STARS

Electronic Theses and Dissertations, 2004-2019

2007

Genetic And Phenotypic Evolution In The Ornate Chorus Frog (pseudacris Ornata): Testing The Relative Roles Of Natural Selection,

Jacob Degner
University of Central Florida

Part of the Biology Commons

Find similar works at: https://stars.library.ucf.edu/etd

University of Central Florida Libraries http://library.ucf.edu

Showcase of Text, Archives, Research & Scholarship

GENETIC AND PHENOTYPIC EVOLUTION IN THE ORNATE CHORUS FROG
(*PSEUDACRIS ORNATA*): TESTING THE RELATIVE ROLES OF NATURAL SELECTION,
MIGRATION, AND GENETIC DRIFT

by

JACOB FRANK DEGNER
B.S. The University of Illinois at Urbana-Champaign, 2003

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science in the
Department of Biology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Summer Term
2007

Major Professor: Eric A. Hoffman

ABSTRACT

Understanding how migration, genetic drift, and natural selection interact to maintain the genetic and phenotypic variation we observe in natural populations is a central goal of population genetics. Amphibians provide excellent model organisms for investigating the interplay between these evolutionary forces because amphibians are generally characterized by limited dispersal abilities, high philopatry, and are obligately associated with the areas around suitable habitats (e.g. breeding ponds). Thus, on relatively small geographic scales, the relative effects of all of these evolutionary forces can be studied together. Here, we study the interaction of migration, genetic drift, natural selection, and historical process in the ornate chorus frog (*Pseudacris ornata*). We report the development and characterization of 10 polymorphic microsatellite genetic markers. Number of alleles per locus ranged from 2 to 21 averaging 9.2 and expected heterozygosities ranged from 0.10 to 0.97 averaging 0.52. However, in an analysis of two populations, three locus-by-population comparisons exhibited significant heterozygote deficiencies and indicated that null alleles may be present some loci. Furthermore, we characterized genetic structure and historical biogeographic patterns in *P. ornata* using these microsatellite markers along with mitochondrial DNA sequence data. Our data indicate that in these frogs, migration may play a large role in determining population structure as pairwise estimates of F_{ST} were relatively small ranging from 0.04 to 0.12 (global $F_{ST} = 0.083$). Additionally, we observed an overall pattern of isolation-by-distance in neutral genetic markers across the species range. Moreover, our data suggest that the Apalachicola River basin does not impede gene flow in *P. ornata* as it does in many vertebrate taxa. Interestingly, we identified significant genetic structure between populations separated by only 6 km. However, this fine

scale genetic structure was only present in the more urbanized of two widespread sampling localities. Finally, in this study, we demonstrated that there was a significant correlation between the frequency of green frogs and latitude. There was a higher frequency of green frogs in southern samples and a lower frequency of green frogs in northern samples. However, when we interpreted this phenotypic cline in light of the overall pattern of isolation-by-distance, it was apparent that the neutral evolutionary forces of genetic drift and migration could explain the cline, and the invocation of natural selection was not necessary.

This thesis is dedicated to my grandmother Thelma Hershey and to the memory of her late husband Frank Hershey. My earliest interest in biology was inspired by their guided tours of the Northwoods of Wisconsin, and their generous support has made my higher education possible.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 –GENERAL INTRODUCTION

Understanding population genetic structure in nature is a central goal of population genetics (Hartl and Clark 1997). This includes describing the amount of genetic and phenotypic variation, describing how existing variation is partitioned geographically, and most importantly explaining the mechanisms maintaining and structuring natural variation. Population genetics theory predicts how migration, mutation, random genetic drift, and natural selection act to maintain variation and structure it within populations. Mutation is the only source of new genetic variation and alone would continue to increase variability within a population. However, in natural, sexually-reproducing populations, random genetic drift limits the amount of variation within a population. Genetic drift causes the random loss of genetic variation created by mutation, through the stochastic allele frequency changes after each generation of random mating.

Natural selection can act in various ways to increase, decrease, or maintain constant levels of natural variation. Natural selection can limit genetic variability by disproportionate survival of certain genetic variants causing other variants to be purged from the population over time. Alternatively, natural selection can maintain genetic variation by favoring multiple allelic variants, by favoring the heterozygous state of two or more allelic variants, or by favoring alleles with low frequency in a population (Hartl and Clark 1997). Despite the presumed theoretical importance of natural selection to the maintenance of genetic diversity and patterns of genetic structure, relatively few studies have directly assessed the extent that selection has acted on a particular trait (Chevillon *et al.* 1995, Mithon *et al.* 1995, Taylor *et al.* 1995, Hoffman *et al.* 2006). One limitation of studying natural selection on phenotypic diversity in wild populations

is the rarity of naturally occurring phenotypic polymorphisms that can be attributed to and scored as alleles from a single genetic locus, and hence treated by much of population genetic theory.

Population substructure, a virtually universal property of living systems, affects patterns of variation (Hartl and Clark 1997). Natural populations are essentially never truly panmictic and usually exist in schools, colonies, or other aggregations. Further, suitable habitat for natural populations occurs in patches separated by some geographic distance. Habitat patchiness and the aggregation of individuals within populations limit the probability of individuals encountering and mating with individuals occupying different habitat patches or aggregations. When subpopulations are separated from one another by these or other isolating mechanisms, genetic differentiation (differing frequencies of genetic variants) occurs among subpopulations by the micro-evolutionary forces of mutation, random genetic drift, and selection acting independently in different subpopulations. Migration rates between subpopulations limit the degree to which subpopulations evolve independently and hence the level of population substructure.

Amphibians make ideal model organisms for studying naturally occurring population-genetic structure because amphibians are often characterized by limited dispersal and high levels of philopatry leading to the potential for genetic isolation over small geographic distances (Beebee 1996, 2005). Over time, this can lead to the development of independent evolutionary units within the range of a particular species (Beebee 2005). Further, amphibian populations in North America are declining more rapidly than many other terrestrial vertebrate taxa (e.g., birds and mammals; Stuart *et al*. 2004). Hence, the need for genetic studies to examine basic population biology (e.g., migration rates, effective population sizes, historical occurrence of population crashes) and to define appropriate management units and strategies is heightened.

However, the conservation genetics of few species have been studied or even had molecular tools (e.g. microsatellites) developed to initiate such studies (Jehle and Arntzen 2002).

Microsatellites [also called simple sequence repeats (SSRs) or simple tandem repeats (STRs)] are ubiquitous, abundant, and highly polymorphic in eukaryotic nuclear genomes. Perfect microsatellites are composed of uninterrupted repeated sequences of DNA where the repeat unit is from two to six base pairs (e.g. GATA), while imperfect microsatellites may be a series of repeat units interrupted by non-repetitive DNA. Since their discovery (Hamada *et al.* 1982) and first use as molecular markers (Weber and May 1989, Litt and Luty 1989, Tautz 1989) microsatellites have gained widespread use in many biological disciplines. Their high polymorphism, abundance in eukaryotic genomes, ease of isolation, and ease of scoring make them ideal for use in genome mapping, forensics, population and evolutionary biology, and conservation genetics. However, because microsatellites occur in non-coding regions of DNA that have high effective mutation rates compared to coding DNA, they are generally species-specific, although cross-species amplification and use in sister taxa is not uncommon.

While microsatellites are ideal for determining current population structure their high mutation rates limit evolutionary inference of deeply divergent evolutionary lineages because homology cannot be distinguished from recurrent mutation. High confidence in deeper evolutionary events can be obtained from gene sequence data. For this reason, I use mitochondrial 12S ribosomal RNA and Cytochrome-b gene sequence data to investigate the evolutionary relationship among populations throughout the extant range of *P. ornata*. Further, this mitochondrial sequence data is an independent line of evidence that can support inferences about current population-genetic structure made from analysis of microsatellite data.

<u>Study Species</u>

The Ornate Chorus Frog (*Pseudacris ornata*) is a small terrestrial frog (30-40 mm SVL) occurring in the coastal plains of the southeastern United States from Louisiana to North Carolina south to central Florida (Fig. 1). Sexual maturity is reached within the first year of life, and adult frogs breed during the winter (November-March) in temporary wetlands of Longleaf Pine ecosystems, mixed hardwood forests, Cypress swamps, flooded fields, and roadside ditches (Caldwell 1987). Males enter ponds before females and begin to call, attracting gravid females. Males call most frequently for several nights following rainfall when temperatures are between 3 and 18 °C (Gerhardt 1973). The total sex ratio of males and females entering breeding ponds is not significantly different than 1:1 although males are more common early and females more common late (Caldwell 1987). Juvenile frogs emerge from breeding ponds in April and May at 18-23 mm and immigrate to surrounding forest habitat. The majority of frogs breed for only one year, but occasionally frogs have returned to a breeding pond in consecutive years (Caldwell 1987).

Populations of the Ornate Chorus Frog isolated by relatively short geographic distances show a large degree of mitochondrial DNA divergence (Moriarty and Cannatella 2004). Single specimens from Liberty County in the Florida Panhandle and Barbour County in southeast Alabama formed a clade and showed 0.7% uncorrected sequence divergence in the stretch of DNA mentioned above. These specimens showed 1.5% sequence divergence with a single specimen from Aiken South Carolina. However, since data from only three samples were

4

presented, little inference about overall genetic structure or phylogeographic patterns can be made beyond the fact that a large degree of variation may exist within this species.

In addition, a post metamorphic skin color polymorphism exists in the Ornate Chorus Frog. This polymorphism is controlled by a single gene showing complete dominance with green dominant to brown/grey/copper (Blouin 1986). Although non-green color variants may change through the lifetime of a frog or under differing environmental conditions, green color morphs do not change to non-green and vice versa (Harkey and Semlitsch 1988, Blouin 1986). Thus, this system offers a unique opportunity to study a trait that potentially is under selection and which is readily observed in wild populations.

References

Beebee T (1996) *Ecology and Conservation of Amphibians* Chapman and Hall, London.

Beebee T (2005) Conservation genetics of amphibians. *Heredity* **95**, 423-427.

Blouin M (1989) Inheritance of a naturally occurring color polymorphism in the Ornate Chorus Frog. *Copeia* **4**, 1056-1056.

Caldwell J (1987) Demography and life history of two species of chorus frog (Anura: Hylidae) in South Carolina. *Copeia* **2**, 114-127.

Chevillo C, Paseur N, Marquine M, Hayse D, Raymond M (1995) Population structure and dynamics of selected genes in the mosquito (*Culex pipiens*). *Evolution* **49**, 997-1007.

Gerdhardt H (1973) Reproductive interactions between *Hyla crucifer* and *Pseudacris ornata*. *American Midland Naturalist* **89**, 81-88.

Hamada H, Petrino G, Kakunga T (1982) A novel repeated element with Z-DNA forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proceedings of the National Academy of Sciences* **79**, 6465-6469.

Harkey G, Semlitsch R (1988) Effects of temperature on growth, development and color polymorphism in the Ornate Chorus Frog. *Copeia*, 1001-1007.

Hartl DL, Clark AG (1997) *Principles of population genetics*, 3rd edn. Sinauer Associates, Sunderland MA.

Jehle R, Artzen JW (2002) Review: Microsatellite Markers in Amphibian Conservation Genetics. *Herpetological Journal* **12**, 1-9.

Litt M, Luty J (1989) A hypervariable microsatellite revealed by in vitro amplification of a

    dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human

    Genetics* **44**, 397-401.

Mithon R, Raybould A, Giamoustaris A (1995) Divergent selection for secondary metabolites

    between wild populations of *Brassicola oleracea* and its implications for plant-herbivore

    interactions. *Heredity* **75**, 478-484.

Moriarty EC, Cannatella DC (2004) Phylogenetic relationships of the North American chorus

    frogs (*Pseudacris*: Hylidae). *Molecular Phylogenetics and Evolution* **30**, 409-420.

Stuart SN, Chanson JS, Cox NA*, et al.* (2004) Status and trends of amphibian declines and

    extinctions worldwide. *Science* **306**, 1783 - 1786.

Tautz R (1989) Hypervariability of simple sequences as a general source for polymorphic DNA

    markers. *Nucleic Acids Research* **17**, 6463-6471.

Tayler M, Shen Y, Kreitman M (1995) A population genetic test of selection at the molecular

    level. *Science* **270**, 1497-1499.

Weber J, May P (1989) Abundant class of human DNA polymorphisms which can be typed

    using the polymerase chain reaction. *American Journal of Human Genetics* **44**, 388-396.

CHAPTER 2 – DEVELOPMENT OF 10 POLYMORHPIC MICROSATELLITE LOCI IN THE ORNATE CHORUS FROG (*PSEUDACRIS ORNATA*)

Introduction

Amphibian populations in North America are declining more rapidly than many other terrestrial vertebrate taxa (e.g., birds and mammals; Stuart *et al*. 2004), and yet their biology remains comparatively understudied. Hence, there is a need for genetic studies to examine basic population biology (e.g., migration rates, effective population sizes, historical occurrence of population crashes) and to define appropriate management units and strategies. However, the conservation genetics of few amphibian species has been studied and even fewer species have had molecular tools (e.g., microsatellites) developed to initiate such studies (Jehle and Arntzen 2002). The Ornate Chorus Frog (*Pseudacris ornata*) is a small terrestrial frog (30-40 mm SVL) occurring in the coastal plains of the southeastern United States (Fig. 1) from Louisiana to North Carolina south to central Florida. Initial evidence suggests some degree of phylogeographic structure in this species (Moriarty and Cannatella 2004) although the extent of this structure has yet to be investigated. Additionally, this frog exhibits a conspicuous Mendelian color polymorphism, offering a unique opportunity to study the characteristics of natural selection in wild populations (Hoffman and Blouin 2000). In this study, we characterize 10 polymorphic microsatellite loci in the ornate chorus frog. These loci were derived from a biotin enrichment protocol and from cross-species amplification.

8

Methods and Results

Five novel microsatellite loci were developed from a whole-genome enrichment for (GATA)$_n$ repeats following the protocol of Hoffman *et al.* 2003. Whole genomic DNA was isolated using DNeasy DNA isolation kits (Quiagen, Irvine CA) according to the manufacturer's protocols. Random DNA fragments (~200-2500 bp) were generated using degenerate oligonucleotide-primed polymerase chain reaction (DOP-PCR), using the K6-MW primer developed by Macas *et al*. 1996. DOP-PCR was performed in 50 µL reactions having the following final concentrations: 3 mM MgCl$_2$, 1X PCR buffer (10 mM Tris-HCL, 50 mM KCL and 0.1% Triton X-100), 0.2 uM each dNTP, 1 uM K6-MW primer, and 1 Unit *Taq* polymerase, 10-100 ng template DNA. DOP-PCR Temperature cycling parameters were 95 $^{\circ}$C for 2:00, 5 × (95 $^{\circ}$C for 30 s. 30 $^{\circ}$C to 72 $^{\circ}$C at 1 $^{\circ}$C /s, 72 for 48 sec) 29 × (95 $^{\circ}$C for 30 s. 56 $^{\circ}$C for 1.5 min, 72 $^{\circ}$C for 3 min), 72 $^{\circ}$C for 20 min. Successful DOP-PCR was verified using gel electrophoresis and ethidium bromide staining.

DOP-PCR fragments were enriched for microsatellites using a (GATA)$_8$ repeat motif biotinylated on the 5' end and with a terminating loop on the 3' end (Integrated DNA Technologies, Coralville, IA). Biotin probes were hybridized to DOP-PCR fragments containing (GATA)$_n$ repeats in 65 µL 6X standard saline citrate (SSC). The hybridization was performed under the following temperature conditions: 5 minutes at 98 $^{\circ}$C. The temperature was decreased at a rate of 0.1 $^{\circ}$C/s to 50 $^{\circ}$C and held for 25 minutes. The hybridization mixture was immediately combined with streptavidin-coated magnetic beads (Promega Magnasphere, Madison, WI) in 100 µL total 6X SSC with 20 minutes gentle shaking at room temperature.

Hybrid-bead complexes were washed 6 times total. Wash 1 and 2 consisted of 5 minutes gentle mixing at room temp with 1 mL 2X SSC, 0.1% sodium dodecyl sulfate (SDS), wash 3 and 4 were 5 minutes gentle mixing at room temp with 1 mL 1X SSC, wash 5 was 2 minutes at 72 $^{\circ}$C with 1 mL 1X SSC, and wash 6 was 5 minutes at 72 $^{\circ}$C with 1 mL 1X SSC. Single stranded enriched DNA was eluted in a final wash of 65 µL Tris-EDTA buffer (10 mM Tris HCL pH 8.0, 0.1 mM EDTA) at 97 $^{\circ}$C. Repeat enriched DNA was made double stranded by re-running the DOP-PCR as described above using 2 µL of elution as template. Enriched DOP-PCR product was cloned using TOPO TA Cloning kits according to manufacturer's protocols (Invitrogen, Carlsbad, CA) and plated on LB plates containing 100 mg/L ampicillin. Clones were grown overnight, and all colonies were individually plucked into 100 µL Milli-Q H_2O. DNA was immobilized by rupturing cells via heating the sample for 10 minutes at 100 $^{\circ}$C followed by 30s of vortex mixing.

PCR screening followed the procedure of Cabe and Marshal (2001). Briefly, two PCR reactions were carried out per sample. One PCR included T3 and T7 primers only while the second PCR reaction included T3, T7 and a $(GATA)_8$ oligonucleotide primer. The products of these two PCR reactions were visualized side-by-side on a 2.0% agarose gel. Both PCRs from clones without a $(GATA)_n$ repeat region produced bands equal to the size of the insert. However, for clones containing a $(GATA)_n$ repeat region, the PCR containing a $GATA_8$ oligonucleotide gave a distinctive smear in contrast to the band produced in the T3-T7 reaction. 183 of 1056 clones indicated the presence of a $(GATA)_n$ repeat (17%) and were sent to Nevada Genomics Center (Reno Nevada) for sequencing on an ABI automated sequencer (Applied Biosystems, Foster City CA). The sequences obtained were aligned together and edited with

SEQUENCHER *v.* 4.1 (Gene Codes Corporation, Ann Arbor, MI) to correct sequencing errors and to identify and avoid redundancy among microsatellite loci.

Primers flanking the repeat motif were developed for 45 unique loci using Primer 3 (Rosen and Skaletsky 2000). To each forward primer (except Por 105, 106, and 111), an M13 (-21) tail (5'-TGTAAAACGACGGCCAGT-3') was added for fluorescent labeling of PCR product (Schuelke 2000). Of 45 unique primer pairs flanking repeat regions, the 5 primers described here were polymorphic and gave PCR amplification consistently (Table 1). Additionally, 26 primer pairs for loci developed in other *Pseudacris* species (Van Buskirk *et al.* unpublished manuscript) were screened for amplification in *P. ornata.* Of these, 5 more were polymorphic and amplified product consistently (Table 1). Final PCR reactions were performed in 20 μL volumes and contained ~10 ng template and 1 unit *Taq* polymerase. Final concentrations of other reaction components were 2.5 mM $MgCl_2$ (Por 105, Por 106, Por 111) or 2 mM $MgCl_2$ (all other loci), 1X PCR buffer, 0.2 uM each dNTP, and 0.3 uM each primer (Por 105, Por 106 Por 111), or 0.5 uM reverse primer, 0.5 uM M13(-21) fluorescently labeled primer and 0.13 uM forward M13(-21) tagged primer (all other loci). Thermocycling parameters were, for Por 105 and Por 106, 95 ^{o}C for 4:00, 30 × (95 ^{o}C for 30 s. T_a for 30 sec, 72 ^{o}C for 30 sec), 72 ^{o}C for 7 min with T_a for Por 105 and Por 106 being 52 ^{o}C and 62 ^{o}C respectively, for Por 111, 95 ^{o}C for 4:00, 48 ^{o}C for 45 sec, 69 ^{o}C for 1:00, 35 × (95 ^{o}C for 30 s. 53°C for 30 sec (decreasing at 0.1 ^{o}C/cycle), 72 ^{o}C for 30 sec), 72 ^{o}C for 7 min, and for all other loci, 95 ^{o}C for 5:00, 30 × (95 ^{o}C for 45 sec 54 ^{o}C for 45 sec, 72 ^{o}C for 45 sec), 8 × (95 ^{o}C for 45 s. 52°C for 45 sec , 72 ^{o}C for 45 sec), 72 ^{o}C for 20 min. Amplified PCR products were sized in groups of 2-3 loci per run using

the CEQ 8000 genetic analysis system and software according to manufacturers protocols

(Beckman-Coulter, Fullerton, CA).

Table 1. Basic information for 10 microsatellite loci isolated and/or optimized in this study. The locus name, GeneBank accession number, species cloned, repeat motif, and primer sequence are shown. Population data (number of diploid genotypes obtained (n), observed and expected heterozygosity (H_o and H_E respectively), allele size range, and number of alleles (N_a) for *Pseudacris ornata* individuals collected from each of Colleton County, SC and Baker County, GA are given. Allele sizes do not include 21 bp M13 primer. Significance was assessed in Arlequin *v.* 3.0 using default MCMC parameters. Loci cloned from *P. crucifer* and *P. triseriata* were developed in Van Buskirk *et al.* (unpublished manuscript).

Locus	GeneBank accession no.	Species cloned	Repeat motif in clone	Primer sequence	Allele size range	Colleton Co., SC				Baker Co., GA			
						n	N_a	H_o	H_e	n	N_a	H_o	H_e
Por105	XXXX	*P. ornata*	(GATA)$_{22}$ imperfect	F-ACGACCATAGTGGGCTATCA	162-294	30	17	0.87	0.88	15	5	0.33	0.89*
				R-TGGTTAGGCCATTGTCAAGC									
Por106	XXXX	*P. ornata*	(GATA)$_{93}$ imperfect	F-ATGTCGGTAAGAAACTCTTAATGT	274-686	21	17	0.86	0.94	5	5	0.20	0.97
				R-TTGCTGTCCTTGGATTAGGC									
Por111	XXXX	*P. ornata*	(GATA)$_{58}$ imperfect	F-AAAGGAATTATTACATTCTGGCATTC	252-624	19	11	0.68	0.90	18	21	0.61	0.97*
				R-TGTGTGAACTAGTTTAAAAAACACA									
Por151	XXXX	*P. ornata*	(GATA)$_{11}$ imperfect	F-ACCAGGACAGTCAAACAGAG	87-177	32	19	0.81	0.94	21	18	0.76	0.94
				R-CCGATGTGGATAGATGGA									
Por165	XXXX	*P. ornata*	(CA)$_7$(TA)$_5$(GATA)$_{13}$ imperfect	F-CTCTGTCCTGTGCATTGAA	199-269	32	14	0.63	0.88*	21	19	0.86	0.95
				R-TAGGTAGGAAAAACCAAGCA									
Pcru09	EF190895	*P. crucifer*	(CA)$_{15}$	F-ACAATTAACACAATTGGCTCTCC	120-132	28	3	0.64	0.64	20	5	0.58	0.50
				R-AGTAAGGGAGCGTGCAAGAG									
Pcru10	EF190896	*P. crucifer*	(CA)$_{13}$	F-CATTTGGAACAAAGCAGCAAG	307-478	26	5	0.31	0.51	21	5	0.19	0.31
				R-GTGGAGTTCCAGCCTAATGC									
Pcru14	EF190899	*P. crucifer*	(GT)$_{13}$	F-GATCAGACAGTCTACAGTAATGAGGAG	158-162	29	3	0.31	0.36	21	2	0.24	0.51
				R-CATAACACAGGGCAACCAAG									
Pcru24	EF190902	*P. crucifer*	(CA)$_{17}$	F-TGCCATGGGGATGTTATATG	201-248	32	4	0.34	0.32	20	5	0.40	0.39
				R-CGAGCTATAGGAAAAGGCAGAG									
Ptri29	EF190917	*P. triseriata*	(GT)$_{14}$	F-CACCATCTGGGAGGTGCTAC	133-147	30	3	0.07	0.10	21	3	0.19	0.22
				R-ATGTCAAAGCCGACCAAATC									

We amplified product and scored allele sizes at these loci in individuals from two natural populations of *P. ornata* collected in 2006 and 2007 (n=21 from Baker County GA., and n=32 from Colleton County, SC). Number of alleles per locus per population ranged from 2 to 21 averaging 9.2 and expected heterozygosities ranged from 0.10 to 0.97 averaging 0.52 (Table 1). Tests for departures from Hardy-Weinberg Equilibrium (HWE) proportions and linkage disequilibrium were performed in Arlequin *v.* 3.0 using the default Markov chain parameters (Schneider *et al.* 2000). Three of the 20 locus by population comparisons may contain null alleles as indicated by a significant heterozygote deficiency after Bonferroni correction. However, no locus was significantly out of HWE for both populations. After Bonferroni correction, no loci showed significant linkage disequilibrium in both populations and thus all loci can be treated as independent.

Future studies using these markers will contribute to our understanding of the complex biogeographic history of the southeastern United States through the use of this model organism, the Ornate Chorus Frog. Further, future studies will improve our understanding of amphibian population dynamics such as dispersal rates, effective population sizes, geographic subdivision, and the occurrence of historical population crashes. For the specific study species, these markers will help to define appropriate evolutionary significant units for prioritizing populations or regions for conservation. These loci are, to our knowledge, the first microsatellite markers developed for use in this species. The microsatellite markers we developed and optimized here will be important tools to further investigate the evolutionary history, population structure, and natural selection in *P. ornata.*

References

Cabe PR, Marshall KE (2001) Microsatellite loci from the house wren (*Troglodytes aedon*). *Molecular Ecology Notes* **1**, 155-156.

Hoffman EA, Blouin MS (2000) A review of colour and pattern polymorphisms in anurans. *Biological Journal of the Linnean Society* **70**, 633-665.

Hoffman EA, Ardren WR, Blouin MS (2003) Nine polymorphic microsatellite loci for the northern leopard frog (*Rana pipiens*). *Molecular Ecology Notes* **3**, 115-116.

Jehle R, Artzen JW (2002) Review: Microsatellite Markers in Amphibian Conservation Genetics. *Herpetological Journal* **12**, 1-9.

Macas J, Gualberti G, Nouzová M*, et al.* (1996) Construction of chromosome-specific DNA libraries covering the whole genome fo field bean (*Vicia faba L.*). *Chromosome Research* **4**, 531-539.

Moriarty EC, Cannatella DC (2004) Phylogenetic relationships of the North American chorus frogs (*Pseudacris*: Hylidae). *Molecular Phylogenetics and Evolution* **30**, 409-420.

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds. Krawetz S, Misener S), pp. 365-386. Humana Press, Totowa, NJ.

Schneider S, Roessli D, Excoffier L (2000) Arlequin: A software for population genetics data analysis. Ver 3.000. Genetics and Biometry Lab, Department of Anthropology, University of Geneva.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* **18**, 233-234.

Stuart SN, Chanson JS, Cox NA*, et al.* (2004) Status and trends of amphibian declines and extinctions worldwide. *Science* **306**, 1783 - 1786.

Van Buskirk A, Pesaro M, Smith D (unpublished manuscript) Polymorphic microsatellite markers for two North American frogs (spring peeper *Pseudacris crucifer* and chorus frog *P. triseriata*.

CHAPTER 3 – FAT FROGS, MOBILE GENES– UNEXPECTEDLY HIGH LEVELS OF GENE FLOW AMONG POPULATIONS OF *PSEUDACRIS ORNATA*, A MEMBER OF THE 'FAT FROGS' SPECIES GROUP

<u>Introduction</u>

A central goal of population genetics is to illustrate how genetic variation is partitioned within and among natural populations of a species (Wright 1931, Templeton *et al.* 1995, Rosenberg *et al.* 2002). Population substructure is virtually a universal property of living systems, and has major effects on patterns of variation. Natural species are essentially never truly panmictic and usually exist in schools, colonies, or other aggregations. Further, suitable habitat for natural populations often occurs in geographically separated patches. Habitat patchiness and the aggregation of individuals within populations limit the probability of individuals encountering and mating with individuals occupying different habitat patches or aggregations. When subpopulations are separated from one another, genetic differentiation occurs among subpopulations owing to the micro-evolutionary forces of mutation, random genetic drift, and selection acting independently in different subpopulations. In contrast, migration acts to limit the degree to which subpopulations evolve independently and hence the level of population substructure (Hartl and Clark 1983).

One important consequence of natural levels of genetic structure is the expected consequence of habitat fragmentation and increased inbreeding on genetic load. Genetic load is formally defined as the proportional decrease in average fitness in a population relative to the optimal genotype (Wallace 1970). Thus, there is a markedly different theoretical expectation for the genetic load effect of increased inbreeding in naturally structured versus unstructured

populations. Interconnected populations with high gene flow tend to have high effective

population size and low background levels of inbreeding. Hence, the probability of rare

deleterious recessive alleles forming homozygotes is low. Consequently, in large populations,

deleterious recessive alleles accumulate in the genome, masked from selection in the

heterozygous state. In contrast, naturally subdivided populations tend to have low baseline

effective population sizes and deleterious recessives are continually purged from the population

by natural selection. Thus, species characterized by high gene flow and large effective

population sizes would be more severely affected by genetic load under a scenario of sudden

fragmentation due to the increased probability of homozygosity for deleterious recessive alleles.

As such, our understanding of the natural degree of population structure in amphibians is very

important for predicting the consequences of habitat fragmentation.

Amphibians provide excellent model organisms for investigating the interplay between

genetic drift and migration and the potential consequences of genetic load. Amphibians are

generally characterized by limited dispersal abilities, high philopatry, and are obligately

associated with the areas around suitable habitats (e.g. breeding ponds). Thus, amphibians often

occur in naturally highly fragmented populations, and show a high degree of genetic structure

(Beebee 1996, 2005, Palo *et al.* 2004, Newman and Squire 2002). Owing to this limited

dispersal ability, major geographic barriers known to restrict gene flow in other species (e.g.

rivers or mountain ranges) would be expected to truncate the range of amphibians. Indeed, both

high genetic structure due to distance and the efficiency of geographic features for restricting

gene flow has been repeatedly demonstrated in amphibian species (Newman and Squire 2002,

Shaffer *et al.* 2000, Funk *et al.* 2005). However, the ubiquity of large-scale genetic structure in

18

amphibian species has not been assessed. Here, we investigate the relative roles of historical

geography and climate, genetic drift, and migration on the partitioning of genetic variation in an

amphibian species, the Ornate Chorus frog (*Pseudacris ornata*), that occupies a region with a

well characterized history – the coastal plains of the Southeastern United States.

A large number of molecular-genetics studies spanning many types of vertebrates have

explored the biogeography of the southeastern United States (Avise *et al.* 1979, 1987;

Bermingham and Avise 1986; Avise and Nelson 1989; Donovan *et al.* 2000; Hayes and Harrison

1992; Vogler and DeSalle 1993; Ellsworth *et al.* 1994; Phillips and Sexton 1994; Osentoski and

Lamb 1995; Walker *et al.* 1995). Generally, these studies reveal strong east-west divergence

separated by the Apalachicola river basin and the Appalachian Mountains (Donovan *et al.* 2000).

From the late Pliocene through the Pleistocene, glacial-interglacial cycles caused major

fluctuations in sea levels. At the height of interglacial periods (188, 120, 94, and 72 thousand

years ago), sea levels in the southeastern US were 10 to 25 feet higher than they are today

(Cronin *et al.* 1981). During these periods of highest sea levels, the Apalachicola river basin was

a large saltwater embayment extending north through Alabama and Georgia to the Appalachian

mountains creating a saltwater channel acting as a barrier to many species and originating the

deep genetic divergences now seen (Cooke 1945, Blainey 1971).

Combining the expectation that amphibians have limited dispersal abilities with the well

characterized biogeographic history of the Southeastern United States, we were able to test

specific hypotheses regarding the interaction of migration and genetic drift in forming patterns of

population genetic structure in *P. ornata*. First, we hypothesized that we would find an overall

pattern of isolation by distance. Such a pattern is common in other anuran species (Hoffman *et*

al. 2004, Monsen and Blouin 2005, Shaffer *et al.* 2000), and is often attributed to the inability of individuals to disperse long distances (Wright 1942). Second, because we expect amphibians to be highly philopatric and exhibit limited dispersal abilities, we hypothesized that populations sampled in distant geographic regions would be genetically differentiated (as measured by F_{ST}). Furthermore, we expected to observe patterns of population structure on a small geographic scale (i.e. several kilometers) such as have been observed in other small anurans (Newman and Squire 2001, Monsen and Blouin 2005). Lastly, we hypothesized that barriers to gene flow for other vertebrate taxa (i.e. the Apalachicola River) would also act as a barrier for this species.

To test these three hypotheses, we used 1304 bp of mitochondrial DNA sequence in addition to 10 unlinked microsatellite loci to characterize the patterns of population genetic structure of 211 individuals collected from 8 populations sampled throughout the species range. We tested Hypothesis 1 (isolation by distance) using both Nested Clade Phylogeographic Analysis (NCPA, Templeton *et al.* 1995, 2004) of mitochondrial haplotypes, and by testing for a correlation of genetic distance (F_{ST} at microsatellite loci) with geographic distance. We tested Hypothesis 2 (significant genetic structure on small spatial scales) using a genetic clustering algorithm implemented in STRUCTURE (Pritchard *et al.* 2000) to estimate the spatial scale at which individuals were assignable to discrete genetic clusters. Further, we characterized the degree of genetic divergence among populations with F_{ST} and compared these statistics to the degree of divergence among populations of other amphibians. To test Hypothesis 3 (that the Apalachicola River acts as a significant barrier to gene flow in *P. ornata*) we compared the patterns of genetic structure (as determined in NCPA and genetic clustering analysis) among populations spanning the river to the population structure observed among populations not

divided by this river. Moreover, we reconstructed the phylogenetic relationships among sampled

haplotypes to determine if samples from east and west of the Apalachicola River showed

evidence for a deep phylogenetic divergence. Finally, we discuss all of these results with regards

to the consequences of fragmentation in this species as well as general patterns of amphibian

evolution.

Materials and Methods

Sampling – To test predictions made about the expected genetic structure in this species,

we collected samples from throughout the range of *P. ornata*. A special effort was made to

obtain samples from both east and west of the Apalachicola River to determine the extent that

this river acts as a barrier to gene flow. Overall, we obtained samples from 8 discrete geographic

locations ranging from the Florida panhandle west of the Apalachicola River to east-central

South Carolina (Fig. 1, Table 2). Sample sizes at each collecting locality ranged from 1 to 85

(Table 2).

Individual frogs were located in the winter and spring (November through May) of 2006

and 2007. Frogs were primarily captured either by road cruising on rainy nights, or by hand-

capturing calling males in breeding choruses. However, the samples from Barbour County,

Alabama were captured in pitfall traps set along a drift fence surrounding an ephemeral pond.

We scored all captures for sex, and we recorded the precise location of capture. For DNA

analysis, we removed the longest toe on the right hind-foot with sterilized surgical scissors. We

placed toe clippings in vials filed with anhydrous calcium sulfate (a desiccant) and transported

them to the University of Central Florida for genetic analysis. Additionally, we obtained 13

samples from the University of Texas, Austin herpetological collections (TNHC #s 62178 -

62189).
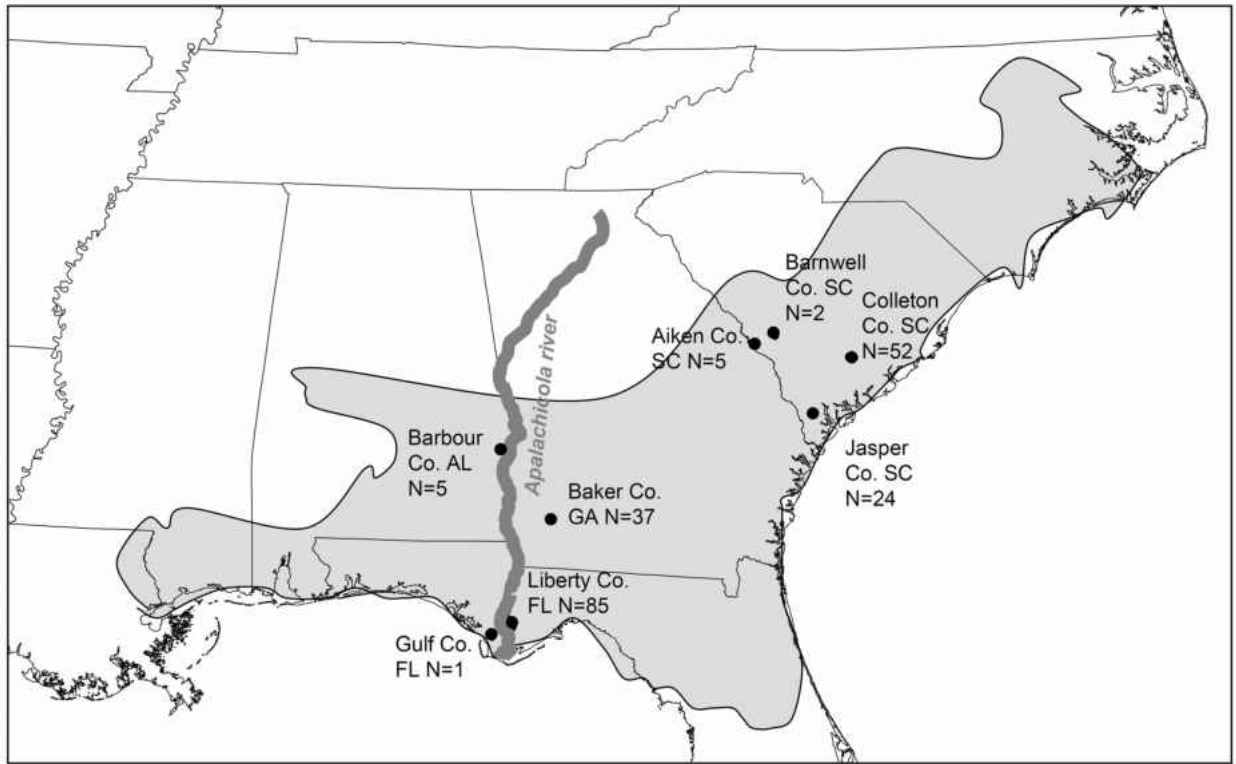
Figure 1. Native range of *P. ornata* and sampling locations. The natural range of *P. ornata* is shaded in grey. The location of sampling sites and the total sample sizes at each site are also shown.

Table 2. Sample sizes obtained from each sampling locality and the number individuals sequenced for 1304 bp of mitochondrial DNA.

Location	Total sample size	MtDNA sample size
Baker County Georgia	37	15
Aiken County South Carolina	5	5
Liberty County Florida	85	9
Jasper County South Carolina	24	14
Colleton County South Carolina	52	11
Barnwell County South Carolina	2	2
Barbour County Alabama	5	3
Gulf County Florida	1	1

DNA isolation and mitochondrial DNA amplification and sequencing – We isolated DNA from all toe clips obtained using DNeasy DNA extraction kits (Quiagen, Valencia CA) according to the manufactures' "Animal Tissue Protocol." However, the elution of DNA was done using a 1:10 dilution of the DNeasy Kit "elution solution." Two elutions were performed and we used the second elute at full strength or a 1:10 dilution of the first elute (both having final concentrations of about 5-10 ng per μL.) for PCR reactions.

We used mitochondrial DNA haplotypes along with NCPA and phylogenetic analysis to test the prediction of an overall pattern of isolation by distance as well as to assess whether the Apalachicola River acts as a barrier to gene flow. We sequenced two disjointed sections of mitochondrial DNA (mtDNA) in both directions for up to 15 samples per sampling locality (Table 2). We sequenced a 690 bp section that spans partial 12S ribosomal RNA (rRNA; 466 bp), complete tRNA-Val (69 bp), and part of 16S rRNA (155 bp). We also sequenced a 612 bp

section of the Cytochrome-b (*Cyt-b*) gene. We amplified and sequenced the fragment coding for

RNA genes using primers 12L1 (5'-AAAAAGCTTCAAACTGGGATTAGATACCCCACTAT-

3') and 16Sh (5'-GCTAGACCATKATGCAAAAGGTA-3') which correspond to primer

numbers 46 and 76, respectively, in Goebel *et al.* (1999). We amplified and sequenced the

section of *Cyt-b* using primers MVZ 15 (5'- GAACTAATGGCCCACACWWTACGNAA-3')

and MVZ 18 (5'-GTCTTTGTATGAGAAGTATG-3') after Moritz *et al.* 1992. PCR reactions

contained 5-10 ng template DNA, 1 unit *Taq* polymerase, 2.25 mM $MgCl_2$), 1X PCR buffer (10

mM Tris-HCL, 50 mM KCL and 0.1% Triton X-100), 0.2 μM each dNTP, and 0.5 μM each

primer. PCR thermocycling parameters were 95 ^{o}C for 4:00, 41 ^{o}C (*Cyt-b* primers) or 46 ^{o}C

(rRNA) for 45 sec, 69 ^{o}C for 1:00, 35 × (95 ^{o}C for 45 s. 45 ^{o}C (*Cyt-b* primers) or 50 ^{o}C (rRNA)

for 45 sec, 72 ^{o}C for 45 sec), 72 ^{o}C for 7 min.

We electrophoresed PCR reactions in a 1% agarose-TAE gel, and excised and purified

positive PCR bands using GENECLEAN III Kits according to manufacturer's protocols (Q-

Biogen Inc., Irvine, CA). We quantified purified PCR product by comparing band brightness to

DNA quantification ladder. Fragments were sequenced at the Nevada Genomics Center (Reno,

NV). We aligned multiple fragments for each individual and reviewed base-calls for accuracy

using SEQUENCHER *v.* 4.7 (Gene Codes Corp, Ann Arbor, MI). We aligned complete, edited

sequences using CLUSTAL W. We concatenated disjoint alignments using BIOEDIT *v.* 7.07 (Ibis

Biosciences, Carlsbad, CA) creating a single alignment consisting of complete sequences from

58 individuals and nearly complete sequences from 2 individuals (Table 2).

Statistical analysis of mtDNA data – To determine the historic consequence of the

Apalachicola River and to identify any deep divergence between existing populations, we

estimated phylogenetic relationships among unique haplotypes. This phylogenetic

reconstruction used Bayesian Marcov Chain Monte Carlo (MCMC) phylogenetic methods

implemented in MR. BAYES *v.* 3.2.1 (Ronquist and Huelsenbeck 2003). In addition to 33

complete sequences of unique haplotypes, we included 2 partial sequences from Barbour County,

AL. A *Pseudacris regilla* sequence obtained from genbank (Accession numbers AY364542 and

DQ195169 for RNA genes and *Cyt b*, respectively) and a *P. crucifer* individual sequenced in this

study were used as outgroups.

We used MRMODELTEST *v.* 2.2 (Nylander 2004) using Akaiki information criteria to

determine the model of DNA evolution most appropriate for different partitions of the entire

dataset. These partitions were 1) the unpartitioned data set, 2) *Cyt-b* gene alone 3) RNA genes

alone 4) first codon position of *Cyt-b* 5) second codon position of *Cyt-b* 5) first and 2nd codon

position of *Cyt-b*, and 6) third codon position of *Cyt-b*. We used the best models for each

partition in four Bayesian MCMC analysis implemented in MR. BAYES. MCMC chains were run

for 4 million generations and the first 1,000,000 generations were discarded as burn-in. Bayes

Factors according to Kass and Raftery (1995) were used to choose an appropriately complex

number of partitions, and the phylogeny obtained with this partitioned dataset was reported.

We used Nested Cladistic Phylogeographical Analysis (NCPA) *sensu* Templeton (2004)

to formally distinguish between possible biological hypotheses that could explain statistically

significant geographical associations of haplotypic variation (e.g. isolation by distance, allopatric

fragmentation, restricted gene flow with long distance dispersal, range expansion, etc.,

Templeton 2004) . This method uses statistics based on an intraspecific haplotype network with

hierarchically nested clades and the geographic distribution of these nested clades to 1)

determine if the null hypothesis of no geographic association of haplotypes can be rejected and 2) to select from a set of possible biological mechanisms (using an inference key) which could lead to the observed geographic pattern. We created the intraspecific haplotype network using the algorithm of Templeton *et al.* (1992) implemented in the program TCS *v.* 1.21 (Clement *et al.* 2000). Hierarchically nested clades were determined according to the rules of Templeton *et al.* (1987, 1992). GEODIS 2.0 (Posada *et al.* 2000) was used to determine NCPA distance statistics, and these were interpreted according to the inference key given in Templeton (2004). Statistical significance was determined based on 100,000 permutations of clades. The automation of this entire process was facilitated by the program ANECA *v.* 1.1 (Panchal 2007), with the results of this automated process confirmed by hand.

Microsatellite genotyping – We further characterized the genetic structure in this species using ten unlinked microsatellite markers (Chapter 2). Specifically, the microsatellite markers addressed Hypothesis 2, that due to limited dispersal abilities and high rates of philopatry, the ornate chorus frog would show high genetic divergence among sampling regions (high F_{ST}), and there would be detectable, fine-scale genetic structure over relatively short geographic distances. We genotyped all frogs (n=211) for ten unlinked microsatellite loci (Table 1) following the methods described in Chapter 2.

Statistical analysis of microsatellite data – To determine the variability and data-quality of the microsatellite dataset, we estimated basic information describing the characteristics of variation at ten microsatellite loci in each discrete sampling locality with reasonable sampling (n>10). We estimated observed (H_O) and expected (H_E) heterozygosities and performed an exact test for significant deviations from Hardy-Weinberg equilibrium (HWE) based on 100,000

permutations in ARLEQUIN *v.* 3.0 (Schneider *et al.* 2000). We estimated allelic diversities

compensating for sample size differences in FSTAT *v.* 2.9.3.2 (Table 3, Goudet 1995).

To test the prediction that genetic structure would be detectable even over small spatial

scales, we used a genetic clustering algorithm (STRUCTURE *v.* 2.2; Pritchard *et al.* 2000). This

method estimates the number of genetically distinct clusters (K) in an allelic data set. We

applied this method to both the entire microsatellite dataset as well as separately to sampling

localities that were geographically spread-out (i.e. Liberty County, Fl, and Colleton County, SC).

The algorithm in STRUCTURE introduces population structure to the data that minimizes Hardy

Weinberg disequilibrium and linkage disequilibrium, and produces an estimate of the log

probability of the data Pr(X|K) for a specified value of K. Often, the value of Pr(X|K) estimated

by STRUCTURE continues to increase with increasing values of K, rendering its maximum an

inconsistent criterion for determining the ''best'' estimate of K. An alternative measure based on

Pr(X|K), ΔK, has been shown to successfully identify the highest level of meaningful population

structure under a wide variety of simulation scenarios (Evanno *et al.* 2005). Thus, we used the

maximum value(s) of ΔK to determine the number of clusters that best explain the highest level

of population structure in our data.

In STRUCTURE, we set most parameters to their default values except where noted below.

We chose a model allowing admixture and correlated allele frequencies between populations.

We let ά, the degree of admixture, be inferred from the data. The parameter of the distribution of

allele frequencies (k) was set to one. The first 100,000 generations of data were discarded as

burn-in, and data were collected for 1,000,000 (full data-set) or 100,000 (Liberty County and

Colleton County data-sets) generations thereafter. A visual inspection of Pr(X|K) plotted against

the number of generations, and consistency across runs, supported these parameters as sufficient.

For each value of K (1–8), 5-10 independent STRUCTURE runs were conducted to confirm

convergence across runs and to obtain estimates of the variance among runs (as these pooled data

were used to estimate ΔK.)

Finally, to address the hypotheses that there would be high divergence among

populations we estimated both global and all pairwise estimates of F_{ST}. For populations that

were supported in the STRUCTURE analyses, divergence among populations was assessed by

pairwise F_{ST} estimated in ARLEQUIN *v. 3.0*. Only F_{ST}'s that were significantly different than zero

($\alpha=0.05$) were reported. Evidence for isolation by distance was assessed with a Mantel test for

correlation between the pairwise F_{ST} matrix and the pairwise geographic distance matrix. The

Mantel test was implemented in IBDWS (Jensen 2005) with 1000 permutations of the data.

<center>Results</center>

Phylogenetic analysis –We reconstructed phylogenetic relationships among unique

mtDNA haplotypes to identify the presence of deep genetic divergences among populations.

Specifically, we addressed the hypothesis that the Apalachicola River created an impassible

barrier throughout the Pleistocene epoch and divided populations of *P. ornata* into eastern and

western population groups. Of the 60 individuals from which we obtained nearly complete

sequences of the 1304 bp of mtDNA, there were 35 unique haplotypes. The greatest pairwise

uncorrected sequence divergence among these haplotypes was 2.8%. Harmonic means of Ln

likelihoods (LnL) of four different partitionings of the dataset obtained in Bayesian MCMC were

compared to determine the extent that increasing complex partitions improved the dataset. Increased partitioning continued to increase the likelihood values as expected. Following the recommendations of Kass and Raftery (1995), we took increases in LnL values of $2LnB_{10} > 10$ as strong evidence for using the more complex partition. The final model chosen contained 3 partitions: *Cyt-b* codon positions 1 and 2 under a HKY+G model of DNA evolution, *Cyt-b* position 3 under a HKY+G model of DNA evolution, and RNA genes under a separate GTR+G model of DNA evolution. Bayesian MCMC analysis using this overall model yielded a relatively well resolved phylogeny of haplotypes (Fig. 2). Most notably, there was a well resolved "Eastern clade" (Posterior probability = 0.86) consisting only of haplotypes sampled from South Carolina. Samples typed from both east and west of the Apalachicola River were included in a "Western group", but samples from east and west of the Apalachicola were not members of reciprocally monophyletic groups. Thus, this analysis did not support the hypothesis that the Apalachicala River acted as a historically continuous partition in this species.

Nested Cladistic Phylogeographic Analysis (NCPA) - We used NCPA to identify significant geographic associations of haplotypes, and to choose from a set of *a priori* biological hypotheses which could potentially explain significant geographic associations (See Templeton 2004). An intraspecific haplotype network was constructed wherein all haplotypes were connected by <22 steps, the 90% confidence limit of validity of parsimony estimated in the program TCS. Hierarchical nesting of haplotypes yielded 21, 1-step clades nested into 11, 2-step clades, nested into 6, 3-step clades in 3, 4-Step clades which were grouped together in a single 5-step clade (Fig. 3). Clades in the intra-specific haplotype network corresponded to some clades identified in the phylogenetic analysis. All 4-step clades were supported in the phylogenetic

analysis except for the placement of haplotype AB (Fig. 3). Haplotype AB occurs west of the

Apalachicola River, but groups, with weak nodal support, as basal to the eastern haplotype

group. The geographic distribution of the higher order (3 step and 4-step) clades is displayed in

Fig. 3. The null hypothesis of no geographical association could be rejected for 3 clades in the

nested cladogram. Restricted gene flow with isolation by distance was inferred for the total

cladogram (clade 5-1) and for clade 4-1 while there was insufficient sampling to make an

inference for clade 4-2.
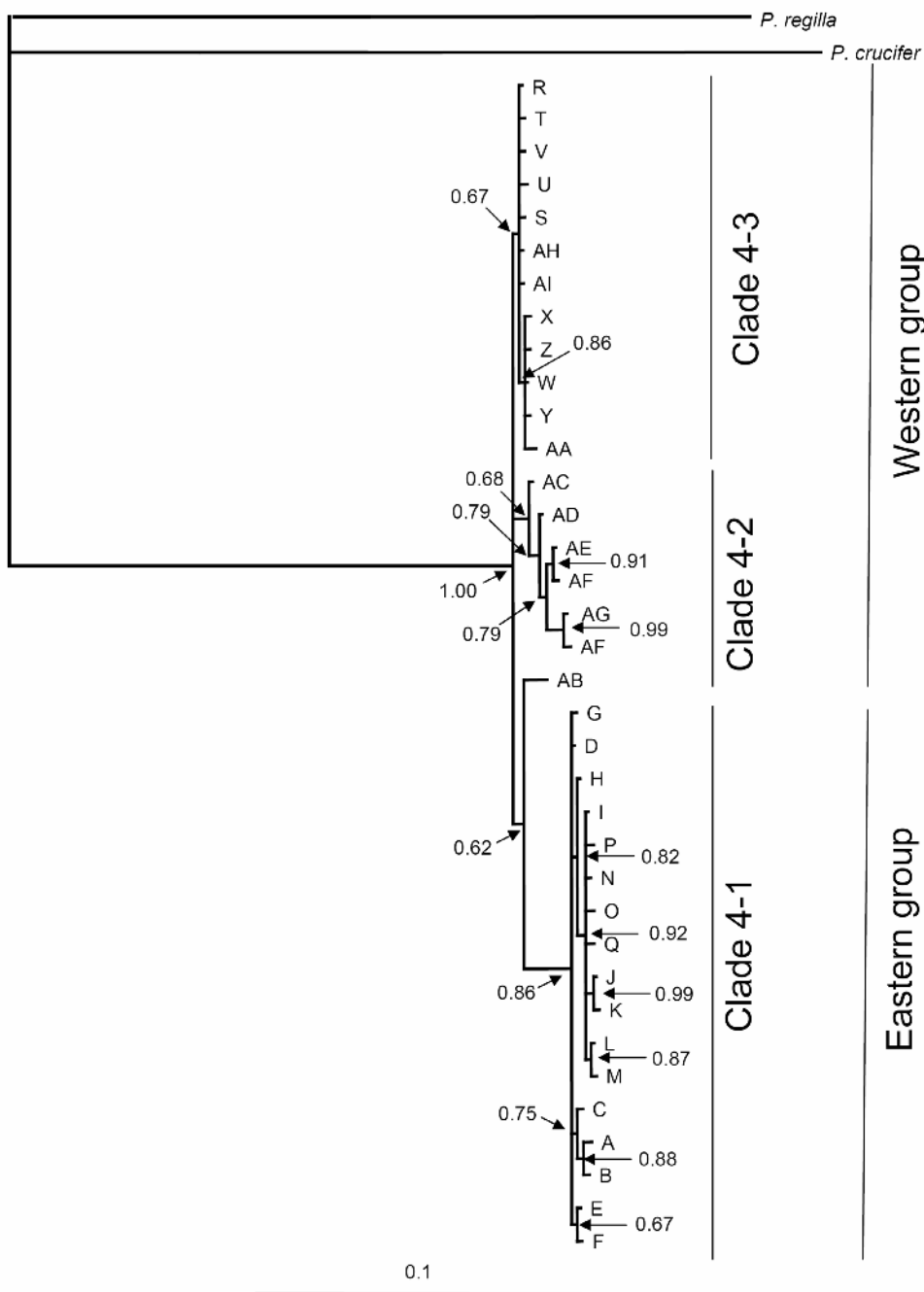
Figure 2. Phylogenetic relationships among haplotypes sampled in *P. ornata*. Tips are labeled with letters representing unique haplotypes. Clades which correspond to nesting levels in NCPA are also labeled. Nodal support is displayed as Bayesian posterior probabilities. The phylogeny was rooted using homologous *P. regilla* and *P. crucifer* DNA sequences.

Figure 3. Haplotype network of 33 unique (58 total) haplotypes created using the algorithm of Templeton *et al.* (1992) implemented in the computer program TCS. Individual haplotypes are represented by circles. The size of each circle represents the relative frequency of that haplotype in the total sample (n=58). The smallest circles represent unsampled inferred haplotypes. Neighboring haplotypes are distinguished by 1 mutation in 1304 bp of mtDNA. Hierarchically nested clades as determined by the method of Templeton *et al.* (1988, 1993) are indicated.

33

Figure 4. Map showing the geographic distribution of clades identified in nested clade analysis at the 3-, 4-, and 5-step clades. In NCPA clades 4-2 and 5-1 showed significant geographic association and this association was characteristic of isolation by distance with restricted gene flow. NCPA for other clades was inconclusive either due to sampling inadequacies or inability to reject a null hypothesis of no geographic association.

Microsatellite analysis- We characterized levels of variation and data quality in all samples using standard population genetics statistics. Table 3 shows basic information for the ten microsatellite loci used in this study for each population with reasonable sample size (n>10). Expected heterozygosity (H_E) ranged from 0 to 0.99 with a multi-locus H_E estimate of 0.65. After Bonforoni correction for multiple comparisons, there remained 9 out of 50 significant deviations from HWE (Corrected P = 0.001). Colleton County South Carolina Population A and Liberty County Florida had two loci significantly out of HWE while Baker County Georgia had

4 loci that were out of HWE. Por 151 was out of HWE for three of five populations and Por 111 and Por 105 were out of HWE for two populations each. All significant deviations from HWE were due to heterozygote deficiencies and thus may be due to the presence of null alleles in some populations. Allelic richness compensating for sample size averaged 7.6, but ranged widely from 1 to 17.4 (based on the smallest sample size of 11 diploid individuals). Thus, the microsatellite loci ranged from moderate to high variability. Because no locus or population was consistently out of HWE, all loci and populations were utilized in future analyses.

Table 3. Basic information for microsatellite loci for each major sampling locality (n>10). For each locus at each sampling locality diploid sample size (n) allelic richness compensating for sample size (A), observed (H_O) and expected (H_E) heterozygosities are shown. Significant deviations from HWE are marked with an asterisk.

Locus	Colleton Co., SC A				Colleton Co., SC B				Jasper Co., SC				Baker Co., GA				Liberty Co., Fl			
	n	A	H_o	H_E	n	A	H_o	H_E	n	A	H_o	H_E	n	A	H_o	H_E	n	A	H_o	H_E
por106	21	12.93	0.86	0.95	12	17.74	0.92	0.99	22	15.82	0.77	0.96	11	9.00	0.27*	0.87	51	17.31	0.45	0.98
por111	19	9.40	0.68	0.90	14	12.71	0.71*	0.96	23	14.21	0.78	0.96	28	15.45	0.61*	0.97	64	16.47	0.78	0.97
pcru09	28	3.00	0.64	0.64	16	3.82	0.56	0.65	20	3.72	0.45	0.57	35	4.59	0.54	0.60	82	4.51	0.50	0.74
pcru14	29	2.75	0.31	0.36	16	2.98	0.75	0.57	20	2.00	0.80	0.52	36	2.00	0.19	0.41	82	3.06	0.27	0.27
pcru24	32	2.67	0.34	0.32	19	1.93	0.16	0.15	24	1.00	0.00	0.00	33	3.77	0.36	0.41	81	3.69	0.49	0.52
por105	30	9.87	0.87	0.88	18	8.94	0.83	0.90	24	12.01	0.88	0.93	28	8.47	0.32*	0.83	79	8.63	0.43*	0.86
ptri29	30	1.73	0.07	0.10	19	1.00	0.00	0.00	24	2.88	0.13	0.24	37	2.26	0.27	0.24	83	1.63	0.04	0.07
pcru10	26	3.52	0.31	0.51	19	2.74	0.16	0.20	23	3.47	0.22	0.35	37	3.28	0.32	0.35	40	5.79	0.28	0.52
por151	32	12.23	0.81*	0.94	19	10.40	0.95	0.91	24	13.19	0.83	0.94	37	11.79	0.76*	0.92	78	11.02	0.49*	0.92
por165	32	9.20	0.63*	0.88	18	12.86	0.83	0.94	25	12.52	0.80	0.93	37	11.62	0.86	0.92	81	14.63	0.70	0.96

*significant deviation from expectation after bonferroni correction for multiple comparisons

.

36

We used a genetic clustering method implemented in STRUCTURE to identify discrete breeding populations represented in the dataset and to determine the geographic scale on which there was significant genetic structure. We did this in two phases: first on the entire dataset together (combined dataset) and second on the two geographically spread-out sampling localities individually (separate dataset). For the combined dataset, going from K=1 to K=2 accounted for the most genetic structure and separated the eastern (South Carolina) samples from the western (Florida, Alabama, Georgia) samples. However, increases in K thereafter continued to separate populations according to sampling location and produce substantial increases in Pr(X|K) up to K = 6 (Figs. 5, 6). While the method of Evanno (2005) indicated that the highest level of population structure existed at K=2, this method also showed evidence that K=6 was the true number of genetic clusters, as this was the second highest value of ΔK. Interestingly, samples collected from two discrete ponds (A and B) separated by only 5.7 km in Colleton County, SC were assigned to two separate genetic clusters at K=6 (Figs. 7, 8). Samples from populations with smaller sample sizes generally fell into the clusters one would expect. For example, the samples from Aiken and Barnwell Counties, SC fell into the "eastern" cluster at K=2. However, unexpectedly, 3/5 individuals from Barbour County, AL also fell clearly into this "eastern" cluster despite this cluster otherwise comprising almost exclusively samples from SC. This is likely due to the small sample size of this population. For K > 2, it became more difficult to assign samples from under-sampled populations to any one genetic cluster.

Individual STRUCTURE analyses on the geographically separate sampling localities (i.e. Colleton County, SC and Liberty County, Fl) revealed differing results. While both of these sampling localities were spread out geographically (Colleton County over 5.8 km, and Liberty

37

County over 14.1 km), Colleton County SC samples clearly clustered into 2 discrete clusters while samples from Liberty County, Fl were not separable into distinct clusters, and were indicative of a single panmictic population (Figs. 9, 10, Pritchard et al 2000, STRUCTURE documentation). We checked whether this was an artifact of possible null alleles by re-running the STRUCTURE analysis excluding the two loci that were significantly out of HWE (Por 105 and Por 151; Table 3) in the Liberty County, Florida samples. However, this did not significantly alter our results, and thus the Liberty County samples were treated as coming from a single breeding population.
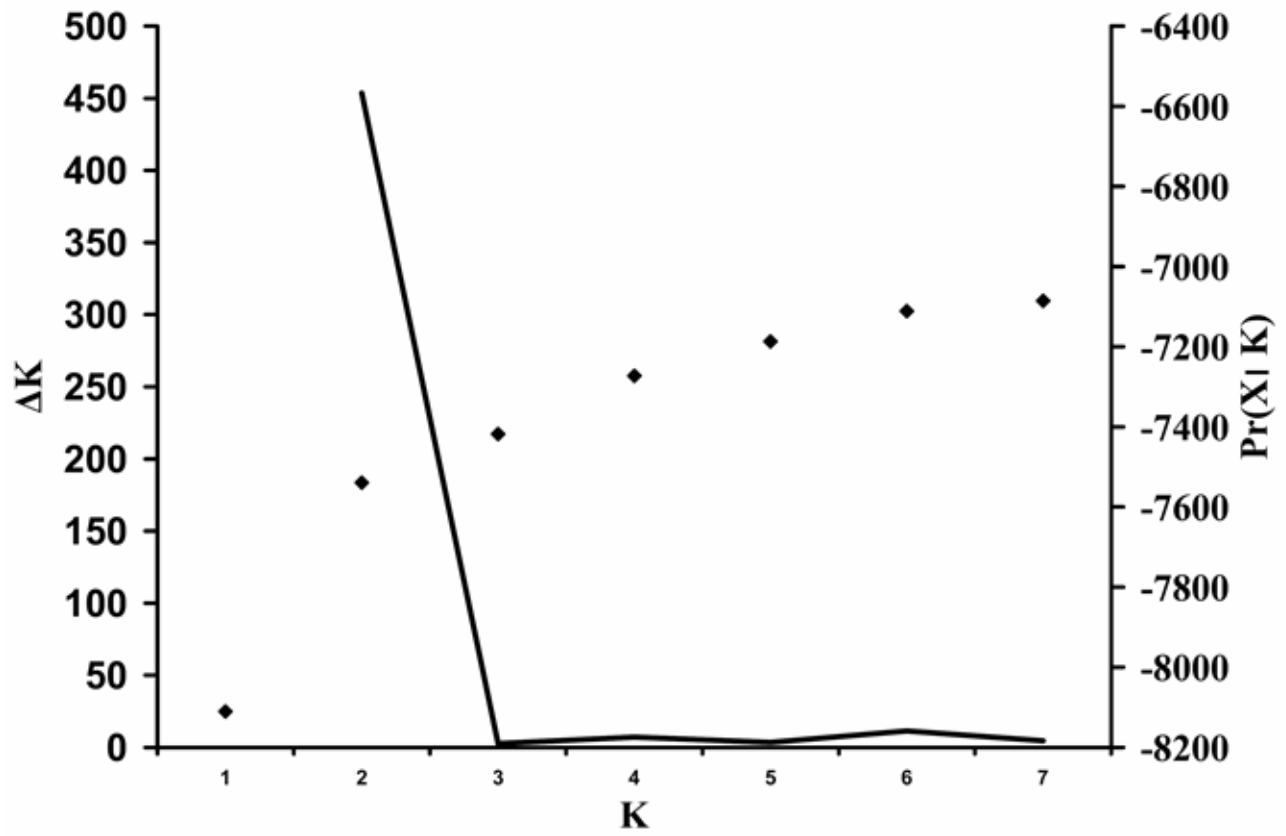
Figure 5. Two methods for estimating the true number of clusters in the entire dataset (n=211). Delta-K (Evanno 2005) for values of K from 1 to 7 is scaled to the left-hand Y-axis. Likelihood of data given K (Pritchard *et al.*1999) is scaled to the right-hand Y-axis. For both of these criteria, the maxima indicate the estimate of the true value of K.
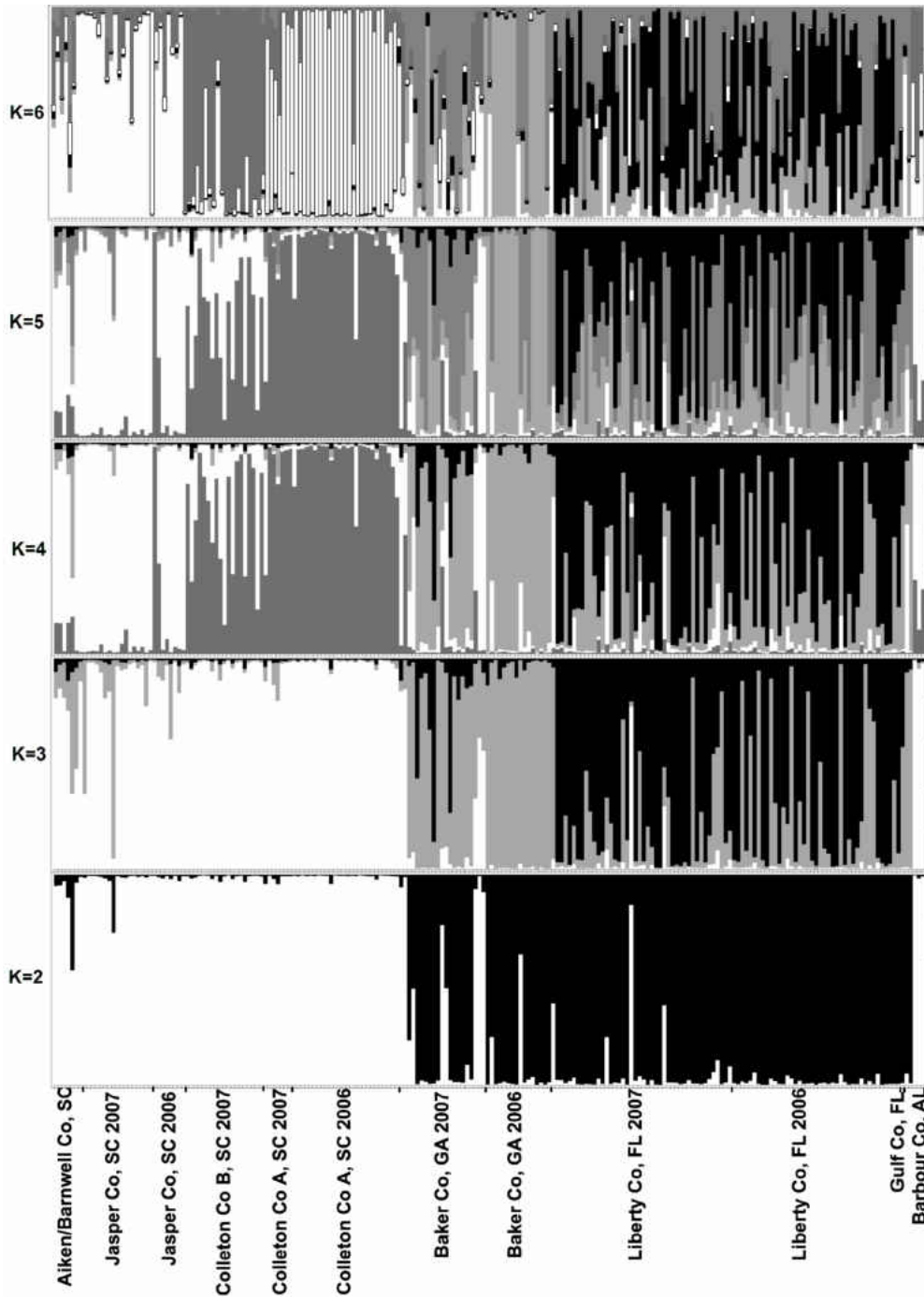
Figure 6. Results of clustering analysis as implemented in STRUCTURE. Y-axes are membership coefficient in each of K clusters (K=2-K=6 are displayed). All individuals collected in this study (n=211) were included in a single analysis, and their population of origin is displayed on the X-axis.
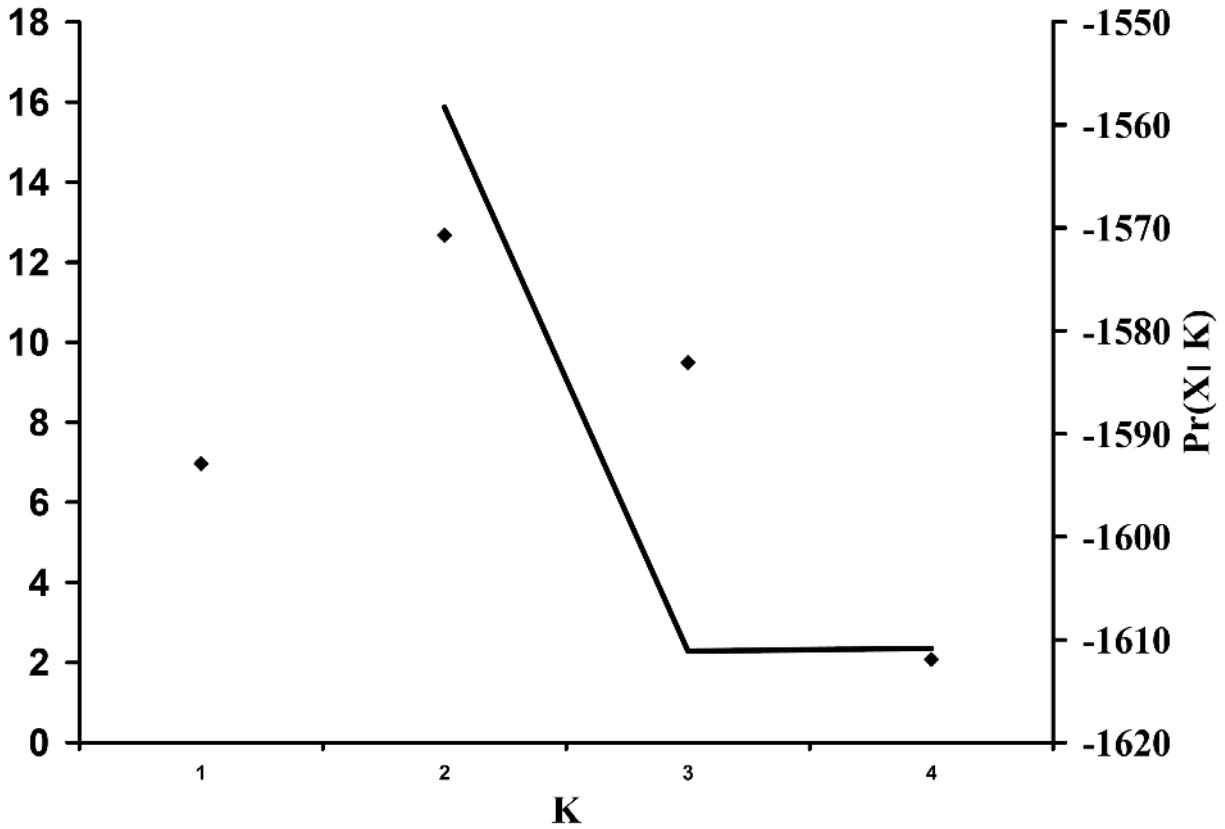
Figure 7. Two methods for estimating the number of clusters in microsatellite dataset comprising samples from Colleton County, South Carolina (n=52). Delta-K (Evanno 2005) for values of K from 1 to 4 is scaled to the left-hand Y-axis. Likelihood of data given K (Pritchard *et al.*1999) is scaled to the right-hand Y-axis. For both of these criteria, the maxima indicate the estimate of the true value of K.
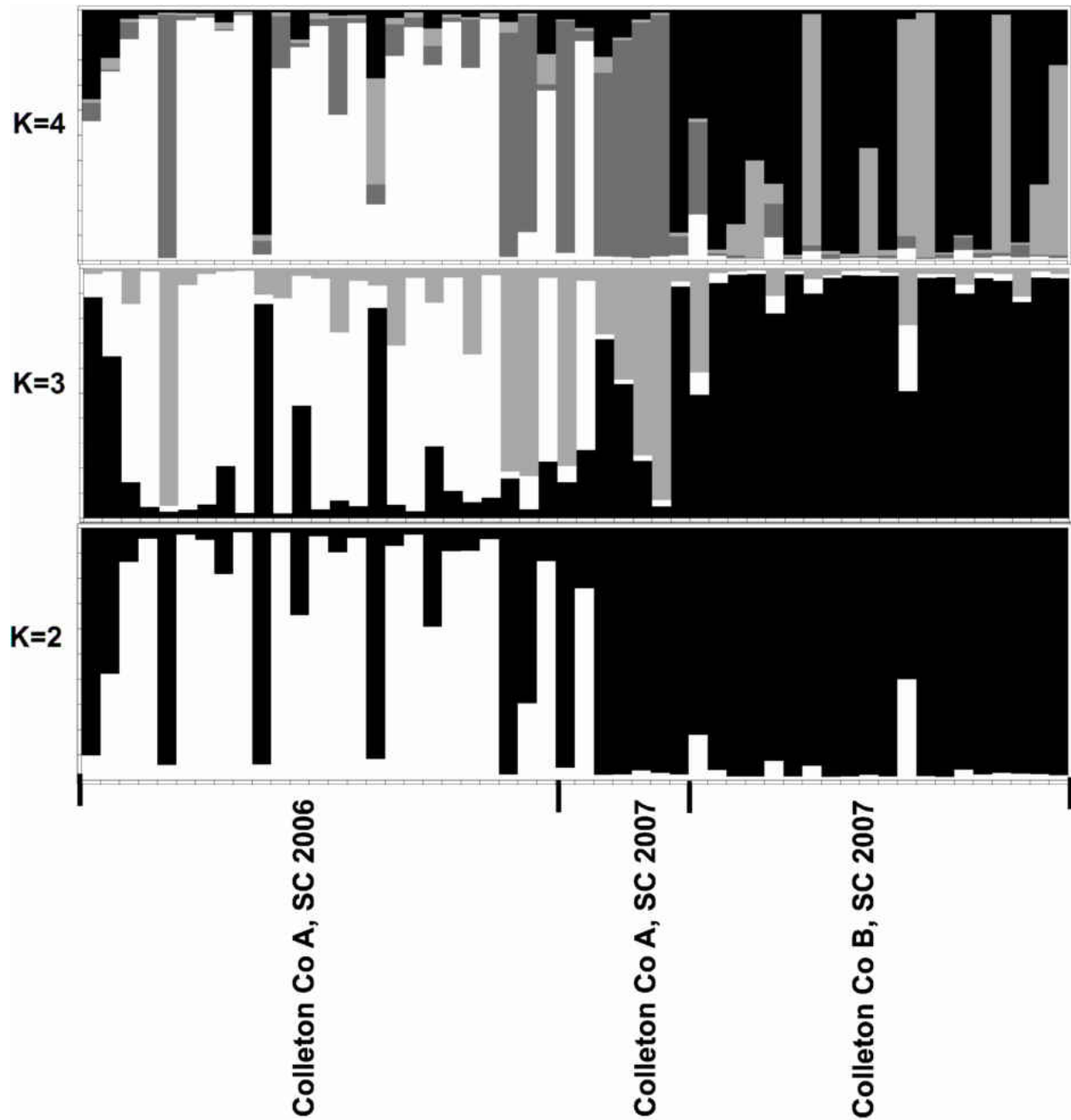
Figure 8. Results of STRUCTURE clustering analysis of only Colleton County, South Carolina samples. Y-axes are membership coefficient in each of K clusters (K=2-K=4 are displayed). All individuals collected in Colleton County, SC in 2006 and 2007 (n=52) were included in a single analysis, and their population of origin is displayed on the X-axis.
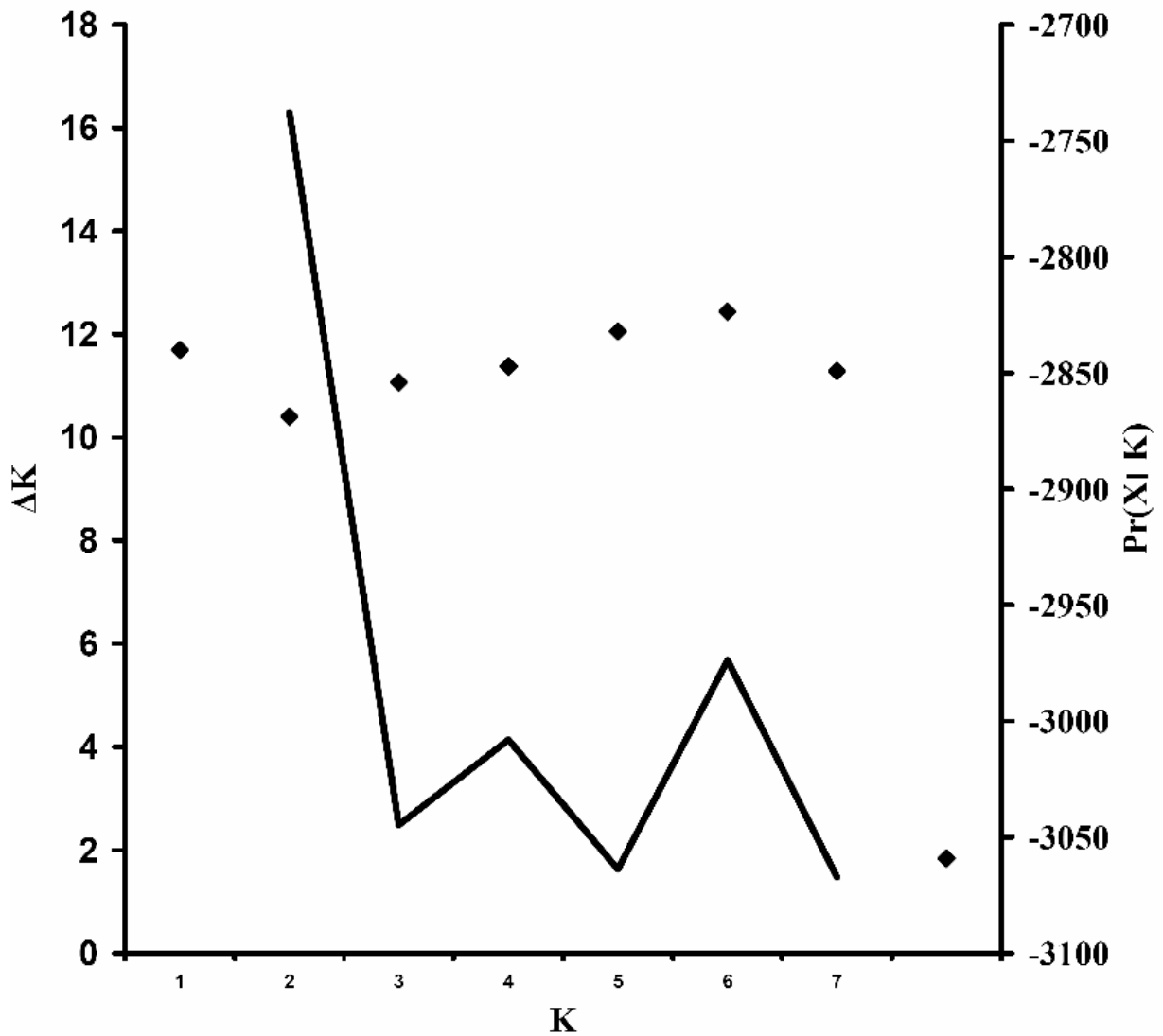
Figure 9. Two methods for estimating the true number of clusters in microsatellite dataset comprising samples from Liberty County, Florida (n=85). Delta-K (Evanno 2005) for values of K from 1 to 7 is scaled to the lefthand Y-axis. Likelihood of data given K (Pritchard *et al.*1999) is scaled to the righthand Y-axis. For both of these criteria, the maxima indicate the estimate of the true value of K.
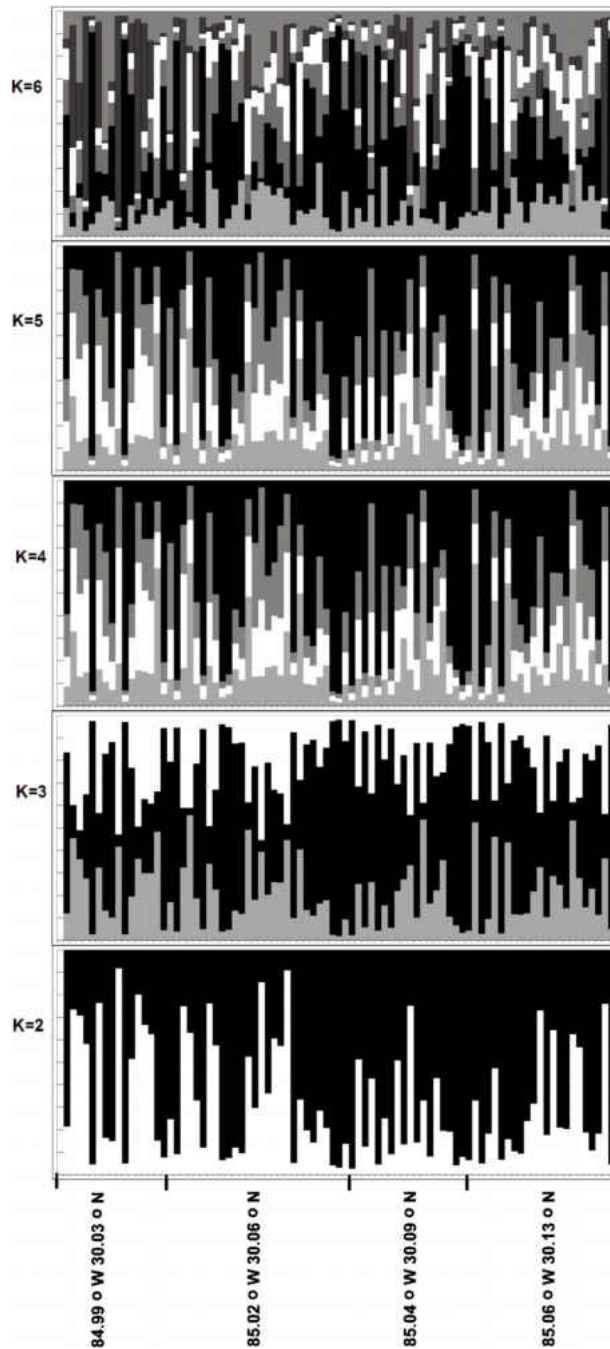
Figure 10. Results of STRUCTURE clustering analysis of only Liberty County, Florida samples. Y-axes are membership coefficient in each of K clusters (K=2-K=6 are displayed). All individuals collected in Liberty County, Fl in 2006 and 2007 (n=85) were included in a single analysis. Because these samples were collected haphazardly along a stretch of road and not in discrete localities, X axis is organized from left to right by GPS coordinate along this road from southeast to northwest.

44

Population differentiation and patterns of isolation by distance in microsatellite loci – To

estimate the level of population structure and to estimate the importance of geographic distance

on genetic isolation, we estimated global F_{ST} as well as F_{ST} among all pairwise population

comparisons. Global F_{ST} was highly significant, but quite low ($F_{ST} = 0.083$, 95% CI [0.04, 0.16])

All pairwise F_{ST} estimates save one were significantly different than zero after Bonforonni

correction (Table 4). However, these estimates were also relatively small ranging from 0.04 to

0.12. F_{ST} estimates among the South Carolina samples ranged from 0.04 to 0.05. Pairwise F_{ST}

between the Florida and Georgia samples was 0.04. F_{ST} between "eastern" and "western"

samples were slightly larger ranging from 0.07 to 0.13 and averaging 0.10. A Mantel test for

correlation between genetic and geographic distance showed a significant association (P=0.02,

1000 randomizations) indicating an overall trend of isolation by distance (Fig. 11).

Table 4. Genetic and geographic distances among populations of *P. ornata* sampled for this study. Below diagonal are values of pairwise F_{ST} (Weir and Cockerham 1984). Straight line distance among populations in kilometers is display above diagonal. Note that there is a significant association among these variables (see Fig. 11).

	Colleton Co A, SC	Colleton Co B, SC	Jasper Co, SC	Liberty Co, FL	Baker Co, GA
Colleton Co A, SC	-	5.7	85	542	429
Colleton Co B, SC	0.046	-	87	547	433
Jasper Co, SC	0.048	0.044	-	463	353
Liberty Co, FL	0.125	0.111	0.087	-	139
Baker Co, GA	0.118*	0.104	0.066	0.042	-

* Not significant after bonforroni correction
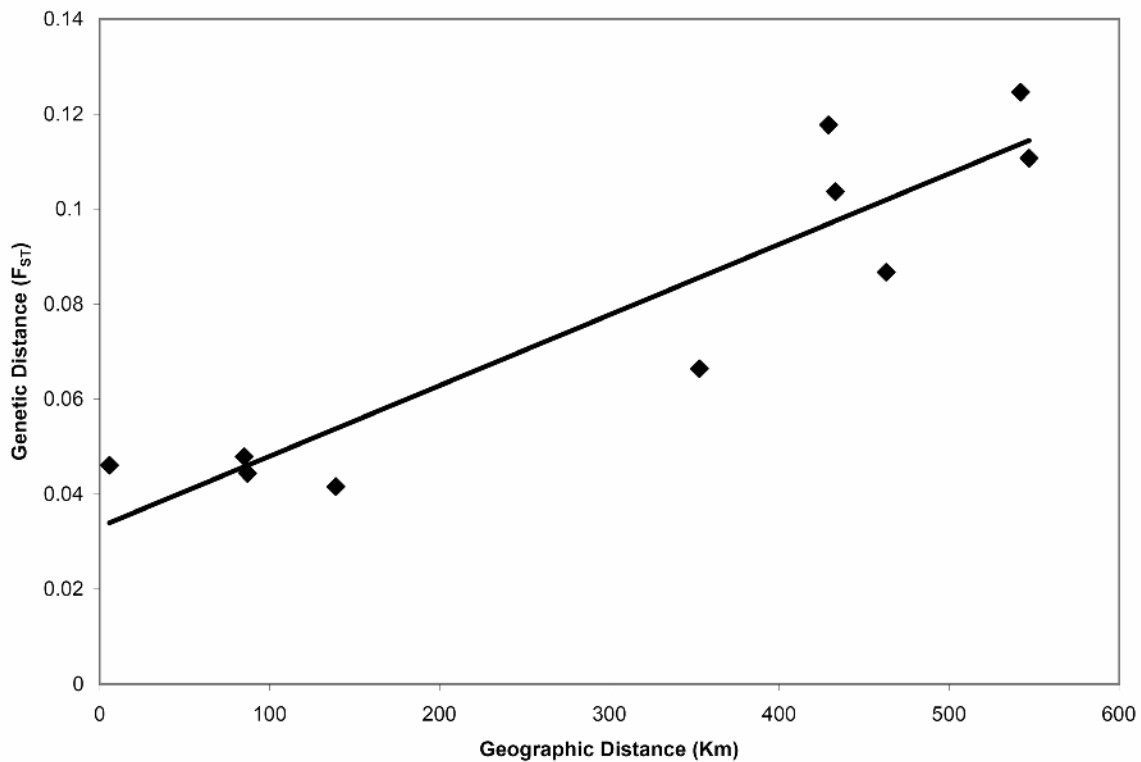
Figure 11. Plot of F_{ST} values (based on microsatellite data) versus geographic distance among all sampling localities indicating evidence for isolation by distance. F_{ST} values are plotted against corresponding straight-line geographic distances (d) between sites. The equation of the best fit line (shown) is $F_{ST} = 0.000149d + 0.033$ (Mantel Test, one-sided P = 0.02 from 1000 randomizations)

Discussion

The striking feature of these results was the relative lack of large scale genetic structure

in this species. All evidence from microsatellite markers and mitochondrial DNA sequence data

suggests a simple pattern of isolation by distance. Moreover, we were surprised to see that this

trend was even true across major geographic features such as the Apalachicola River (Fig. 3).

47

There is no evidence for distinct breaks in the distribution of *P. ornata* and gene flow seems relatively constant throughout the sampled range.

Do these patterns fit our expectations? We predicted three outcomes based on previous data. Hypothesis 1 predicted an overall pattern of isolation by distance due to the limitations of amphibian dispersal. We did observe this overall pattern of isolation by distance; however, we were surprised by the large geographic scale on which this pattern was evident. Using microsatellite data, we found no pattern of IBD in samples separated by up to 150 km (Fig. 11, Table 4). Furthermore, F_{ST} for samples separated by as much as 150 km was low ($F_{ST} < 0.05$, Table 4).

Our second hypothesis predicted high differentiation among populations as measured by F_{ST} and detectable genetic structure on small spatial scales. However, this was not the pattern we observed. Differentiation among populations was at most moderate (maximum pairwise $F_{ST} = 0.125$), and this was between samples separated by 550 km. This result is markedly different than the expectation that amphibian species have limited gene flow among subpopulations and are characterized by a high degree of natural genetic structure.

Our second hypothesis also predicted that genetic structure would be significant even on geographic scales of several kilometers. Interestingly, there was some evidence for fine scale population genetic structure in this species. However, this fine scale genetic structure was only present in the more urbanized of two widespread sampling localities. In Colleton County, SC, samples separated by 5.8 km were distinguishable based on assignment to different genetic clusters in STRUCTURE analysis and significant values of F_{ST} between discrete sampling sites ($F_{ST} = 0.046$). However, samples from Liberty County, Florida collected over an even larger

geographic range (14.1 km) were not assignable to discrete genetic clusters and appeared to be from a single population (Figs. 9, 10). Several previous studies in amphibians have found differences in genetic structure, migration and dispersal, or density and abundance in rural versus urban (Rubbo and Kiesecker 2005), developed versus undeveloped (Carr and Fahrig 2001, Vos *et al.* 2001), and fragmented versus non-fragmented habitats (Gibbs 1998). Thus, it is likely that the differences in patterns of genetic structure observed within these two sampling sites are due at least in part to urbanization and landscape fragmentation.

Third, we hypothesized a sharp division in genetic structure corresponding to crossing the Apalachicola River. However, our data suggest that the Apalachicola River basin does not impede gene flow in *P. ornata*. The STRUCTURE analysis clearly assigned all 6 individuals sampled from west of the Apalachicola River to genetic clusters comprised of individuals east of the Apalachicola River rather to their own unique genetic cluster. Furthermore, using NCPA, both the entire haplotype network and sub-clades comprising individuals sampled from east and west of the Apalachicola River showed evidence of isolation by distance (Fig. 3). Phylogenetic analysis supported the monophyly of clades which included individuals from both east and west of the Apalachicola River. Taken together, our data do not support a historic split that has been contemporaneously maintained in *P. ornata* attributable to the Apalachicola River. Either *P. ornata* was relatively unaffected by historical changes that caused the Apalachicola River to divide so many other species, or *P. ornata* underwent a range expansion from one side of the Apalachicola River to both after the Pleistocene. Alternatively, the Apalachicola River may indeed act as a barrier to dispersal. However, our sampling scheme failed to detect this pattern. Our populations west of the Apalachicola River were located close to the river and lack of

genetic divergence may indicate that frogs traverse within the river's floodplain. Further testing of genetic variation in populations further west could address this explanation.

As a member of the genus *Pseudacris*, we do not suppose that the high basal levels of gene-flow in *P. ornata*, are due to an extraordinary ability to disperse long distances. Indeed, other studies have found that individual frogs usually travel little. For example, Kramer (1973) studied movement for a five month period in *P. trisariata* and found that most (76%) frogs stayed within 100 m of breeding ponds while only 2% moved more than 200 m. Instead, we hypothesize that this species was historically abundant and continuously distributed throughout the species range. This continuity would allow for migration among neighboring populations and gene flow between geographically distant populations via many intermediate populations (i.e. a stepping stone model of migration).

Amphibians are often thought to naturally occur in highly structured metapopulations (Beebee 1996, 2005, Palo *et al.* 2004, Newman and Squire 2002). Thus, some authors have concluded that genetic load due to *outbreeding* depression may be of great consequence in amphibian species while genetic load due to *inbreeding* may be comparatively unimportant (Sagvic *et al.* 2005). However, our results demonstrate the opposite may be true in some amphibian species. This may have profound conservation implications. Populations of the Ornate Chorus Frog appear to be historically highly interconnected by gene-flow. Hence, genetic effective sizes were historically high, and the species is expected to carry a comparatively large number of deleterious recessive alleles. Whereas, this would not have been a concern within populations of large effective population size, suddenly fragmented populations would be expected to suffer reduced fitness owing to the cost of exhibited deleterious recessive

genes due to inbreeding. This study highlights the importance of understanding the historical evolutionary history of a species to determine the influence of fragmentation on that species. Indeed, the genetic load caused by sudden habitat fragmentation and increased inbreeding may be of greater consequence than was previously supposed in historically unstructured populations such as was found in the Ornate Chorus Frog.

References

Avise J, Arnold J, Ball R, *et al.* (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* **18**, 489-522.

Avise J, Giblin-Davidson J, Laerm J, Patton J, Lansman R (1979) Mitochondrial DNA clones and patriarchal phylogeny within and among geographic populations of the pocket gopher, Geomys pinetis. *Proceedings of the National Academy of Sciences* **76**, 6694-6698.

Avise J, Nelson WS (1989) Molecular genetic relationships of the extinct dusky seaside sparrow. *Science* **243**, 646-648.

Beebee T (1996) *Ecology and Conservation of Amphibians* Chapman and Hall, London.

Beebee T (2005) Conservation genetics of amphibians. *Heredity* **95**, 423-427.

Bermingham E, Avise J (1986) Molecular zoogeography of freshwater fishes in the southeastern United States. *Genetics* **113**, 939-965.

Blainey R (1971) An annotated check list and biogeographic analysis of insular herpetofauna of the Apalachicola region, Florida. *Herpetologica* **27**, 406-430.

Carr LW, Fahrig L (2001) Effect of road traffic on two amphibian species of differing vagility. *Conservation Biology* **15**, 1071-1078.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**, 1657-1659.

Cooke C (1945) Geology of Florida. *Florida Geological Survey Bulletin*, 1-339.

Cronin T, Szabo B, Ager T, Hazel J, Owens J (1981) Quaternary climates and sea levels of the
U.S. Atlantic Coastal Plain. *Science* **211**, 233-240.

Donovan MF, Semlitsch RD, Routman EJ (2000) Biogeography of the southeastern United
States: a comparison of salamander phylogeographic studies. *Evolution* **54**, 1449 - 1456.

Ellsworth D, Honeycutt R, Silvey N, Vickham J, Klimstra W (1994) Historical biogeography
and contemporary patterns of mitochondrial DNA variation in white tailed deer from the
southeastern United States. *Evolution* **48**, 122-136.

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

Funk WC, Blouin MS, Corn PS*, et al.* (2005) Population structure of Columbia spotted frogs
(*Rana luteiventris*) is strongly affected by the landscape. *Molecular Ecology* **14**, 483-496.

Gibbs JP (1998) Distribution of woodland amphibians along a forest fragmentation gradient.
Landscape Ecology **13**, 263-268.

Goebel A, Donnelly JM, Atz ME (1999) PCR primers and amplification methods for 12S
ribosomal DNA, the control region, cytochrome oxidase I, and cytochrome b in bufonids
and other frogs, and an overview of PCR primers which have amplified DNA in
amphibians successfully. *Molecular Phylogenetics and Evolution* **11**, 163-199.

Goudet (1995) FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. *Journal of
Heredity* **86**, 485-486.

Hartl DL, Clark AG (1997) *Principles of population genetics*, 3rd edn. Sinauer Associates,
Sunderland MA.

Hayes J, Harrison R (1992) Variation in mitochondrial DNA and biogeographic history of woodrats (Neotoma) of the eastern United States. *Systematic Biology* **41**, 331-344.

Hoffman EA, Schueler FW, Blouin MS (2004) Effective population sizes and temporal stability of genetic structure in *Rana pipiens*, the northern leopard frog. *Evolution* **58**, 2536-2545.

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. *BMC Genetics* **6**.

Jorgen S, Tobias U, Olsson M (2005) Outbreeding depression in the common frog, *Rana temporaria*. *Conservation Genetics* **6**, 205-211.

Kass AC, Raftery AE (1995) Bayes factors. *Journal of the American Statistical Association*. 90:733-795

Kramer DC (1973) Movements of western chorus frogs *Pseudacris triseriata triseriata* taged with Co^{60}. *Journal of Herpetology* **7**, 231-235.

Monsen KJ, Blouin MS (2003) Genetic structure in a montane ranid frog: restricted gene flow and nuclear–mitochondrial discordance. *Molecular Ecology* **12**, 3275-3286.

Moritz C, Schneider CJ, B WD (1992) Evolutionary relationships within the *Ensatina eschscholtzii* complex confirm the ring species interpretation. *Systematic Biology* **41**, 273-291.

Newman RA, Squire T (2001) Microsatellite variation and fine-scale population structure in the wood frog (*Rana sylvatica*). *Molecular Ecology* **10**, 1087-1100.

Nylander JAA (2004) MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Osentoski MF, Lamb T (1995) Intraspecific phylogeography of the gopher tortoise, *Gopherus polyphemus*: RFLP analysis of amplified mtDNA segments. *Molecular Ecology* **4**, 709-718.

Palo JU, Schmeller DS, Laurila A*, et al.* (2004) High degree of population subdivision in a widespread amphibian. *Molecular Ecology* **13**, 2631-2644.

Panchal M (2007) The automation of nested clade phylogeographic analysis. *Bioinformatics* **23**, 509-510.

Phillips C, Sexton O (1994) Geographic distribution of mitochondrial DNA variants and the historical biogeography of the spotted salmander, *Ambystoma maculatum*. *Evolution* **48**, 597-607.

Posada D, Crandall KA, Templeton A (2000) GeoDis: A program for the Cladistic Nested Analysis of the Geographical Distribution of Genetic Haplotypes. *Molecular Ecology* **9**, 487-488.

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

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574.

Rosenberg NA, Pritchard JK, Weber JL*, et al.* (2002) Genetic structure of human populations. *Science* **298**, 2381-2385.

Rubbo MJ, Kiesecker JM (2005) Amphibian breeding distribution in an urbanized landscape. *Conservation Biology* **19**, 504-511.

Schneider S, Roessli D, Excoffier L (2000) Arlequin: A software for population genetics data
analysis. Ver 3.000. Genetics and Biometry Lab, Department of Anthropology,
University of Geneva.

Shaffer HB, Fellers GM, Magee A, Voss SR (2000) The genetics of amphibian declines:
population substructure and molecular differentiation in the Yosemite Toad, *Bufo
canorus* (Anura, Bufonidae) based on single-strand conformation polymorphism analysis
(SSCP) and mitochondrial DNA sequence data. *Molecular Ecology* **9**, 245-257.

Templeton A (2004) Statistical phylogeography: methods of evaluating and minimizing
inference errors. *Molecular Ecology* **13**, 789-809.

Templeton A, Boerwinkle E, Sing C (1987) A cladistic analysisof phenotypic associations with
haplotypes inferred fromrestriction endonuclease mapping. I. Basic theory and an
analysis of Alcohol Dehydrogenase activity in *Drosophila*. *Genetics* **117**, 343-351.

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

Templeton A, Routman E, Phillips C (1995) Separating population structure from population
history: a cladistic analysis of the geographical distribution of mitochondrial DNA
haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics* **140**, 767-782.

Vogler A, Desalle R (1993) Phylogeographic patterns in the coastal North American tiger beetles
(*Cicindela dorsalis*) inferred from mitochondrial DNA sequences. *Evolution* **47**, 1192-
1202.

Vos CC, Antonisse-De Jong AG, Goedhart PW, Smulders MJM (2001) Genetic similarity as a measure for connectivity between fragmented populations of the moor frog (*Rana arvalis*). *Heredity* **86**, 598 - 608.

Walker D, Burkr V, Barak I, Avise J (1995) A comparison of mtDNA restriction sites versus control region sequences in phylogeographic assessment of the musk turtle (*Sternotherus minor*). *Molecular Ecology* **4**, 365-373.

Wallace B (1970) *Genetic load. Its biological and conceptual aspects* Prentice-Hall Inc., Englewood Cliffs, NJ.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the analysis of popultion structure. *Evolution* **38**, 1358-1370.

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

Wright S (1942) Isolation by Distance. *Genetics*, 114-138.

CHAPTER 4 –LATITUDINAL CLINE IN SKIN COLOR OF THE ORNATE CHORUS FROG (*PSEUDACRIS ORNATA*): NATURAL SELECTION OR NEUTRAL EVOLUTIONARY FORCES.

Introduction

Genetic and phenotypic clines are observed across a vast array of taxa (Szymura and Barton 1991, Berry and Kreitman 1993, Long and Singh, 1995, Sotka et al. 2004, Toju and Sota 2006). These phenotypic clines can be due to spatially varying natural selection (Haldane 1948), and are often taken as evidence of such. However, the neutral evolutionary forces of genetic drift and migration can equally act to maintain clinal genetic or phenotypic variation (Gould and Johnston 1972). Distinguishing the relative roles of selection versus neutral evolutionary forces in maintaining clinal variation remains an important goal of population biology (Endler 1977, Storz 2002, Vasemagi 2006).

Population genetics and genomics are perhaps our most powerful tools for explaining the role of natural selection in the maintenance of clinal variation. Because natural selection leaves predictable signatures on population genetic variation, population genetics theory can be used to statistically reject the null hypothesis of no selection (i.e. neutral evolution). Moreover, population genetics can be used to accurately distinguish classes of natural selection acting on a locus or trait (e.g. balancing selection or selective sweeps). Nevertheless, population genetic theory also suffers from limitations and biases. First, the proportion of phenotypic traits that have known single-locus effectors is small. Thus, much of population genetic theory is not directly applicable to the majority of traits, even those in model organisms. Furthermore, the

proportion of organisms for which even the simplest genetic tools have been developed (e.g. molecular marker) is likewise very small. Even when it is possible to use population genetic theory to test for evidence of natural selection, there are biases toward assuming a natural selective explanation without due process.

For example, single-locus environmental or latitudinal clines in allelic variation have been taken as evidence for natural selection despite the ability of neutral evolution via isolation by distance (IBD) to explain such a pattern (Vasemagi 2006). Using population genetic simulations Vasemagi (2006) was able to show that significant associations between latitude and single locus allele frequencies were common for species that showed patterns of isolation by distance. To identify loci that truly fell outside of expectations of neutrality, Vasemagi (2006) used estimates of the joint distribution of correlation coefficient and slope of allele frequency versus latitude. Because this approach identifies particular loci as statistical outliers, it can be used as a general approach to identify loci influenced by natural selection.

Color polymorphisms are often shown to be under natural selection in many species, such as the peppered moth (Kettlewell 1955, 1956), the Rock Pocket Mouse (Hoekstra, et al. 2005) and the Tiger Beetle (Hadley et al. 1988). Therefore, it is sometimes presumed that all color polymorphism are under selection. However, some studies that have statistically tested for natural selection have found that neutral genetic drift or very weak selection appears to maintain phenotypic polymophism (e.g., O'Hara 2005, Oxford 2005, Hoffman et al. 2006). These negative results demonstrate the importance of initially showing evidence that natural selection maintains a color polymorphism before jumping directly to explaining the mechanism of natural selection.

The Ornate Chorus Frog offers a unique opportunity to study this potentially selectively important trait in a natural system. Blouin (1989) showed that a common green/brown color polymorphism in the Ornate Chorus Frog is controlled by a simple Mendelian-inherited gene where the green allele shows complete dominance over the brown allele. Thus, using Hardy-Weinberg equilibrium assumptions, allele frequencies for the color locus can be estimated from observed phenotype frequencies. Obtaining these frequencies enables us to estimate genetic variation and population genetic structure occurring at the color locus and compare this variation directly to that at selectively neutral loci (e.g. microsatellites). This allows us to test the null hypothesis of selective neutrality at the color locus by rejecting a hypothesis of similar genetic structure to that estimated from neutral genetic markers. To our knowledge, this is the first study describing any geographic patterns of this polymorphism or attempting to determine the role of natural selection in this unique and interesting system.

Three methods were used to test the ultimate hypothesis that natural selection maintains the relative frequencies of alleles determining skin color in the ornate chorus frog. First, we tested for a correlation of allele frequency at the color locus with latitude and compared the strength of this correlation to ~300 microsatellite alleles to determine if natural selection was necessary to explain the strength of any observed cline. Second, we simulated bi-allelic genetic loci under parameters that produced the same levels of IBD as that seen in the ornate chorus frog. This was done because microsatellite loci tend to have many rare alleles that would not be directly comparable to a bi-allelic system. We compared the strength of the correlation of color locus and latitude to the strength of this same correlation for 1000 simulated neutral loci. Finally, we compared overall allelic divergence among populations of the ornate chorus frog using the

methods described in Beamont and Nichols (1996). This method relies on the fact that divergent or disruptive selection at a genetic locus will cause increased divergence among populations as compared to neutral loci while balancing selection will cause decreased divergence among populations compared to neutral loci. This method has been successful in detecting divergent selection (Storz and Nachman 2003), balancing selection (Gillespie and Oxford 1998, Andres et al. 2000, 2002) and coupled with computer simulations to determine statistical power, the lack of all but weak or no selection (Hoffman et al 2006). Finally, we discuss our results with regards to natural selection and the maintenance of color polymorphisms.

<u>Materials and Methods</u>

Sample collection and scoring of green and brown morphs - The overall strategy used to determine if natural selection was needed to explain observed patterns of variation in skin color was to compare patterns at neutral genetic loci to patterns observed at the color locus. Samples were collected from 5 discrete sampling localities (Fig. 12). Individual frogs were located in the winter and spring (November through May) of 2006 and 2007. Frogs were captured primarily by road cruising on rainy nights, or by hand-capturing calling males in breeding choruses. However, the samples from Barbour County, Alabama were captured in pitfall traps set along a drift fence surrounding an ephemeral pond. Captured frogs were scored for color (green or non-green), sex, and a precise location were recorded. Additionally, color frequency data for collections in Leon County Florida were obtained from Blouin (1986). However, because genetic samples were not available from this historic sample, these data were only included in

analyses where genetic marker data were not used directly. For DNA analysis, the longest toe on the right hind-foot was removed with sterilized surgical scissors. Toe clippings were placed in vials filed with anhydrous calcium sulfate (a desiccant) and transported to the University of Central Florida for genetic analysis. Ten independent microsatellite loci were scored using standard capillary electrophoresis techniques (Chapter 2). Patterns of allelic divergence and isolation by distance were characterized for putatively neutral microsatellite loci with a mantel test for correlation between genetic distance (F_{ST}) and geographic distance (See Chapter 3 for details.) Allele frequencies at the color locus were calculated as $q^2 = B$, where q represents the frequency of the recessive (brown allele) and B represents the frequency of the recessive phenotype (brown). These did not differ meaningfully from estimates using Lynch and Milligan's (1994) unbiased estimate of allele frequencies for dominant markers (i.e. all frequency estimates differed by less than 0.002) and so estimates from $q^2 = B$ were used in all future analyses.

Testing for the presence of a cline at the color locus and comparison to clines at microsatellite alleles. An ANOVA of the least squares linear regression of frequency of the green allele versus latitude was used to test the statistical significance of the observed differences in color frequencies among samples collected at different latitudes (Fig. 13). This test was implemented in the statistics package R. We performed the regression analysis on both raw allele frequency data and on arcsine transformed allele frequency data (Sokal and Rohlf 1995). We also used Aikaiki Information Criterion (AIC) to determine if the additional free parameter in a quadratic regression was justified. Because a significant correlation of a portion of allele

frequencies and latitude is expected even for neutral loci under a pattern of isolation by distance (IBD, Vasemagi 2006), we compared the strength of the correlation with latitude between putatively neutral microsatellite alleles (n=312 and the color locus alleles. For each individual microsatellite allele, the-least-squares linear regression of allele frequency verus latitude was estimated. The absolute slope and r^2 value for all 312 microsatellite alleles were plotted together. The 90% and 95% confidence interval for the joint distribution of r^2 and slope (scaled to take a maximum of 1) under neutrality were approximated by ellipses drawn around the 282 and 298 (90% and 95% respectively) points falling closest to the origin (Fig. 14).

Simulation of neutral bi-allelic datasets showing equivalent patterns of IBD-To further test the role of selection in maintaining observed patterns of variation at the color locus, we used population genetic simulations to recreate datasets that reflect the 2-allele system acting at the color locus. Simulations were performed in EASYPOP (Balloux 2001) under a stepping stone migration model with 5 populations (2000 generations) followed by 25, 50, or 75 generations without migration. This method allowed us to vary the slope of the IBD correlation and the overall divergence among populations by changing the migration rate. We were also able to vary the strength of the linear correlation (r^2) by changing the number of final generations that populations drifted without migration. We first used a series of exploratory analyses that mimicked the empirical microsatellite dataset to identified a migration rate and number of no-migration generations that consistently produced patterns of isolation-by-distance equivalent to that seen in the microsatellite data-set (in terms of F_{ST} and the slope and correlation coefficient (r^2) of latitudinal distance-and genetic distance). For these exploratory analyses, we assumed

that each step in the stepping-stone model represented an equal latitudinal distance $1/4^{th}$ of the largest difference in latitude between actual sampled populations. We used 10 alleles with mutation parameters similar to microsatellite loci (i.e. 0.0001 mutations/gamete/generation, 70% 1-step mutations, 30% multi-step mutations, 100 total allelic states). Ten simulations for each of the following migration rates and final no-migration generations were performed: 0.1, 0.05, 0.025, 0.01, 0.005, and 0.00025 each combined with 25, 50, and 75 final no-migration generations. The mean value of global F_{ST} and the slope and r^2 of the correlation between genetic and latitudinal distance were plotted for all combinations of parameters (Fig. 15). The simulation parameters that gave statistics closest to the empirical dataset were used for simulations of bi-allelic neutral loci.

When we arrived at an appropriate migration rate and number of final no-migration generations, we used this overall model to simulate datasets which mimic the color locus evolving under only selectively neutral forces. The model that produced patterns of isolation by distance and overall divergence most similar to that observed in the empirical microsatellite dataset was a stepping stone model of migration with a migration rate of 1% of individuals migrating per generation between neighboring populations followed by 50 generations of no-migration. We performed simulations for both 5 and 6 populations (corresponding to the inclusion and exclusion of historical green allele frequency data). Simulations were run for 2000 generations followed by 50 generations without any migration. Breeding population sizes in all simulated populations were 1000 with equal numbers of males and females. Mutation rate was set to 10^{-4} with only 2 possible allelic states. Simulations were performed for 1000 independent loci.

As with the microsatellite alleles, we compared the strength of the correlation with latitude between simulated bialleleic neutral loci (n=1000) and the color locus. Latitudes were assigned to the simulated populations such that the first population in the stepping stone scheme corresponded to the lowest latitude in the empirical dataset, the last population corresponded to the highest latitude. Other populations were assigned such that each step in the stepping stone model corresponded to equal change in latitude. Because we performed the exploratory analysis using latitudinal distance, we did not need to correct for imperfect correlation between latitudinal and geographic distance in the empirical dataset. The absolute slope and r^2 value of the least squares regression for 1000 simulated alleles were plotted together. The 90% and 95% confidence interval for the joint distribution of r^2 and slope (scaled to take a maximum of 1) under neutrality were approximated by ellipses drawn around the 900 and 950 (90% and 95% respectively) points falling closest to the origin (Fig. 16 and 17).

Analysis of allelic divergence among populations – We used the statistical methods of Beaumont and Nichols (1996) to determine if the overall allelic divergence at the color locus was significantly different than would be expected under neutrality. Briefly, null distributions and 95% confidence intervals of global F_{ST}, conditional on heterozygosity, were built via coalescent simulations that mimic the properties of the 10 neutral microsatellite markers described in Chapter 2 and 3 (in terms of median F_{ST}, number of populations sampled, and sample size at each population) to obtain confidence limits on F_{ST} under neutrality. We followed the recommendations of Beaumont and Nichols (1996) and used a model with 100 islands and a sample size of 50 individuals per sampling locality when true sample sizes exceeded 50. We

65

then determined if F_{ST} at the color locus fell outside the confidence limits generated by this method, and thus if natural selection was necessary to explain overall allelic divergence at the color locus. We used this technique separately on both the entire sample and on only the "western" population-group where the frequency of the green allele and thus our ability to detect selection was highest (Fig. 18, 19). Simulations were implemented in the software FDIST2 (Beaumont and Nichols 1996.)

<div align="center">Results</div>

Establishment of the presence of a latitudinal cline in skin color frequencies – There was a clear trend toward a higher frequency of the "green allele" in southern samples as compared to northern samples. The frequency of green frogs in samples from the Florida panhandle ranged from 13% to 27% (corresponding to an estimated frequency of the green allele of 0.06-0.15). The frequency of green frogs in the Baker County, Georgia sample was 5% (green allele 3%) while the frequency of green frogs in the South Carolina samples ranged from 0% to 5% (green allele 0% – 3%.) With an analysis of variance (ANOVA), we tested the null hypothesis that the slope of the least squares linear regression of green allele frequency against latitude was zero (F=8.08, P=0.047). This null hypothesis was rejected and thus, there is a significant linear correlation of the frequency of the green allele and latitude. This was slightly more significant after arcsine transformation of allele frequency data ($r^2 = 0.69$, F = 10.3 P =0.033; Sokal and Rohlf 1995.) A quadratic regression produced a higher value of r^2 (r^2=0.89), but AIC indicated

that the better fit did not justify the additional model complexity. Thus, linear regression was

used in subsequent analyses.
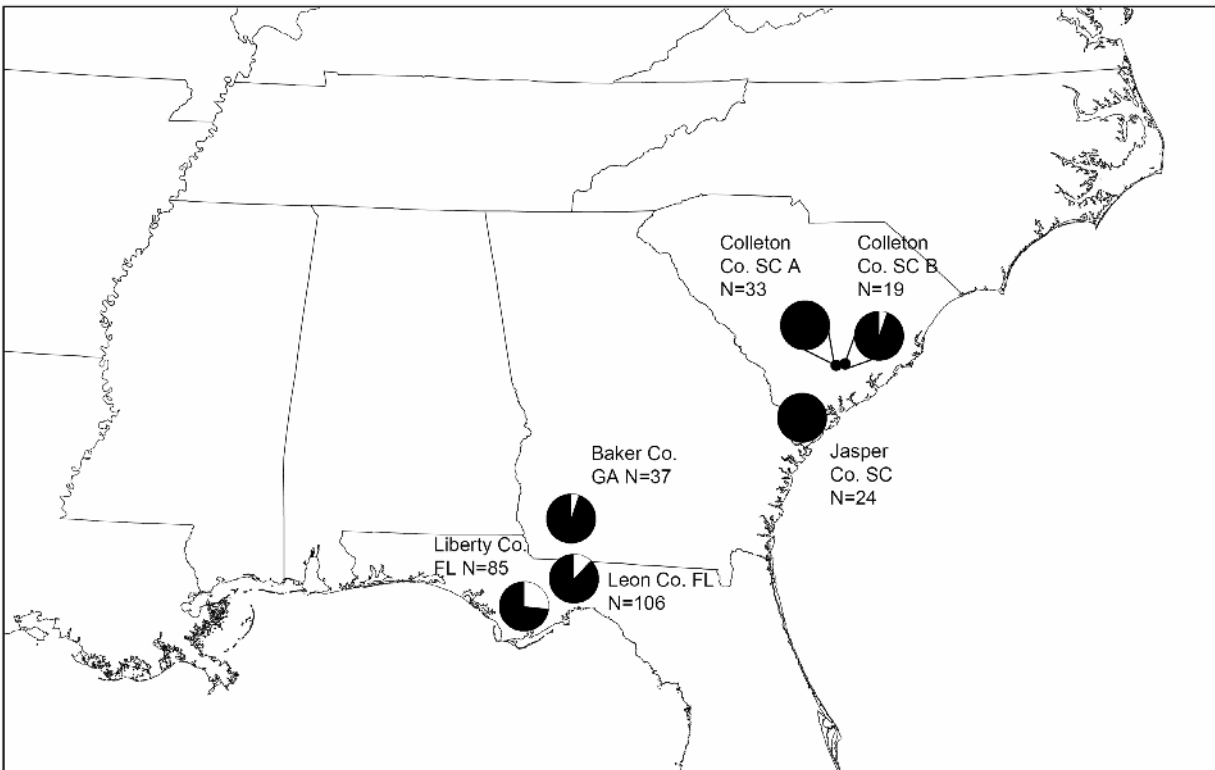
Figure 12. Frequency of color morphs throughout the native range of P. ornata. The frequency of brown frogs is represented by black in pie charts while the frequency of green frogs is represented by white. Leon County, Florida frequencies were taken from Blouin (1989) while all other frequencies were estimated from samples collected in 2006 and 2007.
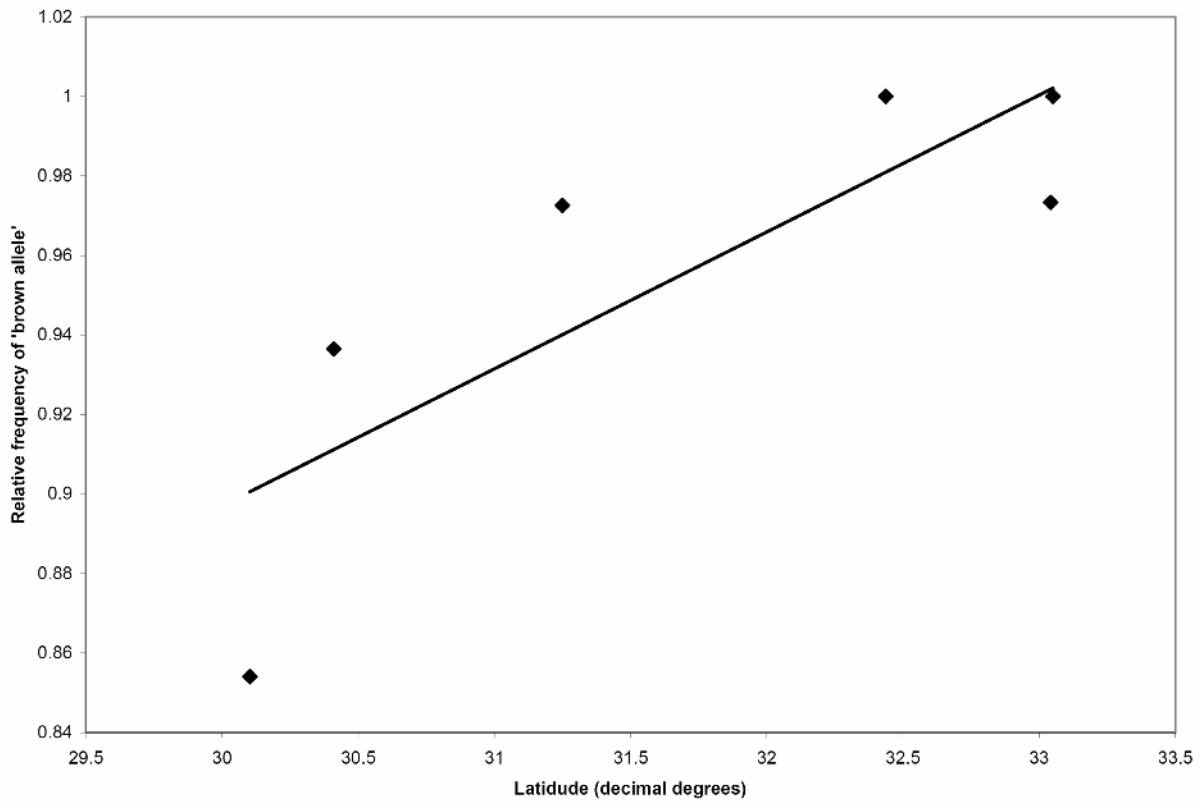
Figure 13. Linear regression analysis of the frequency of "brown allele" against latitude. The best fit line corresponds to $y = 0.0344x - 0.136$ with an r^2 value of 0.6688. An ANOVA indicates that the regression model was significantly better than the mean at explaining the variance in brown allele frequency ($P = .047$). Similar results were obtained for arcsine transformed allele frequencies ($r^2 = 0.69$, $P = 0.033$)

Comparison of cline at color locus to clines at microsatellite alleles – Plots of absolute

slope (*b*) against the coefficient of determination of least-squares linear regression (r^2) showed

that a relatively high proportion of microsatellite alleles were also correlated with latitude (Fig.

14). Presumably, these correlations were due to the overall pattern of isolation by distance

caused by the neutral forces of genetic drift and migration, and not due to differential selection at

microsatellite loci or at closely linked loci. One method used to determine if natural selection

was necessary to explain the observed cline was a direct comparison of *b* and r^2 at the color locus

alleles to the observed cline at neutral microsatellite alleles. The estimates of *b* and r^2 at the

color locus were 4% change in allele frequency per latitudinal degree, and 0.73, respectively.

These estimates fell just inside the 95% confidence interval created from the microsatellite

alleles, but outside the 90% confidence interval. Thus, this analysis gave marginal evidence that

natural selection was acting to create a steeper more linear cline in color allele frequencies than

in microsatellite allele frequencies. However, we must recognize that 1) microsatellite alleles at

a single locus are not independent and 2) there are many rare microsatellite alleles that could not

possibly show slopes as high as the color locus owing to their rarity in the total sample. Thus, we

tested these results by conducting computer simulations of bi-allelic datasets.
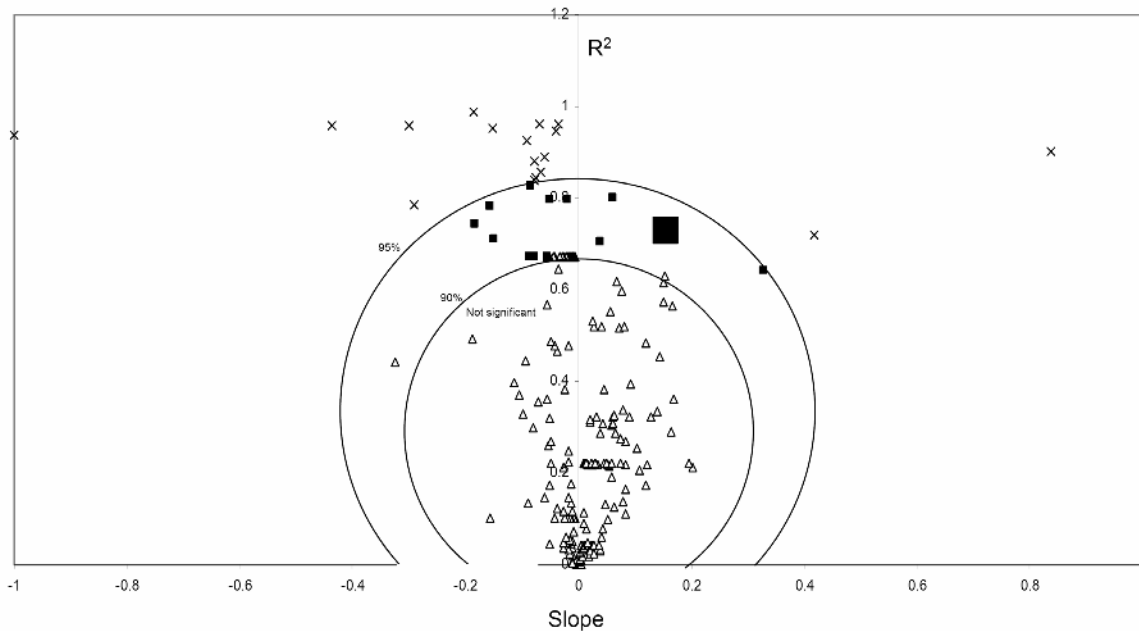
Figure 14. Slope (scaled to a maximum of 1) versus coefficient of determination (r^2) for the best fit line of allele frequency versus latitude for all microsatellite alleles and for the "brown allele." The 90% and 95% significance levels are approximated by semi-circles drawn around the least extreme 90% and 95% of microsatellite allele points. The color locus is represented by the larger black square falling outside the 90% significance curve.

Comparison of cline at color locus to clines in simulated neutral bi-allelic loci – We

conducted exploratory analyses to find appropriate simulation parameters for the level of IBD

observed in the empirical microsatellite dataset. These analyses gave a wide range of the

relevant statistics (global F_{ST}, and r^2 and slope of pairwise F_{ST} versus latitudinal distance, Fig.

15). However, for any one set of simulation parameters (migration rate and number of final

generations without migration), the variance among 10 replicates for these same statistics was

low (< 0.4% of mean value for global F_{ST} and slope, and < 27% of mean for r^2). Thus,

differences in observed IBD statistics among simulations describe differences in the simulation parameters rather than stochastic variation among replicates.

The set of simulation parameters that produced patterns of IBD most similar to the patterns observed in the empirical dataset were a migration rate of 0.01 for 2000 generations then 50 generations without any migration. Mean values of F_{ST}, slope (with units of change in pairwise F_{ST} per latitudinal degree), and r^2 for simulations using these parameters were 0.070, 0.031, and 0.81 respectively. These were very similar to the values obtained for the empirical dataset; that is 0.083, 0.029, and 0.75, respectively.

Plots of least squares regression statistics (b and r^2) of latitude versus allele frequency for simulated datasets including (Fig. 16) and excluding (Fig. 17) the historic samples are shown. Each of these plots represents 1000 independent realizations (i.e. loci) of the described stepping-stone simulations. Like the microsatellite alleles, a large portion of the simulated allele frequencies show a significant correlation with latitude. For analysis including the historic samples (N=6), the regression of the "brown allele" versus latitude gave a slope of 3.4% change per latitudinal degree with an r^2 value of 0.67. These statistics fell within both the 90% confidence interval created from simulated datasets. For the analysis excluding the historic samples, the slope and r^2 of the "brown allele" with latitude were 4.1% change per latitudinal degree with an r^2 value of 0.73. These values also fell within the 90% confidence interval created from simulated neutral datasets. Thus, overall using these simulations we found no evidence that natural selection was necessary to explain the observed cline in allele frequencies at the color locus.
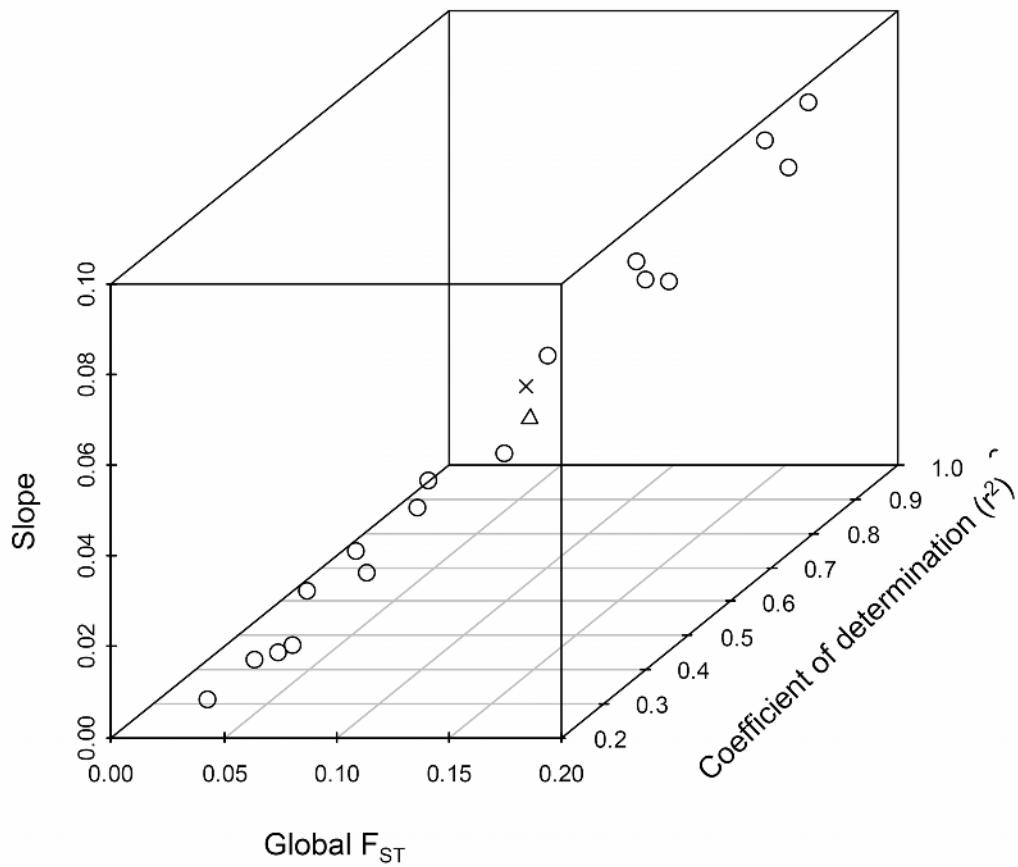
Figure 15. Results of exploratory simulations where we determined appropriate parameters to match the pattern of isolation by distance observed in the empirical microsatellite dataset. Values of global F_{ST}, and slope and coefficient of determination (r^2) for regression of pairwise F_{ST} versus pairwise latitudinal distance are shown. Values for empirical microsatellite dataset are displayed as a hollow triangle while values for the nearest set of simulation parameters and those chosen for future simulations are displayed as an X. All other simulation results are displayed as hollow circles.

Figure 16. Slope (scaled to a maximum of 1) versus coefficient of determination (r^2) for the best fit line of allele frequency versus latitude for all simulated alleles and for the "brown allele." This analysis excludes historical color-allele frequency data and thus the number of populations for all regressions was 5. The 90% and 95% significance levels are approximated by semi-circles drawn around the least extreme 90% and 95% of simulated allele points. The color locus is represented by the larger black triangle falling inside the 90% significance curve.
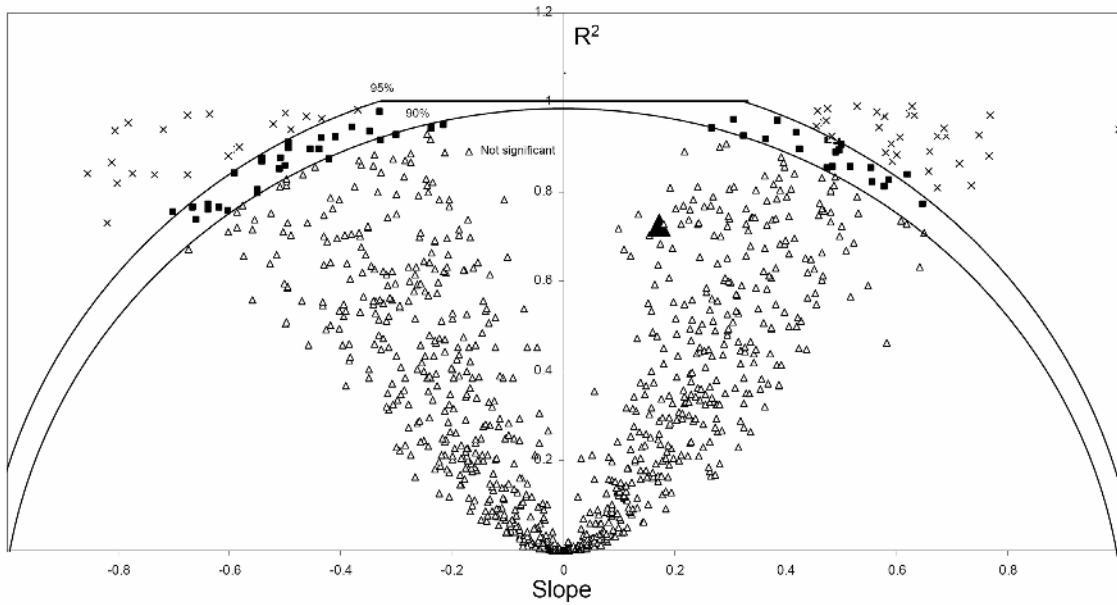
Figure 17. Slope (scaled to a maximum of 1) versus coefficient of determination (r^2) for the best fit line of allele frequency versus latitude for all simulated alleles and for the "brown allele." This analysis includes historical color-allele frequency data thus the number of populations for all regressions was 6. The 90% and 95% significance levels are approximated by semi-circles drawn around the least extreme 90% and 95% of simulated allele points. The color locus is represented by the larger black triangle falling inside the 90% significance curve.
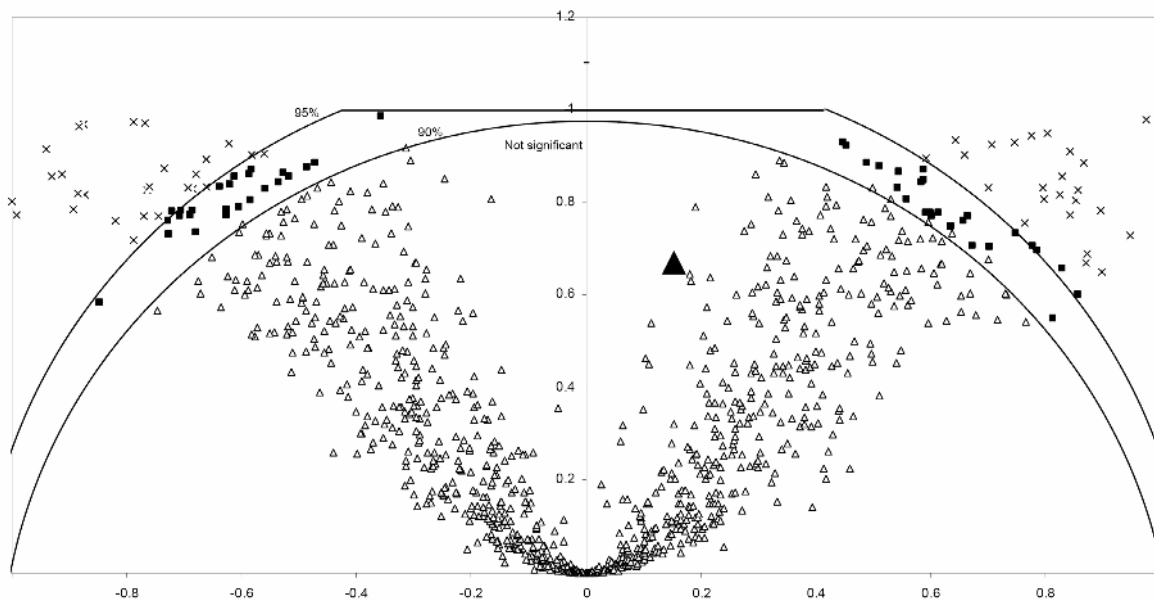
Overall allelic divergence among populations for microsatellite loci and the color locus –

In the total sample, global F_{ST} was calculated for the neutral microsatellite loci as 0.083. 95% confidence intervals of F_{ST} with heterozygosity estimated in FDIST2 were reasonably narrow such that if strong selection was acting to cause high divergence among populations at the color locus, it would be possible to detect (Fig. 18). Seven of 10 individual estimates of F_{ST} at microsatellite loci fell within the 95% confidence interval (Fig. 18). This was only slightly less than the 9.5 expected to fall within this interval due to chance alone. The single locus global F_{ST} estimate at the color locus was 0.086 and also fell well within the 95% confidence limits under neutrality

(95% CI [0.0, 0.18], Fig. 18). Thus, we found no evidence that natural selection was acting to disrupt allele frequencies at the color locus in the entire sample.

The same analysis was performed on a more local scale for the Florida and Georgia samples. Using only this "western group," global F_{ST} at microsatellite markers was 0.046 while F_{ST} at the color locus was 0.080. Nine of ten individual microsatellite loci estimates and the color locus fell well within the 95% confidence interval expected under neutrality (95% CI [0.0, 0.17, Fig. 19]. Again, there was no evidence for the selective maintenance of allele frequency differences among populations at the color locus.
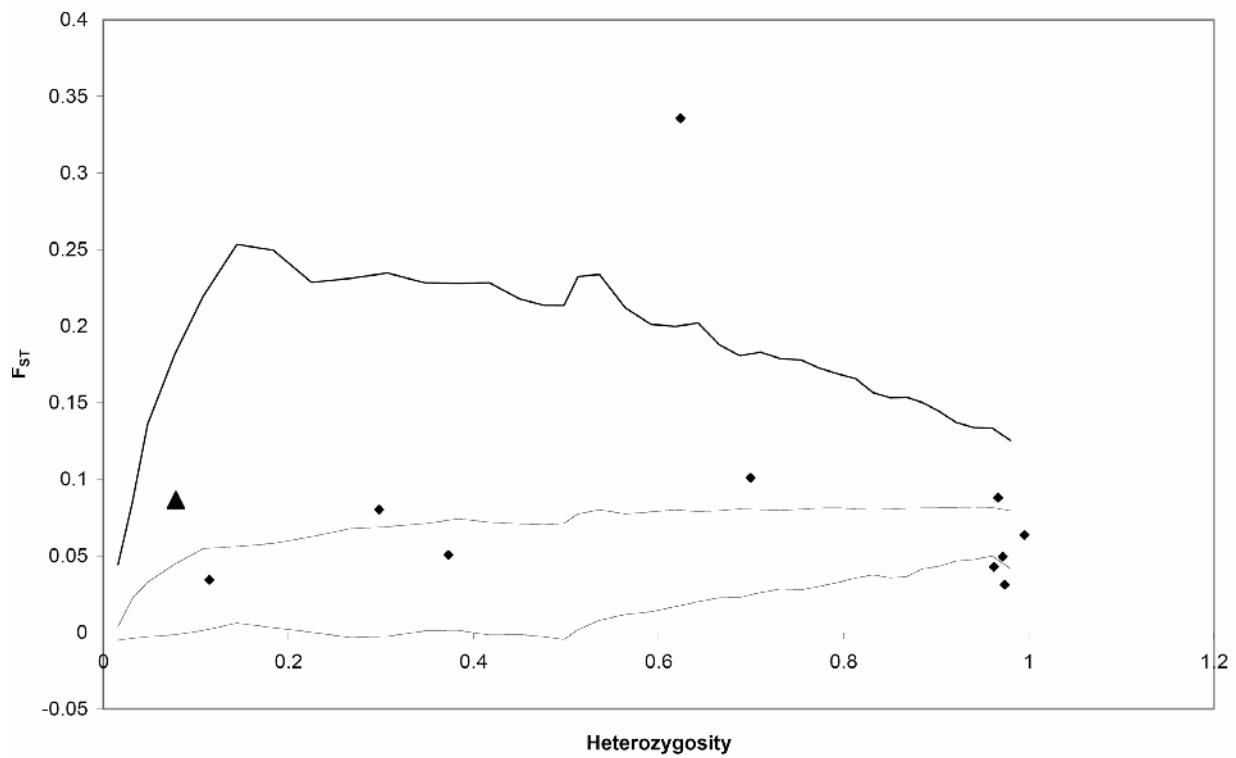
Figure 18. FDIST2 analysis where F_{ST} conditional on heterozygosity is plotted. This analysis included all sampled populations. 95% confidence intervals under neutrality are approximated by the area between 0.025 and 0.975 quantiles (solid line). The median (0.5 quantile) is also shown. Estimates of F_{ST} for microsatellite loci are represented by black diamonds while the estimate at the color locus is represented by a triangle.
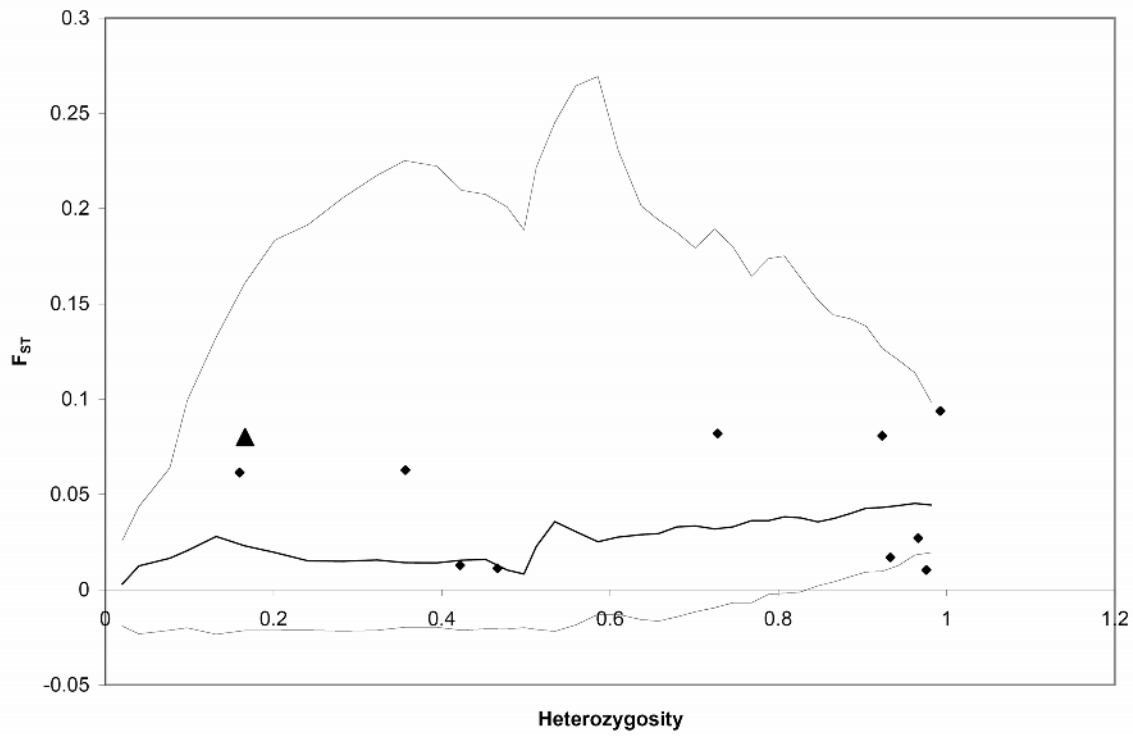
Figure 19. FDIST2 analysis where F_{ST} conditional on heterozygosity is plotted. This analysis includes only Liberty County, Florida samples (N=85) and Baker County, Georgia samples (N=38). 95% confidence intervals under neutrality are approximated by the area between 0.025 and 0.975 quantiles (solid line). The median (0.5 quantile) is also shown. Estimates of F_{ST} for microsatellite loci are represented by black diamonds while the estimate at the color locus is represented by a triangle.

Latitudinal cline in genetically controlled color frequencies of the ornate chorus frog – In this study, we demonstrated that there was a significant correlation between the frequency of green frogs and latitude (Fig. 13). There was a higher frequency of green frogs in southern samples and a lower frequency of green frogs in northern samples (Fig. 12). Blouin (1986) showed that the green-brown polymorphism in *P. ornata* is controlled by a single locus 2-allele system where the "green allele" shows complete dominance to the "brown allele." Therefore, we know that this phenotypic variation is heritable and hence meets the minimum requirements for adaptive evolution via natural selection. Further, in the Pacific Treefrog, *Pseudacris regilla*, studies have repeatedly shown that green and brown morphs showed habitat matching preference both in the presence and absence of a predator (Morey 1990, Wente and Phillips 2005). Coupling these behavioral observations with the fact that winters are longer and harsher in northern latitudes (and hence foliage browner longer), this system could arguably be explained by selection acting along a geographic cline. However, because this was a simple 1-locus 2 allele system, we were able to actually characterize the cline in light of an overall pattern of isolation by distance, and show that neutral evolutionary forces would not give a significantly different pattern than the observed latitudinal cline.

Correlations of environmental variables such as latitude with putatively heritable phenotypes have been considered one of the best evidences for the action of natural selection. Classic evidence for industrial melanism in the peppered moth is in part an observed cline in the frequency of dark-colored moth across a region of varying industrialization (Bishop 1972).

Acceptance of the selective control of sickle-cell alleles in humans is partly due to a cline in the frequency of the sickle-cell allele that corresponds to the presence/absence of malaria across the African continent (Livingstone 1958). Clinal genotypic or phenotypic variation continues to be taken as evidence for natural selection. For example, the correlation of flowering time with latitude was recently taken as strong evidence for local adaptation in the model plant *Arabidopsis thaliana* (Stinchcombe et al. 2004). Sezgin (2004) recently proposed 3 adaptive latitudinal clines for functional alleles in *Drosophila melanogaster* although this interpretation has been criticized (Vasemagi 2006). We do not necessarily contest the action of selection in these or other examples where clinal variation is taken as evidence for spatially varying selection. We only point out that without successfully rejecting a null hypothesis that the cline can be explained by neutral evolutionary forces a selective hypothesis cannot be substantiated.

This study adds to a growing number of studies that rigorously test for the signature of natural selection on presumably selectively important traits, but do not find it (e.g. Hoffman et al. 2006, Oxford 2005). These negative results actually have very important implications. First, they caution biologists against presuming that natural selection is the primary cause of observed phenotypic variation without sufficient evidence. Second, these studies suggest that natural selection on major phenotypic variation may not be as strong as was once thought.

It is important to point out that we have not ruled out the selective maintenance of this phenotypic character. Several reasons could explain why we failed to detect selection in this study. First the number of populations we were able to sample was relatively low (N=6). This certainly could affect the regression parameters (r^2 and *b*) we used to assess the significance of this cline. Second, using the methods we describe, it may be impossible to detect a selection

gradient even for strong selection and large sample sizes if 1) the relationship between 'ideal color frequency' and latitude is inherently non-linear and r^2 does not reflect the relative strength of selection over genetic drift, and 2) if the true slope of the relationship between 'ideal color frequency' and latitude is smaller than relationships commonly produced by the overall pattern of isolation by distance. Third, more complex models of selection than we address may be acting at this color locus. These models of selection (e.g. negative frequency-dependent selection occurring only in southern populations) make it difficult to detect selection using molecular techniques (see Oxford 2005).

We suggest that two approaches are necessary to determine the relative importance of natural selection in maintaining this color polymorphism. First, ecological experiments in natural environments where the relative frequencies of green and brown frogs are artificially manipulated could determine current strength and patterns of selection. Second, identifying the physical chromosomal location of the color locus and characterizing its gene products would allow for comparative genomics studies across a diverse group of anurans that show a similar green-brown color polymorphism (Hoffman et al. 2000, 2004). These could determine if this color polymorphism has been maintained by selection over thousands or millions of years of evolution.

There remains a tendency in biology to explain phenomenon in terms of adaptive evolution via natural selection. Rarely are alternatives to selective control sufficiently addressed. Here we demonstrate a latitudinal cline in phenotype frequencies, but also demonstrate that natural selection is not necessary to explain this cline. We caution that more rigorous testing

rather than environment-genotype/phenotype correlations are necessary to establish the

importance of natural selection in the evolution of traits, populations, and of species.

References

Andres JA, Sanchex-Guillen RA, Rivera AC (2000) Molecular evidence for selection on female color plymorphism in the damselfly *Ischnura graellsii*. *Evolution* **54**, 2156-2161.

Andres JA, Sanchex-Guillen RA, Rivera AC (2002) Evolution of female colour polymorphism in the damselflies: testing the hypothesis. *Animal Behavior* **63**, 677-685.

Balloux F (2001) EASYPOP (version 1.7): a computer program for population genetics simulations. *Journal of Heredity* **92**, 301-302.

Beaumont M, Nichols R (1996) Evaluating loci for use in genetic analysis of population structure. *Proceedings of the Royal Society of London* **263**, 1616-1626.

Berry A, Kreitman M (1993) Molecular analysis of an allozyme cline - alchohol dehydrogenase in *Drosophila melanogaster* on the east coast of North America. *Genetics* **134**, 869-893.

Bishop JA (1972) An experimental study of the cline of industrial melanism in *Biston betularia (L.)* (Lepidoptera) between urban Liverpool and rural North Wales. *The Journal of Animal Ecology,* **41**, 209-243.

Blouin MS (1986) Inheritance of a naturally occurring color polymorphism in the Ornate Chorus Frog. *Copeia* **4**, 1056-1056.

Endler J (1977) *Geographic variation, speciation, and clines* Prinston University Press, Prinston, NJ.

Gillespie R, Oxford GS (1998) Selection on the color polymorphism in Hawaiian happy-face spiders: evidence from genetic structure and temporal fluctuations. *Evolution* **52**, 775-783.

Gould S, Johnston R (1972) Geographic variation. *Annual Review of Ecology and Systematics* **3**, 457-498.

Hadley NF, Schultz TD, Savil A (1988) Spectral reflectances of three tiger beetle subspecies (*Neocicindela perhispida*): Correlations with their habitat substrate. **15**, 343-346.

Haldane J (1948) The theory of a cline. *Journal of Genetics* **48**, 277-284.

Hoekstra H, Krenz J, Nachman MW (2005) Local adaptation in the rock pocket mouse (Chaetodipus intermedius): natural selection and phylogenetic history of populations. *Heredity* **94**, 217-228.

Hoffman EA, Blouin MS (2000) A review of colour and pattern polymorphisms in anurans. *Biological Journal of the Linnean Society* **70**, 633-665.

Hoffman EA, Schueler FW, Jones AG, Blouin MS (2006) An analysis of selection on a colour polymorphism in the northern leopard frog. *Molecular Ecology* **15**, 2627-2641.

Kettlewell H (1955) Selection experiments on industrial melanism in the Lepidoptera. *Heredity* **9**, 323-342.

Kettlewell H (1956) Further selection experiments on industrial melansim in the Lepidoptera. *Heredity* **10**, 287-301.

Livingstone F (1958) Anthropological implications of sickle cell gene distribution in West Africa. *American Anthropologist* **60**, 533-562.

Long A, Singh R (1995) Molecules versus morphology - detection of selection acting on morphological characters along a cline in *Drosophila melanogaster. Heredity* **74**, 569-581.

Lynch M, MIlligan B (1994) Analysis of population genetic structure with RAPD markers. *Molecular Ecology* **3**, 91-99.

O'Hara R (2005) Comparing the effects of genetic drift and fluctuating selection on genotype frequency changes in the scarlet tiger moth. *Proceedings of the Royal Society of London* **272**, 211-217.

Oxford GS (2005) Genetic drift within a protected polymorphism: enigmatic variation in color-morph frequencies in the candy-stripe spider, *Enoplognatha ovata*. *Evolution* **59**, 2170-2184.

Sezgin E, Duvernell DD, Matzkin LM*, et al.* (2004) Single-locus latitudinal clines and their relationship to temperate adaptation in metabolic genes and derived alleles in *Drosophila melanogaster*. *Genetics* **168**, 923-931.

Smymura JM, Barton NH (1991) The genetic-structure of the hybrid zone between the fire bellied toads *Bombina bombina* and *B. Variegata* - comparisons between transects and between loci. *Evolution* **45**, 237-261.

Sokal R, Rohlf F (1981) *Biometry* WH Freeman, New York.

Sotka EE, Wares JP, Barth JA, Grosberg RK, Palumbi SR (2004) Strong genetic clines and geographical variation in gene flow in the rocky intertidal barnacle *Balanus glandula*. *Molecular Ecology* **13**, 2143-2156.

Stinchcombe J, Weinig C, Ungerer M*, et al.* (2004) A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene FRIGIDA. *Proceedings of the National Academy of Scieces USA* **101**, 4712-4717.

Storz JF (2002) Contrasting patterns of divergence in quantitative traits and neutral DNA

    markers: analysis of clinal variation. *Molecular Ecology* **2002**, 2537-2551.

Storz JF, Nachman MW (2003) Natural selection on protein polymorphism in the rodent genus

    Peromyscus: evidence from interlocus contrasts. *Evolution* **57**, 2628-2635.

Tuju H, Sota T (2006) Phylogeography and the geographic cline in the armament of a seed-

    predatory weevil: effects of historical events versus natural selection from the host plant.

    Molecular Ecology **15**, 4161-4173.

Vasemagi (2006) The adaptive hypothesis of clinal variation revisited: Single-locus clines as a

    result of spatially restricted gene flow. *Genetics* **173**, 2411-2414.

CHAPTER 5 –GENERAL CONCLUSION

In 1972, Gould and Johnston wrote. "The foundation of most evolutionary theory rests upon inference drawn from geographic variation or upon the verification of predictions made about it." This describes essentially the rationale behind this research. Using patterns of genotypic and phenotypic geographic variation in *Pseudacris ornata*, we were able to test specific predictions about the evolutionary processes shaping variation in this species in order to ultimately improve our understanding of evolution.

We predicted three outcomes for patterns of neutral genetic variation based on previous data. Hypothesis 1 predicted an overall pattern of isolation by distance due to the limitations of amphibian dispersal. We did observe this overall pattern of isolation IBD; however, we were surprised by the large geographic scale on which this pattern was evident. Hypothesis 2 predicted high differentiation among populations as measured by F_{ST} and detectable genetic structure on small spatial scales. However, this was not entirely the pattern we observed. Differentiation among populations was at most moderate (maximum pairwise $F_{ST} = 0.125$), and this was between samples separated by 550 km. Hypothesis 2 also predicted that genetic structure would be significant even on geographic scales of several kilometers, and there was evidence for fine scale population genetic structure in this species. However, this fine scale genetic structure was only present in the more urbanized of two widespread sampling localities. Hypothesis 3 predicted a sharp division in genetic structure corresponding to crossing the Apalachicola River. However, our data suggest that the Apalachicola River basin does not impede gene flow in *P. ornata*.

Finally, we predicted that natural selection would be the principle evolutionary force shaping variation in skin color in *P. ornata.* In this study, we demonstrated that there was a significant correlation between the frequency of green frogs and latitude. There was a higher frequency of green frogs in southern populations and a lower frequency of green frogs in northern populations. However, when we interpreted this phynotypic cline in light of the overall pattern of isolation by distance derived from neutral molecular markers, it was apparent that the neutral evolutionary forces of genetic drift and migration could equally explain the pattern of phenotypic variation we observed.

This study contributed to our understanding of the complex biogeographic history of the southeastern United States. Furthermore, this study contributed to our limited understanding of the strength and mechanisms of natural selection in natural environments.