success may be the result of divergences in the respective gametes' ability to tolerate different environmental conditions (Able & Palmer 1988). It may also be due to the divergence of gamete recognition proteins or to differences in the chemoattractants produced by eggs in order to attract sperm (Palumbi 2009; Toro *et al*. 2006). Dramatic differences at the egg surfaces of northern and southern *F. heteroclitus* subspecies may hint at underlying differences either at the level of gamete recognition proteins or chemoattractants released by the eggs to attract consubspecific sperm (Able & Costagna 1975; Brummett 1966; Brummett & Dumont 1981).

In order to begin to distinguish among the possible pre-zygotic factors maintaining this hybrid zone we collected fish from marshes associated with a single creek in the Metedeconk River system, New Jersey (NJ), and genotyped them at nine microsatellite loci and 31 single nucleotide polymorphisms (SNPs; one mitochondrial and 30 nuclear). We used these data from individuals collected across several sites and two seasons to ask whether the coastal and Chesapeake Bay clines in *F. heteroclitus* are replicated at small spatial scales suggesting a role for habitat specialization as a premating pre-zygotic reproductive barrier. Next, we used individuals from the extremes of the species' distribution to ask the following questions: i) is there evidence of positive assortative mating (premating pre-zygotic barrier) among these individuals and ii) are there differences in fertilization success between the subspecies suggestive of a post-mating pre-zygotic barrier (such as gametic incompatibilities)?

Materials and Methods

Fish Collection

Fundulus heteroclitus (457 individuals) were collected from various locations along the shoreline of Beaverdam Creek in north central New Jersey (Table 4-1; Figure 4-1). Beaverdam Creek is located at the centre of an mtDNA cline in the coastal hybrid zone of *F. heteroclitus* (see Chapters Two and Three) and is approximately 3.18 km long. Samples were collected during two seasons. The first collection took place in the Summer of 2008 (n = 114) from three locations at the end, middle, and mouth of the creek (Cherokee Lane, Midstream, and Osprey, respectively; Table 4-1). The creek was revisited in the Fall of 2009 (n = 343) when fish were sampled from eight locations along the length of the creek including the three locations from the previous summer (Table 4-1). The purpose of collecting fish across two different seasons was to attempt to disentangle salinity and temperature as abiotic factors that might influence habitat choice in this species. We first visited the marsh in June 2008, however, the effects of temperature and salinity in a small tidally influenced creek are confounded with respect to predicted killifish behaviour in the warm summer months. For example, northern killifish are more tolerant of freshwater conditions than their southern counterparts and less tolerant of high temperature; however, the landward end of the creek, where salinity is the lowest, is also the area where temperatures are the highest. Thus, the creek was revisited in November 2009 when temperature was constant across the length of the creek. Minnow traps were used and trap set time ranged from two to six hours. Fish length was recorded and a fin clip was taken from each individual and preserved in 95% ethanol. The sex of individuals collected during summer 2008 was also recorded. It was not possible to unambiguously determine the sex of the individuals

collected in the fall because gonads and sexually dimorphic traits regress in this species outside of the breeding season. Additional samples from sites located at the northern and southern edges of the hybrid zone (RUMFS, NJ and Cheesequake, NJ; see Chapter Two, Table 2-1 and Figure 2-1 for collection information) were used as the parental populations for all analyses involving the calculation of hybrid index.

Fish Husbandry for Breeding Experiments

Samples representative of the pure northern and southern populations of the two subspecies (Table 4-1) were collected near Hampton, New Hampshire (42.937055°N, 70.839844°W; Aquatic Research Organisms, Inc.) and Beaufort, North Carolina, USA (34.714224°N, 76.680450°W; provided by Dr. L. Campbell, University of North Carolina), respectively and were held in separate tanks. All fish used in both the fertilization and "choice" breeding experiments were initially housed in 20 gallon tanks at 15°C, in 20ppt seawater made using Instant Ocean for one month. Temperature in the tanks was then gradually increased over the course of a month until a final water temperature of 25°C was reached. Fish were fed a diet of blood worms to satiation twice daily. Each fish was given a unique tattoo using a series of different coloured Visible Implant Elastomer tags (Northwest Marine Technology, Inc. Shaw Island, WA) injected subcutaneously, and a fin-clip from each individual was stored in 95% ethanol for use in later DNA analysis. Animals were treated in accordance with the University of British Columbia animal care protocol A07-0288.

Genotyping

Qiagen's DNeasy® Blood and Tissue Kit was used for the extraction of DNA from all fish used in the microhabitat portion of this study as well as from the adults used in the breeding

experiment, while Fermentas' Fast Tissue-to-PCR Kit® was used to extract DNA from eggs and hatchlings that resulted from the breeding experiment. All polymerase chain reactions were performed using a MJ Research PTC-200 Peltier Thermal Cycler. The mitochondrial type of all adult samples was determined using a PCR-RFLP assay targeting a fixed single nucleotide polymorphism known to be diagnostic between the two subspecies of killifish as previously described (Chapter Two). Briefly, 1100bp of the mitochondrial dloop was amplified using primers K (forward) and G (reverse) (Lee *et al*. 1995) and digested using the enzyme *Sca*I following the manufacturer's instructions (Fermentas; Thermo Fisher Scientific Canada). An undigested fragment was characteristic of the northern mitochondrial type while the presence of the digested 900 bp fragment was diagnostic of the southern type. Microsatellite genotyping of fish used in the microhabitat portion of this study was conducted as described in Chapter Two, with samples being sent to the Boston Children's Hospital Molecular Genetics Core Facility for genotyping on an Applied Biosystems 3730 DNA Analyzer. However, parents and offspring from the breeding experiments were genotyped using seven microsatellite markers (FhATG2, FhATG18, FhATG20, FhATG4, FhATGB128, FhATGB103, and FhATGB101) with the same reaction mixtures and under the same PCR conditions as described in Chapter Two with the exception of FhATGB101 which was not used in the microhabitat portion of the study. PCR conditions for this locus were the same as those previously described for FhATG2 and FhATG4. Samples from the breeding experiment were genotyped at the University of British Columbia's Nucleic Acid Protein Service (NAPS) Unit (Michael Smith Laboratories) on an Applied Biosystems 3730 DNA Analyzer. Alleles were scored using Peak Scanner™ Software v1.0 (Applied Biosystems). Lastly, fish used in the microhabitat portion of this study were further genotyped at 30 nuclear SNPs known to be diagnostic or semi-diagnostic for "northern" or

"southern" individuals (see Chapter Three). 30µl of genomic DNA (20 ng/µl) of each sample was sent on dry ice to the McGill University and Génome Québec Innovation Centre for genotyping using Sequenom® iPLEX®Gold Genotyping Technology.

Fertilization Success

In order to determine the point at which fertilization success begins to decline among pure crosses, sperm was extracted from fish using a capillary tube, concentration estimated using a hemocytometer, and five 2,000-fold serial dilutions were prepared. 1 ml of each of these five stocks was added to one of five Petri dishes each containing 23 ml of seawater (20 ppt) and 12 eggs. The eggs were incubated at 20°C for three hours after which each egg was examined for the presence of a fertilization envelope (Armstrong & Child 1965). Percent fertilization success was plotted against sperm concentration and the concentration of sperm at which 60% fertilization success (the point after which fertilization success among pure crosses rapidly declines; Levitan *et al*. 1991) was achieved was used for all subsequent experiments in order to ensure that super-saturating levels of sperm were not being applied to the eggs (Appendix D Figure D.1; Levitan 2002; Palumbi 2009).

For the fertilization experiments, approximately 100 - 180 eggs from several females (the number of females varied daily depending on the number it took to obtain this number of eggs) of each subspecies were stripped daily by applying slight pressure to the abdomen, divided equally among four Petri dishes, and sperm from two northern males and two southern males (100 sperm/0.1 µl) was added to one of each of the dishes containing the eggs. The same four males were used for each trial. The gametes were then gently mixed and allowed to stand undisturbed for three hours after which the number of fertilized and unfertilized eggs resulting

from each cross were counted and recorded. On a single day this procedure was performed for pooled eggs from both southern and northern females and four trials were performed in total.

Mating Trial

Forty-five fish originating from the northern (13 males and 14 females) and southern (10 males and eight females) subspecies were placed in a single 75 gallon tank (3 feet x 18 inches x 19 inches) at 25°C and 20ppt with five "nests". "Nests", constructed by Dr. Rashpal Dhillon, consisted of two pieces of fitted polyvinyl chloride (PVC) pipe. The smaller half (diameter = 7.62 cm; height = 5.08 cm) was sealed on one end with window screen size mesh (18 x 16 openings per cm^2) and formed the bottom of the nest. The top half (diameter = 10.16 cm; height = 5.08 cm) of the nest was covered at one end by larger nylon mesh (5 mm mesh size) which allowed for the passage of eggs into the nest but prevented adults from entering and eating the fertilized eggs. The nests were checked daily for eggs, which were then removed and examined for the presence of a fertilization envelope. Fertilized eggs were then placed in petri dishes containing a shallow layer of water and diluted (0.03%) methylene blue (as a fungicide) was added. Checking of nests and sampling of eggs was repeated each day for 32 days. Eggs that contained embryos that died prior to hatch were preserved in 95% ethanol. Hatchlings that resulted were also placed in 95% ethanol upon hatch.

Analyses

Microsatellite data were screened for the presence of null alleles, large allele dropout, and scoring errors using the program Microchecker V2.2.3 (Van Oosterhout *et al.* 2004). All groups of fish, whether genotyped at microsatellite loci, SNPs, or both, were tested for significant departures from Hardy Weinberg Equilibrium (HWE) using the exact test as performed in

Genepop 4.0.10 (Raymond & Rousset 1995). Each locus in each population was tested for heterozygote excess and heterozygote deficiency using the following Markov chain parameters: dememorization number = 1000; 100 batches; 1000 iterations per batch. Tests of linkage disequilibrium were performed for all locus pairs (microsatellite and SNPs) in all populations of fish used in the microhabitat study using the Black and Krafsur procedure in Genetix 4.05.2 (Cockerham & Weir 1977; Belkhir *et al.*1999). This program was also used to test each pair of microsatellite loci among the eggs and hatchlings resulting from the "choice" breeding experiment for evidence of linkage disequilibrium. Individuals in the microhabitat portion of this study were tested for evidence of cytonuclear disequilibrium among mitochondrial type and SNP markers using the program CNDd (available at http://statgen.ncsu.edu/cnd/CNDd.php; Asmussen & Basten 1996; Basten & Asmussen 1997). The Black and Krafsur procedure, as implemented in Genetix 4.05.2, was used to (Cockerham & Weir 1977; Belkhir *et al.*1999).

Pairwise F_{ST}s among sites within each sampling season, between sampling sites visited in both seasons, and between sexes at different locations for the Summer 2008 data set were calculated using Arlequin version 3.5.1.2 (Excoffier *et al*. 2005).

A hybrid index for each individual used in the microhabitat phase of this study was calculated using the program INTROGRESS (Gompert & Buerkle 2009, 2010) for both multilocus microsatellite and SNP genotypes, separately. Individuals from RUMFS and Cheesequake (Chapter Two; Table 2-1) were used as parental types in training the program. Hybrid index values for each individual from each location sampled for both Summer 2008 and Fall 2009 sampling trips were then binned in 0.10 increments and compiled in histograms representing the SNP and microsatellite hybrid indices present at each location in this marsh, for both seasons. The program MCLUST was used to estimate the number of clusters present in the

frequency distributions of hybrid index values (Fraley *et al.* 2012). The number of clusters (either 1, 2, or 3) best-fitting the shape of the hybrid index frequency distributions was decided based on the one producing the highest Bayesian Information Criterion (BIC).

JMP 9.0 (SAS Institute) was used to test for deviation of sex ratios and mitochondrial proportions from expected values as well as to test for independence of sex and mitochondrial type among individuals from the three sites visited in the Summer 2008. The program R (R Development Core Team 2008) was used to perform the Fisher's exact test for all sex vs. cytonuclear type combinations between the mitochondrial (M, m) and SNP (A, a) genotypes (AAM, AaM, aaM, AAm, Aam, aam, where all uppercase represents a pure southern cytonuclear type and all lowercase indicate a pure northern individual). JMP 9.0 was also used to perform the split-plot ANOVA testing for differences in mean fertilization success between pure and reciprocal crosses.

The program PAPA 2.0 (Duchesne *et al*. 2002) was used for parentage assignment of the offspring generated in the "choice" breeding experiment. Briefly, this program uses likelihood to assign the most likely candidate parent pair to each offspring. Unambiguous parentage assignments, where the likelihood of one pair parenting a particular offspring exceeded that of all other parental pairs, were retained for further analyses. Approximately 95% of all offspring could be unambiguously assigned to parents using this approach. Mitochondrial type for assigned parents was determined and individual parents were subsequently scored based on percent of preferred matings with respect to consubspecific ("like") and heterosubspecific ("unlike") mates, based on mitochondrial type, which is diagnostic of subspecies origin. The mean percentage of matings among like and unlike types was calculated for each of four groups (northern females, southern females, northern males, and southern males), and four Wilcoxan

Signed Rank tests (JMP 9.0) were then conducted in order to determine if matings occurred more often than would be expected by chance with consubspecific ("like") mates.

Results

Do the Coastal and Chesapeake Bay Clines Also Occur on a Microhabitat Scale?

We found no evidence of clines such as those observed among populations of killifish distributed along the Atlantic Coast or the rivers of the Chesapeake Bay system occurring in a small tidally influenced creek. Among both Summer 2008 and Fall 2009 samples, we found no significant effect of sampling location on mitochondrial type ($\chi^2 = 0.927$, df = 2, p = 0.6289; $\chi^2 =$ 3.786, df = 7, p = 0.8040, respectively). There were no significant pairwise F_{ST}s among any of the three locations sampled in the Summer 2008, at either the microsatellite or SNP loci. The Fall 2009 sampling locations, however, yielded several instances of significant pairwise F_{ST}s (Table 4-2). However, following an adjustment for false discovery rate (FDR; Pike 2010) none of these results remained significant.

Are Deviations From HWE and Evidence of Linkage Disequilibrium Consistent Across Locations and Seasons?

Among the Summer 2008 samples, four of the nine microsatellite loci for which the fish were genotyped exhibited significant deviations from HWE (Table 4-3A). Four of the five observed deviations had positive values of F_{IS}, indicating heterozygote deficit. Following FDR-adjustment, all occurrences of heterozygote deficit remained significant while the single occurrence of heterozygote excess did not. In contrast, the Fall 2009 samples exhibited roughly equal numbers of significant positive and negative values of F_{IS} (Table 4- 3B). Following FDR-

adjustment only two instances of heterozygote deficit and one of heterozygote excess remained significant among these samples.

The Summer 2008 samples all exhibited low levels of linkage disequilibrium among the microsatellite markers with the exception of a single location (Midstream; Table 4-3A). Similarly, there was little evidence of excessive linkage disequilibrium among the microsatellite markers at any of the eight locations sampled in this marsh during the Fall 2009 (Table 4-3B).

The fish collected in the Summer 2008 and the Fall 2009 were also genotyped at 30 nuclear SNPs. Among the Summer 2008 samples, ten loci showed significant deviations from HWE; 12 of the deviations were heterozygote deficits while the remaining two were heterozygote deficits. Following FDR-adjustment, all occurrences of heterozygote deficit remained significant while the two instances of heterozygote excess did not (Table 4-4A). Most notably, three loci showed significant heterozygote deficit at two or more locations. The nuclear encoded mitochondrial gene HD domain containing isoform 2 (HDDC2) and an actin binding LIM family gene showed significant heterozygote deficits at two of the three locations (Cherokee Lane and Osprey, and Midstream and Osprey, respectively) and the locus 14-3-3 episilon (signaling protein) exhibited high and significant levels of heterozygote deficit at all three locations. A total of 47 significant deviations from HWE were detected among 16 of the 30 the SNP markers for the Fall 2009 samples (Table 4-4B). Of the 47 significant deviations observed, 46 represented heterozygote deficits. Following FDR-adjustment, none of the instances of heterozygote excess remained significant, however, 20 of the occurrences of heterozygote deficit remained significant. Eight of the SNPs exhibited significant heterozygote deficit at two or more locations. Most notably, 40S ribosomal protein S17 showed significant heterozygote deficit at five locations, locus 1176_169 showed significant heterozygote deficit at

six locations, the actin binding LIM family gene was in significant heterozygote deficit at seven locations, and locus 14-3-3 epsilon showed significant heterozygote deficit at all eight of the locations sampled during the Fall 2009.

Of the 435 possible pairwise linkage disequilibria among the 30 SNPs at each location, 23 significant occurrences of linkage disequilibria were observed at the end location (Cherokee Lane) while 33 were observed at the middle location (Midstream) and 29 at the mouth (Osprey) locations among the Summer 2008 samples (Table 4-4A; Appendix D Table D.1). All three locations have more instances of linkage disequilibrium than would be expected by chance alone (435 * 0.05 = 22); however, few remained significant following FDR-adjustment. There was no clear evidence of a bias towards positive values of Rij, which would indicate a significant association of alleles from like parents. Overall, levels of linkage disequilibrium were similar among the SNP markers for the Fall 2009 collections when compared to the Summer 2008 collections, with all locations except one (Hidden Harbour) exhibiting more incidences of pairwise linkage disequilibrium among the SNP markers than the roughly 22 pairs of significant linkage disequilibria that would be expected based on chance alone (Table 4-4B; Appendix D, Table D.2). Similar to the observations in the summer, these observed linkage disequilibria exhibited both positive and negative Rij values, but the majority of the associations were deemed non-significant after adjustment for FDR. This adjustment may be too conservative, however, as the physical associations known to exist among two pairs of SNP markers (LDHB 1033 and 654; hemoglobin a1 and a2) were often initially deemed significant but then rejected following this adjustment.

Is the Trimodal Hybrid Zone Consistent Across Sampling Locations and Between Seasons?

We found strong evidence that the previously observed trimodal hybrid zone separating the subspecies of Atlantic killifish is consistent both across sampling locations and between seasons. Hybrid index was calculated for individuals sampled across all sites and in both seasons. All nine microsatellite loci were used in this calculation. However, hybrid index as calculated in INTROGRESS depends on the differences in allele frequencies at each locus between the parental populations (Buerkle 2005). Inclusion of markers exhibiting low levels of differentiation (δ) between the parental populations can result in unreliable estimates of hybrid index (Buerkle 2005). Shriver *et al.* (1997) have suggested that an appropriate cut-off value when selecting "population-specific alleles" is $\delta > 0.5$. Thus, hybrid index was calculated using the SNP loci with allele frequency differentials that exceeded this threshold (actin binding LIM family gene and 40S ribosomal protein S17; Appendix D Table D.3). Following either a logit transformation or the addition of a bit of random noise to the hybrid indices, the number of clusters present in each frequency distribution of hybrid indices (for both the SNP and microsatellite data) was tested. For the Summer 2008 collection locations, a trimodal distribution of hybrid indices was supported (highest BIC value) for both types of markers at both the Midstream and Osprey locations (Table 4-5). These patterns are depicted in Figure 4-2B and Figure 4-2C. In all figures pertaining to hybrid index, a value of 0 indicates a pure southern individual while a hybrid index value of 1 indicates a pure northern individual. At Cherokee Lane, the trimodal pattern was also supported among the hybrid indices calculated for the SNP data (Table 4-5). The microsatellites displayed a bimodal pattern (Table 4-5; Figure 4-2A).

Among the Fall 2009 collections, both markers types exhibited a trimodal pattern of distribution of hybrid indices at all locations except Crescent Park Woods and Osprey (Table 4-

6; Figure 4-3A-D, F, and G). At these two locations the SNP data was suggested a trimodal distribution, whereas the microsatellite data showed a bimodal fit to the distribution of hybrid indices (Table 4-6; Figure 4-3E and H).

Pairwise F_{ST}s were calculated in order to estimate the amount of genetic distance among locations sampled in the Summer 2008 and re-visited in the Fall 2009. All comparisons of the mouth, middle, and end locations (Osprey, Midstream, and Cherokee Lane, respectively) resulted in no significant pairwise F_{ST} across seasons using both nuclear microsatellite and SNP markers, except for the comparison of the mouth location which did result in a significant F_{ST} for the microsatellites between the Summer 2008 and Fall 2009 samples (Table 4-7). This comparison was no longer significant after FDR-adjustment.

Patterns of Cytonuclear Disequilibrium within the Hybrid Zone

To test the hypothesis that cytonuclear incompatibilities could be playing a role in the maintenance of the trimodal hybrid zone, we estimated cytonuclear disequilibrium for both nuclear SNP and microsatellite markers. Among the Summer 2008 and Fall 2009 samples we saw no evidence of cytonuclear disequilibrium among the microsatellite markers. However, the SNP markers showed significant cytonuclear disequilibria across most locations and both seasons (Table 4-8). All three locations visited in the Summer 2008 exhibited some degree of cytonuclear disequilibrium. In addition, seven of the eight locations sampled during Fall 2009 exhibited some degree of cytonuclear disequilibrium. However, detection of an excess of "matched" cytonuclear types was no more frequent than was detection of an excess of "unlike" cytonuclear types, when all SNPs and all locations were considered (Table 4-8). No SNP showed a consistent pattern of an excess of "matched" cytonuclear types at all locations, and only the

SNP LDH-B 654 showed a consistent pattern of an excess of "matched" cytonuclear types at more than one location.

Certain patterns of cytonuclear disequilibrium, wherein excess cytonuclear types are observed relative to the expected amount under conditions of random mating, can suggest sex-based directionality in mating within a hybrid zone (Avise 2000). Such patterns were observed at the locus cytochrome p450 (significant excess of AAM and Aam in Cherokee Lane, Summer 2008), as well as glyceraldehyde 3 phosphate dehydrogenase, and atrial natriuretic peptide (both showing significant excess of AaM and aam at Cherokee Lane and Crescent Park Woods in Fall 2009, respectively: Appendix D Table D.4). At the LDHB654 locus we observed a significant excess of MAA and maa, at Midstream during Summer 2008, which is consistent with introgression occurring in both species and with selection against hybrid types. However, the fact that various loci showed different patterns argues against a simple pattern of sex-based directionality in mating.

We observed an interesting relationship between sex, geography and mitochondrial type among individuals collected in the Summer 2008. There was a higher proportion of males (of both mitochondrial types) than females at the freshwater, warmer end of the marsh (furthest from the ocean; Table 4-9; Figure 4-4A). At the central location, characterized by intermediate temperatures and salinity, we found a significant excess of females (irrespective of mitochondrial type) and very few males (Table 4-9; Figure 4-4B). At the mouth location, characterized by cooler more salty water, we found equal proportions of both sexes and mitochondrial types. However, females possessing a southern mitochondrial type and males harbouring a northern mitochondrial type were in significant excess (Table 4-9; Figure 4-4C).

If the summer locations are further broken down by sex, several locations exhibit significant F_{ST} values at both the microsatellite and SNP markers (Table 4-10). For both the microsatellite and SNP markers, the males from the mouth location (predominantly northern mitotype; Figure 4-4C) were found to be significantly different from the females at the middle location and from the females at the mouth location (primarily southern mitotype; Figure 4-4C). Based on the F_{ST}s calculated from the 30 SNPs, the females at the middle location were found to differ significantly from the males at the middle location. Lastly, the males at the middle location were significantly different from the males at the mouth location (majority northern mitochondrial type; Figure 4-4C). Following FDR-adjustment, the only pairwise comparisons that remained significant were those between females from the middle location and males from the mouth location, as calculated for both the SNP and microsatellite data, suggesting that there is nonrandom geographical segregation of sexes of divergent genetic backgrounds.

Eighteen SNPs were found to be significantly associated with sex when linked to mitochondrial type among individuals collected in the Summer 2008. When hybrid index was recalculated using these 18 SNPs, there was no significant difference in mean hybrid index among males and females at the end (Cherokee Lane; Figure 4-5A) location or at the middle location (Midstream; Figure 4-5B). However, there was a clear pattern of females sampled from the mouth of the marsh (Osprey) having a more southern multilocus nuclear SNP genotype and males from the same location tending to have a more northern nuclear SNP genotype (Figure 4-5C). On average, males in this region of the marsh had a significantly higher mean hybrid index calculated using the 18 SNPs (0.614) than did females (0.355; $t = 3.365$, $df = 47$, $p = 0.0015$; two-sample t-test comparing mean hybrid index), where a hybrid index of 0 represents a pure southern genotype (RUMFS-type) while a value of 1 indicates a pure northern (Cheesequake-

type) genotype. Together with the observation that males collected from this mouth location more frequently possess a northern mitochondrial type while the southern mitochondrial type is significantly more common among females, these results suggest an association between mitochondrial type, nuclear type, and sex. It should be noted that this is a phenomenon observed with the hybrid index calculated for SNP genotypes only; there was no significant difference in average hybrid index based on microsatellite genotype between males and females at this site (χ^2 = 0.6946, df = 1, p = 0.6946; two-sample Wilcoxan sign-rank test).

Is There Evidence of Gametic Incompatibility?

The results of the analysis of the split-plot design of the fertilization experiments indicated that there was no significant effect of female type (F = 1.73, p = 0.2364), male type (F = 0.123, p = 0.9453), or cross type (F = 0.5246, p = 0.6709) on fertilization success among pure and reciprocal crosses (Table 4-11).

Is There Evidence of Positive Assortative Mating?

To assess the possibility for positive assortative mating we allowed a population composed of approximately equal numbers of each sex of each subspecies to breed freely in a community tank, and then collected and genotyped the resulting fertilized embryos and hatchlings. All of the parents used in the "choice" breeding experiment were successfully genotyped at seven microsatellite loci and a single SNP in the mitochondrial Dloop. Of the 370 embryos that died prior to hatch and 309 hatchlings collected and preserved, 77 (20.8%) of the embryos and 252 (81.6%) of the hatchlings were successfully genotyped using the microsatellite loci.

While tests for deviations from HWE showed no evidence of significant heterozygote excess, several instances of significant heterozygote deficit occurred (Table 4-12). Most notable were the significant deficits of heterozygotes detected at all loci among the embryo genotypes.

In the parentage assignment analysis, 73 (94.8%) of the embryos and 238 (94.4%) of the hatchlings were unambiguously assigned parents. Mitochondrial type of these parents was then used to calculate the proportion of offspring produced by each type of pure and reciprocal cross (Figure 4-6). An excessive number of pure type offspring were produced by matings in this single tank than would be expected if mating and developmental success was random. In order to determine if positive assortative mating played a role in this observed deficit of hybrid offspring, each parent that participated in mating(s) was scored based on the proportion of matings that occurred with individuals of like mitochondrial type (and hence with the consubspecific type). The mean of this value was taken for all southern females, northern females, southern males and northern males (Figure 4-7), and the Wilcoxon Signed Rank Tests were performed to determine if the observed mean proportion of matings with "like" types deviated significantly from what was expected based on the number of males and females of both types present in the tank. Although southern females and northern males showed a significant preference for mating with consubspecifics ($W = 30.00$, $df = 10$, $p = 0.0049$; $W = 33.00$, $df = 10$, $p = 0.001$, respectively), northern females and southern males did not ($W = 29.50$, $df = 17$, $p = 0.2056$; $W = 5.50$, $df = 5$, $p = 0.2813$, respectively).

Discussion

In this study we attempted to discern if one or more pre- or post-zygotic reproductive barriers were operating or could potentially be operating among divergent, but naturally

hybridizing, subspecies of the killifish *Fundulus heteroclitus*. Our data provide evidence suggesting that a combination of assortative mating and selection against backcross genotypes might be operating to maintain this hybrid zone.

Persistence of the Trimodal Hybrid Zone Across Sampling Locations and Seasons Suggests the Zone is Stable

We found that the previously described trimodal hybrid zone between northern and southern forms of *F. heteroclitus* (Chapter Two) persists across all sampling locations and between both seasons. Consistent with the results presented in Chapters Two and Three, in both seasons and across all sampling sites, this zone is characterized by high levels of heterozygote deficit, particularly among the SNP markers, suggesting the presence of a barrier to reproduction between these divergent subspecies (Jiggins & Mallet 2000), either due to assortative mating by phenotype or to hybrid failure of particular backcross genotypes.

The hybrid indices generated from the SNP data were consistently found to fit a trimodal pattern. The microsatellite hybrid indices fit a trimodal pattern the majority of the time; however, one location in the Summer 2008 samples and two from the Fall 2009 sampling trip showed a pattern consistent with a bimodal distribution. The fact that the bimodal pattern was not picked up at the same locations across the two seasons sampled suggests that the microsatellites are showing a short-lived contemporary pattern of population structure perhaps restricted to that one sampling time, and that the SNPs are providing a more clear and consistent view into the long-term evolutionary patterns responsible for the maintenance of the hybrid zone.

Adult Habitat Preference Cannot Account for the Persistence of the Trimodal Hybrid Zone

There was no population genetic differentiation between sites that differed in abiotic characteristics within Beaverdam Creek in either season, suggesting that fish are generally

randomly distributed within the marsh with respect to genotype. Thus, our data provide no evidence that strong habitat preferences related to temperature or salinity result in population substructure across a small spatial scale. However, many other subspecific differences leading to spatial and/or temporal divergence at the time of reproduction could be playing a significant role in maintaining the trimodal pattern in hybrid indices observed. Among these are differences in preferred spawning substrate, as well as differences in spawning cues and duration of the spawning season between the subspecies (Able 1984; Marteinsdottir & Able 1992; Taylor 1986). The presence of such differences in Beaverdam Creek would have been difficult to detect based on our sampling strategy which only took into consideration the variation in adult habitat preference that might be occurring along the length of the creek. Future studies should look for more cryptic population structure at the time of reproduction, such as that which might be present among eggs laid on proximal but differentially preferred spawning substrates. Similarly, such a sampling strategy would be able to detect the potential for temporal variation in spawning activity. Such variation could occur due to reliance on different spawning cues or as the result of partially non-overlapping spawning seasons.

Cytonuclear Disequilibria and Sex May Play a Role in the Maintenance of the Hybrid Zone

One of the features of the trimodal hybrid zone of *F. heteroclitus* is the reduced presence of backcross types. The presence of cytonuclear disequilibrium provides a plausible explanation for this deficit of backcrosses, because such phenomena have been shown to result in decreased fitness in backcross types in a variety of species (Liepins & Hennen 1977; Edmands & Burton 2008, Ellison & Burton 2008). In these cases, successful F1 hybrids could be produced among divergent parental types, but backcross types suffer a decrease in fitness, particularly with respect to mitochondrial function. The success of the F1s was attributed to the mitochondrion

retaining its ability to function properly in the presence of at least a haploid set of conspecific alleles. The failure of the backcross types is thus explained by the increase occurrence of mismatched nuclear genotypes and cytotypes. We propose that these types of phenomena may be occurring among backcross types in *F. heteroclitus* and may be partially responsible for the maintenance of the trimodal hybrid zone and thus the two subspecies of Atlantic killifish. However, we did not observe a clear and consistent pattern of an excess of "matched" cytonuclear types (or a deficit of "mismatched" types at any SNP locus), suggesting that, if such a mechanism is occurring, it is either not operating strongly at the loci we surveyed, or that additional factors may be involved.

The fish collected in the Summer 2008 showed an interesting pattern with respect to the distribution of sex, mitochondrial type and sampling location in the creek. We observed an excess of males at the upstream end location (Cherokee Lane), an excess of females at the middle location (Midstream), and an excess of southern mitochondrial type females and northern mitochondrial type males at the downstream mouth location (Osprey). In addition, males at this mouth location were also significantly more northern with respect to their nuclear genotypes, while females were more southern. This pattern suggests the possibility that an interaction between sex, nuclear genotype, and mitochondrial type could be playing a role in maintaining this hybrid zone.

Significant associations between gene expression and sex (such that different sexes exhibit different responses to hypoxia at the level of gene expression) have been reported in this species among individuals collected from a nearby marsh (Mantoloking NJ, 40.039014°N, -74.050598°W; Flight *et al*. 2011). This observation opens the possibility of sex-specific differences in phenotype such that males might prefer the conditions at the end of the marsh and

females might prefer conditions at the middle of the marsh. Flight *et al.* (2011) also found one example of a significant effect of mitochondrial type on gene expression in response to hypoxia such that individuals of the northern mitochondrial type showed significantly elevated transcription of a mitochondrial elongation factor (EF-Ts_{mt}) when compared to males and females possessing the southern mitochondrial type. The Flight *et al.* (2011) study did not, however, report evidence of an interaction between sex and mitochondrial type on gene expression. Thus, while this study presents a reasonable explanation for the significant deviations from a 50:50 sex ratio observed at the downstream end and middle of the creek it is more difficult to explain the observed interaction between sex, mitochondrial type, and nuclear genotype observed at the upstream mouth site. It is possible that differences in microhabitat preferences between the subspecies coupled with sexual selection could lead to sexes of different genetic background preferring different locations within this creek.

A second possible explanation for the observed association between sex, mitotype, and location is that there is an environment-dependent cytonuclear effect involving the sex chromosomes such that male and female genotypes are more successful in opposite cytoplasmic backgrounds in certain environments (Rand *et al.* 2001). If the two subspecies are mating randomly or even semi-assortatively (i.e.: southern females choosing southern or heterozygous males and northern females choosing northern or heterozygous males), all possible combinations of male and female embryos should be produced, but pre-zygotic or post-zygotic barriers might prevent the production of southern male and northern female types for reasons unknown at this time. Laboratory experiments involving subjecting all combinations of males and females with respect to their cytonuclear types to varying temperatures, salinities, and oxygen levels could help determine if such a mechanism is operating in this species.

We also observed four SNP loci showing interesting patterns with respect to cytonuclear disequilibrium that are consistent with sex based directionality among subspecific matings (Avise 2000). However, sex-based directionality in mating should result in a whole-genome signature, which is absent in this case. Perhaps the locus-specific nature of this phenomenon suggests that a specific phenotype, controlled by a small number of loci, might be involved in a sex-specific pattern of assortative mating. Thus, while these findings are difficult to interpret they are of interest because the cline centres of two of these loci (Cytochrome p450 and Atrial Natriuretic Peptide) are among the most northerly (cline centre 1474 km north of Sapelo Island, GA) and most southerly (cline centre 738 km north of Sapelo Island, GA) of all of the cline centres for loci with known gene function (Chapter Three).

Assortative Mating Could Be Playing a Role in Maintenance of the Hybrid Zone

In this study, we found no evidence of differential fertilization success among laboratory-generated pure and reciprocal crosses. It is possible that if we had performed sperm competition experiments we would have observed fewer eggs being fertilized by heterospecific sperm (Geyer & Palumbi 2003). However, killifish are not broadcast spawners but instead form close associations during mating such that mating is virtually monogamous. Thus, sperm competition is unlikely in killifish in nature. It is also possible that reproductive character displacement in the form of gametic incompatibilities would be favoured in areas of overlap if there is strong selection against hybrid offspring, but may not be evident among individuals from allopatric populations, since these taxa never have the opportunity to mate (Palumbi 2009). Thus, while we saw no evidence of gametic incompatibilities among reciprocal crosses of pure types, it is possible that such incompatibilities would be more strongly favoured to prevent the production

of less fit hybrid offspring in the hybrid zone than among individuals from distant regions of the species range.

Unpublished results from "no-choice" experiments involving pure and reciprocal crosses of *Fundulus heteroclitus* suggest that when not given a choice, males and females of both subspecies will mate indiscriminately with members of the other subspecies (Bucking unpublished data). However, the results from the "choice" experiment presented here show that the deviation from random mating observed among the parental types suggests that assortative mating or post-zygotic hybrid breakdown could be playing a strong role in the maintenance of the hybrid zone. We cannot formally distinguish between these two possibilities because we were unable to successfully genotype all fertilized embryos. In particular, only 20.8% of the embryos that died before hatch could be genotyped. Thus, the low number of hybrid offspring we observed could be consistent with random mating followed by very early death of hybrid offspring that occurred before enough DNA had accumulated in the eggs for successful amplification of microsatellite loci. We believe that this explanation is unlikely, however, based on the ability of previous studies using forced crosses to produce fully viable hybrid offspring (Powers & Place 1978; DiMichele & Westerman 1997). Instead, we suggest that our results are more consistent with positive assortative mating. Specifically, we found that southern females show a significant preference for southern males, and likewise that northern males exhibit a significant preference for northern females. This pattern is interesting because it is in accordance with the observed excesses of southern female and northern male types at the upstream mouth of Metedeconk Creek and may provide a partial explanation for the role of positive assortative mating in the maintenance of the hybrid zone. The size and differences in meristic traits including the number of predorsal, lateral line, and caudal peduncle scales as well as anal and

caudal fin rays and gill rakers between the subspecies could help individuals recognize more suitable mates when in sympatry (Able & Felley 1986).

Conclusions

In summary, gene frequency clines in *F. heteroclitus* that have been observed along the coast and in the Chesapeake Bay were not present within a small marsh creek at the center of the coastal mtDNA cline, suggesting that genotype at the loci surveyed does not influence habitat choice in response to these abiotic variables at this spatial scale. Our data clearly indicate, however, that the previously reported trimodal hybrid zone between subspecies of *F. heteroclitus* is stable across multiple sites within this marsh and between seasons. Microhabitat partitioning based on salinity and temperature is not likely to be responsible for the maintenance of this trimodal zone. Instead, complex interactions between nuclear genotype, mitochondrial genotype and sex may be playing a role. In addition, we found some evidence in support of positive assortative mating, suggesting that reduced gene flow between the subspecies, likely due to assortative mating by phenotype, may, in part, act to maintain this hybrid zone. An extensive list of significant subspecific differences in traits relating to all aspects of the biology of these fish has been compiled and can be found in Appendix A Table A.1. It is clear, that during their time spent in allopatry the two subspecies of *Fundulus heteroclitus* accumulated numerous differences with respect to physiology, behaviour, genetics, biochemistry, embryonic and larval traits, egg morphology and nutrient allocation, all of the traits routinely used to designate the existence of species pairs. Taken together, these data imply that these various differences dependent on various combinations of traits at mitochondrial and nuclear loci and even the sex chromosomes may be reducing the frequency of successful hybridization between the two subspecies in the

region where they come into contact, thus leading to the persistence of a stable trimodal hybrid zone.

Figure 4-1. Collection locations for *F. heteroclitus* in Beaverdam Creek located between the towns of Brick and Point Pleasant in north central New Jersey, USA. Numbers correspond to locations listed in Table 4-1.

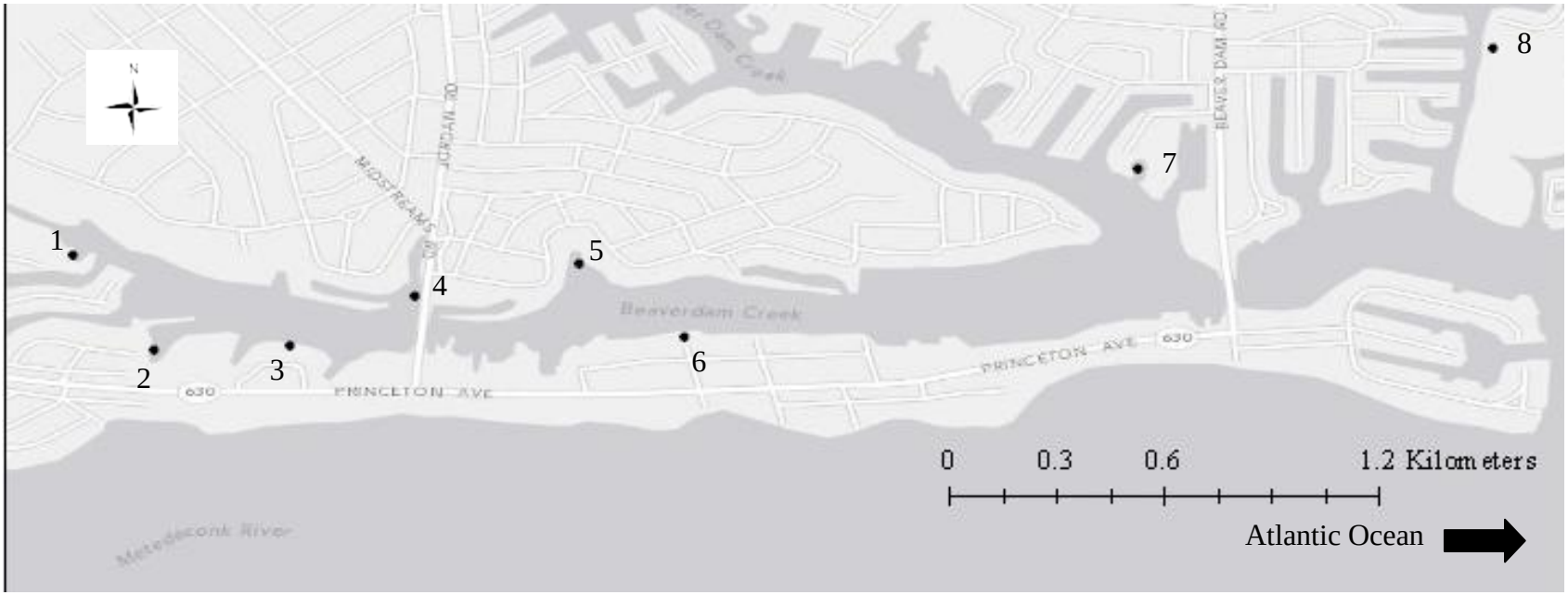


Figure 4-2. Frequency distributions of hybrid indices calculated from microsatellite (white bars) and SNP (black bars) markers for *Fundulus heteroclitus* collected in the Summer of 2008 from three locations in Beaverdam Creek, New Jersey: A) Cherokee Lane (site 1), B) Midstream (site 4), and C) Osprey (site 8).

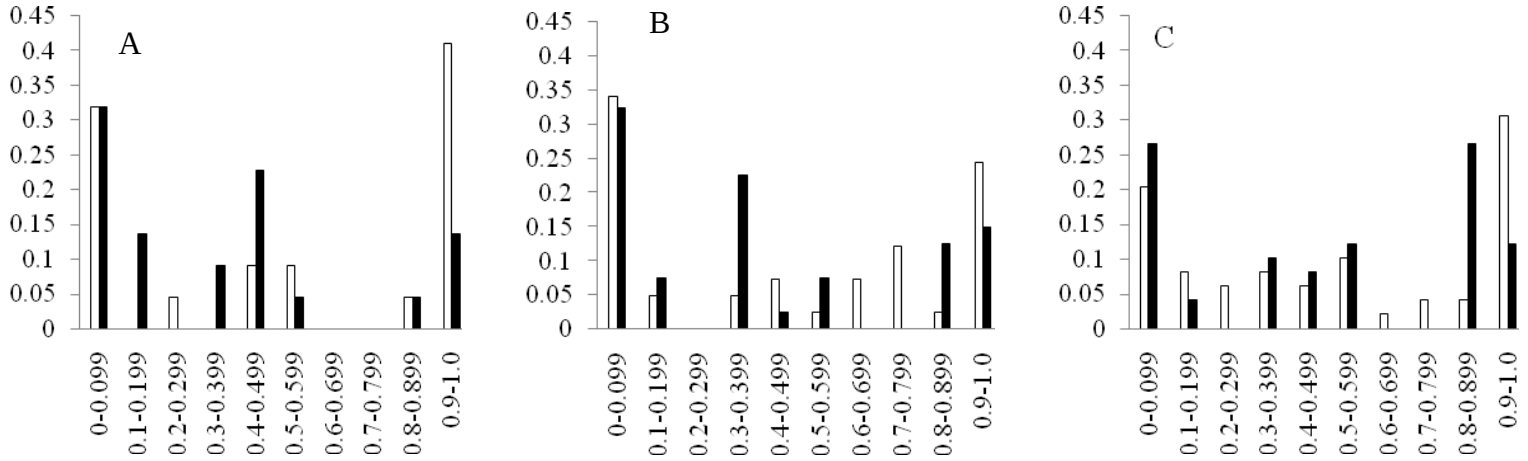


Figure 4-3. Frequency distributions of hybrid indices calculated for microsatellite (white bars) and SNP (black bars) markers for *Fundulus heteroclitus* collected in the Fall of 2009 from eight locations in Beaverdam Creek, New Jersey: A) Cherokee Lane (site 1), B) Rancosas (site 2), C) Margarita (site 3), D) Midstream (site 4), E) Crescent Park Woods (CPW; site 5), F) Parker & 1st, (site 6) G) Hidden Harbour (site 7), and H) Osprey (site 8).

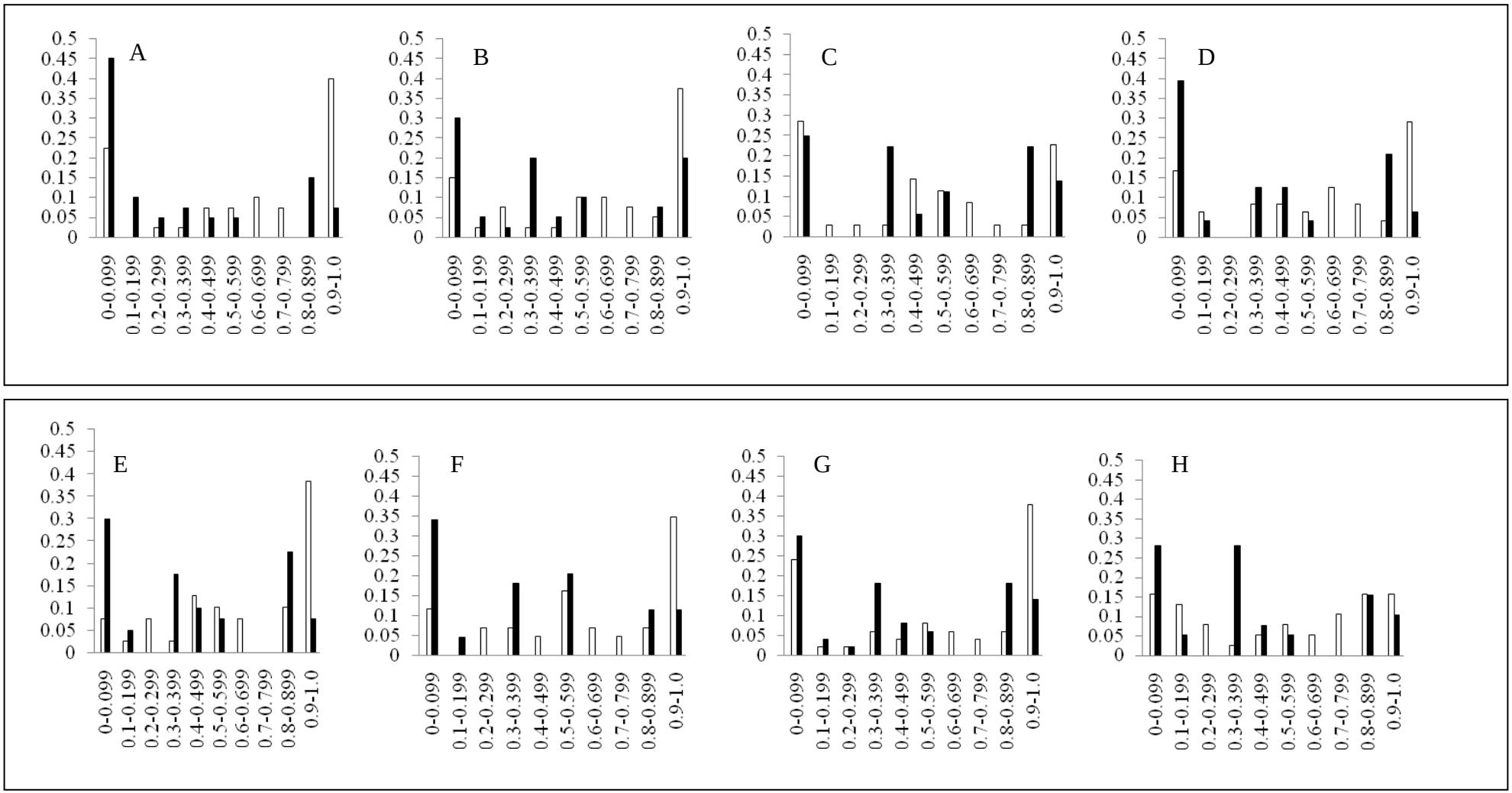


Figure 4-4. Sex of *Fundulus heteroclitus* versus mitochondrial type at three locations sampled along Beaverdam Creek, NJ during the Summer of 2008: A) Cherokee Lane (site 1), B) Midstream (site 4), and C) Osprey (site 8).

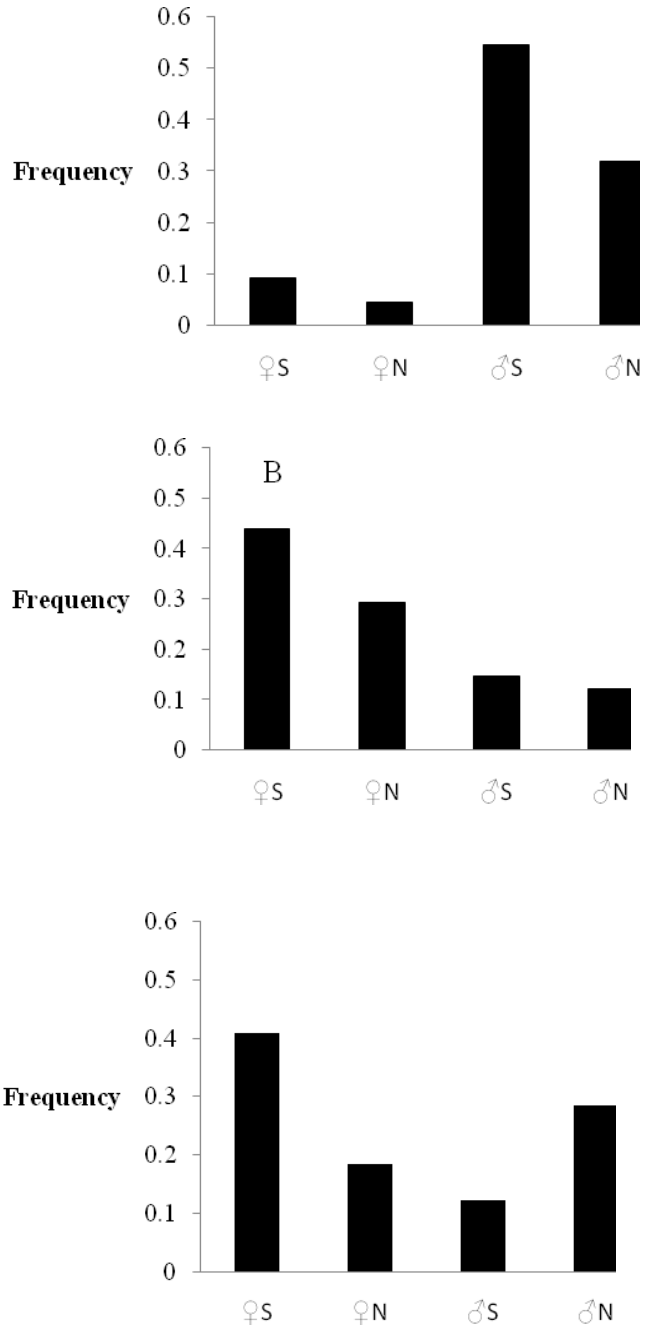


Figure 4-5. Hybrid index as calculated from 18 SNPs significantly associated with sex and mitochondrial type for *Fundulus heteroclitus* collected from three locations along Beaverdam Creek, NJ in the Summer of 2008: A) Cherokee Lane (site 1), B) Midstream(site 4), C) Osprey (site 8). White bars represent females, black bars represent males.

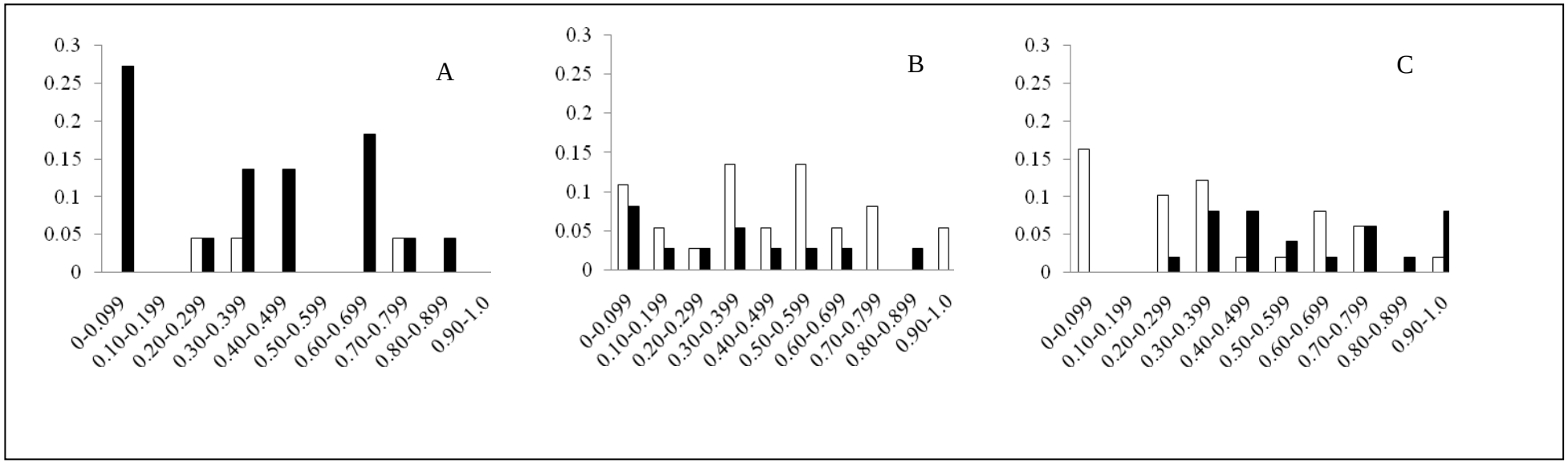


Figure 4-6. Frequency of pure and reciprocal hybrid offspring resulting from "uncontrolled" crosses involving the two subspecies of *Fundulus heteroclitus*. Embryos in light grey, hatchlings in dark grey.

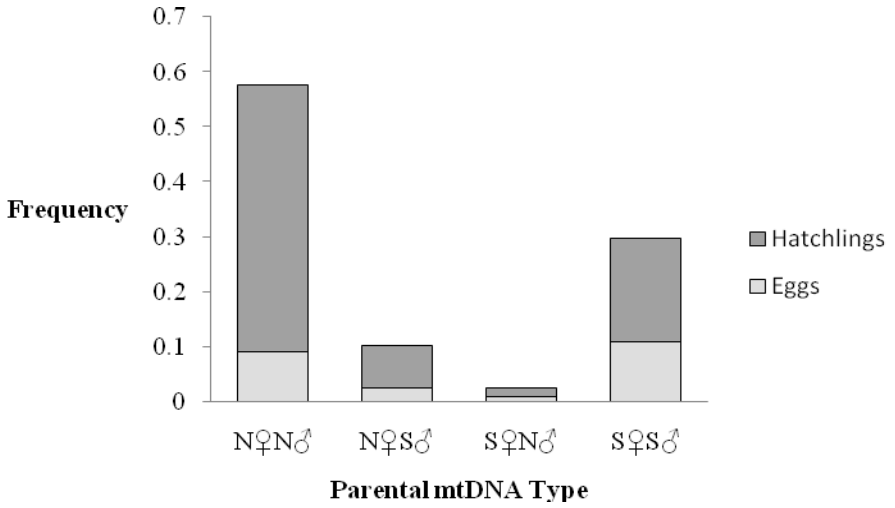


Figure 4-7. Mating preference in *F. heteroclitus*. Black circles represent the mean (±SE) of the observed frequency of matings of each parental type with "like" parental type in uncontrolled crosses in a community tank. Red circles indicate the expected frequency of matings with "like" parental type, if mating was random.

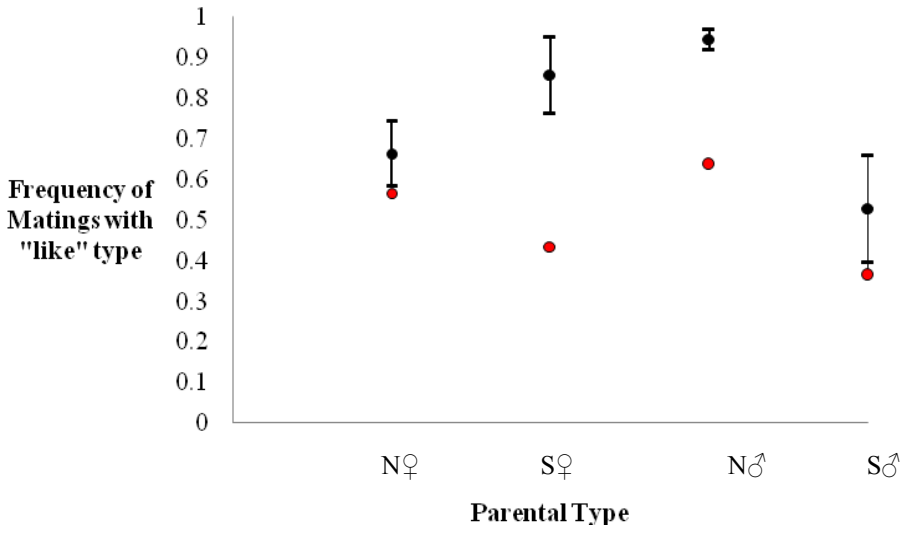


Table 4-1. Collection locations and sample sizes (n) for killifish samples collected in Summer 2008 and Fall 2009. Fall sample sizes (n) are in parentheses. Temperature (°C) values represent the maximum temperature recorded at each location; salinity (ppt) represents minimum salinity recorded. All sampling sites are located in Beaverdam Creek, New Jersey, USA.

Sampling site (Abbreviation)	Latitude	Longitude	n	Date Collected		Summer 2008		Fall 2009	
				Summer 2008	Fall 2009	Temperature	Salinity	Temperature	Salinity
1. Cherokee Lane (CL)	40.06067	-74.101217	22, (40)	✓	✓	32.4	0.8	14.7	6.4
2. Rancosas (R)	40.05828	-74.099178	40		✓			14.2	9.6
3. Margarita Place (M)	40.0584	-74.095767	37		✓			15.2	13.0
4. Midstream (MR)	40.05966	-74.092634	42, (48)	✓	✓	26.2	5.5	14.0	4.5
5. Crescent Park Woods (CPW)	40.06045	-74.088514	40		✓			14.8	8.0
6. Parker & 1st (P)	40.05859	-74.085853	48		✓			15.1	17.4
7. Hidden Harbour (S)	40.06289	-74.074438	50		✓			13.7	17.8
8. Osprey (O)	40.06597	-74.0885	50, (40)	✓	✓	27.4	21.8	13.6	17.6

Table 4-2. Pairwise F_{ST}s for the Fall 2009 sampling locations; F_{ST}s for SNPs above diagonal, microsatellites below; bold values are significant at $\alpha = 0.05$. None remained significant following FDR adjustment; locations are arranged in order along the creek, from end to mouth. Location abbreviations as in Table 4-1.

Location	**CL**	**R**	**M**	**MR**	**CPW**	**P**	**S**	**O**
CL	---	0.00068	0.00133	0.00210	-0.00059	0.00230	0.00017	-0.00030
R	-0.00236	---	0.00552	0.00088	-0.00064	0.00460	-0.00091	0.00329
M	-0.00008	-0.00361	---	**0.01003**	0.00382	0.00062	-0.00118	-0.00330
MR	-0.00073	0.00166	-0.00231	---	0.00219	0.00585	-0.00153	0.00333
CPW	**0.00546**	0.00302	0.00012	**0.00791**	---	-0.00085	0.00021	-0.00142
P	-0.00586	-0.00154	-0.00101	-0.00217	**0.00541**	---	0.00208	0.00169
S	-0.00033	0.00028	0.00073	0.00167	-0.00113	-0.00483	---	0.00233
O	0.00264	0.00467	-0.00116	-0.00768	**0.00946**	-0.00220	-0.00287	---

Table 4-3. Summer 2008 (A) and Fall 2009 (B) deviations from HWE (F_{IS}) and total number of linked loci per location calculated from 9 microsatellite loci for samples of *Fundulus heteroclitus*. Positive values of F_{IS} indicate heterozygote deficit; negative value indicate heterozygote excess. Bold type significant at $\alpha = 0.05$. Values marked with * were significant after FDR-adjustment.

(A)

Location	Locus Name									# pairs of linked loci
	FhATG18	FhATG20	FhATG17	FhATGB101	FhATGB128	FhATG2	FhATG4	FhATG6	FhCA-1	
CL	0.090	0.070	0.314	0.032	0.162	0.132	0.003	-0.139	-0.136	3 (1)
MR	**0.099***	**0.115***	-0.0284	**0.197***	0.197	-0.099	0.102	0.089	**0.123***	27 (27)
O	0.041	-0057	0.123	0.068	0.134	0.038	-0.099	0.044	**-0.112**	5 (2)

(B)

Location	Locus Name									# linked loci
	FhATG18	FhATG20	FhATG17	FhATGB101	FhATGB128	FhATG2	FhATG4	FhATG6	FhCA-1	
CL	-0.058	-0.096	0.158	-0.073	-0.032	-0.155	0.047	0.114	-0.012	5 (2)
R	0.095	-0.074	0.169	-0.124	0.083	-0.195	0.089	0.058	0.048	10 (7)
M	**0.095***	-0.059	-0.055	0.006	**0.080***	-0.067	-0.088	0.046	-0.088	5 (2)
MR	-0.128	0.003	0.051	0.035	-0.083	0.107	-0.024	-0.053	0.025	2 (0)
CPW	-0.069	0.083	**0.425**	-0.042	-0.049	**-0.343**	0.105	-0.048	-0.066	4 (0)
P	-0.130	0.065	-0.011	-0.046	-0.118	-0.112	-0.052	-0.166	**0.054**	9 (8)
S	0.050	0.041	0.177	0.104	0.137	**-0.245***	**0.137**	0.095	0.034	8 (1)
O	**0.169**	0.117	0.113	0.107	-0.009	**-0.426**	**-0.155**	**-0.247**	0.088	7 (2)

Table 4-4. Deviations from HWE (F_{IS}) and total number of linked loci per locations calculated from 30 nuclear SNPs for samples of *Fundulus heteroclitus* for Summer 2008 (A) and Fall 2009 (B). Positive values of F_{IS} indicate heterozygote deficit while negative values represent heterozygote excess. F_{IS} values in bold are significant at $\alpha = 0.05$; values marked with * are significant following FDR-adjustment. Locus numbers are: 1. 1173_9, 2. 1176_169, 3. HDDC2, 4. actin binding LIM family gene, 5. LDHB1033, 6. cytochrome p450, 7. myoglobin, 8. ribosomal protein S2, 9. 60Sribosomal protein L6, 10. activated protein kinase, 11. atrial natriuretic peptide, 12. warm-acclimation related protein, 13. 14-3-3 epsilon, 14. 40S ribosomal protein S17, 15. 60S ribosomal protein L35, 16.translationally controlled tumor protein, 17. phosphate carrier protein mitochondrial precursor, 18. ribosomal protein, 19. tropomyosin. "+" represents the number of linked pairs of loci with positive Rij values, while "-" indicates those with negative Rij values in the column labeled FDR are the number of significant pairwise comparisons following FDR-adjustment. Values in bold are significant at $\alpha = 0.05$; value marked with an * are significant after FDR-adjustment; NA = analysis not applicable because locus represented by single allele.

(A)

Location	Locus Number										
	1	2	3	4	5	6	7	8	9	10	11
CL	0.135	0.270	**0.596**	0.091	-0.327	-0.024	0.052	0.313	0.070	-0.346	0.173
MR	0.188	0.011	0.290	**0.638***	-0.083	0.127	0.092	0.183	0.157	-0.075	0.019
O	0.094	**0.466***	**0.481***	**0.547***	-0.219	**0.299**	0.299	0.227	**0.311**	-0.194	0.085

(A) continued.

Location	Locus Number								# linked loci		
	12	13	14	15	16	17	18	19	+	-	FDR
CL	0.162	**0.908***	0.203	0.004	NA	-0.235	0.25	0.041	15	8	0
MR	0.279	**0.375**	0.302	-0.054	-0.059	0.094	**-0.329**	-0.014	18	15	1
O	**0.309**	**0.668***	0.147	0.077	**0.793***	-0.121	-0.084	**-0.313**	20	9	0

(B)

Location	Locus Number										
	1	2	3	4	5	6	7	8	9	10	11
CL	-0.067	0.010	**0.552***	0.318	-0.022	0.044	0.003	**0.627***	0.149	-0.029	0.312
R	0.037	**0.556***	0.192	**0.734***	-0.156	-0.258	-0.156	0.169	0.261	0.179	-0.134
M	**0.421**	**0.454**	**0.382**	**0.481**	-0.015	0.314	**0.397**	0.303	**0.389**	**0.339**	0.167
MR	0.027	**0.342**	**0.317**	**0.513***	**-0.362**	0.041	0.255	0.153	0.011	-0.075	**0.338**
CPW	0.037	**0.607***	0.194	**0.599***	0.090	-0.025	0.110	**0.412**	0.162	0.149	0.245
P	0.007	**0.249**	0.165	**0.674***	0.111	**0.363**	-0.167	0.141	0.235	0.235	-0.143
S	**0.361**	**0.421***	0.272	**0.437***	**0.297**	-0.196	0.059	**0.370**	0.208	0.129	0.265
O	0.037	0.141	0.235	**0.409**	-0.205	0.002	0.036	0.217	0.046	0.103	0.221

(B) continued.

Location	Locus Number								# linked loci		
	12	13	14	15	16	17	18	19	+	-	FDR
CL	0.101	**0.557***	0.039	0.189	NA	-0.085	0.146	-0.116	15	13	0
R	-0.002	**0.842***	**0.504***	-0.094	NA	**0.368**	-0.085	-0.054	15	14	2
M	0.219	**0.318**	**0.429**	-0.038	-0.030	0.167	0.209	-0.094	10	13	0
MR	0.147	**0.500***	0.052	-0.277	-0.022	-0.040	-0.089	0.032	13	10	0
CPW	**0.339**	**0.639***	0.092	0.060	-0.026	-0.093	0	0.101	14	12	1
P	0.198	**0.642***	**0.531***	0.078	-0.035	-0.012	-0.052	-0.057	14	21	2
S	**0.541***	**0.666***	**0.361**	-0.060	**0.651***	-0.039	-0.037	-0.038	9	12	2
O	**0.382**	**0.645***	**0.508***	**0.434**	NA	0.166	0.100	0.108	17	8	1

Table 4-5. Summary of output from MCLUST for Summer 2008 collection locations.

Location/marker	Transformation	Number of Clusters		
		1	2	3
Cherokee Lane microsatellites	Logit	-30.73512	-24.61771	-30.79979
Cherokee Lane SNPs	Noise	-21.184938	-7.774056	-5.710004
Midstream microsatellites	Logit	-47.053487	-21.376388	-6.882592
Midstream SNPs	Logit	-39.600248	-26.950261	-6.417191
Osprey microsatellites	Noise	-53.719100	-9.509720	4.245583
Osprey SNPs	Noise	-48.584947	-36.352313	9.262901

Table 4-6. Summary of output from MCLUST for Fall 2009 collection locations.

Location/marker	Transformation	Number of Clusters		
		1	2	3
Cherokee microsatellites	Logit	-44.261433	-3.258341	24.110287
Cherokee SNPs	Logit	-36.81409	24.30078	28.85191
Rancosas microsatellites	Logit	-36.78286	-17.90805	-10.45267
Rancosas SNPs	Noise	-37.88811	-17.09798	-19.12959
Margarita microsatellites	Noise	-39.01564	-23.78540	-7.53136
Margarita SNPs	Logit	-35.4922440	-26.3629337	-4.1644031
Midstream microsatellites	Logit	-44.967166	-1.485234	4.036609
Midstream SNPs	Noise	-43.29987	1.56740	10.89095
Parker microsatellites	Logit	-35.116660	-1.641666	5.502901
Parker SNPs	Noise	-39.747210	-9.905465	-5.676620
CPW microsatellites	Logit	-27.4726843	2.8097491	0.7279944
CPW SNPs	Noise	-36.791359	-40.041471	-3.008797
Hidden Harbor microsatellites	Logit	-55.829968	-15.776235	4.112093
Hidden Harbour SNPs	Logit	-48.671687	-33.226701	-1.408832
Osprey microsatellites	Noise	-38.18114	-36.50655	-43.30832
Osprey SNPs	Noise	-35.410845	-26.100453	-6.505458

Table 4-7. Summer vs. Fall pairwise F_{ST} for *Fundulus heteroclitus* from the three locations in Beaverdam Creek, NJ sampled in both seasons. SNPs above diagonal, microsatellites below; bold values are significant at $\alpha = 0.05$. None remained significant following FDR adjustment.

	Cherokee Lane Summer	**Midstream Summer**	**Osprey Summer**	**Cherokee Lane Fall**	**Midstream Fall**	**Osprey Fall**
Cherokee Lane Summer	---	0.00256	0.00524	0.00425	-0.00115	0.00918
Midstream Summer	-0.00850	---	0.00627	-0.00365	0.00180	0.00163
Osprey Summer	-0.01526	-0.00098	---	0.00253	0.00367	0.00051
Cherokee Lane Fall	-0.01643	-0.00313	-0.00175	---	0.00210	-0.00030
Midstream Fall	-0.01180	0.00127	0.00315	-0.00073	---	0.00333
Osprey Fall	-0.01414	-0.00144	**0.00616**	0.00264	-0.00768	---

Table 4-8. Summary of cytonuclear disequilibria occurring among individuals collected in Summer 2008 (-S) and Fall 2009 (-F). M = southern mitochondrial type, m = northern mitochondrial type; A = southern nuclear allele, a = northern nuclear allele; inferred signatures of cytonuclear disequilibria as in Avise 2000. $^{\delta}$ = southern females less discriminatory; * = northern females less discriminatory; ** = pattern consistent with introgression into both species.

Location	Locus	Pattern (excess cytonuclear type(s) relative to expected value) with p-values following in parentheses
Cherokee Lane-S	Cytochrome p450	Excess MAA (0.03096), mAa (0.006192)$^{\delta}$
	60S ribosomal protein L6	Excess Maa (0.04556)
Midstream-S	LDHB654	Excess MAA (0.001981), maa (0.03609)**
	Elastase	Excess mAa (0.01892)
	Glyceraldehyde 3 phosphate dehydrogenase	Excess mAA (0.01699)
	40S ribosomal protein S17	Excess mAa (0.02861)
Osprey-S	Phosphate Carrier Protein	Excess mAA (0.03866), MAa (0.0164)
	LDHB654	Excess MAa (0.04507)
	60S ribosomal protein L35	Excess maa (0.04107)
Cherokee Lane-F	LDHB654	Excess MAA (0.03574)
	Nucleotide diphosphate kinase 1	Excess MAa (0.04885)
	Glyceraldehyde 3 phosphate dehydrogenase	Excess MAa (0.00255), maa (0.01781)*
Rancosas-F	Ribosomal Protein S2	Excess Maa (0.01124)
	1176	Excess MAa (0.01248)
	Nucleotide diphosphate kinase 1	Excess maa (0.04231)
Margarita-F	Myoglobin	Excess mAa (0.02749)
	Ribosomal protein S2	Excess MAa (0.01918)
Midstream-F	Glyceraldehyde 3 phosphate dehydrogenase	Excess mAa (0.03703)
	Parvalbumin	Excess mAA (0.0008474), MAa (0.04869)
	14-3-3 epsilon	Excess Maa (0.01883), mAA (0.004092)
Crescent Park Woods-F	Elastase	Excess mAa (0.02658)
	Atrial Natriuretic Peptide	Excess MAa (0.008377), maa (0.009324)*
	Titin Cap	Excess MAa (0.01887)
Hidden Harbour-F	Elastase	Excess mAA (0.03599), MAa (0.03404)
Osprey-F	Elastase	Excess mAA (0.01279), MAa (0.01279)
	40S ribosomal protein S17	Excess MAa (0.01164)

Table 4-9. Summary of deviations of sex ratios and mtDNA type ratios from 50:50, and interaction between sex of killifish and mtDNA type for the Summer 2008 samples of *Fundulus heteroclitus*. Values are χ^2, df, p-value; Pearson values from JMP

Location	♀	♂	Sex Ratio 50:50?	S	N	mt Ratio 50:50?	♀S	♀N	♂S	♂N	Sex & mt type independent?
Cherokee Lane	3	19	No; 11.6364, 1, 0.0006	14	8	Yes; 1.6364, 1, 0.2008	2	1	12	7	Yes; p = 1.0000*
Midstream	30	11	No; 8.8049, 1, 0.003	24	17	Yes; 1.1951, 1, 0.2743	18	12	6	5	Yes; p = 1.0000*
Osprey	29	20	Yes; 1.6531, 1, 0.1985	26	23	Yes; 0.1837, 1, 0.6682	20	9	6	14	No; 7.216, 1, 0.0072

*Values from Fisher's Exact Test; expected frequency of >20% cells less than 5.

Table 4-10. F_{ST} between sexes of *Fundulus heteroclitus* and locations sampled during the Summer of 2008. SNPs above diagonal, microsatellites below; values in bold significant at $\alpha = 0.05$; * indicates values significant after FDR-adjustment.

	♀ **Cherokee Lane**	♀**Midstream**	♀**Osprey**	♂ **Cherokee Lane**	♂ **Midstream**	♂ **Osprey**
♀ **Cherokee Lane**	---	0.03661	0.04570	0.01150	0.00969	0.04106
♀ **Midstream**	-0.01871	---	0.00623	0.00730	**0.02560**	**0.02567***
♀ **Osprey**	-0.02982	-0.00663	---	-0.00268	0.01244	**0.01704**
♂ **Cherokee Lane**	-0.00088	-0.00682	-0.01318	---	0.00498	0.01786
♂ **Midstream**	-0.02584	0.00980	0.00913	-0.00180	---	**0.03042**
♂**Osprey**	-0.02327	**0.02448***	**0.01773**	-0.00395	-0.00764	---

Table 4-11. Mean fertilization success for *in vitro* pure and reciprocal crosses of *Fundulus heteroclitus,* mean fertilization success rate for each cross, and the pooled mean fertilization success rate for pooled males.

Pooled Females	N♀				S♀			
Individual Males	N♂1	N♂2	S♂1	S♂2	N♂1	N♂2	S♂1	S♂2
	0.90	0.89	0.93	0.84	0.95	0.88	0.92	0.88
	0.80	0.81	0.84	0.89	0.81	0.87	0.76	0.80
	0.81	0.88	0.84	0.77	0.82	0.65	0.76	0.82
	0.94	0.93	0.97	0.91	0.63	0.87	0.77	0.87
Mean fertilization success rate	0.86	0.88	0.89	0.85	0.80	0.82	0.80	0.84
Mean fertilization success for pooled males	0.87		0.87		0.81		0.82	

Table 4-12. F_{IS} values from tests for deviations from Hardy Weinberg Equilibrium of offspring resulting from "choice" breeding experiment among subspecies of *Fundulus heteroclitus*. Positive values of F_{IS} indicate heterozygote deficit while negative values indicate heterozygote excess. Values significant at $\alpha = 0.05$ are highlighted in bold.

	FhATG2	FhATG18	FhATG20	FhATG4	FhATGB128	FhATGB103	FhATGB101
South Parents	0.0423	-0.0231	-0.0255	**0.232**	**0.1535**	-0.0234	0.15
North Parents	-0.0612	-0.047	0.093	0.0127	0.0805	-0.009	-0.0193
Embryos	**0.3198**	**0.1846**	**0.4054**	**0.281**	**0.2351**	**0.0637**	**0.0564**
Hatchlings	**0.2998**	0.0912	**0.2972**	0.1083	0.1677	-0.0067	0.0236

Chapter Five: Conclusions

For my Ph.D. dissertation I used multi-locus genetic analyses, in conjunction with experimental crosses and studies of mating preference, to provide insight into the various factors that maintain differentiation between two subspecies of the Atlantic killifish, *Fundulus heteroclitus*. More specifically, I performed a genetic characterization of the *F. heteroclitus* hybrid zone using a combination of mitochondrial markers, putatively neutral nuclear microsatellite markers, and potentially non-neutral SNPs in order to determine the shape of the hybrid zone between *F. heteroclitus* subspecies. Next, I performed clinal analyses on data from mitochondrial and nuclear SNPs in coding genes from *F. heteroclitus* populations along the Atlantic coast of North America to ultimately determine if selection or assortative mating might be involved in the maintenance of the hybrid zone. My last objective was to determine whether pre-zygotic factors including habitat specialization, gametic incompatibility or assortative mating are involved in the maintenance of the hybrid zone. In an attempt to implicate habitat specialization as a pre-zygotic barrier operating within the hybrid zone I looked for evidence of fine-scale population structure among fish sampled from a marsh located within the centre of the zone using a combination of mitochondrial, nuclear microsatellite, and nuclear SNPs. I then looked for differences in fertilization success among pure and reciprocal crosses of *F. heteroclitus* in the laboratory in order to determine if gametic incompatibilities could be promoting the maintenance of the zone. Lastly, I performed a choice breeding experiment to determine if positive assortative mating could possibly be involved in the genetic patterns observed within the hybrid zone.

The major findings of my Ph.D. research were:

1) My observation of reduced frequency of backcross genotypes in populations at the centre of the *F. heteroclitus* contact zone suggest that this zone may be, in part, maintained by contemporary forces such as endogenous or exogenous selection or assortative mating reproductive isolation.

2) My observation of concordant and coincident clines in mitochondrial and nuclear SNP markers suggest that this reproductive isolation may involve some form of endogenous selection involving the mitochondrial genome, but that this pattern may have a sex-specific component

3) My studies on fertilization success and mate choice suggest that the northern and southern subspecies of *F. heteroclitus* are not separated by any obvious gametic barrier that prevents fertilization in forced crosses, but instead appear either to mate assortatively when provided with a choice or to produce some non-viable hybrids under these circumstances.

Genetic Evidence for Reproductive Isolation in Marshes Near the Centre of the Contact Zone

Three lines of genetic evidence support the presence of reproductive isolation within the contact zone of *F. heteroclitus*. The high levels of heterozygote deficit observed specifically among the nuclear SNP loci among individuals collected from the centrally located marsh Beaverdam Creek (of the Metedeconk River system) indicates that pre- and/or post-zygotic reproductive barriers are operating in the maintenance of the zone, ultimately resulting in the decreased production of hybrid offspring. Supporting this observation are the similarly high levels of cytonuclear disequilibrium observed among individuals also collected from this creek. These high levels of heterozygote deficit and disequilibrium were evident across sampling sites within the same marsh and across seasons. The third and final piece of genetic evidence supporting reproductive isolation in these marshes was my observation of a trimodal frequency

distribution of hybrid indices in Metedeconk marsh. As with the other forms of evidence, this trimodal frequency distribution was evident across sampling sites and seasons.

Evidence for Endogenous Selection Maintaining the Contact Zone

The coincidence and concordance of steep clines in allele frequency in the mitochondrial genome and at several nuclear loci coupled with high levels of heterozygote deficit at these markers suggests a possible role for endogenous selection in the maintenance of this hybrid zone. Furthermore, the coincidence and concordance of the steepest clines in allele frequency of several nuclear SNPs with the mitochondrial SNPs, in conjunction with observed cytonuclear disequilibrium at several loci in a centrally located marsh, suggest a role for cytonuclear epistasis in the maintenance of the hybrid zone. However, similar patterns could also be produced by exogenous selection or assortative mating for particular aspects of phenotype.

Evidence for Assortative Mating or Hybrid Inviability

My breeding experiments suggest that there is no barrier to fertilization between *F. heteroclitus* subspecies, but nevertheless, the offspring resulting from the choice breeding experiment showed strong evidence of being parented by subconspecifics. These results were further supported by the observation of a deficit of offspring produced from crosses of southern female killifish with northern male killifish. These data suggest that either southern female killifish and northern male killifish showed significant preference for mating with consubspecifics, or that crosses between these two types result in non-viable offspring. The results of the fertilization experiments indicate that gametic incompatibilities do not pose a pre-zygotic reproductive barrier, thus it appears that positive assortative mating or hybrid inviability could be playing a role in the maintenance of the hybrid zone.

Limitations of My Approach

All Ph.D. theses are ultimately limited by time and going hand in hand with this time limitation was the fact that *Fundulus heteroclitus* live on the Atlantic coast of North America, which limits the amount of field collection that is possible for a student based on the west coast. That said, I was fortunate to be able to travel to the east coast on two occasions to collect my own samples. The advantage that accompanies being able to visit the actual location from which your animals came goes a long way when trying to integrate the ecological factors that may be playing a role in the divergence of a species.

I was also very fortunate to be performing research utilizing population genetic techniques at a time when the use of medium-throughput SNP genotyping such as that presented by the Sequenom iPLEX Gold technology was becoming more and more affordable and thus accessible for researchers studying non-model organisms. However, knowing the complete genome of *F. heteroclitus* would have been useful in my study, not only because it would have allowed me to identify my "unknown" SNPs (i.e., 1173, 1176), but also because it would have enabled me to determine any physical linkages occurring among my nuclear markers. The full genome of *F. heteroclitus* is currently being sequenced, thus this is potentially only a short-term limitation.

Another precautionary note when utilizing such medium throughput techniques is that strengths and types of selection can be extremely variable among loci. Analyzing data before distinguishing between loci experiencing the at least differential and potentially opposed effects of drift and selection can obscure the patterns observed and thus the interpretations of the data. In my study, it appears that while clines in allele frequency are being maintained by a selective force related to mitochondrial function, several other allele frequency clines show patterns of not

being related to mitochondrial function but perhaps by an exogenous factor at the edges of the distribution of the subspecies. Ultimately, because selection can act differentially at different markers, researchers still need to investigate the patterns of variation at each locus individually from the perspective of the questions posed in order to really narrow in on traits important to ecology and evolution of a population or species. High-throughput techniques are now becoming more accessible to evolutionary biologists, and these technical advances accompanied by the sequencing of the *F. heteroclitus* genome should soon allow the hypotheses (e.g. about cytonuclear epistasis and heterozygote deficit at specific loci) that I generated in my thesis to be tested at a whole genome scale, bearing in mind that my data suggest that the results are likely to be locus-specific.

A critical underlying assumption of population genetic studies is that the populations under examination are in migration-drift equilibrium (Whitlock 1992). Interpretation of data unknowingly not meeting this criteria may result in historical processes being discounted and the presence of contemporary forces (i.e. natural selection), being deemed responsible for the observed patterns of genetic variation. Duvernell *et al*. (2008) have suggested based on data showing evidence of large population size (ranging from 3500 to 30000 individuals) and low migration rate (95% confidence limits on Bayesian estimates of $N_e m$ ranged from 1 -12 effective migrants per generation) that coastal populations of *F. heteroclitus* are not yet in migration-drift equilibrium and that secondary contact at the time of the last glaciation is responsible for the patterns of genetic variation observed today. They also argue that there results are in agreement with Sweeney *et al.*'s (1998) mark-recapture study who calculated an effective population size of 6800 individuals and $N_e m = 6.3$. However, these calculations are based on within marsh mark-recapture movement estimates and most likely represent an underestimate of dispersal. Our

estimates of dispersal as calculated from average pairwise linkage disequilibrium gave a value of $\sigma = 14.53$ km. Since only a few migrants per year would be needed to effectively homogenize the genetic composition of populations across this distance, it seems reasonable to conclude that while the historical process of secondary contact following the last glaciation was responsible for the initial formation of the hybrid zone of *F. heteroclitus*, a contemporary force such as natural selection is responsible for its maintenance. Indeed, the observation that only a few markers maintain very steep clines across the *F. heteroclitus* hybrid zone (Strand *et al*. 2012 and Chapter Three of this thesis), while most have broad clines or are homogenized, supports this conclusion.

Suggestions for Future Studies

While differences in temperature and salinity tolerance are known to exist among the subspecies of killifish, a host of other differences such as use of spawning substrate, divergence in spawning cues, and subspecific shifts with respect to spawning season are potential candidates to play a role in the persistence of these two distinct subspecies. Fish from the northern part of the distribution show strong preference for spawning on sandy marsh bottoms or algal mats while those from the southern end of the range spawn almost exclusively in shells of the ribbed mussel *Geukensia demissa* (Able 1984). Such differences in preference for spawning substrate could present a premating barrier if they were also occur in the hybrid zone. Also, individuals from the southern end of the range show a semilunar spawning strategy, whereby eggs are deposited at the edge of the spring high tide, incubate aerially for two weeks, and hatch upon immersion by the following spring tide (Taylor *et al*. 1977). Such a pattern has not been observed among northern fish that experience a much shorter spawning season (June through mid-July; Taylor 1986) than do their southern counterparts (May through late August; Taylor 1986) and would thus experience fewer spawning events if restricted to a semilunar pattern.

Instead northern fish seem to spawn continuously throughout the spawning season (Marteinsdottir & Able 1992; Bucking and McKenzie, personal observations in the laboratory). These differences in spawning behaviour are highly correlated to significant differences in egg size and hatching time and as such most likely have a strong underlying genetic component. Northern adults lay significantly smaller eggs than do southern fish (Costello *et al*. 1957; Wallace & Selman 1978; Taylor & DiMichele 1980; Marteinsdottir & Able 1988), and northern embryos require fewer days to hatch than embryos generated from crosses of pure southern types (10.5 - 12.5 days vs. 14.0 - 15.0 days based on *in situ* observations; DiMichele & Westerman 1997). These smaller northern eggs, allocated with less nutrients than the larger eggs of the southern fish, would not survive extended incubation if they were not immersed by the following high tide after which they were laid (Marteinsdottir & Able 1992). Such stranding however, could be survived by the larger southern eggs which showed 100% survival after a 28 day aerial incubation compared to 49% survival seen among the northerns (Marteinsdottir & Able 1992). Such asynchrony in spawning activity could also result in the trimodal pattern of genotype distribution such as that seen here.

In addition to these differences in reproductive traits, the two subspecies of killifish differ in a variety of morphological and physiological traits that could result in further restriction of their contact with one another on a microhabitat scale. Differences in appearance of males and females of the respective subspecies especially during breeding could encourage positive assortative mating. More subtle differences in scale numbers and the anal sheath lengths of females of the different subspecies could provide further cues to the opposite subspecies that they are about to venture down an ill-fated reproductive pathway. These morphological differences are also coupled with differences in meristic traits, most notably a significant

difference in gill raker number with the larger southern subspecies having a slightly higher gill raker count than the smaller northern subspecies. Similar differences could have effects on predator-prey interactions and result in further partitioning of the subspecies in order to reduce competition between them. Such differences in size and gill raker number ultimately leading to differences in habitat use and positive assortative mating has been well-documented among species pairs of stickleback and has been implicated in ecological speciation. Freshwater and marine as well as benthic and limnetic forms of threespine stickleback provide some of the best evidence for ecological speciation in sympatry as a result of differences built-up between species pairs during their time spent in allopatry (Hay & McPhail 1975; Nagel & Schluter 1998; Rundle & Schluter 1998; Vines & Schluter 2006). Among marine and stream species pairs, the marine form is characterized by a larger, terete-shaped body, the lateral sides of which are fully covered by an armour of bony plates, and possess long fine gill rakers best suited for the filter-feeding lifestyle this species assumes in the upper layers of the water column (Hagen 1967). In contrast, the stream form is much smaller and the number of bony plates is significantly reduced, as are the number and length of gill rakers, as a result of this species' divergent adaptation to a life spent feeding on bottom-dwelling organisms. Such adaptations to different habitats is sufficient to prevent the formation of otherwise viable hybrids, which, when generated in the laboratory show no sign of hybrid inviability or inferiority (Hagen 1967). Such ecological speciation may be operating among the subspecies of *Fundulus heteroclitus.*

Future studies should focus on explicitly testing the hypothesis of the role of ecological speciation in the maintenance of separate northern and southern subspecies of the killifish. In addition, studies testing for the role of exogenous selection of loci non-coincident with the main mitochondrial and SNP cline centre should be conducted. Lastly, the effects of the interaction

between sex and mitochondrial type needs to be explored. Below I have listed specific testable hypotheses pertaining to these questions and also give a brief outline of the experimental design.

Tests for Ecological Speciation

1. Spawning site preference.

As discussed above, differences in spawning site preference could be responsible for reducing gene flow between northern and southern *F. heteroclitus* in the contact zone. To assess spawning site preference, I suggest returning to Beaverdam Creek, which is located at the centre of the hybrid zone, and genotyping deposited eggs collected from both mussel shells and algal mats and/or muddy substrate using a panel of microsatellite loci. A significant genetic difference among populations of eggs found at these respective sites would be indicative of differential use of spawning habitat by divergent individuals. Furthermore, mitochondrial-typing of these eggs would indicate if there was a preponderance of northern- and southern-type females utilizing the predicted spawning substrate.

2. Differences in spawning cues

Since allopatric populations of *F. heteroclitus* differ in their spawning cues, it is possible that these differences are also present in the hybrid zone and serve to reduce gene flow between the subspecies. Conducting collections of eggs over an entire spawning season at a site such as Beaverdam Creek would enable the researcher to determine if the subspecies were responding to different spawning cues such as has been seen in allopatric populations. If southern-type individuals were retaining their semi-lunar spawning behaviour in the hybrid zone, then eggs possessing southern mitochondrial types should be laid in peaks corresponding to the occurrence of full moons, while northern mitochondrial types should exhibit no such pattern and simply maintain consistent levels of egg production throughout the spawning season.

3. Assortative Mating

My data suggest that assortative mating may be an important factor in reducing gene flow between the two subspecies of *F. heteroclitus*, but these findings should be confirmed in individuals from populations closer to or within the contact zone. Breeding trials should also be conducted among individuals of differing mitochondrial type collected from Beaverdam Creek in order to see if the positive assortative mating observed among allopatric populations persists within the hybrid zone. In addition, individuals used in these experiments could be provided with alternative spawning substrates to see if this encourages divergent reproductive behaviour.

4. Size as an isolating barrier

Fish of the northern subspecies are generally smaller than those of the southern subspecies, which could provide a barrier reducing gene flow between the subspecies. As has been done previously with different stickleback species pairs, northern type individuals (the smaller of the two subspecies) could be given a choice to mate with a size-matched southern individual or larger southern individual in order to see if size is influencing mate choice. If the small northern individuals chose to mate significantly more often with a size-matched southern individual then it may be concluded that size is playing a role in aiding the fish in selecting "matched" mates.

5. The role of gametic incompatibility

My data suggest that there is no gametic incompatibility between individuals from the pure northern and southern populations, but reinforcement and character displacement within the contact zone could potentially cause a greater degree of gametic incompatibility among sympatric individuals. In order to determine if gametic incompatibilities have been subjected to character displacement within the hybrid zone, gametes could be stripped from divergent adults

(possessing opposite mitochondrial types) collected from the contact zone and *in vitro* crosses could be performed to see if differences in fertilization success occur between pure and reciprocal crosses. The occurrence of a significant excess of successful fertilizations among parents with like mitochondrial types would implicate character displacement with respect to factors involved in gametic incompatibilities.

6. Post-zygotic barriers to gene flow

My data are consistent with the hypothesis that backcross individuals are rare within the hybrid zone, suggesting possible endogenous selection, but this possibility has not yet been tested in a controlled laboratory setting. Adults from the respective edges of the species' range could be crossed in the lab. The resulting advanced generation crosses and backcrosses could then be subjected to various fitness tests (exposed to different thermal and/or salinities regimes, tested for differences in critical swimming speed) to see if decreased fitness occurs among recombinant types relative to the parents. This would be a long-term experiment as killifish require one year to reach sexual maturity.

7. Microhabitat partitioning

Although my genetic data show no strong evidence of microhabitat partitioning, other approaches might provide a more fine-grained view of the *in situ* ecological patterns. A simple stomach content analysis of genetically divergent individuals collected from the hybrid zone could be performed in order to see if differences in diet composition could be promoting ecological speciation among divergent fish from a sympatric population. For example if diet differences between the subspecies were found to be correlated with the subspecies using different foraging habitats, such a finding could be taken as evidence in support of ecological divergence.

Tests for Sex-Mitotype-Environment Effects

All sex and mitochondrial type combinations (FS, FN, MS, MN) should be tested with respect to differential fitness, again by being exposed to different thermal and/or salinity regimes, and examined for differences in critical swimming speed. Any significant differences in these fitness experiments observed among different sex*mitochondrial types could help unravel the apparent relationship between sex and mitochondrial type and environment that we observed in the wild.

Tests for Exogenous Selection

Two SNPs clines (cytochrome p450 and atrial natriuretic peptide) having centres highly non-coincident from the predominant mitochondrial and SNP clines could be being maintained by differential responses of individuals harboring either northern or southern genotypes with respect to these loci to conditions on either side of an environmental barrier. In order to determine if temperature could be implicated in presenting a barrier to dispersal, CTMin and CTMax of individuals possessing respective northern and southern genotypes could be determined. The observation of a higher CTMin for example of fish possessing the southern nuclear genotype at the SNP cytochrome p450 (or conversely a lower CTMax among northern genotype fish), could explain why this cline centre is located to the north of the other cline centres.

Conclusions and Implications

Mayr proposed that reproductive isolation be the most important criterion used to diagnose the existence of species (Mayr 1942). However, strict reproductive isolation such as

that described by the biological species concept often does not exist among even very distantly related taxa and as such may lead to the inappropriate lumping of otherwise distinct evolutionary units. Conversely, oversplitting of taxa often results when one adheres to the phylogenetic species concept which does not include reproductive isolation as one of the main criteria used in defining species (Cracraft 1983; Brabury *et al.* 2012). The recognition of subspecies presents a safe middle ground such that two taxa are not wrongfully grouped as a result of incomplete reproductive isolation, but are also not completely split into distinct species by meeting the less stringent criteria of the phylogenetic species concept. The killifish, *Fundulus heteroclitus,* provides an excellent example of a once continuous group that diverged for some time in allopatry and upon secondary contact some form(s) of reproductive isolation between the two divergent forms (now recognized as subspecies) exist. While at this time the reproductive isolation between these two subspecies is incomplete, the number of divergent traits known to among northern and southern forms definitely warrants the recognition of two subspecies. It is the recognition of two subspecies, initially formalized by Able & Felley (1986), that prompted me to address the questions raised in my Ph.D. dissertation. Thus, while some have recommended that the taxonomic unit of subspecies be abolished (Wilson & Brown 1953; Gillham 1956), as seen here it is still an informative measure for recognizing highly divergent yet not completely reproductively isolated taxa.

While the recognition of separate subspecies of killifish has been supported here, the precise location of the contact zone, while currently appearing to be centred around the mitochondrial cline centre in central New Jersey, may not be stable as a result of the impact of global warming. In their recent study, Hibish *et al.* (2012) found that the current location of a hybrid zone between two species of marine mussel (*Mytilus galloprovincialis* and *M. edulis*), had

shifted approximately 100 km when compared to data sets from almost ten (Bierne *et al*. 2003) to more than 20 years ago, corresponding to hybrid zone movement of between 5 - 10 km per year (Coustau *et al.* 1991). These results were consistent with the authors' predictions, that as a result of global warming, hybrid zones between warm- and cold-adapted species will tend to shift into the territory of the cold-adapted species. While not as dramatic, we too see possible evidence of a shift in the *Fundulus heteroclitus* hybrid zone into the territory of the northern subspecies upon inspection of the cline in allele frequencies at the LDH-B locus. Data from Powers & Place (1978) yielded a cline centre of 1298 km from Sapelo Island, Georgia while our more recent data shows a cline centre of 1328 km. While these values are not significantly different from one another they may have significance in the context of global warming. The larger shift seen in the mussel hybrid zone can be explained by higher dispersal levels for these broadcast spawners (larval dispersal ~ 30 - 50 km) while the smaller shift seen among the killifish may be due to the reduced movement of their non-dispersive larvae. Thus, future hybrid zone studies should attempt to perform genetic assays utilizing at least the same markers as past studies in order to perform comparisons of cline shapes across decades in order to assess how global warming may be affecting the distributions of particular hybridizing taxa.

There is also the potential for ecological speciation among subspecies of *Fundulus heteroclitus* and such potential also exists for the multitude of divergent taxa that find themselves distributed along an environmental gradient such as the one presented by temperature along the Atlantic Coast of North America in this case. As a result of their adaptation to different thermal regimes, the two subspecies of killifish not only exhibit significant differences with respect to thermal tolerance, but also with respect to metabolism, a trait that is directly affected by temperature especially among poikilotherms (Keller & Seehausen 2012). Thus, in accordance

with the definition presented by Rundle and Nosil (2005), all that is missing then, to implicate ecological speciation in the maintenance and further divergence of these two subspecies is a genetic mechanism connecting the 'ecological source of divergent selection' (in this case temperature) and 'the form of reproductive isolation' (potentially assortative mating). The increase in metabolic rate observed among killifish of the northern subspecies is thought to be related to a reduced growing season experienced by northern populations and in turn leads to this subspecies being significantly smaller when compared to members of the southern subspecies, which have slower metabolic rates and reside in an area where the growing season is longer. If an association were to be found between size and positive assortative mating (as previously discussed) then substantial evidence for ecological speciation would be found to exist among the subspecies of *Fundulus heteroclitus*. This final link is one that is missing in the majority of studies assessing ecological speciation among subspecies showing divergent thermal adaptation (Keller and Seehausen 2012). Thus, future studies, involving not only killifish but other subspecies whose distribution spans an environmental gradient, should focus on establishing a link between the environmentally associated divergent trait and assortative mating in an attempt to clarify the traits involved in mate choice and how they relate to adaptation to divergent environments.

Lastly, and also closely associated to adaptive divergence with respect to thermal regime, is divergence of mitochondrial properties. As the 'powerhouse of the cell' the regulation of all cellular processes and ultimately the well-being of the entire animal depend on proper mitochondrial function. A booming area of research, up until recently left largely unexplored, is the association between sex, mitochondrial type, and fitness (Ballard & Melvin 2012). Recently, Innocenti *et al.* (2012) found that mutations in the mitochondrial genome were more likely to

have negative consequences in tissues active only in male *Drosophila melanogaster* (i.e. testes, ejaculatory duct), slipping through a "selective sieve" that would otherwise capture mtDNA mutations adversely affecting female fitness as they are usually (but not always) the sole contributors of the mitochondrial genome to progeny. Thus, any negative fitness effects associated with the mitochondrial genome are only screened with respect to female fitness. While we do not know if either sex is suffering a decrease in fitness due to mutation load in the mitochondrial genome, we do have evidence that sex and mitochondrial type are interacting in a intriguing way in *F. heteroclitus*. Future studies wherein mitochondrial typing has been performed and sexing of individuals was possible should look for preliminary evidence of an association between sex and mitochondrial type. Such a simple inspection could lead to a wide range of new and largely unexplored area of research for a wide range of organisms.

The results presented in my thesis clearly indicate that some combination of pre- and/or post-zygotic reproductive barriers are operating in the maintenance of the hybrid zone and thus the two subspecies, of the Atlantic killifish, *Fundulus heteroclitus*. The presence of a trimodal hybrid zone coupled with high levels of heterozygote deficit and cytonuclear disequilibrium provide the first line of evidence in support of this. The results of the clinal analyses indicated that steep clines in both nuclear and mitochondrial allele frequencies imply the action of endogenous selection and cytonuclear epistasis in the maintenance of this zone. I also found evidence that positive assortative mating, which appears strong among fish collected from allopatric populations, may be playing a role in the maintenance of the hybrid zone. Taken together, the results of my thesis provide the foundation for future studies investigating even further the roles of pre-zygotic and post-zygotic reproductive barriers as they present themselves

within the hybrid zone, the interaction between sex and mitochondrial type and environment, and the physiological consequences of being a recombinant type with respect to fitness.

References

Able KW (1984) Variation in spawning site selection of the mummichog, *Fundulus heteroclitus*. *Copeia*, **2**, 522-525.

Able KW, Castagna M (1975) Aspects of an undescribed reproductive behavior in *Fundulus heteroclitus* (Pisces: Cyprinodontae) from Virginia. *Chesapeake Science*, **16**, 282-284.

Able KW, Felley JD (1986) Geographical variation in *Fundulus heteroclitus*: Tests for concordance between egg and adult morphologies. *American Zoologist*, **26**, 145-157.

Able KW, Palmer RE (1988) Salinity effects on fertilization success and larval mortality of *Fundulus heteroclitus*. *Copeia*, **2**, 345-350.

Able KW, Vivian DN, Petruzzelli G (2012) Connectivity Among Salt Marsh Subhabitats: Residency and Movements of the Mummichog (*Fundulus heteroclitus*). *Estuaries and Coasts*, **35**, 743-753.

Abraham, BJ (1985) Species Profiles: Life histories and environmental requirements of coastal fishes and Invertebrates (Mid-Atlantic): Mummichog and striped killifish. U.S. Fish Wildl Serv Biol Rep 82(11.40). U.S. Army Corps of Engineers, TR EL-82-4. 23 pp.

Adams SM, Lindmeier JB, Duvernell DD (2006) Microsatellite analysis of the phylogeography, Pleistocene history and secondary contact hypotheses for the killifish, *Fundulus heteroclitus*. *Molecular Ecology*, **15**, 1109-1123.

Adams SM, Oleksiak MF, Duvernell DD (2005) Microsatellite primers for the Atlantic coastal killifish, *Fundulus heteroclitus*, with applicability to related *Fundulus* species. *Molecular Ecology Notes*, **5,** 275-277.

Armstrong PB, Child JS (1965) Stages in the normal development of *Fundulus heteroclitus.* *Biological Bulletin*, **128**, 143-168.

Arnqvist G, Dowling DK, Eady P, Gay L, Tregenza T, Tuda M, Hosken DJ (2010) Genetic architecture of metabolic rate: environment specific epistasis between mitochondrial and nuclear genes in an insect. *Evolution*, **64**, 3354-3363.

Asmussen MA, Arnold J, Avise JC (1987) Definition and properties of disequilibrium statistics for associations between nuclear and cytoplasmic genotypes. *Genetics*, **115**, 755-768.

Asmussen MA, Basten CJ (1996) Constraints and normalized measures for cytonuclear disequilibria. *Heredity*, **76**, 207-214.

Avise JC (2000) Cytonuclear genetic signatures of hybridization phenomena: Rationale, utility, and empirical examples from fishes and other aquatic animals. *Reviews in Fish Biology and Fisheries,* **10**, 253-263.

Avise JC (2004) *Molecular Markers, Natural History, and Evolution*, 2nd edn. Sinauer Associates Sunderland, Massachusetts.

Ballard JWO, Melvin RG (2010) Linking the mitochondrial genotype to the organismal phenotype. *Molecular Ecology*, **19**, 1523-1539.

Barton NH (2001) The role of hybridization in evolution. *Molecular Ecology,* **10**, 551-568.

Barton NH, Gale KS (1993) Genetic analysis of hybrid zones. In Harrison, R. G. (ed) Hybrid Zones and the Evolutionary Process. Oxford University Press, New York, pp 13-45.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology, Evolution, and Systematics,* **16**, 113-148.

Barton NH, Hewitt GM (1989) Adaptation, speciation, and hybrid zones. *Nature,* **341**, 497-503.

Basten CJ, Asmussen MA (1997) The exact test for cytonuclear disequilibria. *Genetics,* **146,** 1165-1171.

Becker, W. M., L. J. Kleinsmith, J. Hardin. 2003. The World of the Cell, 5th edition. Benjamin Cummings, San Francisco.

Belkhir K, Borsa P, Goudet J, Bonhomme F (1999) *GENETIX 4.04.* Université de Montpellier II, Montpellier, France.

Bernardi G, Sordino P, Powers DA (1993) Concordant mitochondrial and nuclear-DNA phylogenies for populations of the teleost fish *Fundulus heteroclitus*. *Proceedings of the National Academy of Sciences,* USA, **90,** 9271 - 9274.

Bierne N, Borsa P, Daguin C, Jollivet D, Viard F, Bonhomme F, David P (2003) Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. *Molecular Ecology*, **12**, 447-461.

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology,* **20**, 2044-2072.

Bland JM, Altman DG (1995) Multiple significance tests: the Bonferroni method. *British Medical Journal*, **310**, 170.

Blier PU, Dufresne F, Burton RS (2001) Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation. *Trends in Genetics,* **17**, 400-406.

Boratynski Z, Alves PC, Berto S, Koskela E, Mappes T, Melo-Ferreira J (2011) Introgression of mitochondrial DNA among Myodes voles: consequences for energetics? *BMC Evolutionary Biology*, **11**.

Braby MF, Eastwood R, Murray N (2012) The subspecies concept in butterflies: has its application in taxonomy and conservation biology outlived its usefulness? *Biological Journal of the Linnean Society*, **106**, 699-716.

Bridle JR, Saldamando CI, Koning W, Butlin RK (2006) Assortative preferences and discrimination by females against hybrid male song in the grasshoppers *Chorthippus brunneus* and *Chorthippus jacobsi* (Orthoptera: Acrididae). *Journal of Evolutionary Biology*, **19,** 1248–1256.

Brodin A, Haas F (2009) Hybrid zone maintenance by non-adaptive mate choice. *Evolutionary Ecology*, **23**:17–29.

Brown WL Jr., Wilson EO (1956) Character displacement. *Systematic Zoology*, **5**, 49-64.

Brumfield RT, Jernigan RW, McDonald DB, Braun MJ (2001) Evolutionary implications of divergent clines in an avian (*Mancus*: Aves) hybrid zone. *Evolution*, **55**, 2070 - 2087.

Brummett AR (1966) Observations on the eggs and breeding season of *Fundulus heteroclitus* at Beaufort, North Carolina. *Copeia*, **3**, 616-620.

Brummett AR, Dumont JN (1981) A comparison of chorions from eggs of Northern and Southern populations of *Fundulus heteroclitus*. *Copeia*, **3**, 607-614.

Buerkle CA (2005) Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes*, **5**, 684 - 687.

Burgarella C, Lorenzo Z, Jabbour-Zahab R, Lumaret R, Guichoux E, Petit RJ, Soto Á, Gil L (2009) Detection of hybrids in nature: application to oaks (*Quercus suber* and *Q. ilex*). *Heredity*, **102**, 442 - 452.

Burnett KG, 26 other authors (2007) *Fundulus* as the premier teleost model in environmental biology: Opportunities for new insights using genomics. *Comparative Biochemistry and Physiology Part D*, **2**, 257-286.

Burton RS, Ellison CK, Harrison JS (2006) The sorry state of F_2 hybrids: consequences of rapid mitochondrial DNA evolution in allopatric populations. *American Naturalist*, **168,** S14-S24.

Cashon RE, Van Beneden RJ, Powers DA (1981) Biochemical genetics of *Fundulus heteroclitus* (L.). IV. Spatial variation in gene frequencies of Idh-A, Idh-B, 6-Pgdh-A, and Est-S. *Biochemical Genetics*, **19**, 715-728.

Cheviron ZA, Brumfield RT (2009) Migration-selection balance and local adaptation of mitochondrial haplotypes in rufous-collared sparrows (*Zonotrichia capensis*) along an elevational gradient. *Evolution*, **63**, 1593-1605.

Cockerham CC, Weir BS (1977) Quadratic analyses of reciprocal crosses. *Biometrics*, **33**, 187-203.

Conover DO, Schultz ET (1995) Phenotypic similarity and the evolutionary significance of countergradient variation. *Trends in Ecology and Evolution*, **10**, 248-252.

Coustau C, Renaud F, Delay B (1991) Genetic characterization of the hybridization between *Mytilus edulis* and *M. galloprovincialis* on the Atlantic Coast of France. *Marine Biology*, **111**, 87-93.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, Massachusetts.

Cracraft J (1983) Species concepts and speciation analysis. *Current Ornithology*, **1**, 159-187.

Crawford DL, Constantino HR, Powers DA (1987) Lactate Dehydrogenase-B cDNA from the teleost *Fundulus heteroclitus*: Evolutionary implications. *Molecular Biology and Evolution*, **6**, 369-383.

Crawford DL, Powers DA (1992) Evolutionary adaptation to different thermal environments via transcriptional regulation. *Molecular Biology and Evolution*, **9**, 806-813.

Crawford DL, Powers DA (1989) Molecular basis of evolutionary adaptation at the lactate dehydrogenase-B locus in the fish *Fundulus heteroclitus*. *Proceedings of the National Academy of Science, USA*, **86**, 9365-9369.

Dalziel AC, Vines TH, Schulte PM (2012) Reductions in prolonged swimming capacity following freshwater colonization in multiple Threespine Stickleback populations. *Evolution*, **66**, 1226-1239.

DeKoning ABL, Picard DJ, Bond SR, Schulte PM (2004) Stress and interpopulation variation in glycolytic enzyme activity and expression in a teleost fish *Fundulus heteroclitus*. *Physiological and Biochemical Zoology*, **77**, 18-26.

DeWan A, Klein RJ, Hoh J (2007) Linkage disequilibrium mapping for complex disease genes. In: *Linkage disequilibrium and association mapping: analysis and applications* (ed. Collins AR), pp. 85-108. Humana Press, New Jersey.

Dhillon RS, Schulte PM (2011) Intraspecific variation in the thermal plasticity of mitochondria in killifish. *The Journal of Experimental Biology*, **214**, 3639-3648.

DiMichele L, Paynter KT, Powers DA (1991) Evidence of lactate dehydrogenase-B allozyme effects in the teleost, *Fundulus heteroclitus*. *Science*, **253**, 898-900.

DiMichele L, Powers DA (1982a) Physiological basis for critical swimming endurance differences between LDH-B genotypes of *Fundulus heteroclitus*. *Science,* **216**, 1014-1016.

DiMichele L, Powers DA (1982b) Ldh-B genotype-specific hatching times of *Fundulus heteroclitus* embryos. *Nature,* **296**, 563-564.

DiMichele L, Powers DA (1984) Developmental and oxygen consumption rate differences between lactate dehydrogenase-B genotypes of *Fundulus heteroclitus* and their effect on hatching time. *Physiological Zoology*, **57**, 52-56.

DiMichele L, Powers DA (1991) Allozyme variation, developmental rate, and differential mortality in the teleost *Fundulus heteroclitus*. *Physiological Zoology*, **64**, 1426-1443.

DiMichele L, Powers DA, DiMichele JA (1986) Developmental and physiological consequences of genetic variation at enzyme synthesizing loci in *Fundulus heteroclitus*. *American Zoologist*, **26**, 201-208.

DiMichele L, Westerman ME (1997) Geographic variation in development rate between populations of the teleost *Fundulus heteroclitus*. *Marine Biology*, **128**, 1-7.

Dobzhansky T (1940) Speciation as a stage in evolutionary divergence. *American Naturalist*, **74,** 312-321.

Duchesne P, Godbout MH, Bernatchez L (2002) PAPA (package for the analysis of parental allocation): a computer program for simulated and real parental allocation. *Molecular Ecology Notes*, **2**, 191-193.

Duvernell DD, Lindmeier JB, Faust KE, Whitehead A (2008) Relative influences of historical and contemporary forces shaping the distribution of genetic variation in the Atlantic killifish, *Fundulus heteroclitus*. *Molecular Ecology*, **17**, 1344 - 1360.

Edmands S, Burton RS (1999) Cytochrome c oxidase activity in interpopulation hybrids of a marine copepod: a test for nuclear-nuclear or nuclear-cytoplasmic coadaptation. *Evolution,* **53**, 1972-1978.

Ellison CK, Burton RS (2008) Gentotype-dependent variation of mitochondrial transcriptional profiles in interpopulation hybrids. *Proceedings of the National Academy of Science, USA,* **105,** 15831-15836.

Endler JA (1977) Geographic variation, speciation, and clines. Princeton Academic Press, Princeton.

Endler JA (1982) Problems in distinguishing historical from ecological factors in biogeography. *American Zoologist*, **22**, 441-452.

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

Fangue NA, Hofmeister M, Schulte PM (2006) Intraspecific variation in thermal tolerance and heat shock protein gene expression in common killifish, *Funudulus heteroclits*. *The Journal of Experimental Biology*, **209**, 2859-2872.

Fangue NA, Mandic M, Richards JG, Schulte PM (2008) Swimming performance and energetics as a function of temperature in killifish *Fundulus heteroclitus*. *Physiological and Biochemical Zoology*, **81**, 389-401.

Fangue NA, Podrabsky JE, Crawshaw LI, Schulte PM (2009a) Countergradient variation in temperature preference in populations of killifish *Fundlulus heteroclitus. Physiological and Biochemical Zoology,* **82**, 776-786.

Fangue NA, Richards JG, Schulte PM (2009b) Do mitochondrial properties explain intraspecific variation in thermal tolerance? *Journal of Experimental Biology*, **212**, 514-522.

Flight PA, Nacci D, Champlin D, Whitehead A, Rand DM (2011) The effects of mitochondrial genotype on hypoxic survival and gene expression in a hybrid population of the killifish, *Fundulus heteroclitus*. *Molecular Ecology,* **20**, 4503-4520.

Fraley C, Raftery AE, Murphy TB, Scrucca L (2012) mclust Version 4 for R: Normal mixture modelling for model-based clustering, clarification, and density estimation. *Technical Report 597,* Department of Statistics, University of Washington.

Freeman S, Herron JC (2004) *Evolutionary Analysis*, 3rd edn. Pearson/Prentice Hall, New Jersey.

Fritz ES, Meredith WH, Lotrich VA (1975) Fall and winter movements and activity level of the mummichog, *Fundulus heteroclitus*, in a tidal creek. *Cheasapeake Science,* **16,** 211-215.

Gay L, Crochet P, Bell D, Lenormand T (2008) Comparing genetic and phenotypic clines in hybrid zones: a window on tension zone models. *Evolution*, **62**, 2789-2806.

Geyer LB, Palumbi SR (2003) Reproductive character displacement and the genetics of gamete recognition in tropical sea urchins. *Evolution*, **57**, 1049-1060.

Gillham NW (1956) Geographic variation and the subspecies concept in butterflies. *Systematic Zoology,* **5**, 110-120.

Gompert Z, Buerkle CA (2009) A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology*, **18**, 1207-1224.

Gompert Z, Buerkle CA (2010) INTROGRESS: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, **10**, 378- 384.

Gonzáles-Villasenor LI, Powers DA (1990) Mitochondrial-DNA restriction-site polymorphism in the teleost *Fundulus heteroclitus* support secondary intergradtion. *Evolution*, **44**, 27-37.

Haakons, KL (2010) Oxygen limited thermal tolerance in hybrid killifish, *Fundulus heteroclitus*. MSc Thesis, University of British Columbia.

Haldane JBS (1948) The theory of a cline. *Journal of Genetics*, **7**, 277-284.

Hardy JD Jr. (1978) Development of fishes of the mid-Atlantic bight: An atlas of egg, larval and juvenile stages. In Volume II Anguillidae through Syngnathidae, U.S. Department of the Interior, pp 141-200.

Healy TM, Schulte PM (2012) Thermal acclimation is not necessary to maintain a wide thermal breadth of aerobic scope in the Common Killifish (*Fundulus heteroclitus*). *Physiological and Biochemical Zoology*,**85**, 107-119.

Healy TM, Tymchuk WE, Osborne EJ, Schulte PM (2010) Heat shock response of killifish (*Fundulus heteroclitus*) candidate gene and heterologous microarray approaches. *Physiological Genomics*, **41**, 171-184.

Hewitt GM (1988) Hybrid zones – Natural laboratories for evolutionary studies. *Trends in Ecology and Evolution*,**3**, 158-167.

Hilbish TJ, Lima FP, Brannock PM, Fly EK, Rognstad RL, Wethey DS (2012) Change and stasis in marine hybrid zones in response to climate warming. *Journal of Biogeography*, **39**, 676-687.

Innocenti P, Morrow EH, Dowling DK (2011) Experimental evidence supports a sex-specific selective sieve in mitochondrial genome evolution. *Science*, **332**, 845-848.

Jiggins CD, Mallet J (2000) Bimodal Hybrid Zones and Speciation. *Trends in Ecology and Evolution*, **15**, 250-255.

Keller I, Seehausen O (2012) Thermal adaptation and ecological speciation. *Molecular Ecology*, **21**, 782-799.

Lee W, Conroy J, Howell WH, Kocher TD (1995) Structure and evolution of teleost mitochondrial control regions. *Journal of Molecular Evolution*, **41**, 54-66.

Levitan DR, Sewall MA, Chia FS (1991) Kinetic of fertilization in the sea urchin *Strongylocentrotus franciscanus*: Interaction of gamete dilution, age, and contact time. *Biological Bulletin*, **181**, 371-378.

Levitan DR (2002) The relationship between conspecific fertilization success and reproductive isolation among three congeneric sea urchins. *Evolution*, **56,** 1599-1609.

Liepins A, Hennen S (1977) Cytochrome oxidase deficiency during development of amphibian nucleocytoplasmic hybrids. *Developmental Biology,* **57**: 284-292.

Lotrich VA (1975) Summer home range and movements of *Fundulus heteroclitus* (Pisces: Cyprinodontidae) in a tidal creek. *Ecology,* **56**, 191-198.

Mallet J, Barton N (1989) Inference from clines stabilized by frequency-dependent selection. *Genetics,* **122**, 967-976.

Marteinsdottir G, Able KW (1988) Geographic variation in egg size among populations of the Mummichog, *Fundulus heteroclitus*. *Copeia*, **2,** 471-478.

Marteinsdottir G, Able KW (1992) Influence of egg size on embryos and larvae of *Fundulus heteroclitus* (L.). *Journal of Fish Biology,* **41**, 883-896.

Mayr E (1942) *Systematics and the origin of species*. New York, NY: Columbia University Press.

Mayr E (1970) Populations, species, and evolution. Oxford University Press, London.

Moore WS, Price JT (1993) Nature of selection in the Northern Flicker hybrid zone and its implications for speciation theory. Genetic analysis of hybrid zones. In Harrison, R. G. (ed) Hybrid Zones and the Evolutionary Process. Oxford University Press, New York, pp 196-225.

Morin RP, Able KW (1983) Patterns of geographic variation in the egg morphology of the Fundulid fish, *Fundulus heteroclitus*. *Copeia,* **3,** 726-740.

Nagel L, Schluter D (1998) Body size, natural selection, and speciation in sticklebacks. *Evolution*, **52**, 209-218.

Newman HH (1907) Spawning behaviour and sexual dimorphism in *Fundulus heteroclitus* and allied fish. *Biological Bulletin,* **12**, 314-348.

Nicolae DL, Kurz T, Ober C (2009) Identifying disease genes in association studies. In: *Medical uses of Statistics,* 3rd edn (ed. Bailar JC, Hoaglin DC), pp. 455-466. John Wiley and Sons, New Jersey.

Niehuis O, Judson AK, Gadau J (2008) Cytonuclear genic incompatibilities cause increased mortality in male F2 hybrids of *Nasonia giraulti* and *N. vitripennis*. *Genetics,* **178,** 413-426.

Oleksiak MF, Roach JL, Crawford DL (2005) Natural variation in cardiac metabolism and gene expression in *Fundulus heteroclitus*. *Nature Genetics,* **37**, 67-72.

Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecological Systems*, **25**, 547-72.

Palumbi SR (2009) Speciation and the evolution of gamete recognition genes: Pattern and process. *Heredity*, **102**, 66-76.

Palumbi SR and Metz EC (1991) Strong reproductive isoation between closely related tropical sea urchins (genus *Echinometra*). *Molecular Biology and Evolution*, **8**, 227-239.

Paynter KT, DiMichele L, Hand SC, Powers DA (1991) Metabolic implications of LDH-B genotype during early development in *Fundulus heteroclitus*. *Journal of Experimental Zoology*, **257**, 24-33.

Picard DJ, Schulte PM (2004) Variation in gene expression in response to stress in two population of *Fundulus heteroclitus*. *Comaprative Biochemistry and Physiology Part A*, **137**, 205-216.

Pierce VA, Crawford DL (1997) Phylogenetic analysis of glycolytic enzyme expression. *Science*, **276**, 256-259.

Pike N (2010) Using false discovery rate for multiple comparisons in ecology and evolution. *Methods in Ecology and Evolution*, **2**, 278-282.

Place AR, Powers DA (1979) Genetic variation and relative catalytic efficiencies: Lactate dehydrogenase B allozymes of *Fundulus heteroclitus*. *Proceedings of the National Academy of Science, USA*, **76**, 2354-2358.

Place AR, Powers DA (1984a) Kinetic characterization of the lactate dehydrogenase (LDH-B_4) allozymes of *Fundulus heteroclitus*. *The Journal of Biological Chemistry*, **259**, 1309-1318.

Place AR, Powers DA (1984b) Purification and characterization of the lactate dehydrogenase (LDH-B_4) allozymes of *Fundulus heteroclitus*. *The Journal of Biological Chemistry*, **259**, 1299-1308.

Podrabsky JE, Javillonar C, Hand SC, Crawford DL (2000) Intraspecific variation in aerobic metabloism and glycolytic enzyme expression in heart ventricles. *American Journal of Physiology*, **279**, R2344-R2348.

Porter AH, Wenger R, Geiger H, Shapiro AM (1997) The Pontia daplidice-edusa hybrid zone in Northwestern Italy. *Evolution*, **51**, 1561 - 1573.

Powers DA, Greaney GS, Place AR (1979) Physiological correlation between lactate dehydrogenase genotype and haemoglobin in killifish. *Nature*, **277**, 240-241.

Powers DA, Lauerman T, Crawford D, DiMichele L (1991) Genetic mechanisms for adapting to a changing environment. *Annual Review of Genetics*, **25**, 629-659.

Powers DA, Place AR (1978) Biochemical genetics of *Fundulus heteroclitus* (L.). I. Temporal and spatial variation in gene frequencies of Ldh-B, Mdh-A, Gpi-B, and Pgm-A. *Biochemical Genetics*, **16**, 593-607.

Powers DA, Ropson I, Brown DC, Van Beneden R, Cashon R, Gonzalez-Villasenor LI, DiMichele JA (1986) Genetic variation in *Fundulus heteroclitus*: Geographic distribution. *American Zoologist*, **26**, 131-144.

Powers DA, Schulte PM (1998) Evolutionary adaptation of gene structure and expression in natural populations in relation to a changing environment: A multidisciplinary approach to address a million year saga of a small fish. *Journal of Experimental Zoology*, **282**, 71-94.

R Development Core Team (2008) A language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.

Rand DM, Clark AG, Kann LM (2001) Sexually antagonistic cytonuclear fitness interactions in *Drosophila melanogaster. Genetics*, **159**, 173-187.

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

Reynolds KE (1994) Variation in the growth rates of young-of-the-year *Fundulus heteroclitus* from different latitudes. MS Thesis, State University of New York, Stony Brook.

Ropson IJ, Brown DC, Powers DA (1990) Biochemical genetics of *Fundulus heteroclitus* (L.). VI. Geographical variation in the gene frequencies of 15 loci. *Evolution*, **44**, 16-26.

Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters*, **8**, 336-352.

Schulte P M, Gomez-Chiarri MC, Powers DA (1997) Variation in the structure and function of the promoter and 5' flanking region of the Ldh-B locus between populations of the teleost *Fundulus heteroclitus*. *Genetics*, **145,** 759-769.

Schulte PM, Glémet HC, Fiebig AA, Powers DA (2000) Adaptive variation in lactate dehydrogenase-B gene expression: Role of a stress-responsive regulatory element. *Proceedings of the National Academy of Science, USA*, **97**, 6597-6602.

Schultz ET, Reynolds KE, Conover DO (1996) Countergradient variation in growth among newly hatched *Fundulus heteroclitus*: Geographic differences revealed by common-environment experiments. *Functional Ecology*, **10**, 366-374.

Scott GR, Rogers JT, Richards JG, Wood CM, Schulte PM (2004) Intraspecific divergence of ionoregulatory physiology in the euryhaline teleost *Fundulus heteroclitus*: possible mechanisms of freshwater adaptation. *The Journal of Experimental Biology*, **207**, 3399-3410.

Scott GR, Schulte PM (2005) Intraspecific variation in gene expression after seawater transfer in gills of the euryhaline killifish *Fundulus heteroclitus*. *Comparative Biochemistry and Physiology, Part A,* **141**, 176-182.

Segal JA, Schulte PM, Powers DA, Crawford DL (1996) Descriptive and functional characterization of variation in the *Fundulus heteroclitus* Ldh-B proximal promoter. *The Journal of Experimental Zoology*, **275**, 355-364.

Servedio MR, Noor MAF (2003) The role of reinforcement in speciation: theory and data. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 339-364.

Sha Z, Xu P, Takano T, Liu H, Terhune J, Liu Z (2008) The warm temperature acclimation protein *Wap65* as an immune response gene: Its duplicates are differentially regulated by temperature and bacterial infections. *Molecular Immunology*, **45**, 458-1469.

Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R, Ferrell RE (1997) Ethnic-affiliation estimation by use of population-specific DNA markers. *The American Journal of Human Genetics*, **60**, 957 - 964.

Singhal S, Moritz C (2012) Strong selection against hybrids maintains a narrow contact zone between morphologically cryptic lineages in a rainforest lizard. *Evolution*, 66, 1474-1489.

Sites, J. W. Jr., C. J. Basten, and M. A. Asmussen. 1996. Cytonuclear genetic structure of a hybrid zone in lizards f the *Sceloporus grammicus* complex (Sauria, Phrynosomatidae). *Molecular Ecology,* **5,** 379-392.

Slatkin M (1973) Gene flow and selection in a cline. *Genetics*, **75**, 733-756.

Smith MW, Chapman RW, Powers DA (1998) Mitochondrial DNA analysis of Atlantic Coast, Chesapeake Bay, and Delaware Bay populations of the teleost *Fundulus heteroclitus* indicates temporally unstable distributions over geologic time. *Molecular Marine Biology Biotechnology*, **7**, 79-87.

Sotka EE, Palumbi SR (2006) The use of genetic clines to estimate dispersal distances of marine larvae. *Ecology*, **87**, 1094-1103.

Soularue JP, Kremer A (2012) Assortative mating and gene flow generate clinal phenological variation in trees *BMC Evolutionary Biology*, **12**, 79.

Strand AE, Williams LM, Oleksiak MF, Sotka EE (2012) Can diversifying selectionbe distinguished from history in geographic clines? A population genomic study of killifish (*Funduls heteroclits*). *PLoS One*, **7.**

Szymura JM, Barton NH (1986) Genetic analysis of a hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*, near Cracow in Southern Poland. *Evolution*, **40**, 1141-1159.

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

Taylor MH (1986) Environmental and endocrine influences of reproduction of *Fundulus heteroclitus*. *American Zoologist*, **26**, 159-171.

Teeter KC, Payseur BA, Harris LW, Bakewell MA, Thibodeau LM, O'Brien JE, Krenz JG, San-Fuentes MA, Nachman MW, Tucker PK (2008) Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Research*, **18**, 67-76.

Toro JE, Thompson RJ, Innes DJ (2006) Fertilization success and early survival in pure and hybrid larvae of *Mytilus edulis* (Linnaeus, 1758) and *M. trossulus* (Gould, 1850) from laboratory crosses. *Aquatic Research*, **37**, 1703-1708.

Tranah GJ, Manini TM, Lohman KK, Nalls MA, Kritchevsky S, Newman AB, Harris TB, Miljkovic I, Biffi A, Cummings SR, Liu Y (2011) Mitochondrial DNA variation in human metabolic rate and energy expenditure. *Mitochondrion*, **6**, 855-861.

Van Beneden RJ, Powers DA (1989) Structural and Functional differentiation of two clinally distributed glucosephosphate isomerase allelic isozymes from the teleost *Fundulus heteroclitus*. *Molecular Biology and Evolution*, **6**, 155-170.

Van Oosterhout C, Hutchinson WF, Willis DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535-538.

Vincze T, Posfai J, Roberts RJ (2003) NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Research*, **31**, 3688-3691.

Wallace DC (2005) A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: A dawn for evolutionary medicine. *Annual Review of Genetics*, **39**, 359-407.

Wallace RA, Selman K (1978) Oogenesis in *Fundulus heteroclitus*. I. Preliminary observations on oocyte maturation *in vivo* and *in vitro*. *Developmental Biology*, **62**, 354-369.

Whitehead A (2010) The evolutionary radiation of diverse osmotolerant physiologies in killifish *(Fundulus sp.). Evolution*, **64**, 2070-2085.

Whitehead A, Crawford DL (2006) Neutral and adaptive variation in gene expression. *Proceedings of the National Academy of Science, USA,***103**, 5425-5430.

Whitehead A, Roach JL, Zhang S, Galvez F (2011) Genomic mechanisms of evolved physiological plasticity in killifish distributed along an environmental salinity gradient. *Proceedings of the National Academy of Sciences, USA,* **108**, 6193-6198.

Whitlock MC (1992) Temporal fluctuations in demographic parameters and the genetic variance among populations. *Evolution,* **46**, 608-615.

Whitlock MC, Schluter D (2009) *The Analysis of Biological Data*, 2nd edn. Roberts and Company Publishers, Colorado.

Williams, L. M., X. Ma, A.R. Boyko, C. D. Bustamante, and M. F. Oleksiak. 2010. SNP identification, verification, and utility for population genetics in a non-model genus. *BioMed Central Genetics*, **11**, 32-45.

Williamson EG, DiMichele L (1997) An ecological simulation reveals balancing selection acting on developmental rate in the teleost *Fundulus heteroclitus*. *Marine Biology,* **128**, 9-15.

Wilson EO, Brown WL (1953) The subspecies concept and its taxonomic application. *Systematic Zoology*, **2**, 97-111.

Appendices

Appendix A - Supplementary Table for Chapter One

Table A.1. Subspecific differences distinguishing *Fundulus heteroclitus macrolepidotus* (northerns) from *Fundulus heteroclitus heteroclitus* (southerns). Where separate studies are responsible for discovering a particular subspecific difference, the reference is given in the column of the subspecies for which the study was performed on. Mean ± standard error; means (standard deviation).

General Characteristic	*Fundulus heteroclitus macrolepidotus* (northerns)	*Fundulus heteroclitus heteroclitus* (southerns)	Reference
DISTRIBUTION	Connecticut north to Nova Scotia	Southern New Jersey south to northern Florida	Able & Felley 1986
	Upper, low salinity portions of Chesapeake and Delaware Bays	Lower reaches of Chesapeake and Delaware Bays	Morin & Able 1983; Able & Felley 1986
PHYSIOLOGICAL			
Ldh-B^bB^b phenotype critical swimming speed (body lengths/sec) at 10°C	4.3 ± 0.1	3.6 ± 0.12	DiMichele & Powers 1982a*
Respiration/body mass (pmol O_2/s*g); heart ventricle aerobic respiration at 20°C, 15ppt	4.09 ± 0.320	2.31 ± 0.229	Podrabsky *et al.* 2000
% survival after transfer from brackish water to freshwater	100	82	Scott *et al.* 2004
Relative (to brackish water controls) Na^+/K^+-ATPase α_{1a} mRNA expression 4 days after transfer from brackish to freshwater	6-fold increase	4-fold increase	Scott *et al.* 2004
Fold-changes of NKCC1(basolateral $Na^+/K^+/2Cl^-$ cotransporter) mRNA expression 14 days after transfer from brackish to freshwater	1.34 ± 0.21	0.59 ± 0.04	Scott *et al.* 2004
Fold-changes of CFTR (apical cystic fibrosis transmembrane conductance regulator Cl^- channel) mRNA expression 14 days after transfer from brackish to freshwater	0.96 ± 0.20	0.38 ± 0.08	Scott *et al.* 2004
Extrarenal clearance rate (index of paracellular permeability) of radiolabelled polyethylene glycol 8 days after transfer from brackish to freshwater (ml/kg*h)	1.3 ± 0.2	4.3 ± 0.3	Scott *et al.* 2004
Renal clearance rate (glomerular filtration rate) of radiolabelled	6.3 ± 0.8	3.9 ± 0.5	Scott *et al.* 2004

polyethylene glycol 8 days after transfer from brackish to freshwater (ml/kg*h)			
Urinary frequency 8 days after transfer from brackish to freshwater (bursts/hr)	1.1 ± 0.1	0.7 ± 0.1	Scott *et al*. 2004
Apical crypt density in gills after transfer from brackish to freshwater	15-fold decrease	3-fold decrease	Scott *et al*. 2004
Metabolic rates for glucose and fatty acid	Significantly higher	Lower	Oleksiak *et al.* 2005
Chronic thermal maximum	36.4°C	38.2°C	Fangue *et al.* 2006
Critical thermal maxima (CTMax) for fish acclimated to temperatures ranging from 2.3° - 34°C	Significantly lower	On average 1.5°C significantly higher	Fangue *et al.* 2006
Critical thermal minima (CTMin) for fish acclimated to temperatures >12.4°C	Significantly lower	On average 1.5°C significantly higher	Fangue *et al.* 2006
Range of critical swimming speeds (body lengths/sec) across acclimation temperatures from 5.2° - 32.4°C	~3.2 - 6.8	~1.8 - 5.3	Fangue *et al.* 2008
Mass specific metabolic rate of individuals possessing different mitochondrial types (μmolO2/g/hr)	9.77 ± 1.21	7.25 ± 0.95	Haakons 2010[δ]
ENDOCRINOLOGICAL			
Plasma cortisol levels 7 days after sham injection (ng/mL)	27.9 ± 4.1	146.53 ± 44.2	DeKoning *et al.* 2004
Plasma cortisol levels 7 days after injection of 400 μg cortisol per gram body mass (ng/mL)	749.6 ± 183.8	1 951.6 ± 658.1	DeKoning *et al.* 2004
Plasma cortisol levels 7 days after moderate handling stress (ng/mL)	368.5 ± 75.3	527.0 ± 68.1	DeKoning *et al.* 2004
Molecular response of gills to seawater transfer	Transient increase in Na^+/K^+-ATPase α_{1a} and NKCC1 expression, prolonged increase in CFTR expression	Prolonged increase in Na^+/K^+-ATPase α_{1a} and NKCC1 expression, transient increase in CFTR expression	Scott &Schulte 2005
Liver citrate synthase activity (units/mg protein) in fish acclimated to 5°C	0.06 ± 0.004	0.05 ± 0.003	Fangue *et al.* 2009a
Liver citrate synthase mRNA levels (relative to total RNA) in fish acclimated to 5°C	1.89 ± 0.352 ()	0.85 ± 0.136	Fangue *et al.* 2009a
Muscle citrate synthase activity (units/mg protein) in fish acclimated to 5°C	0.06 ±0.005	0.04± 0.002	Fangue *et al.* 2009a
Muscle citrate synthase mRNA levels (relative to total RNA) in fish acclimated to 5°C	0.84 ± 0.071	0.26 ± 0.053	Fangue *et al.* 2009a
Liver cytochrome *c* oxidase activity (units/mg protein) in fish acclimated to	0.25 ± 0.012	0.17 ± 0.017	Fangue *et al.* 2009a

5°C			
Muscle cytochrome *c* oxidase mRNA levels (relative to total RNA) in fish acclimated to 5°C	1.19 ± 0.222	0.65 ± 0.090	Fangue *et al.* 2009a
Liver citrate synthase mRNA levels (relative to total RNA) in fish acclimated to 15°C	1.09 ± 0.148	0.43 ± 0.112	Fangue *et al.* 2009a
Liver citrate synthase mRNA levels (relative to total RNA) in fish acclimated to 25°C	1.25 ± 0.086	0.35 ± 0.060	Fangue *et al.* 2009a
Muscle citrate synthase mRNA levels (relative to total RNA) in fish acclimated to 25°C	0.52 ± 0.093	0.16 ± 0.023	Fangue *et al.* 2009a
State III respiratory rates of mitochondria isolated from fish acclimated to 5° and 25°C and tested at 2° and 5°C	Significantly higher	Lower	Fangue *et al.* 2009a
State III respiratory rates of mitochondria isolated from fish acclimated to 5° and 15°C and tested at 25° and 30°C	Lower	Significantly higher	Fangue *et al.* 2009a
Whole organism metabolic rate	Significantly higher at all acclimation temperatures	Lower	Fangue *et al.* 2009a
Whole animal metabolic rate of fish acclimated to 5°C and acutely challenged at 5°C	Significantly higher	Lower	Fangue *et al.* 2009a
Routine metabolic rate in fish acclimated to 5°C	Significantly higher	Lower	Healy &Schulte 2012
Routine metabolic rate in fish acclimated to 15° and acutely exposed to 5° and 15°C	Significantly higher	Lower	Healy & Schulte 2012
Maximum metabolic rate in fish acclimated to 5°C	Significantly higher	Lower	Healy &Schulte 2012
Plasma lactate activity in fish acclimated to 33°C	12.30 ± 0.93	7.23 ± 0.90	Healy & Schulte 2012
Change in wet weight in fish acclimated to 25° and 30°C	Negative specific growth rate	Gradual increase with increasing temperature	Healy & Schulte 2012
BIOCHEMICAL			
Second order rate constant k_{cat}/K_m for Ldh-B and pH 6.50-8.00[§]	Ldh-B^bB^b phenotype has highest 10°C	Ldh-B^aB^a phenotype as highest catalytic efficiency at40°C	Place & Powers 1979[§]
ATP/Hb	Ldh-B^bB^b phenotype has high (2.11±0.22 ATP's/Hb)	Ldh-B^aB^a phenotype has low ATP/Hb (1.65±0.12 ATP's/Hb)	Powers *et al.* 1979[§]
hemoglobin-oxygen affinity	Ldh-B^bB^b phenotype low	Ldh-B^aB^a phenotype high	Powers *et al.* 1979[§]

P_{50} at 10°C in exercised fish	16.29 ± 0.79	12.19 ± 1.21	DiMichele & Powers 1982a*
V_{max}/K_m (substrate affinity)	Ldh-B^b allele higher at lower temperatures	Ldh-B^a allele higher at higher temperatures	Place &Powers 1984a[§]
Allozyme comparison	Ldh-B^b allozyme identical among individuals from Maine and Maryland	Ldh-B^a allozyme identical among individuals from Georgia and Maine	Place & Powers 1984b
Ldh-B allozyme stability when subjected to heat, urea, proteases (trypsin and chymotrypsin), and phenolglyoxal	Ldh-B^b more stable	Ldh-B^a allozyme less stable when subjected to denaturants	Place & Powers 1984b
Gpi-B stability when subjected to heat	Gpi-B^c allele less resistant to heat denaturation (T_{50} = 42.3°C)	Gpi-B^b allele more resistant to heat denaturation (T_{50} = 44.6°C)	Van Beneden & Powers 1989
Gpi-B activity after 2 hour incubation at 35°C	Gpi-B^c allele shows no detectable activity	Gpi-B^b allele virtually fully active	Van Beneden & Powers 1989
Inhibition of Gpi-B by 6-phosphogluconate at temperatures >30C°	Gpi-B^c allele sensitive to inhibition	Gpi-B^b allele less sensitive to inhibition at temperatures >30C°	Van Beneden & Powers 1989
Phosphofructokinase levels (µmol/min*mg protein) in heart ventricles of fish at 20°C, 15 ppt	2.681 ± 0.393	2.025 ± 0.332	Podrabsky *et al.* 2000
Aldolase (µmol/min*mg protein) in heart ventricles of fish at 20°C, 15 ppt	5.362 ± 0.708	3.913 ± 0.583	Podrabsky *et al.* 2000
Enolase (µmol/min*mg protein) in heart ventricles of fish at 20°C, 15 ppt	2.670 ± 0.178	3.055 ± 0.195	Podrabsky *et al.* 2000
Lactate dehydrogenase (µmol/min*mg protein) in heart ventricles of fish at 20°C, 15 ppt	93.54 ± 8.38	107.0 ± 18.03	Podrabsky *et al.* 2000
Phosphofructokinase activity (U/mg protein) in fish held at 20°C	Lower	Higher but decreases when exposed to stress	DeKoning *et al.* 2004
Aldolase activity (U/mg protein) in fish held at 20°C	Lower	Higher but decreases when exposed to stress	DeKoning *et al.* 2004
Lactate dehydrogenase activity (U/mg protein) in fish held at 20°C	Higher	Lower but increases when exposed to stress	DeKoning *et al.* 2004
Glucokinase mRNA expression	No change in expression in response to stress	Significantly higher in response to stress	Picard & Schulte 2004
Serine threonine kinase mRNA expression	No change in expression in response to stress	Significantly higher in response to stress	Picard & Schulte 2004
Muscle [glycogen] (µmol/g wet tissue)	2- to 4-fold higher	Significantly lower	Fangue *et al.*

in fish acclimated to 5° and 15°C, at rest	(~17 - 20)	(~5 - 6)	2008
Muscle [lactate] (µmol/g wet tissue) in fish acclimated to 5° and 15°C, after exercise	Significantly higher (~2.8 - 4.8)	Significantly lower (~0.9 - 1.1)	Fangue *et al.* 2008
Muscle [lactate] (µmol/g wet tissue) in fish acclimated 29°C, at rest	Significantly lower (~1)	Significantly higher (~2.5)	Fangue *et al.* 2008
Muscle [creatine phosphate] (µmol/g wet tissue) in fish acclimated to 5°C, at rest	10.48 ±1.15	17.30± 1.31	Fangue *et al.* 2008
Muscle [creatine phosphate] (µmol/g wet tissue) in fish acclimated to 5°C, after exercise	8.04 ±1.89	16.53± 2.65	Fangue *et al.* 2008
Liver [glycogen] (µmol/g wet tissue) in fish acclimated to 5°C, after exercise	703.46± 116.59	369.44 ±58.74	Fangue *et al.* 2008
Liver [glycogen] (µmol/g wet tissue) in fish acclimated to 15°C, at rest	793.94± 92.65	445.43± 13.83	Fangue *et al.* 2008
Muscle [glucose] (µmol/g wet tissue) in fish acclimated to 15°C, after exercise	1.93 ±0.04	1.37 ±0.09	Fangue *et al.* 2008
Muscle [glucose] (µmol/g wet tissue) in fish acclimated to 29°C, at rest	1.35 ±0.22	2.27 ± 0.16	Fangue *et al.* 2008
Muscle [total lipid] (µg/mg tissue) in fish acclimated to 15°C, after exercise	33.68 ± 5.44	20.76 ± 1.74	Fangue *et al.* 2008
Muscle [total lipid] (µg/mg tissue) in fish acclimated to 29°C, at rest	26.27 ± 2.46	18.14 ± 2.08	Fangue *et al.* 2008
Liver [glucose] (µmol/g wet tissue) in fish acclimated to 29°C, after exercise	10.98 ± 2.47	7.49 ± 1.87	Fangue *et al.* 2008
Resting muscle [glycogen] (µmol/g wet tissue) of fish acclimated to 23°C for 1 month, 2004 collections	Mean range 26.35 - 27.33	Mean range 14.62 - 19.98	Fangue *et al.* 2008
Resting liver [glucose] (µmol/g wet tissue) of fish acclimated to 23°C, 2004 collections	Mean range 4.40 - 4.74	Mean range 5.25 - 9.06	Fangue *et al.* 2008
Resting muscle [glycogen] of fish acclimated to 23°C for 2 months, 2004 collections	54.59 ± 3.36	25.10 ± 3.96	Fangue *et al.* 2008
Liver [glycogen] (µmol/g wet tissue) of fish acclimated to 23°C for 2 months following standardized exercise challenge, 2004 collections	613.66 ± 58.66	413.76 ± 48.84	Fangue *et al.* 2008
Resting liver [glucose] of fish acclimated to 23°C for 2 months, 2004 collections	6.18 ±1.08	2.72 ± 0.36	Fangue *et al.* 2008
Hsp70-1, hsp27, hsp30 expression in fish acclimated to 20°C and exposed to heat shock		Up-regulated to significantly greater extent	Healy *et al.* 2010
Hsp70-2 and Hsp90α expression in fish acclimated to 20°C and exposed to heat shock	Up-regulated to significantly greater extent		Healy *et al.* 2010

Parvalbumin-_ α expression	1.485	0.47	Healy *et al.* 2010
Parvalbumin-β expression	1.454	0.375	Healy *et al.* 2010
SERCA2 expression	0.8	1.285	Healy *et al.* 2010
Troponin I fast isoform expression	0.545	1.46	Healy *et al.* 2010
Troponin I slow isoform expression	0.51	1.44	Healy *et al.* 2010
Troponin T slow skeletal expression	0.595	1.415	Healy *et al.* 2010
Myosin light chain 1 expression	0.62	1.67	Healy *et al.* 2010
Myosin light chain 2 expression	0.57	1.425	Healy *et al.* 2010
Myosin heavy chain (slow isoform) expression	0.405	1.54	Healy *et al.* 2010
Creatine kinase (mitochondrial) expression	0.67	1.59	Healy *et al.* 2010
Creatine kinase muscle B chain expression	0.735	1.485	Healy *et al.* 2010
Creatine kinase M chain expression	0.815	1.4	Healy *et al.* 2010
Fructose-1,6-bisphosphatase isozyme 2 expression	1.23	0.55	Healy *et al.* 2010
L-xylulose reductase expression	1.505	0.59	Healy *et al*. 2010
Isocitrate dehydrogenase (mitochondrial) expression	0.785	1.305	Healy *et al*. 2010
Cytochrome c, somatic expression	1.145	0.625	Healy *et al*. 2010
Cytochrome c oxidase subunit 1 expression	0.685	1.14	Healy *et al.* 2010
Cytochrome c oxidase subunit 2 expression	1.15	0.61	Healy *et al.* 2010
Phosphate carrier protein (mitochondrial) expression	0.84	1.435	Healy *et al.* 2010
Ubiquinol-cytochrome c reductase complex core protein 1 expression	0.8	1.645	Healy *et al.* 2010
NADP transhydrogenase (mitochondrial) expression	0.665	1.41	Healy *et al*. 2010
*EF-1*α expression	0.815	1.35	Healy *et al.* 2010
Collagen α2 expression	1.225	0.61	Healy *et al.* 2010
Collagen α1 expression	1.17	0.615	Healy *et al*. 2010
Desmin expression	0.58	1.6	Healy *et al.* 2010
Vimentin expression	0.53	1.585	Healy *et al.* 2010
Golgi vesicular membrane trafficking protein expression	1.535	0.34	Healy *et al.* 2010
Crystallin expression	1.48	0.655	Healy *et al.* 2010
Fatty acid-binding protein expression	1.385	0.41	Healy *et al*. 2010
Apolipoprotein A-IV expression	1.255	0.705	Healy *et al*. 2010
Skeletal muscle LIM protein 1 expression	0.485	1.58	Healy *et al*. 2010
Nucleolar GTP binding protein 1 expression	0.78	1.355	Healy *et al*. 2010
Mediator of RNA polymerase II transcription (subunit SOH1) expression	1.87	0.32	Healy *et al*. 2010

TSR1 20S rRNA accumulation homolog expression	1.395	0.645	Healy *et al*. 2010
14-3-3 protein expression	0.915	1.56	Healy *et al.* 2010
S phase kinase-associated protein 1(ubiquitin ligase complex) expression	0.74	1.535	Healy *et al*. 2010
Citrate synthase enzyme activity (μmol/min*g wet tissue) in white muscle of fish acclimated to 5°C	1.44 ± 0.12	1.05 ± 0.07	Dhillon & Schulte 2011
Citrate synthase enzyme activity (μmol/min*g wet tissue) in white muscle of fish acclimated to 15°C	1.79 ± 0.11	0.98 ± 0.09	Dhillon & Schulte 2011
White muscle [ADP_{free}] (nmol/g wet tissue) in fish acclimated to 5°C	1.98 ± 1.03	7.37 ± 1.75	Dhillon & Schulte 2011
White muscle [ADP_{free}]:[ATP]($x10^3$) in fish acclimated to 15°C	3.66 ± 0.32	5.16 ± 0.49	Dhillon & Schulte 2011
HISTOLOGY			
Oxidative:glycolytic fibre area in 5° and 15°C acclimated fish	Significantly higher	lower	Dhillon & Schulte 2011
Red muscle mitochondrial volume density in 5°C acclimated fish	Significantly higher	Lower	Dhillon & Schulte 2011
Red muscle mitochondrial cristae surface density ($\mu m^2/ \mu m^3$) in 5°C acclimated fish	Significantly higher	lower	Dhillon & Schulte 2011
White muscle mitochondrial volume density in 5°C acclimated fish	Significantly higher	lower	Dhillon & Schulte 2011
Red muscle mitochondrial cristae surface density ($\mu m^2/ \mu m^3$) in 5°C acclimated fish	Significantly higher	lower	Dhillon & Schulte 2011
MERISTIC			
Predorsal scales	17.18 (1.500)	16.00 (0.961)	Able & Felley 1986
Lateral line scales	35.18 (1.678)	33.70 (1.285)	Able & Felley 1986
Caudal peduncle scales	18.98 (1.097)	19.95 (0.959)	Able & Felley 1986
Anal rays	10.62 (0.740)	10.75 (0.494)	Able & Felley 1986
Caudal rays	16.78 (1.143)	16.30 (1.090)	Able & Felley 1986
Gill rakers	10.35 (1.001)	12.32 (0764)	Able & Felley 1986
MORPHOLOGICAL			
Anal sheath length (females only)	0.647 (0.025)	0.959 (0.034)	Able & Felley 1986
	Longer, shallower heads; larger eyes	Shorter, deeper heads; smaller eyes	Able & Felley 1986
Sexually mature females	Bands of melanophores often broken or bent,	Lateral surface has melanophores organized in	Able & Felley 1986

	consisting of series of spots especially near caudal peduncle	straight, vertical bands	
Sexually mature males	Black ocellus present on dorsal fin	Dorsal fin usually lacks black ocellus	Able & Felley 1986
Both sexes	More heavily pigmented dorsal surface	Less pigmented dorsal surface	Able & Felley 1986
Body Mass (g)	7.29 ± 0.620	16.20 ± 2.07	Podrabsky *et al*. 2000
Body mass of individuals possessing different mitochondrial types (g)	4.09 ± 0.364	5.138 ± 0.477	Haakons 2010[δ]
GENETIC			
Frequency of Ldh-B allozymes	Ldh-B[b] = 0.953	Ldh-B[a] = 0.968	Powers & Place 1978
Frequency of Mdh-A allozymes	Mdh-A[a] = 1.000	Mdh-A[b] = 0.967	Powers & Place 1978
Frequency of Gpi-B allozymes	Gpi-B[c] = 0.941	Gpi-B[b] = 0.749	Powers &Place 1978
Frequency of Idh-B allozymes	Idh-B[c] = 0.990	Idh-B[c] = 0.633 (Cashon et al 1981)**	Cashon *et al*. 1981
Frequency of Idh-A allozymes	Idh-A[c] = 1.00	Idh-A[b] = 0.586	Cashon *et al*. 1981
Frequency of Est-S allozymes	Est-S[b] = 0.967	Est-S[c] = 0.550	Cashon *et al*. 1981
Frequency of Mpi-A allozymes	Mpi-A[d] = 0.964	Mpi-A[b] = 0.910	Ropson *et al*. 1990
Frequency of Est-B allozymes	Est-B[b] = 1.000	Est-B[b] = 0.664	Ropson *et al*. 1990
Frequency of Fum-A alloymes	Fum-A[c] = 1.000	Fum-A[c] = 0.691	Ropson *et al*. 1990
Frequency of H6pdh-A alloymes	H6pdh-A[c] = 0.988	H6pdh-A[c] = 0.105	Ropson *et al*. 1990
Frequency of Pgm-B alloymes	Pgm-B[a] = 1.000	Pgm-B[a] = 0.318	Ropson *et al*. 1990
Ldh-B specific activity (µmol reduced NADH/min/mg total protein) in fish acclimated to 20°C	2.48±0.30	1.31±0.08	Crawford & Powers 1989
Ldh-B protein concentration in fish acclimated to 20°C	Higher	Lower	Crawford & Powers 1989
Ldh-B mRNA concentration (dpm/20µg RNA) in fish acclimated to 20°C	248.9±46.7	121.6±18.5	Crawford & Powers 1989
Ldh-B residue #185 (responsible for stability differences between allozymes)	Serine (TCC)	Alanine (GCC)	Powers *et al*. 1991
Ldh-B residue #311 (no obvious functional significance, results in	Alanine (GCC)	Aspartate (GAC)	Powers *et al*. 1991

charge difference between allozymes)			
LDH-B transcription rate (dpm/10^6dpm hybridized) in fish acclimated to 20°C	5.47±0.54	2.21±0.45	Crawford & Powers 1992
Variation at SP1 (specificity protein 1) binding site in LDH-B promoter sequence	6 base pair deletion before first Sp1	6 base pair insertion before first Sp1	Segal *et al*. 1996
Variation at putative TFIID (protein complex required for initiation of transcription) binding site in LDH-B promoter sequence	TTC	TCC	Segal *et al*. 1996
LDH-B specific activity in response to mild handling stress (units/mg total protein)	No significant effect; no handling stress = 1.89 (0.26) vs. 7 days exposure to stress =2.37 (0.30)	Significant effect; no handling stress = 0.91 (0.29) vs. 7 days exposure to stress = 2.16 (0.19)	Schulte *et al*. 2000
Repressor sequence identical to mouse mammary tumor virus glucocorticoid repressor element (MTV-GRE) ~500 bp upstream of transcription start site of LDH-B	Differs by 1 bp; no evidence of transcription increase in response to stress	Present; evidence of transcription increase in response to stress	Schulte *et al.* 2000
Hsc70 mRNA expression (heat shock temperatures ranged from 30° - 36°C)	No induction	Significantly higher expression at all heat shock temperatures	Fangue *et al.* 2006
Hsp70-2 mRNA expression	Significantly higher at 33°, 34°, and 35°C	Lower	Fangue *et al.* 2006
Hsp90α T_{on} (temperature at which expression induced)	32°C	30°C	Fangue *et al.* 2006
Hsp70 residue #98	Threonine	Serine	Fangue *et al.* 2006
Glucose-6-phosphate isomerase, 6-phosphogluconate dehydrogenase, glutathione peroxidase, phosphomannomutase, cytochrome c oxidase, betaine aldehyde dehydrogenase, copper transport protein show significant correlation of expression with habitat temperature	Expression greater than the mean	Expression less than the mean	Whitehead & Crawford 2006
Cytochrome c oxidase polypeptide II, hypoxia-inducible factor 1 alpha, trypsin precursor, inositol oxygenase, Δ-1-pyrroline-5-carboxylate dehydrogenase, asparagine synthetase show significant correlation of expression with habitat temperature	Expression less than the mean	Expression greater than the mean	Whitehead & Crawford 2006
Differential expression of mitochondrial elongation factor (EF-Ts_{mt}) in different mitochondrial types	Induction in hypoxic conditions	Suppression in hypoxic conditions	Flight *et al.* 2011**

HABITAT			
Mean water temperature	~ 7°C	~24°C	Powers & Place 1978
Lowest Average Monthly temperature	~1°C	~14°C	Powers *et al*. 1986
Mean monthly water temperature range	2.2° - 17.2°C	13.3° - 30°C	summarized in Pierce & Crawford 1997, referenced in DeKoning *et al*. 2004
Annual temperature range	-1.4° - 21°C	7° - 31°C	Fangue *et al*. 2006
REPRODUCTIVE			
Breeding season	June - mid-July	May to late August	Brummett 1966
Spawning season and water temperature range	May - June (15°C-25°C)	March - mid-September (15°C - 30°C)	Taylor 1986
Environmental cue initiating spawning	Water temperature (Wallace and Selman 1981; McMullin et al 2009)	Semi-lunar spawning strategy (Taylor et al 1977)	
Spawning substrate	Hide eggs on algal mats or directly on sandy/muddy substrate	Deposit eggs in shells of ribbed mussel, *Geukensia demissa*	Able 1984
Fertilization success in freshwater	Yes	Never	Able & Palmer 1988
BEHAVIOURAL			
Mean preferred temperature (°C)	28.4 ± 0.58	26.5 ± 0.70	Fangue *et al*. 2009b
Consistently preferred temperature (°C)	30.6 ± 0.51	29.0 ± 0.52	Fangue *et al*. 2009b
Low thermal responsiveness of cold-acclimated fish	Present	Absent	Fangue *et al*. 2009b
LARVAL			
Larval mortality	Highest at 15 ppt	Highest in freshwater	Able & Palmer 1988
Larval length (mm)	5.14 (0.13)	5.56 (0.13)	Marteinsdottir & Able 1992
Larval survival (average number of days to 50% mortality for each female's offspring) after 14 day incubation period (range)	12.3 (10-14)	19.5 (17-25)	Marteinsdottir & Able 1992
Larval survival after 21 day incubation period	9.4 (8-12)	16.3 (14-18)	Marteinsdottir & Able 1992
Larval survival after 28 day incubation period	6.0 (5-7)	12.0 (10-13)	Marteinsdottir & Able 1992
Growth as estimated from otoliths from	0.45	0.40	Reynolds 1994

young of the year collected from wild (mm/day)			
Growth (mm/day) of lab-reared larvae at 21°C	0.508	0.424	Schultz *et al*. 1996
Growth (mm/day) of lab-reared larvae at 28C	0.786	0.702	Schultz *et al*. 1996
EGGS			
	Have chorionic filaments (Shanklin 1959)	Chorionic filaments reduced in number and length (Able and Costagna 1975) or absent (Brummett 1966; Able and Costagna 1975)	
Chorionic fibrils	Long, ~1.5µ in diameter, sparsely distributed; chorionic surface dotted with small protuberances; club-shaped at bases	Shorter, ~0.5µ in diameter, densely distributed; lack club-shaped base	Brummett & Dumont 1981
Egg coating	Thinner, foam-like "jelly" surrounds egg	Thicker "jelly" coat surrounds egg	Brummett & Dumont 1981
Egg diameter	1.7-1.8 mm (Costello et al 1957; Wallace and Selman 1978)	2.04 ± 0.02 mm (Taylor and DiMichele 1980)	
Hatch times of offspring with different Ldh-B phenotypes	Ldh-B^bB^b phenotype ranges from 13.0-14.3 days post-fertilization	Ldh-B^aB^a phenotype ranges from 11.5-12.7 days post-fertilization	DiMichele & Powers 1982b [§]
Mean number of oil droplets/egg	12.7-60	109-172	Morin & Able 1983
Chorionic filament diameter (µ)	0.8-2.2	0.7-0.8	Morin & Able 1983
Range of chorionic filament density (filaments/ 10 000 μ^2)	0 - 5	150-220	Morin & Able 1983
Inhibition constants (KI) of purified allozymes of LDH-B	Ldh-B^b = 36.6±3.5mM	Ldh-B^a = 51.8±5.9mM	Place & Powers 1984b
Glucose utilization rates (µmolx$10^{-8}s^{-1}$)	4.3±0.8	6.6±0.5	DiMichele *et al*. 1991 [Ω]
Glucose utilization rates (µmolx$10^{-8}s^{-1}$) of homozygous eggs injected with purified Ldh-B of opposite homozygote	Ldh-B^b injected with $LdhB^a$ = 6.2±0.8	Ldh-B^a injected with Ldh-B^b = 5.15±0.4	DiMichele *et al*. 1991 [Ω]
Dry weight (mg)	0.593 (0.038)	0.662 (0.054)	Marteinsdottir & Able 1988
Lipids/eggs (mg)	0.039 (0.007)	0.067 (0.021)	Marteinsdottir &

			Able 1988
EMBRYONIC			
Hatch times of individuals with different Mdh-A genotypes	Mdh-A^a individuals hatch later; 18.7 ± 0.1 days	Mdh-A^b individuals hatch earlier; 18.4 ±0.1 days	DiMichele *et al*. 1986 §
Hatch times of individuals with different Gpi-B genotypes	Gpi-B^c individuals hatch later; 18.7 ± 0.2 days	Gpi-B^b individuals hatch earlier; 18.4 ±0.3	DiMichele *et al*. 1986 §
Hatch times of individuals with different Ldh-B genotypes	Ldh-B^b individuals hatch later;19.0 ± 0.3	Ldh-B^a individuals hatch earlier; 17.8 ± 0.1	DiMichele *et al*. 1986 §
Metabolic/developmental rate of embryos with different Ldh-B genotypes	Ldh-B^bB^b phenotype slower	Ldh-B^aB^a faster	DiMichele & Powers 1984*
"trigger level" (rate of oxygen consumption that stimulates hatching) of different Ldh-B genotypes	Ldh-B^bB^b phenotype lower	Ldh-B^aB^a ; able to bind oxygen at lower concentration	DiMichele & Powers 1984*
Hatch times at 20°C, 15ppt (days)	Ldh-B^bB^b = 16.40 ± 0.24	Ldh-B^aB^a = 15.82 ± 0.09	DiMichele & Powers 1991*
Hatch time at 20°C, 33ppt (days)	Ldh-B^bB^b = 17.90 ± 0.18, MDH-A^aA^a = 17.80 ± 0.16	Ldh-B^aB^a = 17.50 ± 0.07, MDH-A^bA^b = 17.40 ± 0.08	DiMichele & Powers 1991*
Hatch time at 20°C, 33ppt (days)	Ldh-B^bB^b/ MDH-A^aA^a = 18.2 ± 0.08	LDH-B^aB^a/MDH-A^bA^b = 17.3 ± 0.02	DiMichele & Powers 1991*
Fitness of mutilocus phenotype at 30°C	LDH-B^bB^b/MDH-A^aA^a/GPI-B^cB^c fitness = 0	LDH-B^aB^a/MDH-A^bA^b/GPI-B^bB^b fitness ~ 1	DiMichele & Powers 1991*
% embryo survival after 28 day incubation period	49	100	Marteinsdottir & Able 1992
Lactate concentration at time of fertilization (mM)	LDH-B^bB^b = 40.0±2.0	LDH-B^aB^a = 48.3±2.4	Paynter *et al*. 1991 Ω
Metabolic rate (as measured by oxygen consumption, heat dissipation, and glucose oxidation)	Higher for LDH-B^bB^b individuals	Lower for LDH-B^aB^a individuals	Paynter *et al*. 1991 Ω
Laboratory-rearing	Fail to reach hatch stage when reared at 30°C	Fail to reach hatch stage when reared at 15°C	DiMichele & Westerman 1997
Hatch times when reared at 20°C	10.5 ± 1.0	20.2 ± 6.0	DiMichele & Westerman 1997
Hatch times when reared at 25°C	9.1 ± 1.0	16.3 ± 3.0	DiMichele & Westerman 1997
In situ hatch times (median number of days and temp range across two years)	10.5-12.5 (18°C-20°C)	14.0-15.0 (23°C - 26°C)	DiMichele & Westerman 1997

*fish collected from west shore of Delaware Bay

Ω fish collected from Shady Side, MD (Chesapeake Bay)

§collection location not given

δfish collected from Metedeconk Creek, in central New Jersey

**fish from Manteloking, New Jersey

Appendix B - Supplementary Material for Chapter Two

Table B.1. PCR conditions for microsatellite loci. Values are temperature and duration, respectively. Each reaction began with initial denaturation of 95°C for 2 minutes and ended with a final extension of 72°C for 5 minutes.

Locus	Denaturation 1	Annealing 1	Extension 1	Denaturation 2	Annealing 2	Extension 2
FhCA-1	94°, 30s	52°, 45s	72°, 1 min	94°, 30	56°, 45s	72°, 1 min
FhATG-6	94°, 30s	63°, 45s	72°, 1 min	NA	NA	NA
FhATG-2 FhATG-4 FhATG-B101 FhATG-B128	94°, 30s	58°, 45s	72°, 1 min	NA	NA	NA
FhATG-17 FhATG-18 FhATG-20	94°, 30s	54°, 45s	72°, 1 min	94°, 30s	59°, 45s	72°, 1 min

Appendix C - Supplementary Material for Chapter Three

Table C.1. Details of SNP loci.

SNP name	Putative ID	Variation	Top BLAST HIT
LDHB_654	lactate dehydrogenase-B	[T/G]	>gb|M33969.1|FUNLDHBA Fish (F.heteroclitus) lactate dehydrogenase-B (LDH-B) mRNA, complete cds
LDHB_1033	lactate dehydrogenase-B	[A/C]	>gb|M33969.1|FUNLDHBA Fish (F.heteroclitus) lactate dehydrogenase-B (LDH-B) mRNA, complete cds
1173_79	unknown	[C/T]	none
1176_169	unknown	[C/A]	none
280_83	unknown	[C/T]	none
65_135	unknown	[C/T]	none
TC15246_1153	cytochrome P450 3A	[A/G]	>gb|AY143428.1| Fundulus heteroclitus cytochrome P450 3A56 (CYP3A56) mRNA, complete
TC15306_666	similar to myoglobin	[C/T]	>dbj|AB470257.2| Seriola dumerili Mb mRNA for myoglobin, complete cds;
TC15560_283	elastase	[A/G]	>dbj|AB029757.2| Paralichthys olivaceus mRNA for elastase 3 precursor, complete cds
TC15573_327	chymotrypsinogen 2	[C/T]	>gb|HQ207652.1| Epinephelus coioides chymotrypsinogen 2 mRNA, partial cds
TC15606_316	40S ribsomal protein S2	[C/T]	>ref|XM_003452828.1| PREDICTED: Oreochromis niloticus 40S ribosomal protein S2-like (LOC100693457), mRNA
TC15936_105	60S ribosomal protein L6	[G/C]	>gb|AY190726.1| Pagrus major 60S ribosomal protein L6 mRNA, partial cds
TC16089_129	nucleoside diphosphate kinase	[G/C]	>gb|EU714154.1| Epinephelus coioides nucleoside diphosphate kinase mRNA, partial cds
TC16579_584	RACK1 (receptor for activated protein kinase C)	[C/T]	>gb|AY342000.1| Oreochromis mossambicus receptor for activated protein kinase C mRNA (RACK1), complete cds
TC16765_535	nucleoside diphosphate kinase	[C/T]	>ref|XM_003442550.1| PREDICTED: Oreochromis niloticus nucleoside diphosphate kinase A-like (LOC100697135), mRNA
TC16936_415	atrial naturietic peptide	[G/C]	>dbj|AB087286.1| Fundulus heteroclitus anp mRNA for atrial natriuretic peptide, complete cds
TC17095_363	glyceraldehyde-3-phosphate dehydrogenase	[C/T]	>gb|HM053703.1| Siniperca chuatsi glyceraldehyde 3-phosphate dehydrogenase isoform 1 mRNA, complete cds
TC17549_296	parvalbumin	[C/T]	>gb|GU254024.1| Siniperca chuatsi parvalbumin 3 mRNA, complete cds
TC17783_489	telethonin-like	[G/C]	>ref|XM_003442144.1| PREDICTED: Oreochromis niloticus telethonin-like (LOC100699372),
TC18048_243	hemoglobin beta-A	[C/T]	>ref|XM_003442071.1| PREDICTED: Oreochromis niloticus hemoglobin subunit beta-A-like, transcript variant 1 (LOC100701987), mRNA
TC18081_450	nascent polypeptide-associated complex alpha (naca)	[C/T]	>ref|XM_003438974.1| PREDICTED: Oreochromis niloticus nascent polypeptide-associated complex alpha polypeptide (naca), mRNA
TC18194_242	phosphate carrier protein (mitochondrial)	[G/C]	>ref|XM_003456031.1| PREDICTED: Oreochromis niloticus phosphate carrier protein, mitochondrial-like, transcript variant 2 (LOC100694818), mRNA
TC18378_211	hemoglobin alpha A	[G/C]	>dbj|AB034639.1| Seriola quinqueradiata mRNA for alpha hemoglobin A, complete cds
TC18850_544	tropomyosin	[A/G]	>gb|GU982540.1| Epinephelus coioides tropomyosin mRNA, complete cds
TC18968_395	warm acclimation related protein (hemopexin)	[C/T]	>gb|AY735161.1| Fundulus heteroclitus warm temperature-acclimated 65kDa protein mRNA, partial cds
TC19110_377	similar to 14-3-3 protein	[A/G]	>gb|BT045507.1| Salmo salar clone ssal-rgf-523-227 14-3-3 protein epsilon putative mRNA, complete cds
TC20694_428	40S ribsomal protein S20	[A/G]	>gb|BT083321.1| Anoplopoma fimbria clone afim-evh-512-370 40S ribosomal protein S20 putative mRNA, complete cds
TC20982_97	40S ribosomal protein S17	[C/T]	>gb|HM137256.1| Oryzias melastigma 40S ribosomal protein S17 mRNA, partial cds
TC21476_353	60S ribsomal protein L35	[C/T]	>gb|GU982574.1| Epinephelus coioides 60S ribosomal protein L35 mRNA, complete cds
TC23851_285	translationally controlled tumor protein	[T/G]	>ref|XM_003445310.1| PREDICTED: Oreochromis niloticus translationally-controlled tumor protein homolog (LOC100699383), mRNA

Table C.1. Continued.

SNP name	Fundulus EST if available
LDHB_654	>gb|DR442025.1| Fh_mx0_74c01_SP6 Killifish Multiple Tissue, Normalized Fundulus heteroclitus cDNA clone Fh_mx0_74c01 5' similar to gb|AAA49302.1| lactate dehydrogenase B.
LDHB_1033	>gb|DR442026.1| Fh_mx0_74c02_SP6 Killifish Multiple Tissue, Normalized Fundulus heteroclitus cDNA clone Fh_mx0_74c02 5' similar to gb|AAA49302.1 lactate dehydrogenase B.
1173_79	none
1176_169	none
280_83	none
65_135	none
TC15246_1153	>gb|CN986759.1| 62417_125_126_D03 Fundulus Heteroclitus Liver Fundulus heteroclitus cDNA similar to Cytochrome P450 3A30 (EC 1.14.14.1) (CYPIIIA30).,
TC15306_666	>gb|CV822765.1| 82317_130_65_A09_1 Fundulus Heteroclitus Heart cDNA, mRNA sequence.
TC15560_283	>gb|EV463259.1| 159-059-H04 Fundulus heteroclitus 159 Fundulus heteroclitus cDNA 5', mRNA sequence.
TC15573_327	>gb|EV459599.1| 159-021-A08 Fundulus heteroclitus 159 Fundulus heteroclitus cDNA 5', mRNA sequence.
TC15606_316	>gb|CN983994.1| 52709_126_46_E04 Fundulus Heteroclitus Liver Fundulus heteroclitus cDNA similar to 40S ribosomal protein S2., mRNA sequence.
TC15936_105	>gb|EV412400.2| 153-006-B05 Fundulus heteroclitus 153 Fundulus heteroclitus cDNA 5', mRNA sequence.
TC16089_129	>gb|CV817705.1| 75794_130_011_H02_1 Fundulus Heteroclitus Heart Fundulus heteroclitus cDNA similar to Nucleoside diphosphate kinase A, mRNA
TC16579_584	>gb|CN985513.1| 58488_126_106_F11 Fundulus Heteroclitus Liver Fundulus heteroclitus cDNA similar to Guanine nucleotide-binding protein beta
TC16765_535	>gb|CV817049.1| 75138_130_004_G06_1 Fundulus Heteroclitus Heart Fundulus heteroclitus cDNA similar to Nucleoside diphosphate kinase NBR-B,
TC16936_415	>gb|CV817934.1| 76028_130_013_E08_1 Fundulus Heteroclitus Heart Fundulus heteroclitus cDNA, mRNA sequence.
TC17095_363	>gb|CN985317.1| 58278_126_101_G05 Fundulus Heteroclitus Liver Fundulus heteroclitus cDNA similar to Glyceraldehyde 3-phosphate dehydrogenase
TC17549_296	>gb|EV458592.1| 159-010-C09 Fundulus heteroclitus 159 Fundulus heteroclitus cDNA 5', mRNA sequence.
TC17783_489	>gb|CV819877.1| 78168_130_33_C12_1 Fundulus Heteroclitus Heart Fundulus heteroclitus cDNA similar to Telethonin, mRNA sequence.
TC18048_243	>gb|CN972799.1| 20296_124-011-23 Fundulus Heteroclitus Heart Fundulus heteroclitus cDNA similar to Hemoglobin beta chain., mRNA sequence.
TC18081_450	>gb|CN975517.1| 25376_125_024_F02 Fundulus Heteroclitus Liver Fundulus heteroclitus cDNA similar to EGD2 protein (GAL4 DNA-binding enhancer protein 2), mRNA sequence.
TC18194_242	>gb|CN972363.1| 19860_124-006-34 Fundulus Heteroclitus Heart Fundulus heteroclitus cDNA similar to Phosphate carrier protein, mitochondrial
TC18378_211	>gb|CV817148.1| 75237_130_005_G09_5 Fundulus Heteroclitus Heart Fundulus heteroclitus cDNA similar to Hemoglobin alpha-A chain, mRNA sequence.
TC18850_544	>gb|EV453460.1| 159-001-G11 Fundulus heteroclitus 159 Fundulus heteroclitus cDNA 5', mRNA sequence.
TC18968_395	>gb|CN978088.1| 28836_125_071_F06 Fundulus Heteroclitus Liver Fundulus heteroclitus cDNA similar to Hemopexin precursor (Hyaluronidase) (EC 3.2.1.35)., mRNA sequence.
TC19110_377	>gb|GT102329.1| sb010P0031H22_F.ab1 Mix of cDNA libraries from embryos Fundulus heteroclitus cDNA 5', mRNA sequence.
TC20694_428	>gb|EV462878.1| 159-055-H03 Fundulus heteroclitus 159 Fundulus heteroclitus cDNA 5', mRNA sequence.
TC20982_97	>gb|CN982178.1| 50623_126_024_C03 Fundulus Heteroclitus Liver Fundulus heteroclitus cDNA similar to 40S ribosomal protein S17., mRNA sequence.
TC21476_353	>gb|CN975537.1| 25397_125_024_E11 Fundulus Heteroclitus Liver Fundulus heteroclitus cDNA similar to 60S ribosomal protein L35., mRNA sequence.
TC23851_285	>gb|CV817725.1| 75814_130_011_G10_1 Fundulus Heteroclitus Heart Fundulus heteroclitus cDNA similar to Translationally-controlled tumor protein,

Table C.1. Continued.

SNP name	5' Sequence
LDHB_654	AGCTGGCCTTCACCTCGCTCCGCGGGGCACCACTCAGTCTCTGATCCCAGACTTTTCTCCTGATCTCTGCAATCATGTCCTCAGTCCTGCAGAAGCTGATCACCCCCCTGGCCAGCTCCTCCGCTGAGCCTCCCAGGAACAAGGTGACCGTGGTGGGTGTGGGCCAGGTCGGCATG GCGTGTGCCGTCAGCATCCTGCTGCGGGACCTGTGCGACGAGCTCGCTCTGGTGGACGTGATGGAGGACCGCCTCAAAGGGGAGATGATGGACCTGCAGCACGGCCTCCTCTTCCTGAAGACGTCCAAGGTGGTTGCCGATAAAGACTACGCCGTGACGGCCAACTCCCGCCT GGTGGTGGTGACGGCCGGCGTGCGGCAGCAGGAGGGCGAGAGCCGCCTCAACCTGGTCCAGAGGAACGTCAACGTCTTCAAGTGCATCATCCCGCAGATCATAAAGTACAGCCCCAACTGCACAATCCTGGTGGTGTCCAACCCCGTTGACGTGCTGACCTACGTGACCTGG AAGCTGAGCGGTCTGCCCAAGCACCGGGTCATCGGCAGCGGCACCAACCTGGACTCGGCTCGCTTCCGCTACATGATGGCCGAGCGCCTCGGCATCCACGCCAGT
LDHB_1033	AGCTGGCCTTCACCTCGCTCCGCGGGCACCACTCAGTCTCTGATCCCAGACTTTTCTCCTGATCTCTGCAATCATGTCCTCAGTCCTGCAGAAGCTGATCACCCCCCTGGCCAGCTCCTCCGCTGAGCCTCCCAGGAACAAGGTGACCGTGGTGGGTGTGGGCCAGGTCGGCATG GCGTGTGCCGTCAGCATCCTGCTGCGGGACCTGTGCGACGAGCTCGCTCTGGTGGACGTGATGGAGGACCGCCTCAAAGG
1173_79	AGTGATACACTAGAAATGTAGCAGCCAGCATATAAGTCAGAATAAGTCATCCCGCTGACTGGCTGTCATGACATCA
1176_169	AGGACAGGCAAATTCCACTCAGTTCTAATGAGCTAGCTGTGCTGCTCTTTCTCCTCTGTCTCTCGGGTCCTCCTGCCAGTTCACCACTAAAAGGTTTGAATTAAAATTTTAAACAGAGCTGGGCCATGCTTCTCAGGCTTTTATATCTAAAACATTT
280_83	GGTATTACATTTACACCGCACACTCCTTTGCATGTAGGTCAACAGGTTGCCCCCCTGTACACCATTCCTAAGCAGT
65_135	GTAGTTTAATATGTCATACTGTAAACGTATCACACTAAATACAGATTTTCTATTAGTGAAGCTGTGTGAGCAGCTCATAAACGCCAGTCATCACTTAAAGCACACAAGCCGAAAAAGAGAGTCGAAGT
TC15246_1153	ATGGGTTACTTCTACTTAACCGCTGAGACATGGACCCTGCTGGTGGCCTTTGTCACTCTGCTCCTCATATACGCCTACTGGCCATATGGGACTTTCAAGAGATTAGGGATCTCAGGTCCCAAACCCGTCCCATTTTTTGGTACAATGCTGCATTATCGAAGGGGGTTCTTCACCTT CGACGAAGAATGCAAAAAGAAATACGGGAAGGTATGGGGAATCTACGACGGCCGCCAGCCTGTGCTGTGTGTCACAGAT
TC15306_666	GAAGTCGGATCTGATGTCTGGATACACAGAGGATCATTTCTAAGCTTGCAGACTTTTGTCTCCTCAGACCATGACGGACTATGATATGGTTCTCAAGCACTGGGGTCCAGTGGAAGCGGATTACAACGCCCACGGGAATATGGTTCTGACTCGCCTATTCCAYGAGCACCCAGA TACCCAAAAGCTGTTCCCGAAGTTTGTTGGCATTGCTAAGGGGGACCTGGCCAGTAACGCAGCCGTTTCTGCCCATGGTGC
TC15560_283	TCTGGAACGGATGAAGCTTGTGGTGCTTGCTTTGCTCATTGCCGGGGCCTATGGCTGCGGCAACCCTACCTTCCAGCCTGTTGTGAGCAGAGTGGTGAACGGAGAGGATGTCAGGCCTCACAGCTGGCCCTGGCAGATTTCCCTGCAGTACAACAGTAACGGCGAATGGAGAC ACACTTGCGGAGGCACCCTGATCTCTGACCAGTGGGTTCTAACTGCTGCTCATTGTATCAGCTCAAGCAGGCAGTACAGAGT
TC15573_327	TCCGGGGCTCGAGCATGGCCTTCCTGTGGATCCTCTCCTGCCTCGCCTTCGTCGGCGCCGCCTACGGTTGCGGCACCCCCGCCATCAGTCCCATCATCACGGGTTATTCCCGTATTGTGAACGGTGAGGAGGCGGTGCCCCACTCCTGGCCCTGGCAGGTGTCCCTCCAGGATTA CACMGGCTTCCACTTCTGCGGAGGCTCCCTCATCAACGAGAACTGGGTGGTGACAGCTGCTCACTGCGGTGTCAGKAYCA
TC15606_316	TGAATCACAAATGGCGGACGACGCCGGTGGTAGAGGAGGTTTTCGCGGAGGTTTCGGAGCAGGTGGCCGAGGTCGCGGCCGTGGACGCGGCAGGGGCCGTGGAAGGGGCCGCGGTGCTCGGGGTGGCAAGTCCGAGGACAARGAGTGGGTGCCCGTCACCAAGCTGGGYC GCCTTGTTAAGGACATGAAGATCAAGACCCTGGAGGAGATCTACCTCTACTCTCTTCCTATCAAGGAGTCAGAGATCATTGACTT
TC15936_105	GGGCCGCATTTATTTTCATAGACTACATGAGCAGAGGGAGACAAGAAAAAGTCTGCCAAGAAGAAGCACGGGAGCCGGAACCCGGCCCTGGCCCGGGGCATCGG
TC16089_129	AATCGATCTGATGTCTGGATGTCTGGGTTCTGAAGCGGAGCTCCACTAGTGAAGAAAATGGCCGAATCAGAGGAGCGCACCTTCCTTGCCATCAAGCCCGACGGCGTGCAGCGCGGCATCATCGGCGA
TC16579_584	TTCCCTTTTTTTAATTGGGAAGTGTGCCCCGGTATTGGGCCAAATTTAGGGTTTTCCCCCAGGGTTCCCTGAGGGGGCCATTTTCAAATTTGTTAGGGGACGTGTTCATTTTTTAGATGGGCCCGTTTGCCCTTTCCGAACCCTGGGACGGGACCCTGGGCCTTTGGGACTTGACC ACTGGCGCCACCCCCCGCCAGTTCGTCGGCCACAACAAGGACGTTTTGAGCGTGGCCTTCTCTGCGGACAACCGCCAGA
TC16765_535	AAAGTCGGATCTGATGTCCGGCACTCACCTCACAGTTCTCCTCAGTCCCAGAACTCAGCCTCCTCAGCAGACATGGAGCGTACCTTCATCGCTGTGAAGCCTGATGGCGTCCAGAGAGGACTCTGTGGCGAGATCATCAAGCGCTTCGAGCACCGTGGCTTCAAGCTGGTCGCC GCCAAATTCGTGCAGGCCTCTGAAGACCACATGAAGAAGCACTACCTGGACCTGAAGGACATGCCTTTCTACGCAGGCCTC
TC16936_415	AAAAAGAATAATGCGATCGGATCTGATGTCTGCTGTCCTCGGGAGCGGGCCTCCAGCTGCTTCGGAGCCAGGATGGATCGAATAGGAACGTCAGCGGCCTGGGATGCAACAGTGCAAGAGGGTAGCCTGAATGCTTCTCGGTCCGGATGCTGGATTCACTCTGAGCCTGAAG ATCCAGCTGAGGACATCCATGACTCCGTTGTCCTGCTCTGGGGAAGACTAACTGTTGTGAATGTAAATACGTGTGTCTCACCG
TC17095_363	AGGAATCGGCACGAGGGGCAAACTGACTGGCATGGCCTTCCGCGTCCCCACCCCCAACGTCTCAGTGGTTGACCTGACCGTCCGTCTGGAGAAACCAGCCAAATACGATGACATTAAGAAGGTTGTGAAGGCTGCAGCTGAGGGTCCCATGAAGGGTGTCCTGGGATACACT GAGCACCAGGTCGTCTCCACGGACTTCAACGGTGACAGTCACTCCTCCATCTTTGATGCTGGTGCCGGCATCGCCCTCAATGA
TC17549_296	GTTTTTCCTCGCCTCACGCAGAGCAAGAATAAAAAATGGCCTTCGCTGGTCTCAACGACGCTGATGTCAAAGCTGCCCTGGACGGCTGCGCAGCTGCTGACTCCTTCGACTACAAGAAATTCTTCAAGGCATGCGGCCTGGCTGGCAAGTCCGCCGATGATGTCAAGAAGGCC TTCGCCACCATTGACCAGGACAACAGTGGCTTCATTGAGGAGGAGGAGCTGAAGCTGTTCCTGCAGAACTTCTCTGCCAGTG
TC17783_489	TGAATTAGGCCGAGGGCGAAGAGACGCATCCTGCTGAACCATCCCAGGCTTTTAAAGAACTCGGAGGAGGCAATAAACCCTTCTATAATCTCAGCATAGCGGAGAAGATGCCTTTCTGTAGCGTGCTGGAAAAGAAAAACGGAGTGGTGACGGGGGGCGAGCTGACCTGCA GCGTTCGGGAGGAGAACAAGGCAAAAAGGGAGAGCTACAGCGCCGACTGGCACAGTATTACCCTCAAGACYCARCCTCAGGAGA
TC18048_243	CGGGGATTCGGGCCGAGGACCCAATTATTTTAATTATTATTTGTGGACTAAAGACAAACTTAATTCATCACCATGGTTGAGTGGACAGATGCTGAGCGCAAGGCCATCACAACTCTGTGGGGRCAGATGGATGTGGCTGAGATTGGACCCCAGGCWCTGAYCAGGCTCCTGAT TGTGTATCCCTGGACTCAGAGACATTTCGGATCTTTCGGCAACCTCTCCACCCCTGCTGCCATCTTCAG
TC18081_450	AAGGGAAATTCGGGCAGGGGGATCGATGCCGAATTCGGCACGAGGAAAAAGCTCAGGCTTGCAGCAGCTGCTGAAATAGACGAGGAGCCTGTAAGCAAAGCCAAACAGAGCCGCAGCGAAAAGAAGGCCCGAAAGGCGATGTCAAAGCTCGGACTTAGGCAGGTAACAGG GGTCACTAGAGTCACCATTCGCAAGTCCAAGAACATCTTGTTCGTCATCACCAAACCAGACGTCTACAAGAGCCCTGCGTCAGAT
TC18194_242	GTGAGGATCGGACGAGGGCGGCGGCCGCGACATCAGACGAGGGAGACAGTTGCGAGTTCGGCTCTCAGAAGTATTTAGTCCTGTGTGGTTTTGGTGGGATTCTGAGCTGTGGCACCACCCACACGGCCGTCGTGCCCCTCGACCTGGTCAAATGCAGAATGCAGGTGAATCCC GATAAATACAAGAGCATTGGCAATGGCTTTGCAGTGACCGTCAGGGAGGATGGCATGAGAGGCCTGGC
TC18378_211	TACCAGAGTATTTTCCCTTCAGCTGAGAAACTTGGAAAAAAGGCAAAAATGAGTCTCTCTGGAAAAGACAAGAGTGTGGTGAAGGCCTTCTGGGACAAAGCGTCCTCTAAAAGTGCTGAGATTGGGGCCGAGGCCCTGGGCAGGATGCTGACTTCATACCCGCAAACCAAGA CCTACTTTGCACACTGGTCTGATCTGAGCCCTGAATCT
TC18850_544	AATTGTATGACTCGGCATGCAGGGACGAGAGAGAGATGGAGATCCAGGAGATCCAGCTGATAGAGGCCAAACAATCGCTGAAGGAGGCAGATCGCAATATGAGGAGGTCGCTCGTAAACTGGTCATCATTGAGGGTGACCTGGAACGTACAGAGGAGCGTGCTGAGCTGTC AGAAAGCAAATGCTCTGAACTTGAGGAAGAGTTGAAAACTGTGACCAACAACCTGAAGTCACTGGAGGCCCAGGCAGAGAAGTA
TC18968_395	CAATTGGGCCGAAGCCTCGTGCCGAATTCGGCACGAGGCGACGCTGTTTTCTCCTACCAGGATCACCTCTACATGATCAAGGATGACCAGGTGTTCGTTTACAAAGTTGGCGAGCCACACACCCACCTGGAGGGTTATCCTAAACCCGTGAAAGAGGAGCTGGGCATCGAGGG CCCCATTGATGCCGCGTTTGTCTGCCAGGACCATCACATTGCTCACATCGTCAAAGGTAGAGACATGTACGACGTAGACATG
TC19110_377	CAATCGATCTGATGTCGATACTAGTTCTGGGAGCTATTTTATCTCTGTTTTTAGTGACATGTCGAACAGATTATCCGGCCTTCTGACGGGCACGCTCGGGGAAGCCCCGGCGTCCTTTTTGTTGGGCTGGATTTTCCGAGGAAGTTGGCAGAATGTGCGATATCCACGTGGGGCA AAGAAGTGCTTAATGTTTTACACTACGATTAGTCTCGTGTTTTATAGCTGTTCCTGGACTGTGTTTTGTGCTGAAATACT
TC20694_428	GGGCCACATTTTCTTCCTGTGGAGCTAACCACGTTKGTAGCGTCCGTTCCTGGCTCACCGCTGTTCGCTTGAATTTCTTCCGAATAAGTCGGTCAGGAATAATTAAAATGGCTTTTAAGGACACTGGTAAGGCCCCTGCTGAGACCGAGGTCGCCATCCATCGCATCCGGATCAC CCTCACCAGCCGCAACGTCAARTCTCTGGAGAAAGTSTGTGCTGACCTGATCCGTGGCGCCAAGGAGAAGAACCTGAAGG
TC20982_97	AAATTTTGGGCCGAGCAACATGGRACGAGTCAGGACCAAGACGGTGAAGAAGGCCGCCAGGGTCATCATCGAGAAATACTACACCCGGCTGGGAAG
TC21476_353	ATAAGAATCGCCGAGGCCTCGTGCCGAATTCGGCACGAGGCGACGACGACATGGCCAAGATCAAGGCAAGAGATCTGCGGGGCAAGAAGAAGGAAGAGCTGCTGAAGCAGCTGGATGACCTGAAGAACGAGCTGTCCCAGCTCCGTGTGGCCAAAGTGACCGGCGGAGCA GCCTCCAAGCTCTCCAAGATCCGTGTTGTTCGCAAATCCATCGCCCGAGTCCTGACTGTAATCAACCAGACGCAGAAGGAGAACC
TC23851_285	TGTGGGCGTGTCCAGTGCTATAACAAGCGCTATTGTTGTGGTCGGGTGCGCCAAATGTGTTGTGTGGGGGGGATAAGTTAGTATATCTCCCGAGGAGGCGGCGCGCCCCCCCACTGTGTTTTTATTTTTTTCCTTTCTCCCTCTTCTAAAGTAGAGGTACGTTTTATTTCCAAATT AAACCTGAAGTCTTTGCATTGCAGATGGACTAGTGTGACAACCATAAACCCAAGCATGATGACGGCCGTGGATGGATTG

Table C.2. Summary of locus pairs in significant linkage disequilibrium at each location. Values marked with an * are significant after FDR-adjustment.

Locus 1	Locus 2	Rij	p-value
Wiscasset, ME			
1176	LDHB1033	0.78021	0.0001*
1176	LDHB654	0.78021	0.0001*
1176	nucleotide_diphosphate_kinase2	0.4293	0.0286
1176	Phosphate carrier protein	0.50317	0.0103
280	Elastase	0.40106	0.0338
LDHB1033	LDHB654	0.99999	0.0001*
LDHB1033	Elastase	0.40106	0.0338
LDHB1033	Phosphate carrier protein	0.5763	0.0016
LDHB 654	Elastase	0.40106	0.0338
LDHB 654	Phosphate carrier protein	0.5763	0.0016
Cytochrome p450	60S_ribosomal_protein_L6	0.70432	0.0003*
Cytochrome p450	nucleotide_diphosphate_kinase2	0.45876	0.0135
Cytochrome p450	Parvalbumin	0.44689	0.018
Myoglobin	nucleotide_diphosphate_kinase2	0.42805	0.0235
Elastase	Activated protein kinase	-0.445	0.0185
Elastase	TC19110	-0.39367	0.0408
Activated protein kinase	nucleotide_diphosphate_kinase2	0.36574	0.0451
Nucleotide diphosphate kinase2	Tropomyosin	0.38216	0.0432
G3PD	TC19110	0.48933	0.0084
WAP	60S_ribosomal_protein_L35	0.47863	0.0113
WAP	Translationally controlled tumor protein	0.71916	0.0001*
TC19110	60S_ribosomal_protein_L35	0.70471	0.0001*
TC19110	Translationally controlled tumor protein	0.48933	0.0084
60S_ribosomal_protein_L35	Translationally controlled tumor protein	0.71916	0.0001*
Sandwich, MA			
1173	TC19110	-0.68451	0.0022
1176	Myoglobin	0.49536	0.0267
1176	Parvalbumin	-0.44176	0.0482
280	LDHB654	0.48298	0.0308
280	Cytochrome p450	0.99999	0.0001*
280	ribsomal_protein	0.48298	0.0308
280	Phosphate carrier protein	0.72447	0.0012
LDHB1033	LDHB654	0.91161	0.0001*

LDHB1033	ribosomal_protein	0.60774	0.0066
LDHB1033	Phosphate carrier protein	0.60774	0.0066
LDHB 654	Cytochrome p450	0.48298	0.0308
LDHB 654	nucleotide_diphosphate_kinase1	0.48298	0.0308
LDHB 654	Phosphate carrier protein	0.70175	0.0017
Cytochrome p450	ribosomal_protein	0.48298	0.0308
Cytochrome p450	Phosphate carrier protein	0.72447	0.0012
Myoglobin	Parvalbumin	-0.52459	0.019
Myoglobin	Hemoglobin a2	0.54157	0.0154
Myoglobin	Tropomyosin	0.52988	0.0178
Elastase	Hemoglobin a2	0.52306	0.0193
Parvalbumin	WAP	0.46214	0.0388
Hemoglobin a1	60S_ribosomal_protein_L35	0.50737	0.0313
Hemoglobin a2	Tropomyosin	0.47817	0.0325
Tropomyosin	ribosomal_protein	0.44138	0.0484
Point Judith, RI			
1173	Activated protein kinase	-0.5147	0.0462
1173	Titin cap	0.55759	0.0308
1176	Elastase	0.54226	0.0153
1176	Parvalbumin	0.52893	0.0154
280	65	0.46569	0.0424
280	Activated protein kinase	0.64167	0.0033
65	ribsomal_protein	0.72536	0.0016
LDHB1033	LDHB654	0.5348	0.0168
LDHB 654	Hemoglobin a2	0.43796	0.0447
LDHB 654	Phosphate carrier protein	-0.86155	0.0003
Myoglobin	nucleotide_diphosphate_kinase1	0.53999	0.0157
Myoglobin	60S_ribosomal_protein_L35	0.63445	0.014
Elastase	ribosomal_protein_S2	-0.5727	0.0266
Elastase	nucleotide_diphosphate_kinase2	0.45884	0.0402
Elastase	Parvalbumin	0.52438	0.019
Elastase	Titin cap	0.47607	0.0332
ribosomal_protein_S2	Parvalbumin	-0.75899	0.0024
ribosomal_protein_S2	60S_ribosomal_protein_L35	0.6029	0.0159
Nucleotide diphosphate kinase1	Hemoglobin a1	-0.43743	0.045
Activated protein kinase	WAP	0.51718	0.0178
Nucleotide diphosphate kinase2	ribosomal_protein	0.44896	0.0396
Parvalbumin	Tropomyosin	-0.44819	0.045
Hemoglobin a1	Hemoglobin a2	0.49418	0.0235
Hemoglobin a1	TC19110	-0.51828	0.0205
60S_ribosomal_protein_L35	Phosphate carrier protein	-0.60985	0.0279

Clinton, CT			
1176	60S_ribosomal_protein_L35	0.5822	0.0076
280	nucleotide_diphosphate_kinase2	-0.5528	0.0113
280	G3PD	0.72574	0.0009
65	40S_ribsomal_protein_S17	0.57511	0.0084
65	60S_ribosomal_protein_L35	0.57511	0.0084
LDHB1033	LDHB654	0.97988	0.0001
LDHB1033	Atrial Natriuretic peptide	0.62095	0.0044
LDHB1033	Phosphate carrier protein	0.89261	0.0001
LDHB 654	Atrial natriuretic peptide	0.51983	0.0172
LDHB 654	Phosphate carrier protein	0.88371	0.0001
Cytochrome p450	G3PD	0.48074	0.0276
Myoglobin	ribosomal_protein_S2	0.44191	0.0429
Myoglobin	nucleotide_diphosphate_kinase2	-0.52199	0.0168
Elastase	TC19110	0.65986	0.0025
ribosomal_protein_S2	G3PD	0.46248	0.0341
ribosomal_protein_S2	Parvalbumin	0.46248	0.0341
ribosomal_protein_S2	NACA	0.55613	0.0129
Nucleotide diphosphate kinase1	60S_ribosomal_protein_L35	0.72366	0.0009
Nucleotide diphosphate kinase2	Parvalbumin	-0.5528	0.0113
Hemoglobin a1	Hemoglobin a2	0.64491	0.0031
Hemoglobin a2	WAP	0.71189	0.0011
WAP	TC19110	0.53214	0.0147
1173	ribosomal_protein_S2	0.54798	0.012
Cheesequake, NJ			
1173	nucleotide_diphosphate_kinase1	0.3224	0.0499
1176	G3PD	0.32492	0.0424
1176	60S_ribosomal_protein_L35	0.33277	0.0377
280	Parvalbumin	-0.38408	0.0179
LDHB1033	LDHB654	0.68644	0.0001
LDHB1033	Phosphate carrier protein	0.44928	0.0063
LDHB 654	Phosphate carrier protein	0.46471	0.0042
Cytochrome p450	Translationally controlled tumor protein	-0.52323	0.0015
Myoglobin	Tropomyosin	0.38416	0.0179
Elastase	ribosomal_protein_S2	-0.43138	0.0071
Elastase	NACA	-0.34254	0.0324
Chymotrypsinogen	Parvalbumin	0.33067	0.0389
ribosomal_protein_S2	40S_ribsomal_protein_S17	0.37658	0.0203
Activated protein kinase	Tropomyosin	-0.38113	0.0173
Nucleotide diphosphate	ribosomal_protein	-0.31443	0.0496

kinase2			
Parvalbumin	40S_ribsomal_protein_S17	-0.35814	0.0273
Hemoglobin a1	Hemoglobin a2	0.7798	0.0001
Hemoglobin a2	40S_ribsomal_protein_S17	0.41025	0.0114
Belford, NJ			
1173	LDHB654	0.2864	0.0496
1173	Activated protein kinase	0.29335	0.0421
1173	Activated protein kinase	0.3537	0.0153
1176	280	0.35757	0.0165
280	Elastase	0.3685	0.0124
65	Chymotrypsinogen	0.47068	0.0011
LDHB1033	LDHB654	0.69699	0.0001
LDHB1033	Myoglobin	-0.33427	0.0249
LDHB1033	Activated protein kinase	0.36356	0.0127
LDHB1033	G3PD	0.3898	0.0075
LDHB1033	Phosphate carrier protein	0.31159	0.0327
LDHB 654	Activated protein kinase	0.2881	0.0437
LDHB 654	G3PD	0.29005	0.0423
LDHB 654	Phosphate carrier protein	0.34204	0.0167
Cytochrome p450	nucleotide_diphosphate_kinase1	0.30764	0.039
Cytochrome p450	TC19110	-0.31634	0.0268
Elastase	Hemoglobin a1	0.27935	0.0482
Chymotrypsinogen	NACA	0.31621	0.0254
Chymotrypsinogen	40S_ribsomal_protein_S17	0.39961	0.0047
Activated protein kinase	G3PD	0.36988	0.0089
Activated protein kinase	WAP	-0.31938	0.0239
Activated protein kinase	ribosomal_protein	-0.30108	0.0333
Activated protein kinase	Phosphate carrier protein	0.30797	0.0294
G3PD	Translationally controlled tumor protein	-0.28958	0.0448
Parvalbumin	Titin cap	0.2824	0.0458
Titin cap	WAP	-0.31041	0.0282
Hemoglobin a1	Hemoglobin a2	0.81576	0.0001
NACA	40S_ribsomal_protein_S17	0.36343	0.0102
Hemoglobin a2	40S_ribsomal_protein_S17	0.28693	0.0425
Tropomyosin	TC19110	-0.29609	0.0382
Tropomyosin	ribosomal_protein	0.30409	0.0315
WAP	TC19110	0.29855	0.0366
Sandy Hook, NJ			
1173	65	-0.30378	0.0394
1173	G3PD	0.35305	0.0144
1176	LDHB1033	0.34948	0.0135

1176	ribosomal_protein_S2	0.29511	0.0369
280	65	0.30086	0.0391
280	hemoglobin_a2	0.29823	0.0368
65	LDHB1033	0.30391	0.0352
LDHB1033	LDHB654	0.70043	0.0001
LDHB1033	TC19110	-0.30484	0.0311
LDHB1033	Phosphate carrier protein	0.32589	0.0212
LDHB 654	Phosphate carrier protein	0.2936	0.0399
Cytochrome p450	Elastase	0.30136	0.0331
Myoglobin	Activated protein kinase	0.34173	0.0168
Myoglobin	Parvalbumin	0.30442	0.0331
Elastase	60S_ribosomal_protein_L6	0.34871	0.0137
Chymotrypsinogen	Atrial natriuretic peptide	0.28774	0.044
Nucleotide diphosphate kinase1	Hemoglobin a2	-0.44051	0.0028
Parvalbumin	WAP	0.32148	0.0244
Titin cap	40S_ribsomal_protein_S17	-0.28669	0.0426
Hemoglobin a1	Hemoglobin a2	0.67494	0.0001
Hemoglobin a1	40S_ribsomal_protein_S17	0.2962	0.0362
NACA	40S_ribsomal_protein_S17	0.28187	0.0463
Tropomyosin	TC19110	-0.28121	0.0468
Navesink, NJ			
1173	ribosomal_protein_S2	0.61317	0.027
1173	Activated protein kinase	0.60096	0.0462
1176	LDHB1033	0.6222	0.0311
1176	LDHB654	0.58431	0.0351
1176	Myoglobin	-0.58227	0.0358
280	LDHB654	0.7817	0.0134
280	Cytochrome p450	-0.75378	0.0171
280	NACA	0.80416	0.011
65	Parvalbumin	0.54851	0.048
65	Hemoglobin a2	0.58262	0.0357
65	40S_ribsomal_protein_S17	0.76544	0.0058
65	60S_ribosomal_protein_L35	0.5992	0.0307
LDHB1033	LDHB654	0.84501	0.0034
LDHB1033	Cytochrome p450	-0.56971	0.0484
LDHB1033		-0.57854	0.0451
LDHB1033	NACA	0.73798	0.0106
LDHB 654	NACA	0.82447	0.003
Elastase	60S_ribosomal_protein_L6	-0.73342	0.0082
ribosomal_protein_S2	ribosomal_protein	0.67822	0.0145

60S_ribosomal_protein_L6	TC19110	0.59142	0.033
Nucleotide diphosphate kinase1	Parvalbumin	-0.7379	0.0144
Nucleotide diphosphate kinase2	ribosomal_protein	-0.64091	0.0208
Nucleotide diphosphate kinase2	60S_ribosomal_protein_L35	-0.58147	0.036
Atrial natriuretic peptide	Parvalbumin	-0.75889	0.0118
G3PD	Tropomyosin	-0.61664	0.0262
G3PD	ribosomal_protein	0.63292	0.0225
Hemoglobin a1	Hemoglobin a2	0.66499	0.0165
Metedeconk, NJ			
1176	Phosphate carrier protein	0.44192	0.0022
280	60S_ribosomal_protein_L35	0.33405	0.0194
65	Elastase	-0.28323	0.0497
LDHB1033	LDHB654	0.49927	0.0006
LDHB1033	ribosomal_protein	0.36676	0.0129
LDHB 654	60S_ribosomal_protein_L6	-0.37738	0.0083
LDHB 654	Atrial natriuretic peptide	-0.29352	0.0399
LDHB 654	TC19110	0.4709	0.001
LDHB 654	Phosphate carrier protein	0.39704	0.0054
Cytochrome p450	60S_ribosomal_protein_L6	0.28055	0.0495
Cytochrome p450	G3PD	0.33053	0.0207
Cytochrome p450	TC19110	-0.37646	0.0084
Myoglobin	Translationally controlled tumor protein	0.33648	0.0211
Elastase	ribosomal_protein_S2	0.31421	0.0312
Elastase	NACA	0.35522	0.0129
Elastase	WAP	-0.31759	0.0262
ribosomal_protein_S2	G3PD	0.30811	0.0347
60S_ribosomal_protein_L6	Titin cap	0.34429	0.016
60S_ribosomal_protein_L6	Translationally controlled tumor protein	0.33566	0.02
Nucleotide diphosphate kinase1	Translationally controlled tumor protein	0.2942	0.0437
Activated protein kinase	Translationally controlled tumor protein	-0.31445	0.0294
Nucleotide diphosphate kinase2	Parvalbumin	-0.28353	0.0472
Hemoglobin a1	Hemoglobin a2	0.43767	0.0022
Hemoglobin a1	Phosphate carrier protein	-0.29289	0.0403
NACA	60S_ribosomal_protein_L35	-0.32866	0.0214
Hemoglobin a2	WAP	0.30413	0.0333
Tropomyosin	ribosomal_protein	0.49331	0.0006
Tropomyosin	Translationally controlled tumor protein	0.28858	0.0456

TC19110	Phosphate carrier protein	0.33572	0.0188
Laurel, NJ			
280	Activated protein kinase	-0.33475	0.0264
280	Titin cap	0.30604	0.0424
65	Myoglobin	0.43455	0.0044
65	ribosomal_protein_S2	-0.3333	0.0209
65	Atrial natriuretic peptide	-0.32282	0.0238
65	Titin cap	0.29469	0.0391
LDHB1033	LDHB654	0.56767	0.0002
LDHB 654	Myoglobin	-0.33486	0.03
LDHB 654	60S_ribosomal_protein_L6	0.29476	0.0411
Cytochrome p450	60S_ribosomal_protein_L35	0.32228	0.0271
Myoglobin	Atrial natriuretic peptide	-0.36728	0.016
Myoglobin	Hemoglobin a2	-0.34768	0.0226
Elastase	Hemoglobin a1	0.32314	0.0237
Chymotrypsinogen	Activated protein kinase	-0.30124	0.035
Nucleotide diphosphate kinase1	Translationally controlled tumor protein	-0.35958	0.0147
Activated protein kinase	Parvalbumin	-0.41321	0.0038
Nucleotide diphosphate kinase2	Tropomyosin	0.33976	0.0174
G3PD	Hemoglobin a2	-0.33039	0.0207
Hemoglobin a1	Hemoglobin a2	0.74886	0.0001
Hemoglobin a2	Tropomyosin	0.28514	0.0459
Hemoglobin a2	Translationally controlled tumor protein	-0.42127	0.0043
Tropomyosin	WAP	0.29655	0.0399
Tropomyosin	Translationally controlled tumor protein	-0.36572	0.0131
40S_ribsomal_protein_S17	Translationally controlled tumor protein	0.30405	0.0414
Tuckerton, NJ			
1173	TC19110	-0.45859	0.0403
1176	280	0.46712	0.0475
1176	ribosomal_protein_S2	0.59587	0.014
1176	nucleotide_diphosphate_kinase2	-0.79182	0.0011
1176	Phosphate carrier protein	-0.67252	0.0056
280	nucleotide_diphosphate_kinase2	-0.7199	0.0017
280	Phosphate carrier protein	-0.45017	0.0497
65	LDHB1033	0.45426	0.0477
65	WAP	0.69256	0.0025
LDHB1033	LDHB654	0.60527	0.0068
LDHB1033	60S_ribosomal_protein_L6	-0.58851	0.0085

LDHB1033	WAP	0.64749	0.0038
LDHB1033	60S_ribosomal_protein_L35	0.5682	0.0111
LDHB 654	Activated protein kinase	-0.5857	0.0107
LDHB 654	Atrial natriuretic peptide	0.49108	0.0323
LDHB 654	Parvalbumin	0.50968	0.0226
LDHB 654	Titin cap	-0.46214	0.0388
LDHB 654	TC19110	-0.50941	0.0227
LDHB 654	40S_ribsomal_protein_S17	0.486	0.0341
LDHB 654	Translationally controlled tumor protein	-0.46214	0.0388
Cytochrome p450	Phosphate carrier protein	-0.59726	0.0113
Myoglobin	ribosomal_protein_S2	0.51665	0.0284
Myoglobin	40S_ribsomal_protein_S17	0.62859	0.0077
Elastase	Hemoglobin a1	0.5876	0.0086
Elastase	Hemoglobin a2	0.5876	0.0086
Elastase	60S_ribosomal_protein_L35	-0.54737	0.0144
ribosomal_protein_S2	40S_ribsomal_protein_S17	0.62615	0.0079
Nucleotide diphosphate kinase1	Parvalbumin	0.54268	0.0253
Nucleotide diphosphate kinase1	Titin cap	-0.64214	0.0081
Activated protein kinase	Atrial natriuretic peptide	-0.64305	0.0064
Activated protein kinase	Parvalbumin	-0.46821	0.0413
Activated protein kinase	40S_ribsomal_protein_S17	-0.46564	0.0482
Nucleotide diphosphate kinase2	G3PD	0.51538	0.0212
Atrial natriuretic peptide	Phosphate carrier protein	0.45447	0.0476
Parvalbumin	TC19110	-0.45018	0.0441
Titin cap	Tropomyosin	0.49722	0.0262
Hemoglobin a1	Hemoglobin a2	0.99999	0.0001
ribosomal_protein	Phosphate carrier protein	-0.50043	0.0292
40S_ribsomal_protein_S17	Translationally controlled tumor protein	-0.59529	0.0095
RUMFS, NJ			
1173	G3PD	0.37696	0.0106
1176	ribosomal_protein_S2	0.32183	0.0291
1176	TC19110	-0.30852	0.0291
65	Activated protein kinase	0.32908	0.02
LDHB1033	LDHB654	0.36549	0.0098
LDHB1033	WAP	0.43042	0.0026
LDHB 654	Chymptrypsinogen	-0.28182	0.0463

Myoglobin	ribosomal_protein	0.29024	0.0401
Myoglobin	60S_ribosomal_protein_L35	0.32855	0.0202
Elastase	Titin cap	0.39851	0.0048
60S_ribosomal_protein_L6	nucleotide_diphosphate_kinase1	0.28033	0.0475
Nucleotide diphosphate kinase1	Phosphate carrier protein	-0.29943	0.038
Activated protein kinase	Parvalbumin	-0.2911	0.0396
Activated protein kinase	Titin cap	0.28137	0.0466
Nucleotide diphosphate kinase2	Atrial natriuretic peptide	0.41636	0.0032
Atrial natriuretic peptide	Translationally controlled tumor protein	0.28744	0.0442
Parvalbumin	hemoglobin_a2	-0.28137	0.0466
Parvalbumin	TC19110	-0.36905	0.0091
Titin cap	Tropomyosin	-0.30475	0.0312
Hemoglobin a1	Hemoglobin a2	0.7734	0.0001
Hemoglobin a1	60S_ribosomal_protein_L35	0.38319	0.0067
Magotha, VA			
1173	LDHB654	-0.51363	0.0216
1173	ribsomal_protein	0.44457	0.0468
1173	40S_ribsomal_protein_S17	-0.47499	0.0337
1173	Phosphate carrier protein	0.56577	0.0114
1176	65	0.48179	0.0357
65	nucleotide_diphosphate_kinase2	0.5677	0.0111
65	60S_ribosomal_protein_L35	-0.47607	0.0332
LDHB1033	Titin cap	-0.45511	0.0473
LDHB1033	NACA	-0.50293	0.0329
LDHB 654	G3PD	0.59901	0.0074
LDHB 654	60S_ribosomal_protein_L35	-0.48235	0.031
Cytochrome p450	Activated protein kinase	0.56968	0.0157
Cytochrome p450	nucleotide_diphosphate_kinase2	0.84029	0.0002
Elastase	60S_ribosomal_protein_L6	0.73921	0.0013
Elastase	nucleotide_diphosphate_kinase1	0.45984	0.0397
Chymotrypsinogen	TC19110	0.59881	0.009
ribosomal_protein_S2	NACA	0.54999	0.0196
ribosomal_protein_S2	ribosomal_protein	-0.58228	0.0092
ribosomal_protein_S2	Phosphate carrier protein	-0.6769	0.0025
Titin cap	Hemoglobin a2	-0.78864	0.0006
Hemoglobin a2	Phosphate carrier protein	0.46393	0.038
Manteo, NC			
1173	NACA	0.50995	0.0226
1173	40S_ribsomal_protein_S17	0.466	0.0422
1173	40S_ribsomal_protein_S17	-0.45777	0.0406

1176	65	0.70035	0.003
1176	Myoglobin	0.48043	0.0415
1176	Elastase	0.50621	0.0317
1176	nucleotide_diphosphate_kinase2	-0.54907	0.0198
1176	G3PD	-0.4906	0.0374
1176	Hemoglobin a1	0.48043	0.0415
1176	Translationally controlled tumor protein	-0.47523	0.0438
280	Cytochrome p450	0.48179	0.0357
280	Activated protein kinase	0.49029	0.0283
280	nucleotide_diphosphate_kinase2	0.48298	0.0308
65	Myoglobin	0.72447	0.0012
65	Atrial natriuretic peptide	-0.61714	0.0058
LDHB1033	Cytochrome p450	0.50918	0.0265
LDHB1033	Hemoglobin a1	0.53293	0.0202
LDHB 654	Cytochrome p450	0.50918	0.0265
LDHB 654	nucleotide_diphosphate_kinase2	0.44138	0.0484
Cytochrome p450	ribosomal_protein_S2	0.57183	0.0153
Cytochrome p450	nucleotide_diphosphate_kinase2	0.5114	0.0258
Myoglobin	Elastase	0.57486	0.0101
Elastase	Parvalbumin	0.50225	0.0247
Elastase	TC19110	0.479	0.0322
Elastase	Translationally controlled tumor protein	-0.58959	0.0084
Chymotrypsinogen	Translationally controlled tumor protein	0.44138	0.0484
60S_ribosomal_protein_L6	40S_ribsomal_protein_S17	0.49936	0.0341
60S_ribosomal_protein_L6	60S_ribosomal_protein_L35	0.5114	0.0258
Nucleotide diphosphate kinase1	WAP	0.60764	0.0151
Nucleotide diphosphate kinase1	ribosomal_protein	-0.65361	0.0089
Activated protein kinase	Tropomyosin	-0.55446	0.0157
Activated protein kinase	40S_ribsomal_protein_S17	0.56678	0.0135
G3PD	Titin cap	-0.65873	0.0032
Parvalbumin	hemoglobin_a2	0.47788	0.0326
Parvalbumin	TC19110	0.56882	0.011
Titin cap	Phosphate carrier protein	-0.53416	0.0199
Hemoglobin a1	TC19110	0.62529	0.0052
Hemoglobin a2	TC19110	0.57337	0.0103
Hemoglobin a2	60S_ribosomal_protein_L35	0.60774	0.0066
WAP	Phosphate carrier protein	0.56291	0.0141
ribosomal_protein	40S_ribsomal_protein_S17	-0.53091	0.0207
Sapelo Island, GA			

1173	LDHB654	-0.49086	0.0048
1173	ribosomal_protein_S2	0.35487	0.0415
1173	G3PD	0.37955	0.0376
1173	ribsomal_protein	0.51201	0.0033
1176	280	0.60698	0.0005
1176	Elastase	0.49086	0.0048
1176	Tropomyosin	0.35134	0.0436
1176	WAP	0.42208	0.0153
280	Tropomyosin	0.35805	0.0397
LDHB 654	G3PD	-0.49458	0.0068
Cytochrome p450	Myoglobin	0.54245	0.0025
Cytochrome p450	NACA	0.64357	0.0003
Myoglobin	Translationally controlled tumor protein	-0.44297	0.0137
Elastase	60S_ribosomal_protein_L35	0.4445	0.0107
Elastase	Phosphate carrier protien	0.4445	0.0107
Chymotrypsinogen	Tropomyosin	-0.35134	0.0436
ribosomal_protein_S2	Parvalbumin	-0.37943	0.0346
60S_ribosomal_protein_L6	Phosphate carrier protein	0.35487	0.0415
Nucleotide diphosphate kinase1	Parvalbumin	0.46988	0.0079
Nucleotide diphosphate kinase1	WAP	-0.35864	0.0394
Nucleotide diphosphate kinase1	40S_ribsomal_protein_S17	0.73293	0.0001
Nucleotide diphosphate kinase1	60S_ribosomal_protein_L35	0.41044	0.0184
Activated protein kinase	Titin cap	0.99999	0.0001
Activated protein kinase	Tropomyosin	0.3851	0.032
Nucleotide diphosphate kinase2	ribosomal_protein	0.41912	0.0161
G3PD	Phosphate carrier protein	0.37955	0.0376
Parvalbumin	ribosomal_protein	-0.37496	0.0339
40S_ribsomal_protein_S17	60S_ribosomal_protein_L35	0.4445	0.0107

Figure C.1. Five clines having their centres located to the south of the New Jersey border. A) Atrial Naturetic Peptide, B) Chymotrypsinogen, C) Ribosomal protein S2, D) Translationally controlled tumor protein, E) Ribosomal protein

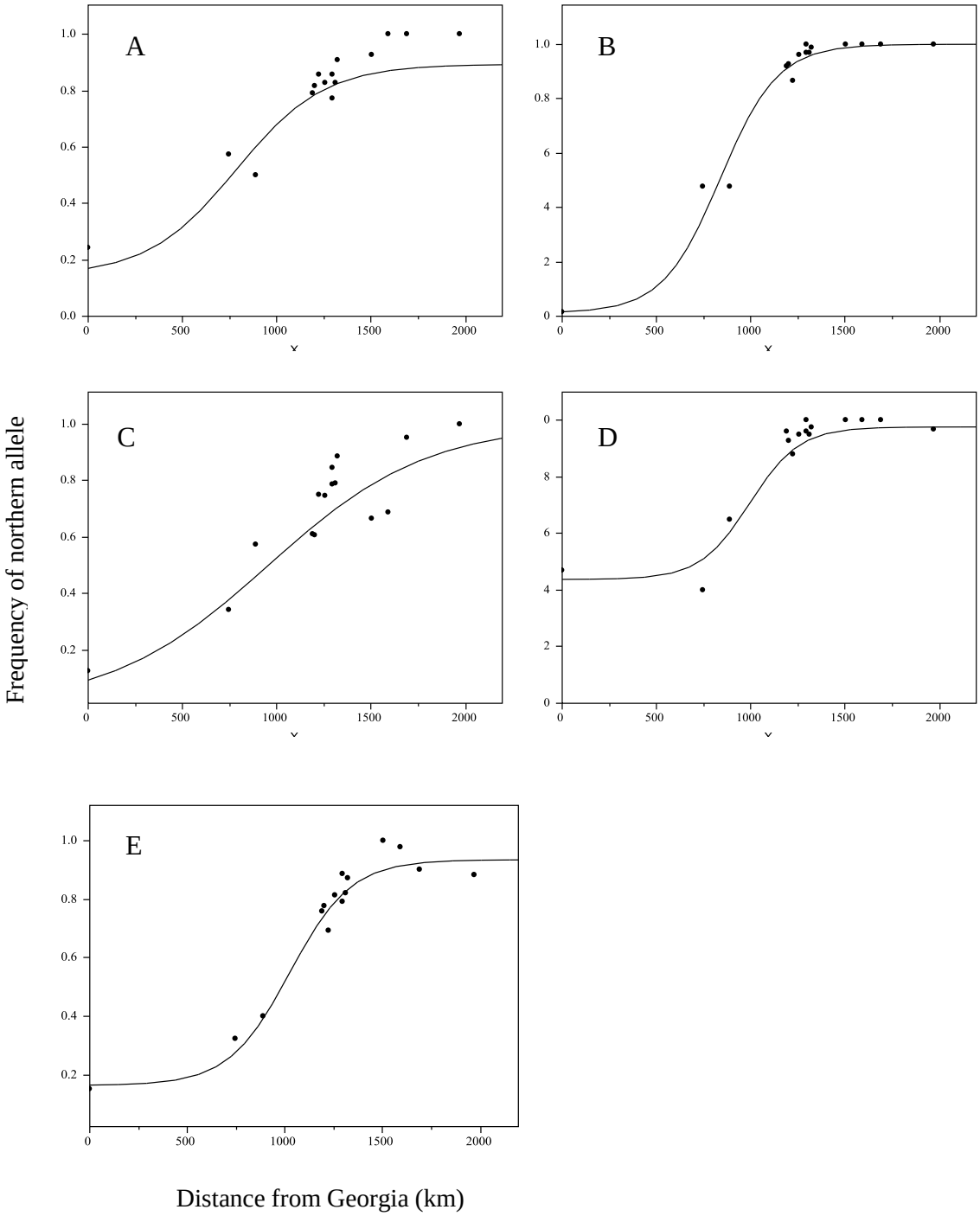
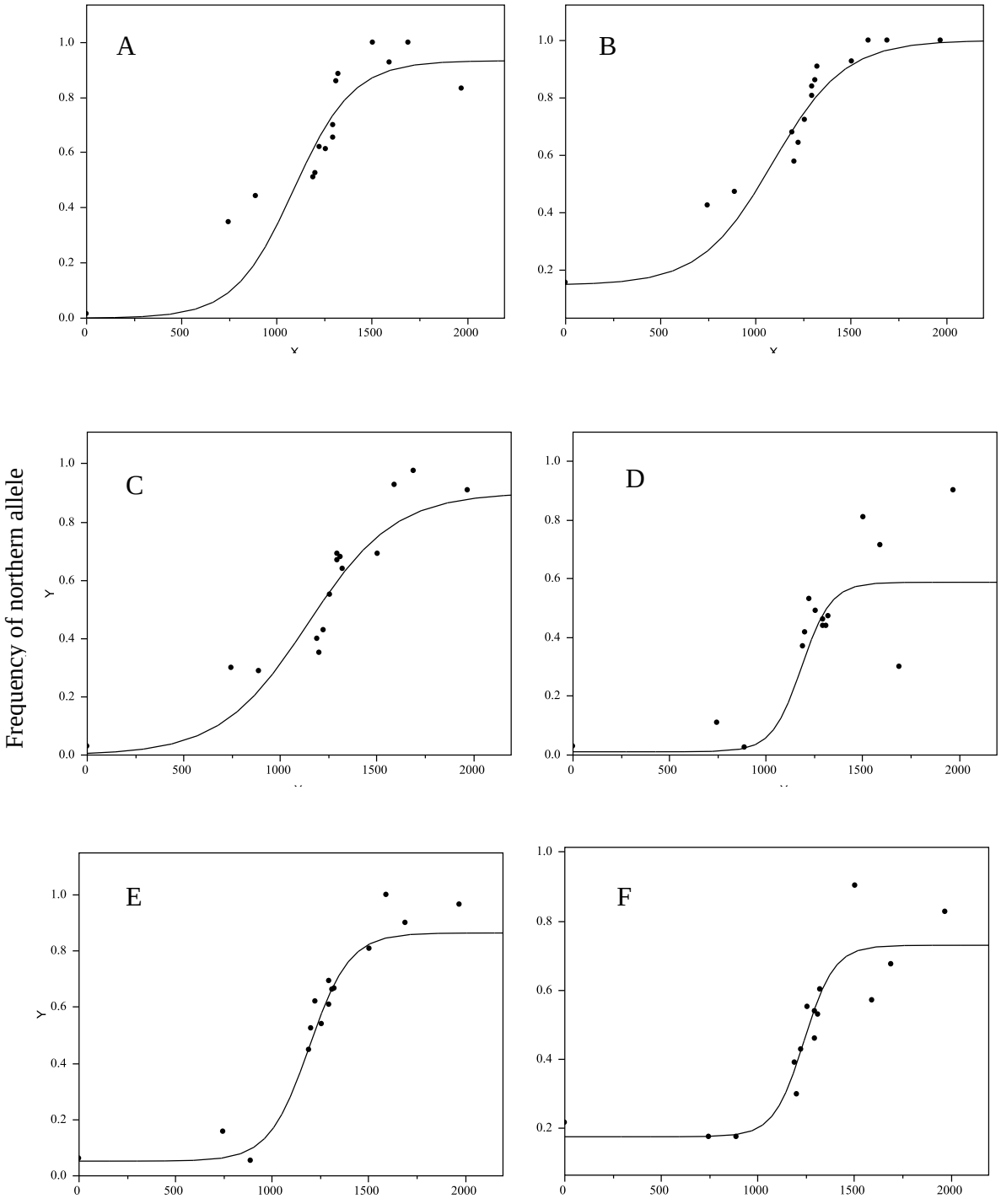
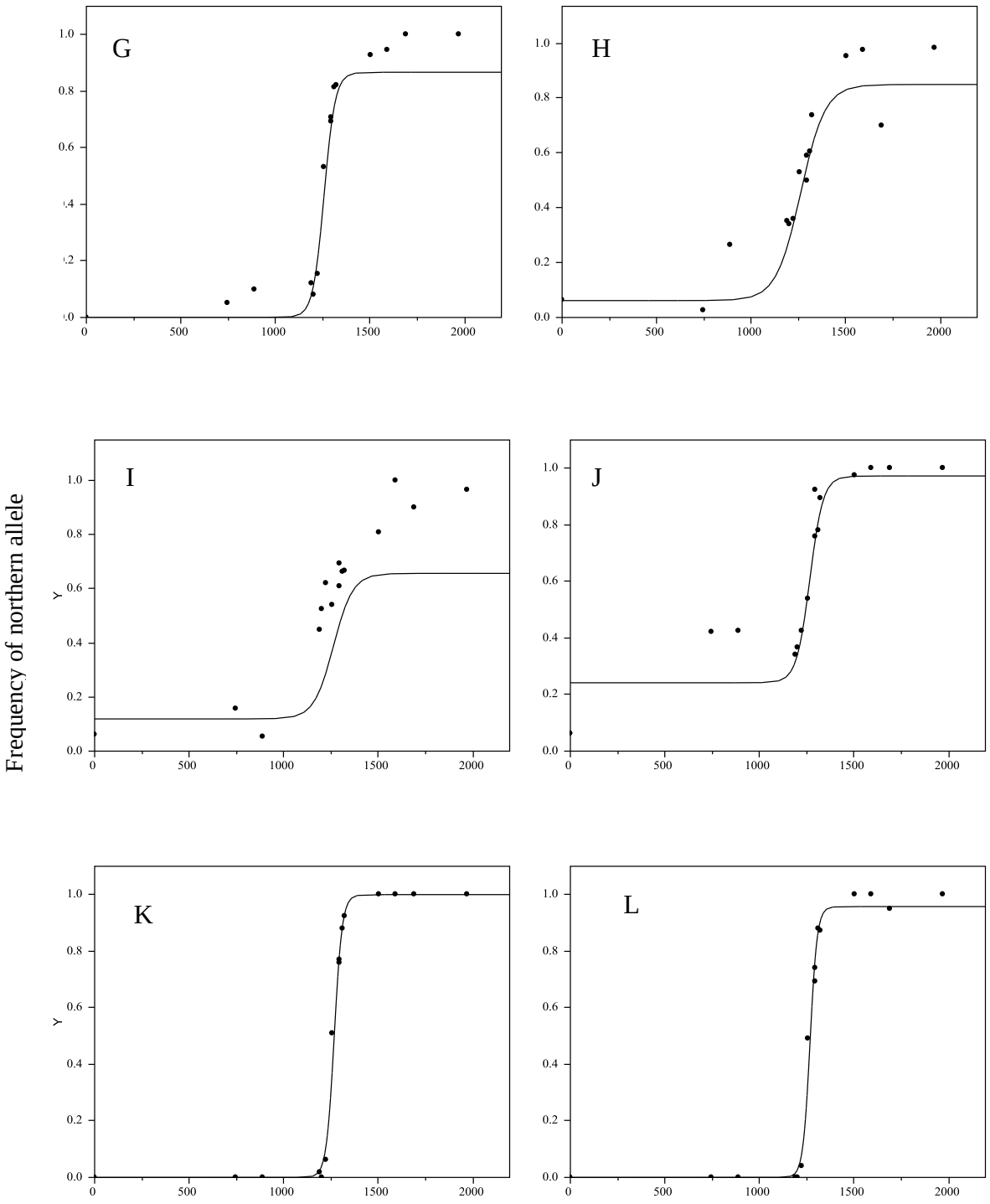


Figure C.2. Seventeen clines having their centres located within the New Jersey border. A) Activated protein kinase, B) NACA, C) Titin cap, D) 1176, E) 60S ribosomal protein L6, F) Parvalbumin, G) 65, H) Myoglobin, I) 60S ribosomal protein L35, J) 40S ribosomal protein S17, K) COXI, L) CytB, M) Phosphate carrier protein, N) 280, O) Warm acclimation related protein, P) Glyceraldehyde 3 phosphate dehydrogenase, Q) Tropomyosin.




G
H
I
J
K
L
Y
Frequency of northern allele
Distance from Georgia (km)

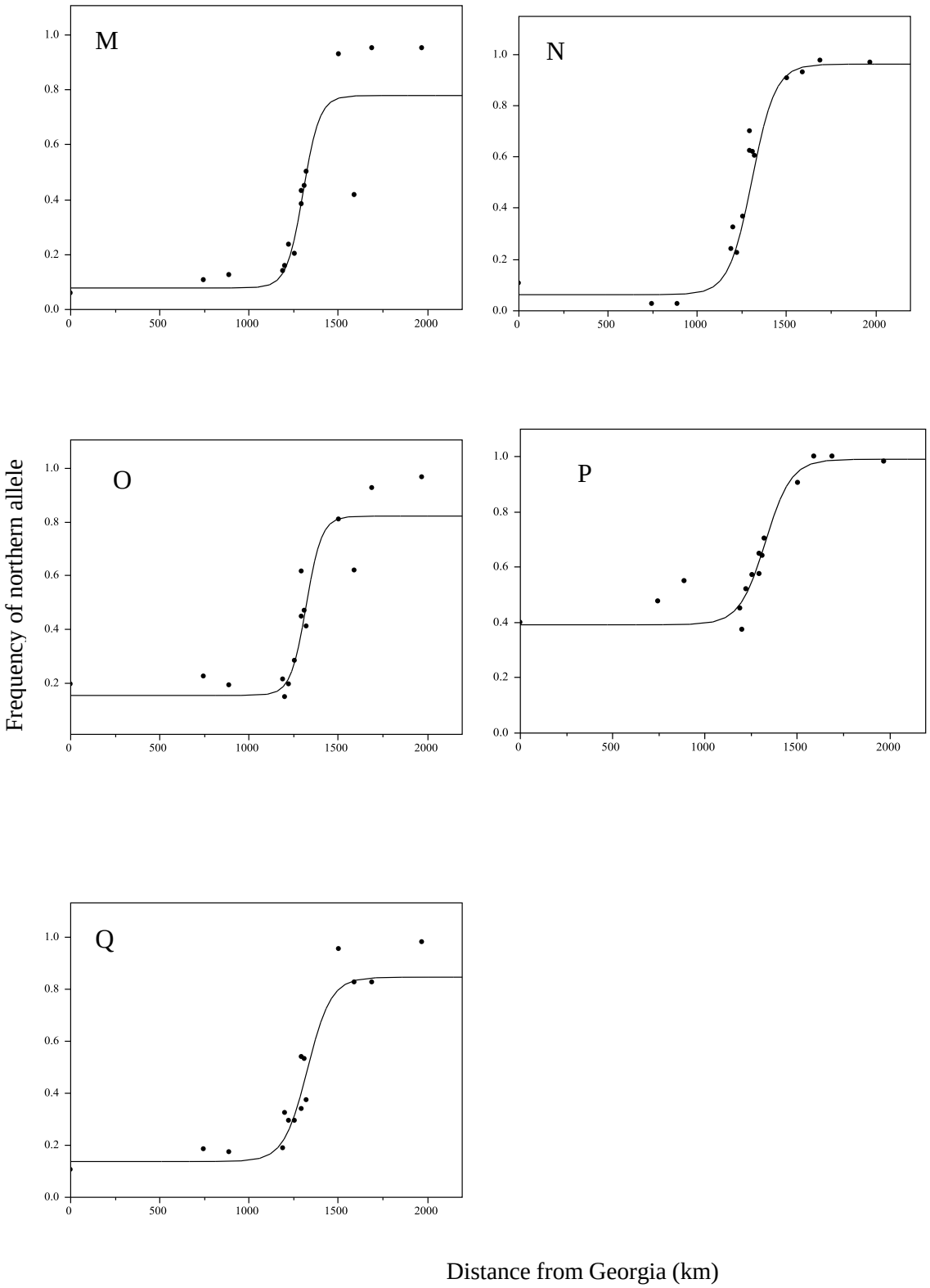
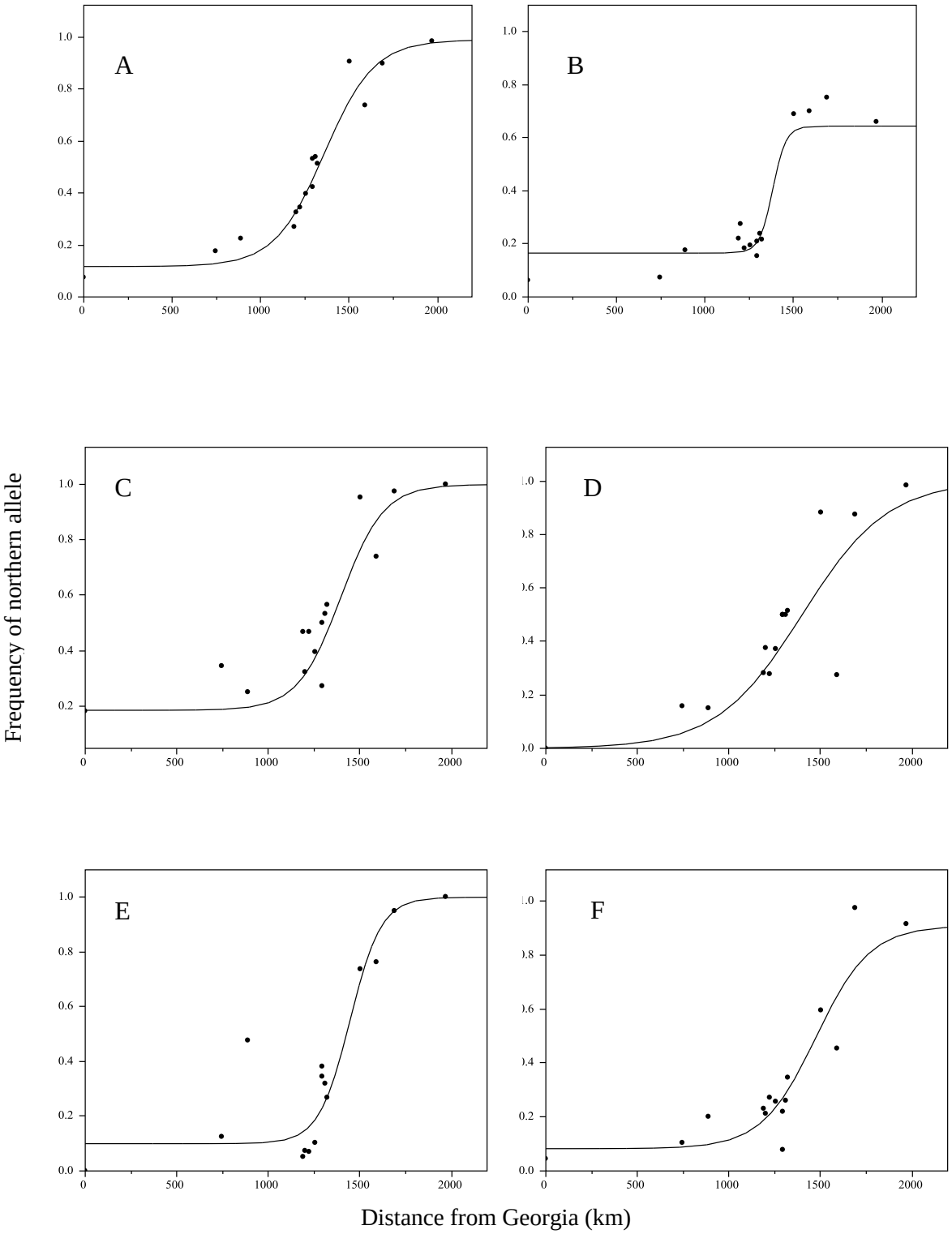
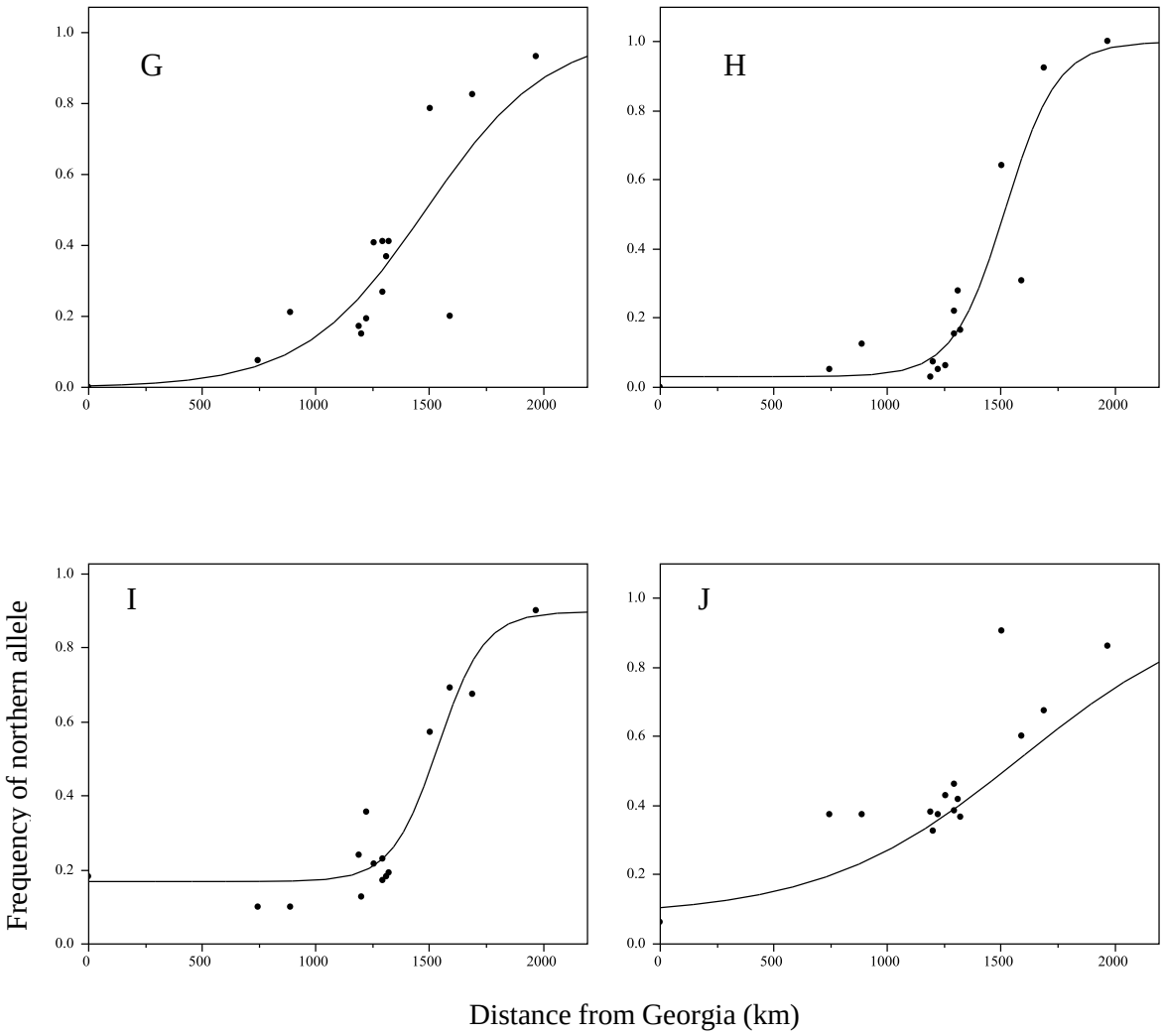

M
N
O
P
Q
Frequency of northern allele
Distance from Georgia (km)

Figure C.3. Ten clines having their centres located to the north of the New Jersey border. A) LDHB 654, B) Elastase, C) Nucleotide diphosphate kinase I, D) LDHB 1033, E) Hemoglobin a2, F) Cytochrome p450, G) 19110, H) Hemoglobin a1, I) Nucleotide diphosphate kinase 2, J) 1173.




G
H
I
J
Frequency of northern allele
Distance from Georgia (km)

Appendix D - Supplementary Material for Chapter Four

Table D.1. Summary of locus pairs in significant linkage disequilibrium at each location visited in the Summer 2008. Values marked with * significant following FDR-adjustment.

Locus 1	Locus 2	Pop	Rij	P
1173	elastase	Cherokee Lane	-0.10714	0.0214
1173	Parvalbumin	Cherokee Lane	-0.45366	0.0376
1176	Hemoglobin a2	Cherokee Lane	0.42	0.0488
280	60S_ribosomal_protein_L6	Cherokee Lane	-0.48759	0.0255
280	G3PD	Cherokee Lane	-0.64583	0.0025
280	Titin Cap	Cherokee Lane	0.43703	0.0404
65	Tropomyosin	Cherokee Lane	0.47786	0.0285
LDHB654	Cytochrome p450	Cherokee Lane	0.59473	0.0053
Cytochrome p450	Activated protein kinase	Cherokee Lane	-0.44317	0.0376
Elastase	ribosomal_protein_S2	Cherokee Lane	0.52307	0.0142
Elastase	60S_ribosomal_protein_L6	Cherokee Lane	0.57661	0.0082
Elastase	nucleotide_diphosphate_kinase2	Cherokee Lane	0.49435	0.0204
ribosomal_protein_S2	nucleotide_diphosphate_kinase2	Cherokee Lane	0.4806	0.0242
ribosomal_protein_S2	TC19110	Cherokee Lane	-0.68461	0.0013
ribosomal_protein_S2	60S_ribosomal_protein_L35	Cherokee Lane	0.44627	0.0363
60S_ribosomal_protein_L6	Titin cap	Cherokee Lane	-0.70232	0.0013
G3PD	TC19110	Cherokee Lane	0.55597	0.0091
Hemoglobin a1	Hemoglobin a2	Cherokee Lane	0.63436	0.0029
NACA	Tropomyosin	Cherokee Lane	0.4818	0.0238
NACA	WAP	Cherokee Lane	0.43595	0.0457
NACA	60S_ribosomal_protein_L35	Cherokee Lane	-0.45412	0.0332
Tropomyosin	WAP	Cherokee Lane	0.55397	0.0111
Tropomyosin	TC19110	Cherokee Lane	0.48456	0.023
1173	Cytochrome p450	Midstream	0.36181	0.0377
1176	Parvalbumin	Midstream	0.59331	0.0004
280	60S_ribosomal_protein_L6	Midstream	0.34442	0.0416
280	ribosomal_protein	Midstream	-0.52307	0.002
280	40S_ribsomal_protein_S17	Midstream	-0.49352	0.0031
65	Cytochrome p450	Midstream	0.37952	0.0228
65	Chymotrypsinogen	Midstream	-0.45544	0.0056
65	Nucleotide diphosphate kinase1	Midstream	-0.3455	0.0472
LDHB1033	G3PD	Midstream	-0.46123	0.0057
LDHB1033	Phosphate carrier protein	Midstream	0.49343	0.0031
LDHB654	Phosphate carrier protein	Midstream	0.55732	0.0007
Cytochrome p450	Activated protein kinase	Midstream	-0.34793	0.0368
Cytochrome p450	Parvalbumin	Midstream	-0.36645	0.0279
Cytochrome p450	WAP	Midstream	-0.39286	0.0184
Elastase	Chymotrypsinogen	Midstream	0.37597	0.0222
Elastase	ribosomal_protein_S2	Midstream	-0.3726	0.0254
Elastase	Tropomyosin	Midstream	-0.34708	0.0373
60S_ribosomal_protein_L6	Activated protein kinase	Midstream	0.33753	0.0458
60S_ribosomal_protein_L6	Tropomyosin	Midstream	0.51586	0.0023
60S_ribosomal_protein_L6	60S_ribosomal_protein_L35	Midstream	-0.48053	0.0045
Nucleotide diphosphate kinase1	Parvalbumin	Midstream	0.36328	0.0369
Nucleotide diphosphate kinase2	Tropomyosin	Midstream	0.40156	0.0146
Nucleotide diphosphate kinase2	WAP	Midstream	0.47303	0.004
Nucleotide diphosphate kinase2	40S_ribsomal_protein_S17	Midstream	-0.33601	0.041
Atrial natriuretic peptide	NACA	Midstream	-0.37438	0.0228
Titin cap	hemoglobin_a2	Midstream	0.37119	0.024
Hemoglobin a1	Hemoglobin a2	Midstream	0.87391	0.0001*
Hemoglobin a1	Translationally controlled tumor protein	Midstream	-0.39826	0.0154
NACA	WAP	Midstream	0.3347	0.0418
Tropomyosin	WAP	Midstream	0.35661	0.0301
Tropomyosin	40S_ribsomal_protein_S17	Midstream	-0.40409	0.014
WAP	Phosphate carrier protein	Midstream	0.3927	0.0169
1176	Phosphate carrier protein	Osprey	0.44192	0.0022
280	60S_ribosomal_protein_L35	Osprey	0.33405	0.0194
65	Elastase	Osprey	-0.28323	0.0497
LDHB1033	LDHB654	Osprey	0.49927	0.0006
LDHB1033	ribosomal_protein	Osprey	0.36676	0.0129
LDHB654	60S_ribosomal_protein_L6	Osprey	-0.37738	0.0083
LDHB654	Atrial natriuretic peptide	Osprey	-0.29352	0.0399

LDHB654	TC19110	Osprey	0.4709	0.001
LDHB654	ribosomal_protein	Osprey	-0.46422	0.0295
LDHB654	Phosphate carrier protein	Osprey	0.39704	0.0054
Cytochrome p450	60S_ribosomal_protein_L6	Osprey	0.28055	0.0495
Cytochrome p450	G3PD	Osprey	0.33053	0.0207
Cytochrome p450	TC19110	Osprey	-0.37646	0.0084
Myoglobin	Translationally controlled tumor protein	Osprey	0.33648	0.0211
Elastase	ribosomal_protein_S2	Osprey	0.31421	0.0312
Elastase	NACA	Osprey	0.35522	0.0129
Elastase	WAP	Osprey	-0.31759	0.0262
Elastase	G3PD	Osprey	0.30811	0.0347
60S_ribosomal_protein_L6	Titin cap	Osprey	0.34429	0.016
60S_ribosomal_protein_L6	Translationally controlled tumor protein	Osprey	0.33566	0.02
Nucleotide diphosphate kinase1	Translationally controlled tumor protein	Osprey	0.2942	0.0437
Activated protein kinase	Translationally controlled tumor protein	Osprey	-0.31445	0.0294
Hemoglobin a1	Hemoglobin a2	Osprey	0.43767	0.0022
Hemoglobin a1	Phosphate carrier protein	Osprey	-0.29289	0.0403
NACA	60S_ribosomal_protein_L35	Osprey	-0.32866	0.0214
Hemoglobin a2	WAP	Osprey	0.30413	0.0333
Tropomyosin	ribosomal_protein	Osprey	0.49331	0.0006
Tropomyosin	Translationally controlled tumor protein	Osprey	0.28858	0.0456
TC19110	Phosphate carrier protein	Osprey	0.33572	0.0188

Table D.2. Summary of locus pairs in significant linkage disequilibrium at each locations visited in the Fall 2009. Values marked with * significant following FDR-adjustment.

Locus 1	Locus 2	Pop	Rij	P
1176	280	Cherokee Lane	-0.32877	0.0401
1176	Parvalbumin	Cherokee Lane	0.36991	0.0209
280	Chymotrypsinogen	Cherokee Lane	-0.34328	0.0299
280	Atrial natriuretic peptide	Cherokee Lane	0.34272	0.0346
280	TC19110	Cherokee Lane	-0.47033	0.0029
65	Tropomyosin	Cherokee Lane	0.4603	0.0051
LDHB 1033	LDHB 654	Cherokee Lane	0.39733	0.0157
LDHB 654	ribosomal_protein	Cherokee Lane	-0.39814	0.0141
LDHB 654	Phosphate carrier protein	Cherokee Lane	0.37789	0.0183
Cytochrome p450	Parvalbumin	Cherokee Lane	-0.40328	0.0108
Cytochrome p450	Hemoglobin a2	Cherokee Lane	-0.3182	0.0442
ribosomal_protein_S2	40S_ribsomal_protein_S17	Cherokee Lane	0.3366	0.0434
60S_ribosomal_protein_L6	Hemoglobin a2	Cherokee Lane	0.42991	0.0065
60S_ribosomal_protein_L6	Phosphate carrier protein	Cherokee Lane	0.31176	0.0486
Nucleotide diphosphate kinase1	Atrial natriuretic peptide	Cherokee Lane	0.40675	0.0122
Activated protein kinase	Atrial natriuretic peptide	Cherokee Lane	0.53004	0.0011
Activated protein kinase	Parvalbumin	Cherokee Lane	-0.3366	0.0333
Activated protein kinase	Titin cap	Cherokee Lane	0.39521	0.0124
Nucleotide diphosphate kinase 2	Parvalbumin	Cherokee Lane	-0.33323	0.04
Atrial natriuretic peptide	WAP	Cherokee Lane	0.32249	0.0468
Parvalbumin	Hemoglobin a2	Cherokee Lane	0.34002	0.0315
Titin Cap	60S_ribosomal_protein_L35	Cherokee Lane	-0.31838	0.0468
Hemoglobin a1	Hemoglobin a2	Cherokee Lane	0.43251	0.0062
Hemoglobin a2	ribosomal_protein	Cherokee Lane	-0.33818	0.0347
Hemoglobin a2	60S_ribosomal_protein_L35	Cherokee Lane	-0.33274	0.0377
Tropomyosin	40S_ribsomal_protein_S17	Cherokee Lane	0.37352	0.0197
40S_ribsomal_protein_S17	Phosphate carrier protein	Cherokee Lane	-0.37212	0.0201
1173	NACA	Crescent Park Woods	-0.41168	0.0112
1173	Hemoglobin a2	Crescent Park Woods	-0.3454	0.0332
1176	LDHB 654	Crescent Park Woods	0.40015	0.0125
1176	Activated protein kinase	Crescent Park Woods	0.45108	0.0043
280	LDHB 654	Crescent Park Woods	-0.31637	0.0482
280	Translationally controlled tumor protein	Crescent Park Woods	0.31385	0.0471
65	Atrial natriuretic peptide	Crescent Park Woods	-0.37957	0.0193
LDHB 1033	LDHB 654	Crescent Park Woods	0.76375	0.0001*
LDHB 654	40S_ribsomal_protein_S17	Crescent Park Woods	-0.3687	0.027
Cytochrome p450	60S_ribosomal_protein_L6	Crescent Park Woods	-0.53639	0.0009
Myoglobin	Nucleotide diphosphate kinase2	Crescent Park Woods	0.53525	0.0008
Myoglobin	Tropomyosin	Crescent Park Woods	0.38614	0.0146
Elastase	Chymotrypsinogen	Crescent Park Woods	-0.36286	0.0217
Elastase	Phosphate carrier protein	Crescent Park Woods	0.34439	0.0294
ribosomal_protein_S2	G3PD	Crescent Park Woods	0.34362	0.0392
ribosomal_protein_S2	ribosomal_protein	Crescent Park Woods	0.37218	0.0277

60S_ribosomal_protein_L6	Tropomyosin	Crescent Park Woods	-0.34039	0.0359
60S_ribosomal_protein_L6	TC19110	Crescent Park Woods	-0.40788	0.0119
Nucleotide diphosphate kinase1	Hemoglobin a1	Crescent Park Woods	0.32882	0.0427
Nucleotide diphosphate kinase1	WAP	Crescent Park Woods	0.32348	0.0434
Atrial natriuretic peptide	60S_ribosomal_protein_L35	Crescent Park Woods	0.35761	0.0275
G3PD	40S_ribsomal_protein_S17	Crescent Park Woods	-0.47037	0.0042
Parvalbumin	NACA	Crescent Park Woods	-0.35189	0.026
Parvalbumin	Hemoglobin a2	Crescent Park Woods	-0.33499	0.0341
Hemoglobin a1	Hemoglobin a2	Crescent Park Woods	0.52792	0.001
Translationally controlled tumor protein	Phosphate carrier protein	Crescent Park Woods	-0.33367	0.0348
1173	G3PD	Hidden Harbour	-0.32415	0.0247
1176	Titin cap	Hidden Harbour	0.342	0.019
280	hemoglobin_a2	Hidden Harbour	-0.31749	0.0263
LDHB 1033	LDHB 654	Hidden Harbour	0.72259	0.0001*
LDHB 654	Tropomyosin	Hidden Harbour	-0.48756	0.0006
Myoglobin	Hemoglobin a2	Hidden Harbour	0.36092	0.0107
Chymotrypsinogen	Tropomyosin	Hidden Harbour	-0.29892	0.0345
Chymotrypsinogen	WAP	Hidden Harbour	-0.31743	0.0248
ribosomal_protein_S2		Hidden Harbour	-0.33935	0.0228
60S_ribosomal_protein_L6	Atrial natriuretic peptide	Hidden Harbour	-0.28878	0.0412
60S_ribosomal_protein_L6	TC19110	Hidden Harbour	-0.28394	0.0447
Nucleotide diphosphate kinase1	Nucleotide diphosphate kinase2	Hidden Harbour	-0.31241	0.0288
Activated protein kinase	G3PD	Hidden Harbour	0.38507	0.0065
Activated protein kinase	60S_ribosomal_protein_L35	Hidden Harbour	0.28062	0.0472
Atrial natriuretic peptide	Hemoglobin a2	Hidden Harbour	-0.29504	0.037
Atrial natriuretic peptide	ribosomal_protein	Hidden Harbour	0.32188	0.0242
Hemoglobin a1	Hemoglobin a2	Hidden Harbour	0.54543	0.0001*
Hemoglobin a1	40S_ribsomal_protein_S17	Hidden Harbour	0.28738	0.0488
ribosomal_protein	60S_ribosomal_protein_L35	Hidden Harbour	-0.31561	0.0272
ribosomal_protein	Translationally controlled tumor protein	Hidden Harbour	0.29998	0.0357
60S_ribosomal_protein_L35	Translationally controlled tumor protein	Hidden Harbour	-0.28736	0.0422
1173	Phosphate carrier protein	Margarita	0.43825	0.0106
1176	Elastase	Margarita	-0.40954	0.014
1176	Nucleotide diphosphate kinase1	Margarita	-0.36348	0.0341
1176	hemoglobin_a1	Margarita	-0.47767	0.0042
1176	Hemoglobin a2	Margarita	-0.36242	0.0297
280	WAP	Margarita	0.35078	0.0408
65	Phosphate carrier protein	Margarita	0.36246	0.032
LDHB 1033	LDHB 654	Margarita	0.67892	0.0003
LDHB 1033	Activated protein kinase	Margarita	-0.36488	0.0309
LDHB 654	Tropomyosin	Margarita	-0.36236	0.0472
Cytochrome p450	Nucleotide diphosphate kinase1	Margarita	0.37652	0.0305
Cytochrome p450	Activated protein kinase	Margarita	-0.37518	0.0264
Myoglobin	60S_ribosomal_protein_L6	Margarita	-0.60459	0.0003
Myoglobin	Hemoglobin a2	Margarita	-0.32906	0.0483
Myoglobin	60S_ribosomal_protein_L35	Margarita	0.34222	0.04
Chymotrypsinogen	Parvalbumin	Margarita	-0.41588	0.0139
ribosomal_protein_S2	Titin cap	Margarita	0.37215	0.0353
Nucleotide diphosphate kinase1	WAP	Margarita	-0.51567	0.0031
Nucleotide diphosphate kinase1	40S_ribsomal_protein_S17	Margarita	0.50741	0.0036
Hemoglobin a1	Hemoglobin a2	Margarita	0.57846	0.0005
WAP	40S_ribsomal_protein_S17	Margarita	-0.40478	0.0166
WAP	Translationally controlled tumor protein	Margarita	-0.37021	0.0309
65	60S_ribosomal_protein_L35	Midstream	0.29168	0.0433
LDHB 1033	LDHB 654	Midstream	0.36669	0.0175
LDHB 1033	Activated protein kinase	Midstream	0.32126	0.0276
LDHB 1033	Phosphate carrier protein	Midstream	0.46555	0.0014
LDHB 654	ribosomal_protein_S2	Midstream	0.32295	0.0386
LDHB 654	Parvalbumin	Midstream	-0.30651	0.0444
LDHB 654	Translationally controlled tumor protein	Midstream	-0.31586	0.0383
Elastase	ribosomal_protein_S2	Midstream	0.3075	0.0391
ribosomal_protein_S2	Hemoglobin a2	Midstream	-0.31129	0.0347
ribosomal_protein_S2	60S_ribosomal_protein_L35	Midstream	0.34722	0.0185
60S_ribosomal_protein_L6	Titin cap	Midstream	-0.28359	0.0494
60S_ribosomal_protein_L6	NACA	Midstream	-0.39264	0.0065
60S_ribosomal_protein_L6	Hemoglobin a2	Midstream	0.28932	0.045
60S_ribosomal_protein_L6	WAP	Midstream	-0.32855	0.0243
Activated protein kinase	NACA	Midstream	0.35952	0.0127
Activated protein kinase	WAP	Midstream	0.34278	0.0188
Nucleotide diphosphate kinase 2	Hemoglobin a2	Midstream	-0.35481	0.015

Parvalbumin	WAP	Midstream	-0.28585	0.05
Titin Cap	Tropomyosin	Midstream	0.32048	0.0264
Hemoglobin a1	Hemoglobin a2	Midstream	0.3557	0.0137
Hemoglobin a2	Translationally controlled tumor protein	Midstream	-0.3275	0.0233
WAP	40S_ribsomal_protein_S17	Midstream	0.32855	0.0243
TC19110	ribosomal_protein	Midstream	-0.34699	0.0174
1173	hemoglobin_a1	Osprey	0.41603	0.0103
1173	Hemoglobin a2	Osprey	0.32945	0.0423
1176	Hemoglobin a2	Osprey	-0.43825	0.0062
280	G3PD	Osprey	0.41193	0.0101
280	NACA	Osprey	-0.3252	0.0423
65	Myoglobin	Osprey	0.39291	0.0141
65	60S_ribosomal_protein_L35	Osprey	0.34299	0.0322
LDHB 1033	LDHB 654	Osprey	0.34219	0.0349
LDHB 1033	Tropomyosin	Osprey	0.33744	0.0375
LDHB 1033	Phosphate carrier protein	Osprey	0.39421	0.0151
LDHB 654	Titin cap	Osprey	-0.32877	0.0427
LDHB 654	Phosphate carrier protein	Osprey	0.3139	0.05
Cytochrome p450	60S_ribosomal_protein_L6	Osprey	0.39082	0.016
Cytochrome p450	Atrial natriuretic peptide	Osprey	0.32793	0.0432
Myoglobin	40S_ribsomal_protein_S17	Osprey	0.33199	0.0407
Chymotrypsinogen	Hemoglobin a1	Osprey	-0.39147	0.0145
60S_ribosomal_protein_L6	Nucleotide diphosphate kinase 2	Osprey	-0.32775	0.0433
Nucleotide diphosphate kinase 2	40S_ribsomal_protein_S17	Osprey	-0.33053	0.0416
Atrial natriuretic peptide	ribosomal_protein	Osprey	0.33234	0.0432
Atrial natriuretic peptide	Phosphate carrier protein	Osprey	-0.3251	0.0451
Hemoglobin a1	Hemoglobin a2	Osprey	0.67659	0.0001*
Hemoglobin a1	Tropomyosin	Osprey	0.37277	0.0199
Hemoglobin a1	60S_ribosomal_protein_L35	Osprey	-0.35847	0.0252
Hemoglobin a2	Tropomyosin	Osprey	0.44721	0.0052
Tropomyosin	Phosphate carrier protein	Osprey	0.0084	0.0093
1173	G3PD	Parker & 1st	-0.3368	0.0291
1176	Myoglobin	Parker & 1st	0.37162	0.0137
1176	Parvalbumin	Parker & 1st	-0.30287	0.0422
280	LDHB 1033	Parker & 1st	0.31607	0.034
280	Nucleotide diphosphate kinase1	Parker & 1st	-0.45066	0.0039
65	Hemoglobin a2	Parker & 1st	-0.31925	0.0342
LDHB 1033	LDHB 654	Parker & 1st	0.65212	0.0001*
LDHB 1033	Myoglobin	Parker & 1st	-0.38264	0.0111
LDHB 1033	G3PD	Parker & 1st	-0.35755	0.0165
LDHB 1033	Phosphate carrier protein	Parker & 1st	0.38239	0.0103
LDHB 654	G3PD	Parker & 1st	-0.45277	0.003
LDHB 654	Translationally controlled tumor protein	Parker & 1st	-0.44437	0.0036
LDHB 654	Phosphate carrier protein	Parker & 1st	0.58056	0.0001*
Cytochrome p450	60S_ribosomal_protein_L6	Parker & 1st	0.32153	0.035
Myoglobin	Chymotrypsinogen	Parker & 1st	-0.31332	0.0377
Myoglobin	Phosphate carrier protein	Parker & 1st	-0.32016	0.0337
Elastase	Chymotrypsinogen	Parker & 1st	-0.32125	0.0331
Elastase	60S_ribosomal_protein_L6	Parker & 1st	-0.32873	0.0292
Elastase	Nucleotide diphosphate kinase1	Parker & 1st	0.36126	0.0223
Elastase	60S_ribosomal_protein_L35	Parker & 1st	-0.31079	0.0392
Chymotrypsinogen	Parvalbumin	Parker & 1st	-0.35016	0.0188
Chymotrypsinogen	Tropomyosin	Parker & 1st	-0.31491	0.0346
ribosomal_protein_S2	40S_ribsomal_protein_S17	Parker & 1st	-0.31221	0.0483
60S_ribosomal_protein_L6	nucleotide_diphosphate_kinase1	Parker & 1st	0.31474	0.0465
60S_ribosomal_protein_L6	ribosomal_protein	Parker & 1st	0.35503	0.0199
Activated protein kinase	Tropomyosin	Parker & 1st	0.41015	0.0059
Nucleotide diphosphate kinase 2	ribosomal_protein	Parker & 1st	-0.30503	0.043
Nucleotide diphosphate kinase 2	Parvalbumin	Parker & 1st	-0.30928	0.038
Atrial natriuretic peptide	Tropomyosin	Parker & 1st	-0.34376	0.0211
G3PD	Parvalbumin	Parker & 1st	0.46606	0.0018
G3PD	WAP	Parker & 1st	0.29823	0.0479
Titin Cap	Hemoglobin a2	Parker & 1st	0.3722	0.0147
Hemoglobin a1	Hemoglobin a2	Parker & 1st	0.48161	0.0012
40S_ribsomal_protein_S17	Phosphate carrier protein	Parker & 1st	-0.31063	0.0441
1173	G3PD	Rancosas	-0.392	0.0157
280	Myoglobin	Rancosas	0.36379	0.0231
280	Elastase	Rancosas	-0.31574	0.0486
280	Nucleotide diphosphate kinase2	Rancosas	0.32954	0.0239
65	60S_ribosomal_protein_L6	Rancosas	-0.43759	0.007

LDHB 1033	LDHB 654	Rancosas	0.65188	0.0001*
LDHB 1033	60S_ribosomal_protein_L6	Rancosas	0.50051	0.002
LDHB 1033	TC19110	Rancosas	-0.32515	0.045
LDHB 654	G3PD	Rancosas	0.42259	0.0083
LDHB 654	NACA	Rancosas	-0.35893	0.025
Cytochrome p450		Rancosas	0.38661	0.0145
Cytochrome p450	WAP	Rancosas	-0.341	0.0332
Myoglobin	Chymotrypsinogen	Rancosas	-0.3567	0.0241
Chymotrypsinogen	TC19110	Rancosas	-0.37311	0.0183
Chymotrypsinogen	Translationally controlled tumor protein	Rancosas	0.33135	0.0361
ribosomal_protein_S2	60S_ribosomal_protein_L6	Rancosas	-0.40614	0.0148
60S_ribosomal_protein_L6	TC19110	Rancosas	-0.41283	0.009
60S_ribosomal_protein_L6	60S_ribosomal_protein_L35	Rancosas	0.31008	0.0499
Activated protein kinase	Atrial natriuretic peptide	Rancosas	-0.32246	0.044
Activated protein kinase	Hemoglobin a2	Rancosas	-0.31388	0.0471
Nucleotide diphosphate kinase 2	NACA	Rancosas	0.3475	0.028
Nucleotide diphosphate kinase 2	40S_ribsomal_protein_S17	Rancosas	0.37703	0.0171
Atrial natriuretic peptide	Titin cap	Rancosas	0.33941	0.034
G3PD	Parvalbumin	Rancosas	0.46099	0.0036
Hemoglobin a1	Hemoglobin a2	Rancosas	0.73362	0.0001*
NACA	WAP	Rancosas	0.39961	0.0126
Tropomyosin	ribosomal_protein	Rancosas	0.0233	0.0369
Tropomyosin	40S_ribsomal_protein_S17	Rancosas	-0.35268	0.0257
Tropomyosin	Translationally controlled tumor protein	Rancosas	-0.41641	0.0084

Table D.3. Allele frequency differentials (δ) between northern and southern parental populations used for calculating hybrid index arranged in order of decreasing δ.

SNP Locus Name	δ
65_135*	0.70
40S ribosomal protein S17*	0.55
myoglobin	0.39
activated protein kinase	0.37
280_83	0.37
phosphate carrier protein mitochondrial precursor	0.36
ribosomal protein S2	0.28
glyceraldehyde 3 phosphate dehydrogenase	0.26
60S ribosomal protein L35	0.25
LDHB 654	0.24
titin cap	0.24
TC19110_377	0.24
LDHB 1033	0.23
naca	0.23
hemoglobin a2	0.22
60S ribosomal protein L6	0.22
parvalbumin 3	0.21
WAP	0.20
tropomyosin	0.18
hemoglobin a1	0.14
atrial naturetic peptide	0.12
cytochrome p450	0.12
ribsomal protein	0.11
1176_169	0.10
nucleotide diphosphate kinase	0.10
chymotrypsinogen	0.07
nucleotide diphosphate kinase	0.05
translationally controlled tumor protein	0.01
1173_79	0.01
elastase	0.00

Table D.4. Patterns of cytonuclear disequilibrium based on significant differences (either excess as indicated by "+" or deficit as indicated by "-" at $\alpha = 0.05$) between observed versus expected (in parentheses) cytonuclear types at A) cytochrome p450, B) Atrial Natriuretic Peptide, C) Glyceraldehyde 3 phosphate dehydrogenase, D) LDHB 654.

A)

Mitochondrial Type	Nuclear Genotype		
	AA	Aa	aa
M	9 (6.36) +	3 (6.36) -	2 (1.27) =
m	1 (3.64) -	7 (3.64) +	0 (0.727) =

B)

Mitochondrial Type	Nuclear Genotype		
	AA	Aa	aa
M	3 (2.85) =	8 (4.57) +	5 (8.57) -
m	2 (2.14) =	0 (3.43) -	10 (6.43) +

C)

Mitochondrial Type	Nuclear Genotype		
	AA	Aa	aa
M	0 (0.86) =	13 (8.23) +	3 (6.92) -
m	2 (1.14) =	6 (10.78) -	13 (9.08) +

D)

Mitochondrial Type	Nuclear Genotype		
	AA	Aa	aa
M	11(6.72) +	9 (10.39) =	2 (4.89) -
m	0 (4.28) -	8 (6.11) =	6 (3.11) +

Figure D.1. Optimization of sperm concentration.