December 2019

Testing the Waters: Diversification and Selection in the *Nerodia fasciata/clarkii* Species Complex

Mark Alexander DiMeo
Clemson University, dimeo38@gmail.com

TESTING THE WATERS: DIVERSIFICATION AND SELECTION IN THE *NERODIA FASCIATA/CLARKII* SPECIES COMPLEX

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Biological Sciences

---

by
Mark Alexander DiMeo
December 2019

Accepted by:
Dr. Christopher Parkinson, Committee Chair
Dr. J. Antonio Baeza
Dr. Vincent Richards

ABSTRACT

Understanding evolution is a key component in trying to decipher the processes generating global species diversity. The strength, direction, and interaction of gene flow and selection often determine diversification patterns and the process of speciation. The *Nerodia fasciata/clarkii* species complex, a lineage of water snakes, is thought to have high levels of both gene flow and selective pressures due to ecological constraint. *Nerodia clarkii* resides in salt marsh and estuarine habitats while *Nerodia fasciata* is typically found in fresh water. Salinity is a strong selective pressure and is thought to play a role in the diversification process. Currently, there are five described subspecies within the complex but their validity is in question, causing concerns about the conservation status of the federally threatened Atlantic Salt Marsh Snake (*N. c. taeniata*). To understand the diversification of the *Nerodia fasciata/clarkii* complex and to resolve the noted taxonomic issues, I generated the first population genomic dataset for this group using double digest restriction site-associated DNA sequencing (ddRADseq). I first used a Discriminate Analysis of Principal Components (DAPC) to identify population structure and SVDQuartets to generate a coalescent based phylogeny. With these data, I identified 4-6 populations that approximate subspecies designations although only one assigned subspecies, *N. c. clarkii*, was monophyletic. Second, I estimated migration among the best supported population clustering (k=5) using Estimated Effective Migration Surfaces (EEMS). EEMS revealed a migration corridor between the mostly *N. fasciata* populations and a reduction in gene flow at the coasts. Third, I used two selection scan analyses and one environmental association analysis to identify genes that

are putatively under selection with an emphasis on local adaptation to saline water. I found 10 candidate genes that may be involved in osmoregulation and multiple correlations to temperature and precipitation. My results indicate that the two species are valid, and that four subspecies are also evolutionary lineages. Although gene flow and population assignment tests provided evidence that *N. c. taeniata* was isolated from other populations I could not unambiguously determine its validity. Both the candidate gene frequencies and EEMS indicate that majority *N. fasciata* populations mostly share a selection regime and gene flow patterns separate them from *N. clarkii*. This may demonstrate how a reduction in gene flow and a change in selection pressures can generate species diversity.

To my family,
Even if I'm in Japan
I know you're with me

ACKNOWLEDGEMENTS

Finishing this thesis was a longer and harder process than I ever expected and I could never have done it without help. First, I want to thank my advisor and committee members. Dr. Parkinson, thank you for taking a chance on me when I had next to no research experience and sticking with me as I struggled to become a scientist. I'm sure it wasn't easy to do so. Additionally, without your grant from USFW I'm not sure I could have funded all the work. I want to thank my committee members Dr. Baeza and Dr. Richards for their patience with me for all the delays to my thesis. I hope I was able to live up to all your expectations.

Second, I want to thank all the members of the Parkinson Lab, past and present, who have provided support to me, without which I could never have finished. In the three years I was a part of the lab I have seen multiple lab mates come and go and all of them have played a part in my growth as a scientist and person. Notably Jason Strickland, Andrew Mason, and Rhett Rautsaw have been with me since I joined and without their guidance this wouldn't be possible. In particular I want to thank Mark Margres who really took me under his wing in the last year and Rooksana Noorai who was always there to listen to me despite the fact that she is not an official member of the lab and was not obligated to do so. Then I have to thank the past Master's students, Greg Territo and Jason Hickson, and the undergrads who laid out the foundation for this project through sample collection and data generation. I cannot forget all the contributions to the research and editing of this project that the Parkinson Lab gave me.

Third, during the course of my research I worked at both the University of Central Florida and Clemson University so I want to thank all the friends I made at both institutions. You may not have directly contributed to my thesis but you helped to keep me sane through the hard times with your friendship.

Finally, I want to thank my family. I don't know why but despite all my setbacks and delays your confidence in me never seemed to waver. This emotional support was worth so much to me, even more than the financial support! Despite the fact that I'm not near you now I know I'll always have your support.

TABLE OF CONTENTS

Page

Table of Contents (Continued)

Page

LIST OF TABLES

LIST OF FIGURES

List of Figures (Continued)

Figure Page

List of Figures (Continued)

Figure Page

CHAPTER ONE
INTRODUCTION

Diversification, Gene Flow, and Selection

Evolution has acted through the array of environmental and genetic factors to develop the staggering amount of organismal diversity we see today. The core of biology is investigating and describing the mechanisms that underlie the diversification of populations and species. The process of diversification begins with the differentiation between populations and whose progression towards speciation is determined by many factors. Two of the most important factors affecting population differentiation are gene flow (migration) and local adaptation (selection) (Tigano & Friesen, 2016). The selection-migration interaction can be difficult to assess as every natural population has its own unique set of behavioral, geographic, environmental, and genetic factors (Räsänen & Hendry, 2008). Biologists have pursued a better understanding of these processes for generations and the improvements to genetics provides a wealth of data for assessing the effects of gene flow and selection.

Gene flow has long been shown to both be a force that can promote the spread of beneficial alleles but often acts as a homogenizing agent among populations (Felsenstein, 1976; Lenormand, 2002). The magnitude of gene flow has a considerable effect on the magnitude of homogenization. At high levels of gene flow, alleles will rapidly spread across populations resulting in the same or similar allele frequencies (Lenormand, 2002). Additionally, gene swamping may occur where the influx of alleles from outside a population are more likely than locally adapted alleles to be inherited due to volume (Lenormand, 2002; Sexton, Hangartner, & Hoffmann, 2014). The reduction in variation among populations and swamping of local alleles may prevent diversification or in some cases reduce fitness of certain populations (Sexton et al., 2014). Alternatively, low levels of migration can allow other processes, such as selection and drift, to potentially drive diversification (Engen & Sæther, 2016; Wolf & Ellegren, 2017). Understanding how gene flow is affecting a population/taxon is an important piece of information to estimate if said population is diversifying.

Local adaptation is the response to differential selection pressures among populations and habitats (Rellstab et al., 2015; Williams, 1966). Local adaptation is often seen as being negatively correlated to the homogenizing effects of gene flow. Positive selection acts through some environmental pressure on alleles that affect a relevant phenotype that could cause population differentiation, potentially leading to speciation (Lenormand, 2012). Just as the rate of migration affects the homogenization of allele frequencies, the strength of selection pressures can rapidly change a population and even restrict gene flow as immigrant fitness is reduced in the local environment (Lenormand,

2012). Locally adapted individuals are likely to outcompete their conspecifics in survival and reproductive success allowing the local adaptation to persist in the population (Balkau & Feldman, 1973). At the genetic level this population is likely to undergo a selective sweep where alleles in close linkage to the adaptive loci get passed on regardless of their fitness (Kaplan, Hudson, & Langley, 1989; Maynard Smith & Haigh, 1973). This causes alleles for traits that otherwise would not get past on to remain in the population serving to reinforce the population's differentiation. This can potentially lead to reproductive isolation and speciation if different traits continue to accumulate. The reinforcement likely only leads to this level of differentiation of the selection pressure is strong and that gene flow is not occurring at a high enough rate to swamp out these alleles.

Uncovering how these evolutionary processes interact in natural populations is a challenge. Directly measuring migration (i.e. radio telemetry, paternity analysis) is often prohibitively difficult while trying understand the process of adaptation through phenotypic differences can be confounded by plasticity and trade-offs (James, 1983; Whitlock & McCauley, 1999). Indirectly measuring migration through examining genetic diversity and scanning for genes under selection can help resolve some of the difficulties inherent in the other methods. The quantity of data modern next generation sequencing techniques generate can to help untangle our understanding of taxa where traditional morphological, ecological, and genetic techniques cannot elucidate their evolutionary history. One group where this is of use is the *Nerodia fasciata/clarkii* species complex found in Florida.

Population Genomics

I will use population genomic data to understand patterns of gene flow, local adaptation, and population differentiation in the *Nerodia fasciata/clarkii* complex. With the advent of Next Generation Sequencing (NGS), improved computing power, and new analytical methods, more accurate inferences can now be made from larger, genomic level datasets. Many programs have been developed to process these data that incorporate Bayesian inference, maximum likelihood statistics, and coalescent theory to model the historical and current population dynamics of a species (Kingman, 2000). Without a historical context it becomes harder to make inferences about speciation or the necessity of conservation. The improvement in genetic data generation and the ability to incorporate these data into better models means estimating population genomics has become increasingly important for academic studies of evolution and conservation.

Population genomic data can also be used for the identification of loci under selection (Dupuis et al., 2017; Savolainen, Lascoux, & Merilä, 2013), potentially those affecting salt tolerance in this species complex. The main method for finding these genes is the identification of outlier loci that are more differentiated between populations than the

average locus. The excessive differentiation is unlikely to be explained by genetic drift alone, indicating that natural selection may be the explanatory factor causing the outlier loci. Unlike in the past where we had few genes or microsatellites, we can now look at thousands of loci from across the genome to identify outliers (Lexer et al., 2013). Improved, cost-efficient techniques make it possible to collect data from non-model wild populations and obtain an understanding of what genes are undergoing selection. We are able to not only identify these outlier loci and the unique populations where they belong but we can sometimes identify the cause for these outliers. The environment acts as a driver of selection and finding gene correlated to environmental factors affects our understanding of taxa. Water salinity, vegetation, and climate are examples of environmental sources of selection that have been known to lead to diversification (Carstens & Knowles, 2007; Fuller, McGhee, & Schrader, 2007; Losos, Warheitt, & Schoener, 1997). Environmental association analyses correlate outlier loci to environmental factors (Harrisson et al., 2014) in an attempt to identify environmental selection pressures. This abundance of information available through population genomics provides us with valuable insight into the genetic and environmental drivers of evolution.

Study Species

The *Nerodia fasciata/clarkii* species complex provides an opportunity to examine diversification and is in need of taxonomic assessment. Within Florida, the complex is exposed to many different environments, biogeographic barriers, and human interference, which provides many opportunities for populations to diverge. However, this complex lacks a thorough investigation into their evolutionary history. *Nerodia fasciata* and *N. clarkii* have a complicated taxonomic history. Currently this complex consists of two sister species with six named subspecies (Clay, 1938; Conant, 1963; Cope, 1895). The Salt Marsh Snake (*Nerodia clarkii*) and the Southern Water Snake (*N. fasciata*) have been historically separated ecologically by a preference for saline habitats (e.g brackish or saltwater estuaries and saltmarshes) in *N. clarkii* and accompanying behavioral and/or physiological adaptations to this novel environment (Dunson, 1980; H.I. Kochman, 1992; Pettus, 1963). *Nerodia clarkii* have been found in salinity levels as high as 73.5 ppt (Territo, 2013). In comparison, oceanic water is only approximately 30 ppt. Although this is an extreme example for this species, it demonstrates their ability to tolerate high salinity environments (Pettus, 1958). However, the ecological differentiation between these species has come under question recently as a comparative study on the physiology and morphology associated with water retention did not find significant differences between the two species' kidneys, cloaca, or colons (Babonis, Miller, & Evans, 2011; Babonis, Womack, & Evans, 2012). Yet, there may still be an unexamined physiological trait or genetic variation in relation to salt tolerance that has yet to be uncovered. My study will take advantage of

modern NGS data unavailable to previous genetic investigations into the *Nerodia fasciata/clarkii* complex to search for adaptive genetic variation in these taxa. Salt tolerance is an interesting topic for investigation due to the rarity of reptiles adapting to saline environments. This trait occurs in approximately 100 out of approximately 10,000 reptilian species in total (Rasmussen et al., 2011; Uetz, 2010). The 100 are not spread evenly among reptile groups as sea snakes comprise about 80 species and the remainder consists of sea turtles, brackish water snakes, the salt water crocodile, and the marine iguana (Rasmussen et al., 2011; Uetz, 2010). The bias towards snakes in the emergence of the otherwise rare trait makes *Nerodia* a interesting candidate for an investigation into the genetics of salt tolerance.

Nerodia clarkii and *N. fasciata* have undergone frequent taxonomic revision. Both species were once considered subspecies of the Northern Water Snake (*N. sipedon*), but later *N. fasciata* was elevated with *N. clarkii* becoming a subspecies of *N. fasciata* followed by the elevation of *N. clarkii* in 1991 (Baird & Girard, 1853; Carr Jr & Goin, 1942; Kennicott, 1860; Kochman, 1977; Lawson et al., 1991). Most of these delimitations were based on scale counts and color patterns, which were later shown to be unreliable. For example, the noted dorsal scale counts of 19–21 for *N. clarkii* vs 23–25 for *N. fasciata* (Clay, 1938) were later found to overlap, as the temperature during pre-natal development affects the development of these characters (Osgood, 1978a). An allozyme study (Lawson, Meier, Frank, & Moler, 1991) was thought to corroborate the species designations, but a reanalysis of those data did not find statistical support for the original results, instead finding no difference at the species level among allozymes (Jansen, 2001). My study will focus on the five subspecies found in Florida: *N. c. clarkii, N. c. compressicauda, N. c. taeniata, N. f. pictiventris,* and *N. f. fasciata* (Fig 1.1). The subspecies are largely described based on color and geographic location (Dunson, 1979; Territo, 2013). However, Jansen (Jansen, Mushinsky, & Karl, 2008) found significant genetic differentiation in microsatellites from *N. c. compressicauda* over small spatial scales (10s km) that is expected over large scales (100s km). The conflicting evidence clouding the current taxonomy is of particular concern given the conservation status of *N. c. taeniata,* the Atlantic Salt Marsh Snake. *Nerodia c. taeniata* is federally listed as a threatened subspecies due to habitat destruction and potential genetic introgression with *N. f. pictiventris* (Brooks, 2008; U.S.F.W Service, 1977, 1993). *Nerodia c. taeniata* was delineated by its two dorsal stripes running down most of the length of the body (Carr Jr & Goin, 1942), yet later studies found no such consistent characteristic and large amounts of overlap in coloration and scale counts with *N. c. compressicauda* (Dunson, 1979; Hebrard & Lee, 1981). The USFWS 5-year Review in 2008 came to the conclusion that there was a lack of genetic information on *N. c. taeniata* and that there is a need to verify its sub specific and conservation status.

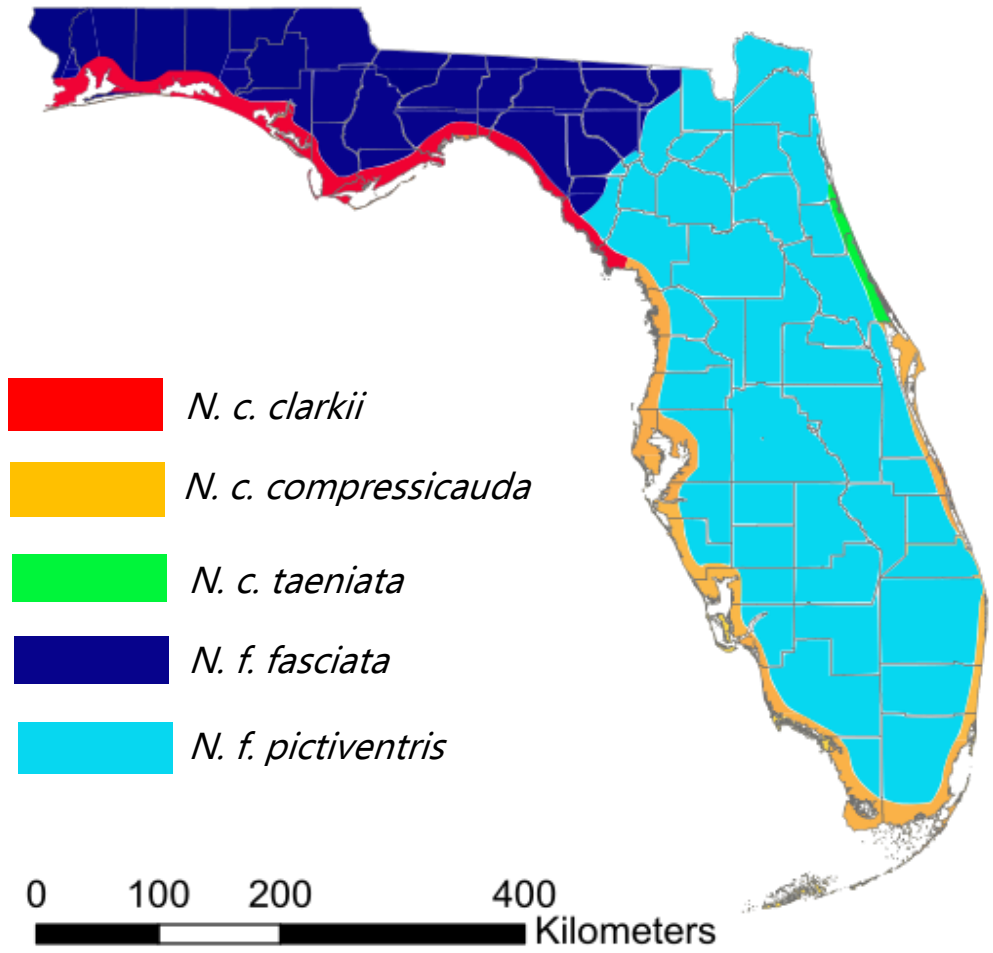

Figure 1.1: Color coded range map for the five subspecies of *N. fasciata* and *N. clarkii* found in Florida.

Objective

In this study, my objective was to infer the evolutionary relationships and diversification patterns in the *Nerodia fasciata/clarkii* species complex. This would allow me to evaluate the validity of the current taxonomy. To obtain this objective I generated the largest genetic dataset of this complex to date. First, I sequenced individual mitochondrial and nuclear genes. Second, I generated a reduced representation library using double digest restriction site-associated DNA sequencing (ddRADseq). Using these data, I estimated the complex`s population structure, phylogeny, and gene flow pattern then

searched for loci putatively under selection. These results allowed me to better evaluate the validity of the current taxonomy within the *Nerodia fasciata/clarkii* species complex

Questions and Predictions

Chapter 2 Question: What is the population structure and gene flow pattern in the *Nerodia fasciata/clarkii* species complex?

1) If the taxonomy is valid then there will be at least five populations correlating to their distribution ranges.
2) Given the difficulty in differentiating subspecies there will be fewer than five populations with few, if any, barriers to gene flow among them.

Chapter 3 Question: Are there different selection regimes between populations or species?

1) If water salinity is acting as a selective pressure then genes associated with osmoregulation will have different frequencies between the coast and inland populations.
2) If water salinity is not acting as a selective pressure then there with either be no osmoregulatory genes found or the allele frequencies will not differ between the inland and coastal populations

CHAPTER TWO

EVOLUTIONARY HISTORY AND POPULATION GENETICS OF THE *NERODIA FASCIATA/CLARKII* SPECIES COMPLEX

Introduction

Gene flow is one of the most important mechanisms for evolutionary change. Gene flow can maintain genetic diversity within populations by avoiding inbreeding depression through the migration of alleles from other populations (Ingvarsson, 2001). There are many factors that influence the direction and strength of gene flow including a taxa's mobility, geographic and other environmental barriers, behavioral differences, and reproductive isolating mechanisms. For example, marine mammals can swim long distances with little resistance from geography allowing them to maintain large population sizes (Quérouil et al., 2007). Geographic barriers can play a much more major role in the diversification of terrestrial species. The Suwannee River serves as one of these barriers for dozens of species which is a contributing factor to the differences between panhandle and peninsular Florida (Bert, 1986). The peninsula itself is a common dividing line between the Atlantic and Gulf coasts that even affects marine organisms (Soltis, Morris, McLachlan, Manos, & Soltis, 2006). These biogeographic barriers have profound effects on the population structure of many species where the restriction of gene flow has allowed populations to diverge. The *Nerodia fasciata/clarkii* complex's range overlaps these Floridian biogeographic barriers along with other environmental barriers such as changing water salinities. These barriers may have factored into the diversification of the complex into the currently recognized five subspecies in Florida, however, there have been few studies that have examined the population structure and gene flow patterns in these taxa.

There has been much debate focused on the validity of the five subspecies in Florida (Baird & Girard, 1853; Carr Jr & Goin, 1942; Jansen, 2001; Kennicott, 1860; Howard Irwin Kochman, 1977; Lawson, Meier, Frank, Moler, et al., 1991) but there has been little genetic investigation into their population structure. The separation of the subspecies lies largely with differences in morphology like color pattern. The *N. clarkii* subspecies' ranges comprise the edges of Florida in coastal wetlands and salt marshes whereas *N.f. fasciata* and *N.f. pictiventris* inhabit the inland of the panhandle and peninsula, respectively (Fig 1.1). Assuming these ranges are the equivalent of populations would be inadvisable considering the problems arising with their morphological descriptions (R. Rautsaw, personal communication) and there might be addition structure within a

subspecies. Dorsal scale row counts are a diagnostic characteristic where *N. clarkii* had 19-21 and *N. fasciata* had 23-25 yet there is evidence that these ranges overlap, possibly due to gene flow and/or temperature during pre-natal development (Osgood, 1978b). The five subspecies descriptions are based mostly on color pattern. Similar to the scale counts, the colors appear to not be discreet but a continuum, such as the overlap in dorsal stripe patters between *N.c compressicauda* and *N.c. taeniata* (Dunson, 1979; Hebrard & Lee, 1981). Morphological data can be hard to interpret, as demonstrated in this complex, and can disagree with genetic data when it becomes available for a taxon. For example, in the land snail *Pyramidula* complex, shell morphology was found to be unable to differentiate between the nine species uncovered using restriction site-associated DNA sequencing (Razkin et al., 2016).

With reliable population assignment, I will be able to characterize signatures of gene flow in the complex for the first time. Without knowledge of migration patterns in these snakes, we are missing an important piece of information about their evolution. The presence of high levels of gene flow is of particular importance to this group because the federally threatened Atlantic Salt Marsh Snake (*N.c. taeniata*) may be facing introgression from *N.f. pictiventris*, posing a risk to the supposed genetic uniqueness of the subspecies. Direct measures of migration (which may not equate to gene flow) by tracking the movement of the animals are often time consuming, expensive, and difficult. What little information we have comes from *Nerodia sipedon*, the sister taxa to the complex, where the mean movement/day was less than 50 meters and was closely associated with aquatic vegetation (Roth & Greene, 2006). Migration data only provides limited inferences because it does not truly show movement between populations, let alone the ability of the migrants influence the allele frequencies of their new population.

To identify population structure and gene flow patterns, I will be utilizing a phylogenetic approach to infer both. Since *N. clarkii* was elevated to species status based on allozymes (Lawson et al., 1991), there has been little genetic work done in the complex since. The allozyme study was later replicated and differences between species was found to be not statistically significant (Jansen, 2001). A phylogeny using the mitochondrial loci *cytb* found little support for more recent nodes but did support a panhandle/peninsula split further in their evolutionary history (Territo, 2013). At the population level *N.c. compressicauda* was examined using microsatellites and found differentiation at levels seen in taxa separated by 100s, not 50, kilometers (Jansen et al., 2008). The previous studies indicate that the genetics of *Nerodia* are still in need of assessment with data that can provide more power than before. Next generation sequencing generates thousands of loci allowing for a resolution unattainable in previous studies. The increased size of the data set will give us an unprecedented insight into the population genetics of the *Nerodia fasciata/clarkii* species complex. With these data we able to more confidently define

populations, identify how alleles are moving between them, and uncover any isolated populations. These data will be an important source of information for resolving the subspecies question, especially for the listed *N.c. taeniata*.

Materials and Methods

Sampling

Snakes from the *Nerodia fasciata/clarkii* complex (Fig 1.1) were captured or sampled between the years of 2012–2015 via road cruising and targeted capture in suitable habitat. All capture locations were marked with a GPS point. Tissue samples were collected via tail clippings or blood draws for captured specimens of *N. c. taeniata*. All other specimens were returned to the lab and euthanized following standard protocols (Conroy et al., 2009). Tissue (blood, liver, muscle) was removed during dissection and stored in 95% ethanol at -20°C. Subspecific designation was based on location within their geographic ranges and recognized morphological characteristics from the literature (Carr Jr & Goin, 1942; Cope, 1895; Territo, 2013).

Sixty-two *Nerodia fasciata/clarkii* individuals were chosen for Sanger sequencing of mitochondrial and nuclear genes, as well as 5 outgroup taxa (Fig 2.1). Samples were selected to encompass the complex's geographic range in Florida, from the farthest west point in the panhandle to the Florida Keys. There was dense sampling in the area where the panhandle and peninsula meet because previous work indicates this area is where the two major populations meet (Territo, 2013). The boarder between these populations may exhibit traces of gene flow. We also chose samples to transect across the central region of the peninsula. No fewer than 6 samples were used for a subspecies.

One hundred and thirty-nine individuals (Fig 2.2) were selected for double digest restriction-site associated DNA sequencing (ddRADseq). Based on coloration and sampling location we chose 26 *N.c. taeniata*, 19 *N.c. clarkii*, 26 *N.c. compressicauda*, 14 *N.f. fasciata*, and 45 *N.f. pictiventris*, along with 6 *N. clarkii* and 1 *N. fasciata* individuals considered intergrades, and one *Nerodia* that could not be distinguished between the species to best represent the distribution of the complex in Florida. Three subspecies' distributions meet in the peninsula potentially creating complex gene flow patterns. An even sampling of the inland and coastal ranges in the peninsula was selected to capture the potential patterns. In the panhandle samples were selected to reach the farthest point west in Florida and to more densely sample where *N.f. fasciata* and *N.c. clarkii* meet. Additionally, *Thamnophis sauritus* was also included as an outgroup.

Mitochondrial and Nuclear Loci Sequencing

DNA was extracted from the 62 *Nerodia* plus five outgroup taxa using Serapure beads following the procedure from Faircloth and Glenn (2014). To create a robust phylogeny, we sequenced 7 total loci, three mitochondrial and four nuclear. Our chosen mitochondrial genes include cytochrome b (*cytb*: Burbrink, Lawson, & Slowinski, 2000), and NADH dehydrogenase subunits 1 (ND1: Jiang et al., 2007) and 4 (ND4: Arevalo, Davis, & Sites Jr, 1994). Four total nuclear genes were amplified, PRLR (Townsend, Alegre, Kelley, Wiens, & Reeder, 2008), anonymous loci M and E (Mcvay, Flores-Villela, & Carstens, 2015), and TATA (Wood, Vandergast, Lemos Espinal, Fisher, & Holycross, 2011). All amplifications were done using polymerase chain reaction (PCR) using the primers from referenced studies with the following conditions: initial denaturing at 94˚C for 3.5 minutes, 35 cycles of denaturation at 94˚C for 30s, annealing (Table 2.1) for 30s, extension at 72˚C for 1 minute, and a final extension at 72˚C for 15 minutes. All sequences were generated using Sanger sequencing by the University of Arizona Genetics Core and Eurofins. *Cytb*, ND1, ND4, and TATA were aligned using MUSCLE in MEGA6 (Tamura et al. 2004) while PRLR, M, and E were done in Geneious v10.1.3 (Biomatters), likewise utilizing MUSCLE. Each sequence was visually inspected. Heterozygous individuals were identified by eye and by the heterozygous plugin for Geneious.

Table 2.1: Locus name, annealing temperature (°C) and alignment length used for Bayesian phylogenetic inference of the *Nerodia fasciata/clarkii* species complex.

Locus	Annealing Temp (˚C)	Alignment Length
cytb	48	1080
ND1	53	950
ND4	53	668
TATA	53	715
PRLR	55	595
M	52	344
E	52	260

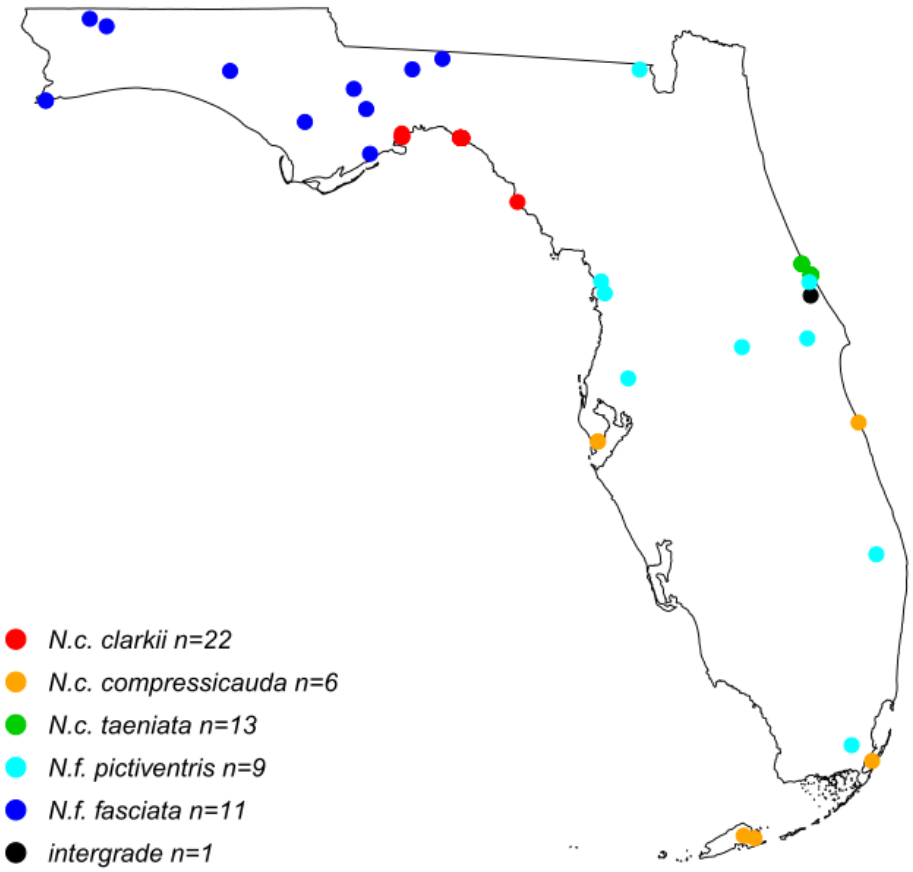

Figure 2.1: Map of sampling location for each individual that mitochondrial and nuclear loci were sequenced. Intergrades were considered any individual who could not be identified at the subspecific level using meristic characters. These samples were used for Bayesian phylogenetic reconstruction in BEAST v2.4.2.

Library Preparation and Sequencing

DNA was extracted from blood, liver, or scale tissue using a standard phenol-chloroform extraction. A Qubit Fluorometer 3 (Thermo Fisher Scientific, Waltham, MA, USA) was used to quantify extractions which required a minimum of 500ng of DNA for ddRADseq with 1000ng preferred. The ddRADseq protocol follows Peterson et al. (2012) with some modifications. Briefly, DNA was digested for a minimum of 8 hours at 37°C using the rare cutting restriction enzyme SbfI-HF (8 bp) and the common cutting Sau3AI (4 bp). The samples were cleaned using AMPure magnetic beads (Beckman Coulter, Inc., Irving, TX, USA) at a 1.1:1 bead-to-DNA ratio and eluted in 42uL of tris-HCl ph8.0 for quantification. Based on DNA concentration they were assigned into groups of 8 samples.

Each group was assigned a standardized amount of DNA, based on the lowest DNA concentration per group, for ligation. One of 16 double-stranded DNA adaptors, or inline barcodes, was ligated to the end of the digested DNA fragments from each individual. These barcodes are necessary for sample identification after sequencing. Barcodes were ligated at 16^{o}C for 5 hours. After ligation, each sample was cleaned using AMPure beads at a 1.1:1 bead-to-DNA ratio and DNA quantified using the Qubit Fluorometer 3. Samples were rearranged. if necessary, into new groups to minimize differences in DNA concentration. The groups were pooled and cleaned with AMPure magnetic beads and eluted with 30uL of TE ph8.0. A Blue Pippin Prep (Sage Science) with a 1.5% agarose gel to size select fragments from 300–700 bp in length. This size-range was selected to recover the maximum number of potential loci for downstream analyses (Schield et al., 2015). The DNA fragments in each group were amplified using PCR following Shield et al. (2015). Each group is assigned a unique primer that adds a second index barcode to each sample during the PCR. This results in 128 unique barcode-index combinations. Samples were cleaned twice at a 0.7:1 bead ratio and quantified on the Qubit. Samples were further quantification using a high-sensitivity DNA chip for the BioAnalyzer 2100 (Agilent Genomics) to confirm the appropriately sized fragments were amplified. Sequencing was performed on an Illumina Nextseq 550 using 150 bp paired-end reads.

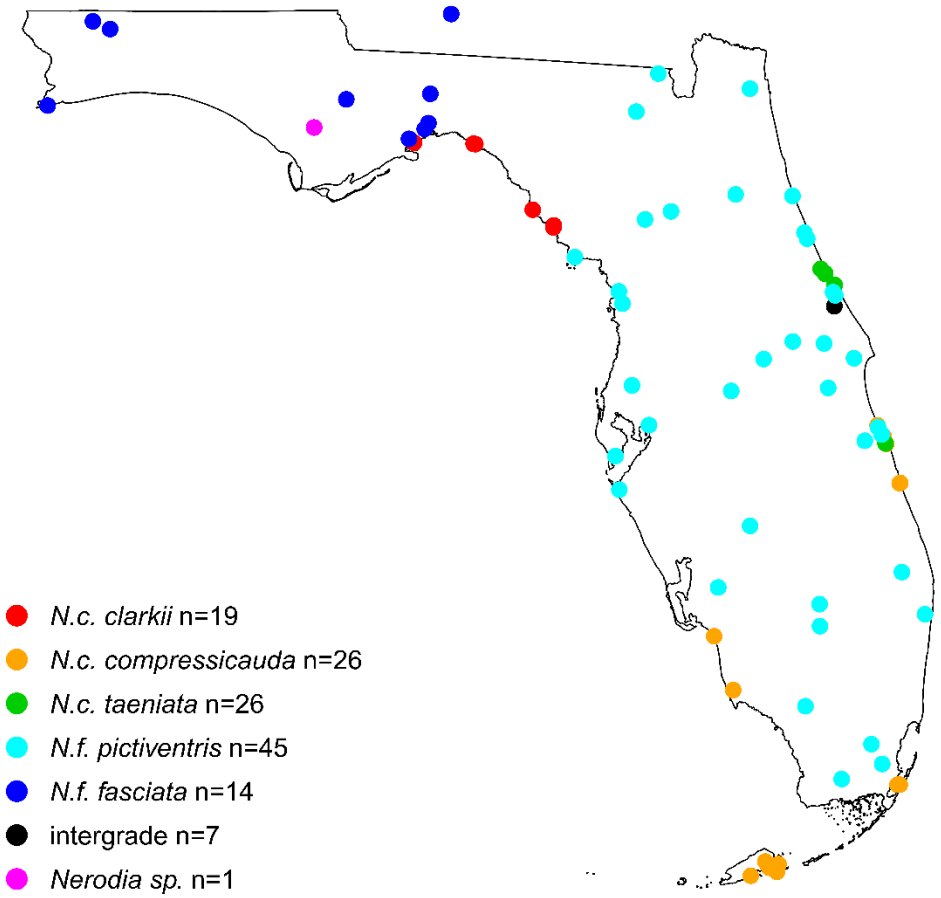

Figure 2.2: Map of all samples for which a ddRADseq library was prepared. Intergrades were considered any sample who could not be identified at the subspecific level using meristic characters. On sample is called *Nerodia sp*. because it was either *Nerodia fasciata* or *clarkii* but we were unable to determine which based on morphology.

Variant Calling

The ipyRAD v0.7.28 (Eaton, 2014) toolkit was used to demultiplex and call SNPs from the raw Illumina reads. Each library had the first 8 base pairs trimmed using Trimmomatic v0.38 (Bolger, Lohse, & Usadel, 2014) prior to demultiplexing. Each library was demultiplexed individually according to the index barcode and then merged for de novo assembly. Reads were filtered if their average phred score offset was less than 33 and if they had 5 or more low quality base calls per read. These were clustered at 90%. This threshold was chosen due to the close relationship of my samples and a pilot run using MiSeq sequences using 85, 90, and 95 percent threshold. Additionally, I trimmed bases off

reads as long as the Qscore was less than 20 but set a minimum length to be 100 bp. All other parameters were left at default settings.

Vcftools v0.1.15 (Danecek et al., 2011) was used to further filter SNPs. Before filtering, individuals were removed if they had fewer than 1000 loci. Clustering analyses showed evidence of sampling bias so individuals were haphazardly removed until there were no more than 7 individuals within a 10-mile radius. Vcftools SNP filtering parameters included removing any non-biallelic SNPs, SNPs that had more than 50% missing data, a mean depth of less than 10x, and minor allele frequency of less than 5%. The –depth flag was used to find the mean depth of each individual across all their reads and individuals below 5x were removed. These steps were repeated but without the outgroup CLP1282 *Thamnophis sauritus* to make separate datasets for phylogenetics and population genetics.

Phylogenetic Reconstruction with BEAST

Bayesian phylogenetic analyses were carried out in BEAST v2.4.2 (Vaughan et al., 2014). Sequences were concatenated and genes were partitioned by codon position except for the introns TATA, M, and E. Models of nucleotide substitution (Table 2.2) were assigned by PartitionFinder v. 2.1.1 (Lanfear, Calcott, Ho, & Guindon, 2012). Each gene was concatenated and the Markov chain Monte Charlo (MCMC) analysis was run for 100 million generations and sampled every 10,000th generation with 10% burnin. We chose a strict clock model for all loci and a Yule-speciation prior. Priors for all other model parameters were left at default. Stationarity was reached if ESS was >200 as shown in TRACER v1.6.

Table 2.2: List of mitochondrial and nuclear genes with the model for each coding position. TATA, M, and E are introns and were not split by codon thus they only have one position.

Gene	position 1	position 2	position 3
cytb	HKY+G	TN93+I+G	HKY+G
ND1	HKY+I+G	TN93+I+G	GTR+G
ND4	HKY+G	TN93+I+G	HKY+G
TATA	HKY+I+G	-	-
PRLR	HKY+I+G	HKY+I+G	HKY+I+G
M	TN93+I+G	-	-
E	HKY+I+G	-	-

Population Structure

Population structure was delimited using Discriminant Analysis of Principal components (DAPC) in adegenet v2.1.0 (Jombart, 2008). DAPC is a machine learning algorithm that maximizes differences between clusters while minimizing within cluster variation using SNP data. The VCF file for population analysis was read into Rstudio v1.1.383 (Rs. Team, 2015) using vcfR (Knaus & Grünwald, 2017) and converted to a genlight format for adegenet. Clusters were assigned using a *k*-means clustering algorithm that determines the best number of clusters based on Bayesian Information Criterion (BIC) where the lowest BIC has the most support. Given that the BIC scores for K=4, 5, and 6 were nearly identical all were used (Fig S1, Figs 5-7). These clusters were given to the DAPC algorithm and the number of principal components retained was chosen by cross validation using the xvalDAPC function. I used vcftools to make separate VCF files for the three most supported DAPC cluster schemes and removed any SNPs that were missing as a result. These were used to check for any additional structure though hierarchal clustering and to generate population genetic statistics. First, Weir and Cockerham's (1984) pairwise F_{st} between population clusters was calculated using vcftools. For comparison, we also calculated two other measures of differentiation: G_{st} and Jost's D. Nei (1973) created the statistic G_{st} as an alternate to F_{st} to better handle multiple alleles however it has been criticized because is often incapable of reaching a value of 1 even when alleles are fixed. Jost (2008) created *D* which uses the fraction of allelic variation among populations to measure differentiation instead of heterozygosity estimates which are used by F_{st} and G_{st}. *D* is more sensitive to high mutation rates though. G_{st} and *D* were both calculated using

vcfR. Hierfstat (Goudet, 2005) was used to calculate H_o, H_e, and the inbreeding coefficient F_{is}.

Phylogenetic estimation using SVDQuartets

An unrooted phylogeny was estimated using SVDquartets (Chifman & Kubatko, 2014) implemented in PAUP* (Swofford, 2003). This coalescent method calculates site pattern probabilities for quartets of taxa and creates a species tree based on the most well supported quartets. SVDQuartets has been shown to be resistant to incomplete lineage sorting and does not necessitate independence among sites. The VCF file including the outgroup *Thamnophis sauritus* was converted into nexus format and loaded into PAUP*. SVDQuartets was run using all possible quartets, of which 500,000 were generated, and 500 nonparametric bootstrap replicates to assess topological support. Each quartet was estimated under the multispecies coalescent approach (expecting matrix-rank 10), and assembled using the QFM algorithm. SumTrees in DendroPy (Sukumaran & Holder, 2010) generated a consensus phylogeny using flags --force-unrooted, --min-clad-freq -0.25, and –set-outgroup CLP1282 *T. sirtalis*. The phylogeny was visualized in FigTree v1.4.3 (Rambaut, 2012).

Estimating Effective Migration Surfaces

To understand how migration is potentially occurring in the Florida *Nerodia fasciata/clarkii* species complex the effective migration surface was estimated in the program EEMS (Petkova, Novembre, & Stephens, 2016). EEMS uses an effective migration model to portray the relationship between population structure as defined by genetics and geography. In essence, geographic areas where genetic similarity decays quickly are described as having a low effective migration rate. The program simulates continuous population structure by imposing a dense regular grid of demes across the landscape and expected genetic dissimilarities are projected across the grid. I used PGDSpider v2.1.1.5 (Lischer & Excoffier, 2011) to convert from VCF into structure format. Then using the R code provided by EEMS, str2diff, a matrix of average pairwise differences was generated. EEMS requires a species range coordinates so an outline of Florida and southern Georgia up to the latitude of my northernmost sample was generated in QGIS 2.0.1 - Dufour (Q. D. Team, 2016). As suggested by Petkova *et al*. (2016) I ran two chains of 1000 demes and 750 demes each and the final migration surface was averaged across all runs (Petkova et al., 2016). Each run was done for 10 million generations with a burn-in of 1 million and was sampled every 10,000 generations. This was visualized in R using rEEMSplots which produces a contour plot of the effective migration and diversity rates.

Results

BEAST Phylogeny

We successfully amplified *cytb*, ND1, and ND4 for 67 individuals (62 + 5 outgroups). TATA, PRLR, M, and E were amplified for 61, 60, 64, and 62 individuals, respectively (Table S2). In total, all genes failed to amplify in fewer than 5% of individuals. The total alignment length was 4612 bp, individual gene lengths can be found in Table 2.1. The best partitioning scheme and models of evolution are found in Table 2.2.

The Bayesian phylogeny generated in BEAST showed strong support at deeper nodes but was unable to resolve recent evolutionary history (Fig 2.3). There are two well supported (posterior probability=1) major clades that roughly correlate with the panhandle and peninsula. The panhandle clade contains all *N.f. fasciata* but is rendered polyphyletic by *N.c. clarkii*. While *N.c. clarkii* are geographically from northern Florida they group with *N.c. compressicauda* from southern Florida. Within the peninsular clade appears two more well supported clades (PP=1) thought there is not a clear break geographically. One of these clades contains another decently supported clade (PP=0.82) but below this there is little support for the placement of other clades. Small clades of 2-5 individuals often show support over 0.7 but where they are in relation to each other is unsupported. The phylogeny shows little evidence for the validity of any subspecies. None form monophyletic clades and *N.c. clarkii* even spans both the peninsular and panhandle clades. *N.c. compressicauda*, *N.c. taeniata*, and *N.f. pictiventris* are constrained to the peninsula while *N.f. fasciata* inhabits the panhandle. Finally, ASMS is notably not genetically unique.

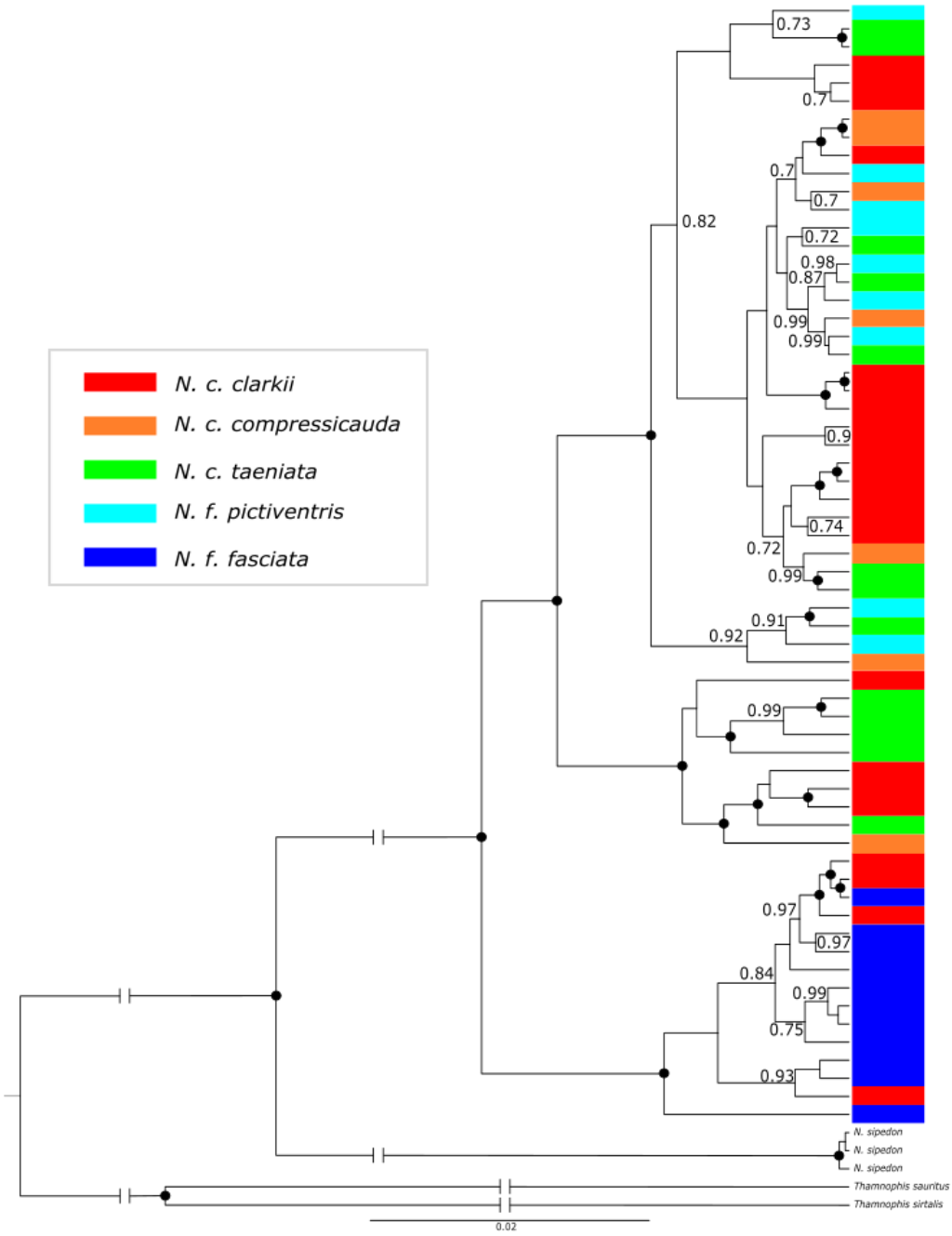

Figure 2.3: Seven gene (Cytb, ND1, ND4, TATA, PRLR, M, E) concatenated Bayesian phylogeny of 62 *Nerodia fasciata/clarkii* samples and 5 outgroup taxa. Posterior probabilities below 0.70 were removed. Black dots represent a posterior probability of 1.

Population Genomics

The ddRAD assembly resulted in 37,804 loci and 104,488 parsimoniously informative sites. A total of 103 samples plus the outgroup *Thamnophis sirtalis* (CLP1282) were kept after filtering for number of loci, depth, and geographic oversampling. After filtering 7,765 SNPs were used for SVDQuartets while 7,336 SNPs passed filter to be used for the population genetics and migration.

The K-means clustering found five population clusters to be most well supported with the lowest BIC score, though the BIC for K of four and six were less than one point higher thus a DAPC was performed for each (Figs 5-7). Hierarchal clustering found two clusters within cluster 1 and cluster 5 though there was no evident pattern to the subpopulations (Fig S2). The 5 clusters fell roughly into the ranges of the 5 subspecies however the individuals' subspecies did not always match the range they were in. *Nerodia c. compressicauda* were mostly concentrated in cluster 5, 4 individuals were found in cluster 1 and one was found in the otherwise *N.f. pictiventris* homogenous cluster 3. Cluster 1 contained all ASMS as well as all individuals labeled as intergrades. It also included one *N.f. pictiventris* that had high levels of admixture with cluster 3. All *N.f. fasciata* were found in cluster 2, one of which (CLP980) had high levels of admixture with cluster 3 and switched clusters at K=4. Cluster 4 was the only one that had a single subspecies, *N.c. clarkii*. *Nerodia f. pictiventris* showed potential introgression with the other subspecies as four of the clusters had at least one *N.f. pictiventris*, the alternative being that the subspecies designations were incorrect. K=4 had the second lowest BIC score and a DAPC with 4 clusters synched the entire western coast of Florida (Fig 2.5). Except for four *N.f. pictiventris* in the mid-west coast this cluster only contained *N.c. clarkii* and *N.c. compressicauda*. The new clade found at K=6 evenly split the west coast into three sections (Fig 2.6) where cluster 4 became *N.c clarkii* only again, cluster 5 was solely *compressicauda*, and cluster 6 contained a mixture of *N.f. pictiventris* and *N.c. compressicauda* on the mid-west coast. What must be considered is that due to the notable difficulty in identifying the subspecies they may have been misidentified. For example, if the *N.f. pictiventris* in population 2 were misidentified then *N.f. fasciata* could be the only members of that population.

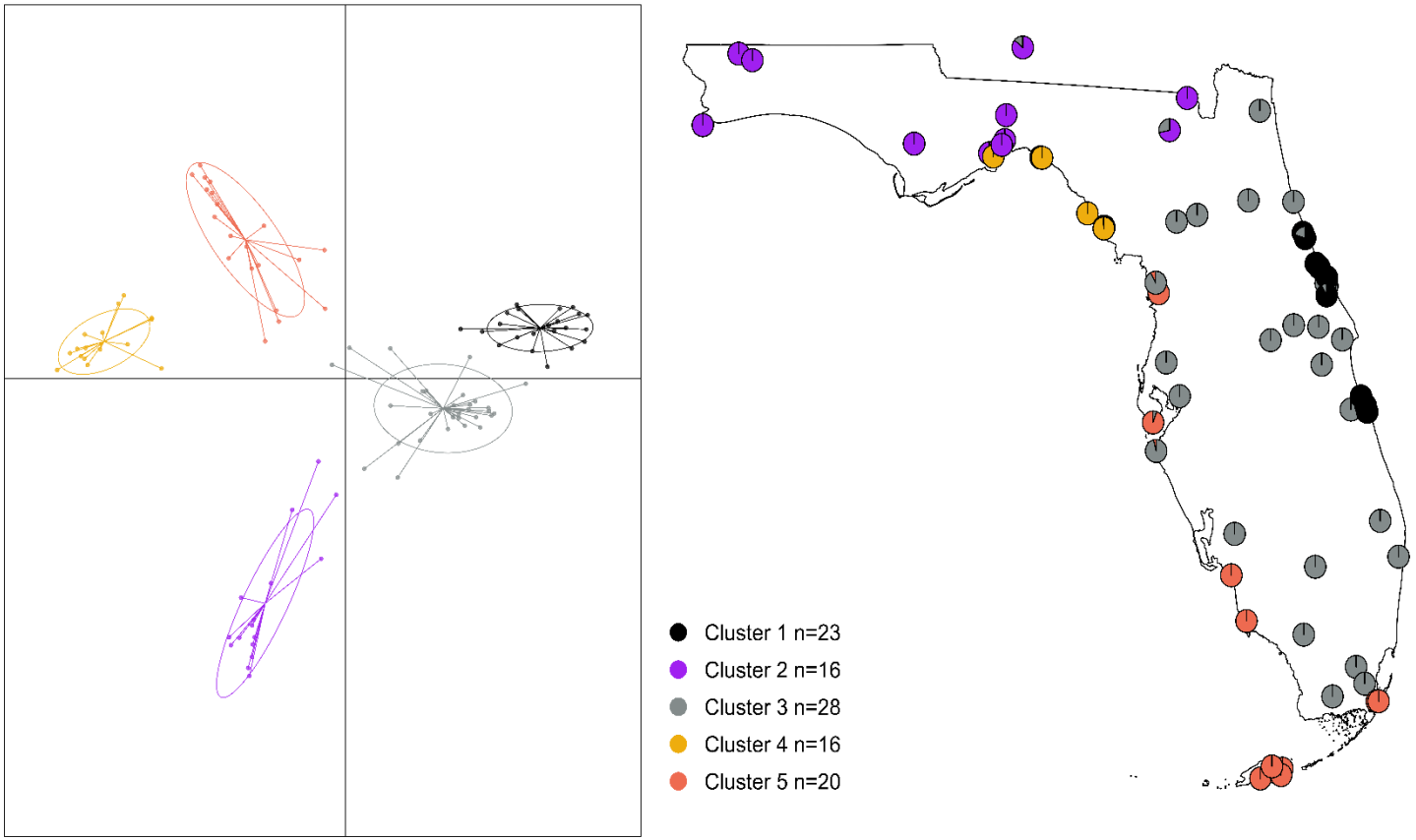

Figure 2.4: DAPC scatter plot (left) of 5 clusters differentiated at 3 PCs (BIC=656.6027). Geographic location of each sample (right) represented by pie charts. Each chart represents the probability of assignment to each cluster, admixture likely occurring in any chart with more than one color.

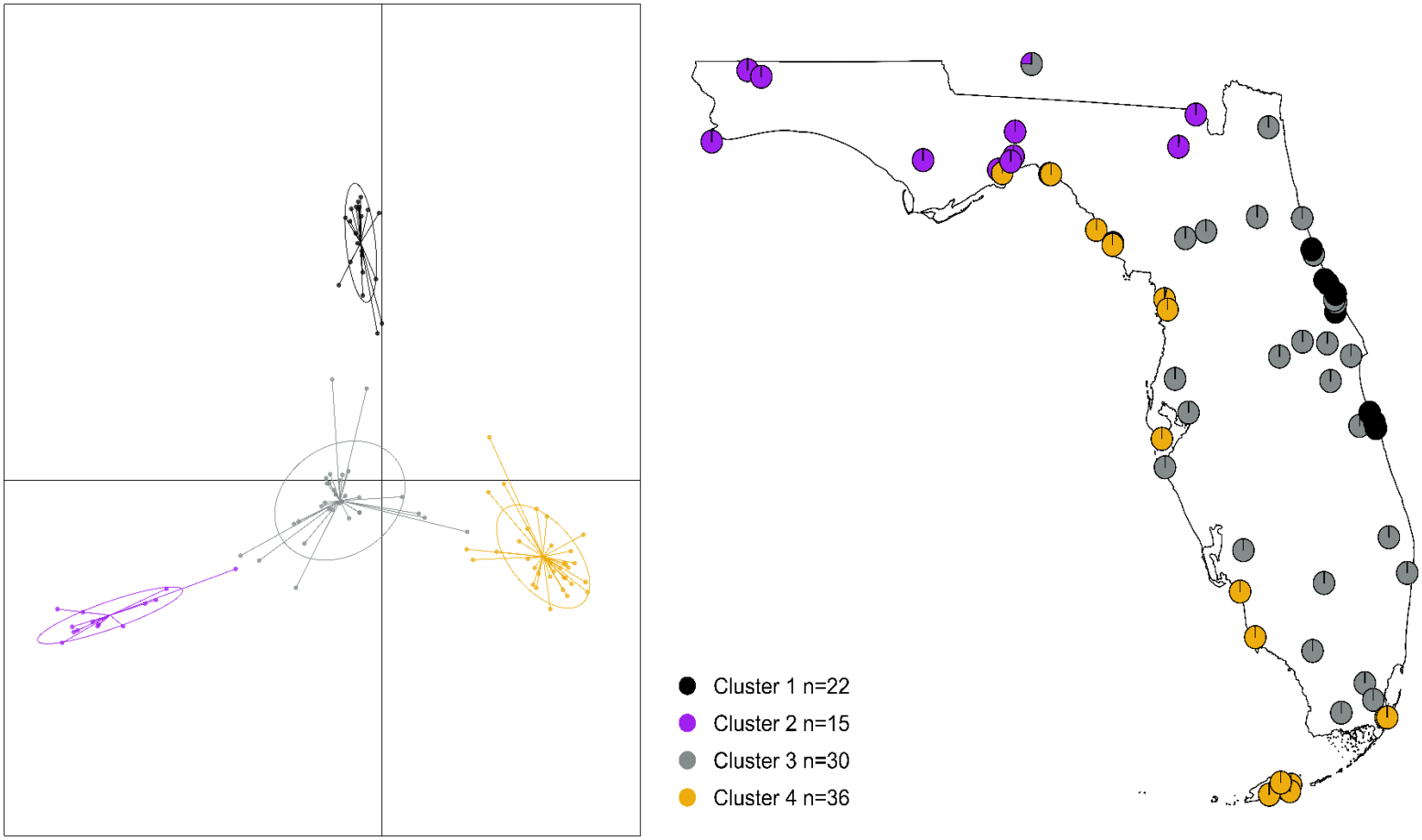

Figure 2.5: DAPC scatter plot (left) of 4 clusters differentiated at 11 PCs (BIC=657.021). Geographic location of each sample (right) represented by pie charts. Each chart represents the probability of assignment to each cluster, admixture likely occurring in any chart with more than one color.

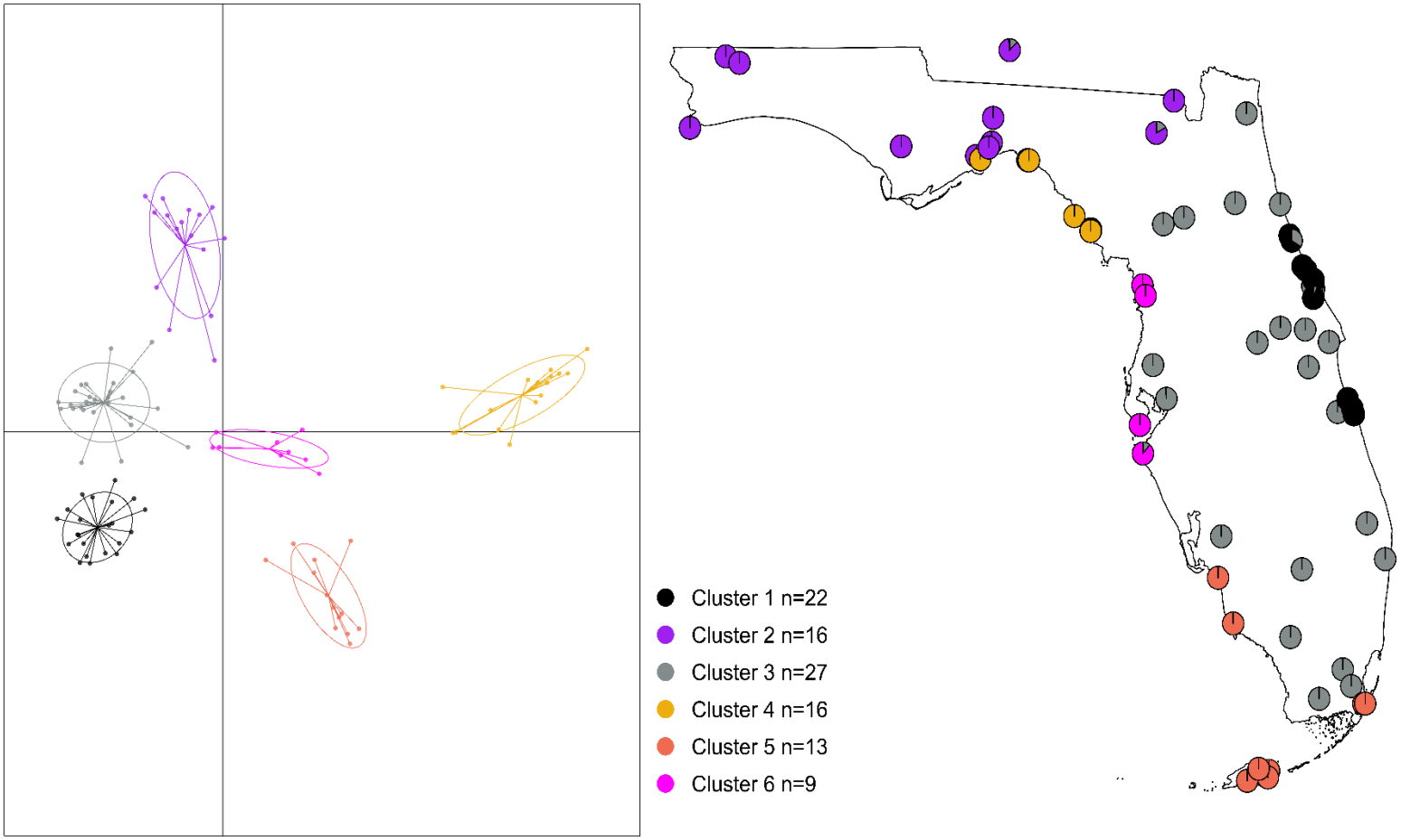

Figure 2.6: DAPC scatter plot (left) of 6 clusters differentiated at 5 PCs (BIC=657.5342). Geographic location of each sample (right) represented by pie charts. Each chart represents the probability of assignment to each cluster, admixture likely occurring in any chart with more than one color.

Population statistics were calculated for K=5 (Table 2.3). Pairwise F_{st} among 5 clusters calculated by vcftools ranged from 0.0856-0.2319. Ranges for G_{st} and Jost's *D* were 0.0721-0.1875 and 0.0563-0.1807, respectively. H_o (0.1306-0.1758) was always lower than H_e (0.2091-0.2604). F_{is} was relatively high at 0.3247-0.4128. When comparing the mostly *N. clarkii* coastal populations and mostly *N. fasciata* inland populations (Table 2.3) F_{st}, G_{st}, and D were 0.1413, 0.0987, and 0.1067. The heterozygosity measures were lower in the coastal population and they had the highest inbreeding coefficient even when compared to the 5 populations.

While the numbers for each measurement had different ranges, the pattern was always the same. Contrary to the DAPC scatter plot (Fig 2.4) clusters 1 and 3 are not most similar and all measures of differentiation found clusters 1 and 5 closest (Fig 2.4; Table 3). Clusters 2 and 5 are unsurprisingly most distantly related given their geographic and species level separation. Mean measures of differentiation across all populations were very similar for both calculations of F_{st} and G_{st} (0.2183, 0.2171). Jost's *D* was quite a bit lower at 0.1232. Overall H_e was 0.2354, H_o was lower at 0.1524, and F_{is} was 0.3523. At the population level cluster 5 has the lowest H_o and highest F_{is}. This may be due to the *N.c. compressicauda* population found in the Keys.

Table 2.3: F_{st}, G_{st}, and Jost's D pairwise measures of differentiation among the five population clusters. Weir and Cockerham's F_{st} was calculated in vcftools (upper table). Nei's G_{st} (middle table) and Jost's D (lower table) were calculated in R with vcfR. Hierfstat was used to calculate observed heterozygosity (H_o), within population gene diversity or expected heterozygosity (H_e), and the inbreeding coefficient (F_{is}).

Fst	pop1	pop2	pop3	pop4	Coastal	Ho	He	Fis
pop1	x					0.1564	0.2381	0.3431
pop2	0.2128	x				0.1619	0.246	0.342
pop3	0.1391	0.1264	x			0.1758	0.2604	0.3247
pop4	0.135	0.2034	0.2125	x		0.1363	0.2091	0.3482
pop5	0.0856	0.2319	0.1778	0.1076		0.1309	0.223	0.4128
Inland					0.1413	0.1723	0.2849	0.3953
Coastal						0.1433	0.2579	0.4443

Gst	pop1	pop2	pop3	pop4	
pop1	x				
pop2	0.1629	x			
pop3	0.1079	0.0989	x		
pop4	0.106	0.1784	0.1679	x	
pop5	0.0721	0.1875	0.1375	0.0938	
Inland					0.0987

D	pop1	pop2	pop3	pop4	
pop1	x				
pop2	0.1641	x			
pop3	0.1057	0.0942	x		
pop4	0.0921	0.1577	0.1704	x	
pop5	0.0563	0.1807	0.1365	1.0524	
Inland					0.1067

Phylogeny by SVDQuartets

SVDQuartets uncovered several clades with high support (>0.90) (Fig 2.7/2.8). The phylogeny showed only *N.c. clarkii* in cluster 4 formed a monophyletic clade (=1). Three well supported clades were uncovered which were monophyletic for a single subspecies, though none contained all of the members like *N.c. clarkii*. These included *N.f. pictiventris* (=0.99), southern *N.c. compressicauda* (0.98), and *N.c. taeniata* (=1). The *taeniata* clade was comprised of only the seven individuals collected from the same location while the three southern members fit into other clades. An additional clade comprised primarily of cluster 3 and one admixed cluster 1 *N.f. pictiventris* was uncovered at moderate support (=0.76). All *N.f. fasciata* individuals belonged to the same clade (0.93) but were rendered paraphyletic because of two *N.f. pictiventris* from the same cluster. One of the *N.f. pictiventris*, CLP980, has one of the highest levels of admixture (Fig 2.4-2.6). SVDQuartets found low support for clades composed of members from clusters 1 and 5 that did not belong to the well supported *N.c compressicauda* and *N.c. taeniata* clades. Without confidence in their placement the cluster 5 individuals from the mid-western coast split up cluster 1. At K=4 the clades of *N.c. clarkii* and south-western *N.c. compressicauda* form one moderately supported clade (=0.72) while K=6 seems to indicate gene flow involving the midwest coast 6th cluster is what's causing the weakly supported pattern in the tree (Fig 2.6).

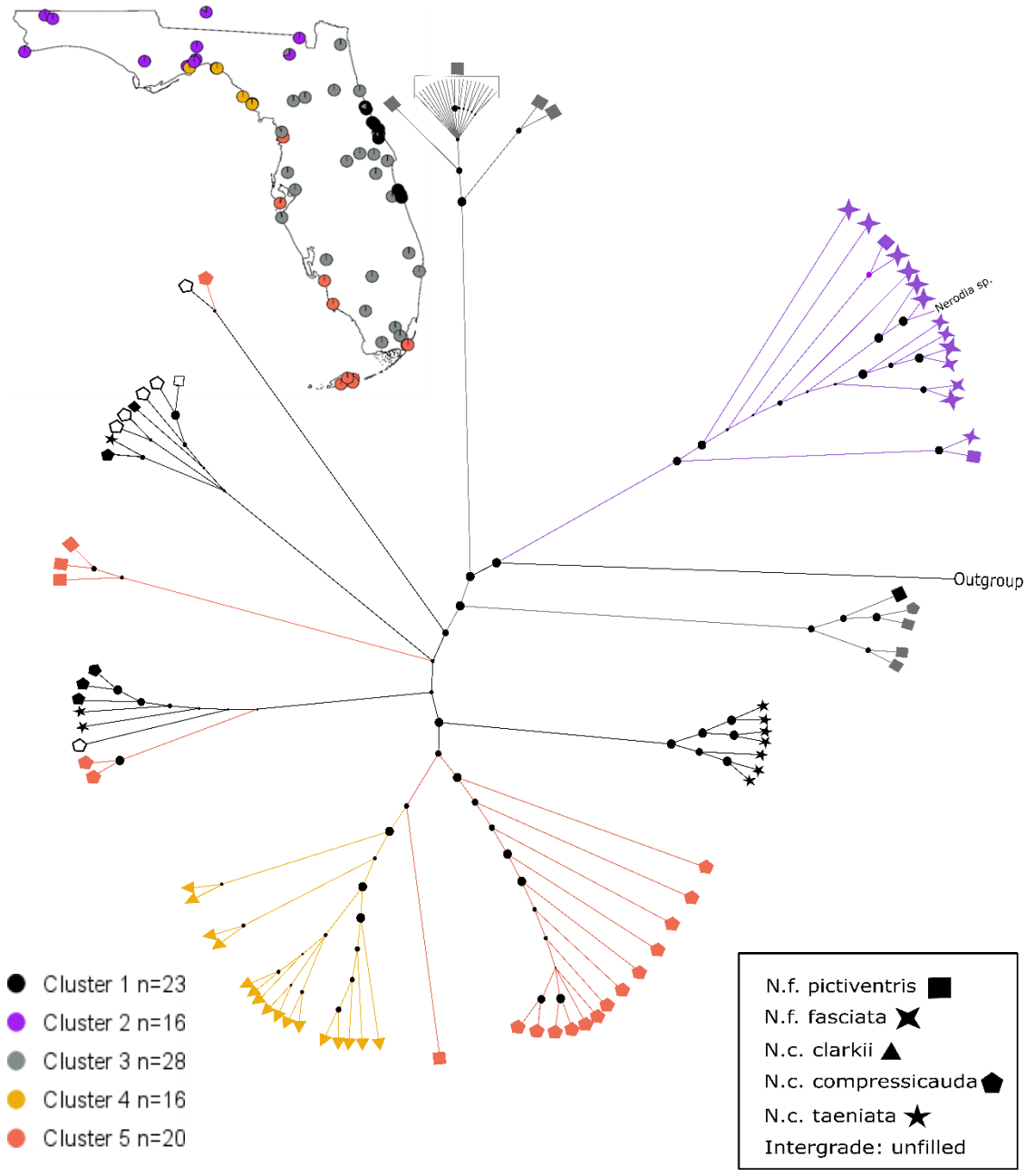

Figure 2.7: SVDQuartet unrooted phylogeny estimated using 500K quartets and 500 bootstrap replicates. The consensus tree was made using the sumtrees function in Dendrophy with min-clade freq of 0.25. Individuals were colored according to their genetic assignment from Fig 2.4 (upper left insert). Their morphological assignment is given by the shape of the tip. Support is given by the dots where their size is proportionate to the level of support. Actual values given in Fig. 2.8.

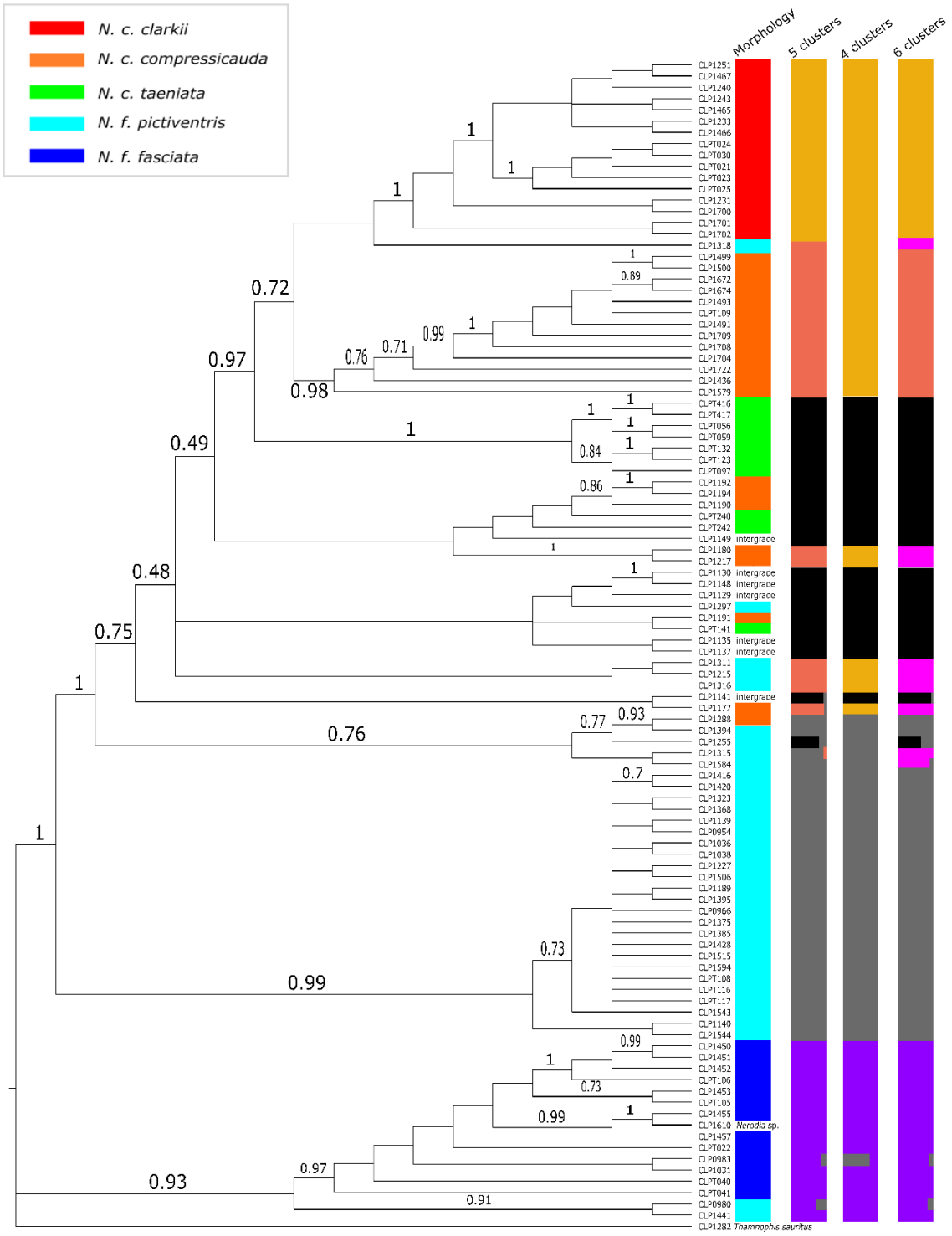

Figure 2.8: Cladogram of the Fig 2.7 phylogeny. Support values over 0.7 shown. The leftmost colored bar represents our subspecies assignments according to their diagnostic characters (color follows Fig. 2.3). The three colored bars to the right represent the population clusters from Figs 2.4-2.6 in order from left to right.

EEMS Migration Estimation

EEMS found several areas of low migration (Fig 2.9). Areas of lowest migration were on opposite coasts in central Florida, around Apalachicola National Forest, and in the northeastern corner of the state. A clear migration corridor was found to run from the panhandle to the tip of peninsular Florida between the coastal migration barriers. The isolated area in Apalachicola National Forest largely matches with *N.c. clarkii* cluster. The migration barrier on the east coast appears to encompass the members of cluster 1 with evidence of a north-south corridor along the coast. The central west coast barrier stretches from Tampa to the Crystal River and includes the individuals who made up the 6th cluster. Unlike its counterpart on the east coast it extends out from the coast more inland. Finally, there appears to be a relatively weaker barrier separating the Keys from the mainland which further indicates that the keys are likely quite interbred.

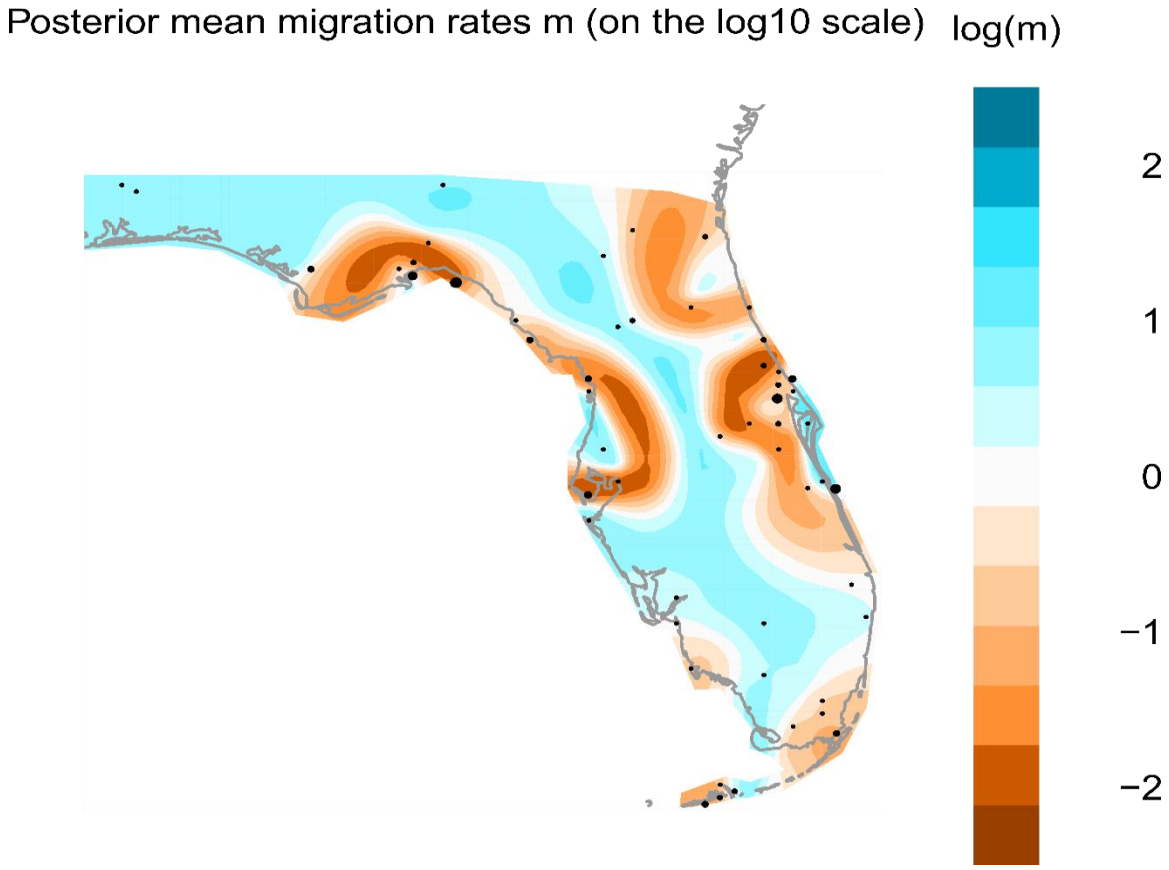

Figure 2.9: Estimated effective migration surface generated in EEMS. White represents the neutral model, isolation by distance. Bands of brown represent barriers to gene flow (where genetic diversity deteriorates) whereas blue areas signify higher rates of gene flow than expected.

Discussion

Despite evidence that delimiting *Nerodia fasciata* and *Nerodia clarkii* and their subspecies using morphological (Dunson, 1979; Hebrard & Lee, 1981; Osgood, 1978b) and genetic (Jansen, 2001; Lawson et al., 1991; Territo, 2013) data is incongruent, this study suggests that the current taxonomy mostly agrees with the genetic populations found in this study. As predicted, five populations were recovered using the ddRADseq data, and they broadly matched the subspecies' distributions (Figs 2.4, 1.1). The BEAST phylogeny was not able to identify any subspecies except for *N. f. fasciata* which was paraphyletic. RADseq data has been shown to be useful in resolving previously low supported phylogenies in taxa with high morphological conservation and low mitochondrial divergence (Pante et al., 2015). I argue that the increased resolution of the thousands of ddRADseq markers is more reliable than the mitochondrial and nuclear sequences. Unlike the BEAST results, each of the DAPC populations is made up almost entirely of a single subspecies, with the possible exception of population 1 (Fig 2.7). Population 1 does not as strongly support *N. c. taeniata* because it includes all the intergrades and several *N. c. compressicauda* and *N. f. pictiventris.* Based on these data, using morphology to delimit the subspecies risks inaccurately assigning individuals to the wrong population. Previous attempts at delimitation were able to identify five populations but correct assignment suffered without the additional data that ddRADseq provides.

Previous population genetic studies in the family *Natricinae* often used microsatellites, not ddRADseq, making it hard to make direct comparisons to this study. The average expected and observed heterozygosity in each study tended to be at least twice that of this study (Jansen et al., 2008; Marshall, Kingsbury, & Minchella, 2009; Tzika et al., 2008; Wood et al., 2015) whereas F_{is} was always lower (J. C. Marshall et al., 2009; Wood et al., 2015). I suspect this is due to the difference in marker as the differences, at least in the heterozygosity measures, holds even in microsatellite studies of *N. c. compressicauda* (Jansen et al., 2008) and the distantly related Jamaican boa, *Epicrates subflavus* (Tzika et al., 2008). F_{st} may be less affected by the difference in markers as F_{st} ranges were found to be 0.01-0.23 and 0-0.297 for the copper belly water snake (Marshall et al., 2009), *Nerodia erythrogaster neglecta*, and the giant garter snake (Wood et al., 2008), *Thmanophis gigas*, respectively. This study found 0.0856-0.2319 which is in the higher end of the previous studies' ranges but not outside the realm of possibility. Additionally, both these snakes inhabit wetlands that have seen a size reduction due to humans, a similar situation to *N. c. taeniata*, which may improve the validity of the comparison. As there was no next generation sequencing results comparable to this study I instead found a ddRADseq study on *Crotalus scutulatus* with F_{st} between 0.026-0.197 (Schield et al., 2018). While there was much overlap in values, their highest was between

populations in the Southwestern deserts of the US and another south of the Central Mexican Plateau, a greater geographic distance than between populations 2&5 from this study.

The SVDquartets phylogeny further supports the validity of the current taxonomy. All subspecies had at least one well supported lineage (Figs 2.7, 2.8) in contrast with the BEAST phylogeny which only had any support for *N. f. fasciata*. Considering both phylogenies separate *N. f. fasciata* there is more support for the split at the Suwannee River. There are more clades however, that show low support. Given the debate over the complex (Carr Jr & Goin, 1942; Cliburn, 1960; Jansen, 2001; Lawson, et al., 1991; Territo, 2013), these low supported clades are unsurprising. If there is hybridization occurring, as has been previously recorded (Cliburn, 1960; Lawson et al., 1991), this could cause issues resolving clade relationships. These clades consist of population 1 and 5 individuals who have the lowest $F_{st.}$ (Table 2.3). Contrary to those populations, population 4 consists entirely of *N. c. clarkii* and is monophyletic. Carr and Goin (1942) hypothesized that *N. c. taeniata* evolved from *N. c. clarkii* but that does not appear to be well supported here.

Finally, EEMS found that there is a high level of migration between populations 2 and 3 and a much-reduced rate towards the coasts (Fig 2.9). Interestingly, these two populations have a relatively high F_{st} at 0.1264 despite the high migration rate. *Crotalus adamanteus* showed a similar pair of populations in Florida but their F_{st} at 0.026 was much lower (Margres et al., 2019). This difference is partially explained by the higher variance expected in ddRADseq data compared to anchored phylogenomics which targets conserved regions (Lemmon, Emme, & Lemmon, 2012) but it is still a substantial difference. Potentially, the sea level fluctuations during the Pleistocene may have provided opportunities for allopatry (Roy, Valentine, Jablonski, & Kidwell, 1996) allowing them to differentiate for a time. The reduction in migration towards the coast matches the distribution of the mostly *N. clarkii* populations 1, 4, and 5. This further supports the validity of the two species and subspecies. Populations 1 and 4 appear to be specifically isolated. Comparing these results to the SVDquartets phylogeny appears to suggest population 1 is less isolated because of is close relationship with the population 5 taxa. This could imply the isolation is relatively recent. With the reduction in gene flow this would allow selection and genetic drift to act upon the populations.

In conclusion, all subspecies are at some level unique but there is evidence of gene flow among them. The five populations geographically appear to mostly follow each subspecies distribution. Despite this overlap the morphological characters used to identify each subspecies fails at correctly assigning individuals to the correct population. Migration patterns suggest that inland populations are experiencing gene flow separate from the coast. *N. f. p*ictiventris, mostly in cluster 3, has the most instances of overlap with other clusters/subspecies suggesting its diagnostic features are not unique. *Nerodia c. clarkii* appears to be the most well supported subspecies through both population genetics and

phylogenetics making it the only subspecies whose subspecies identification is reliable. Additionally, it even seems to be relatively well isolated to gene flow. Of concern is *N. c. taeniata* who all belong in population 1. Population 1 appears to be in an interesting position because it consists of multiple subspecies and is isolated from gene flow but this makes it hard to support the validity of *N. c. taeniata*. This poses problems for its conservation because it is federally threatened. One potential way to resolve this is to assign subspecies based on the geographic ranges of each of the populations from this study, such as assigning the area where population 1 is found as the home range of *N. c. taeniata*. Morphology can be used to supplement this geographic assignment where individuals are on the boarder.

Now that the relationships within the complex are better understood future studies can pursue the potential phylogeographic and/or natural selective origins for the diversification in this group. For example, populations 2 and 3 appear separated by the Suwannee Strait, a phylogeographic break between the mainland US and peninsular Florida (Remington, 1968). This separation has been shown to occur in multiple species found in Florida (Bert, 1986; Carter & McKinney, 1992; Soltis et al., 2006) and specifically in terrestrial reptiles (Burbrink, 2002; Burbrink et al., 2000; Tollis & Boissinot, 2014). Lawson et al. (1991) hypothesized a single origin of *N. clarkii* on the Floridian Pliocene islands that were isolated during periods of high sea levels (Dutton et al., 2015). Also there has appeared multiple cases of taxa forming separate populations on either side of the peninsula (Avise & Nelson, 1989; Saunders, Kessler, & Avise, 1986) similar to the three coastal populations. Likewise the differences in salinity preferences between the two species (Dunson, 1978) can be studied with genetics. The separation between the inland and coast in this study could reflect a difference in adaptation to the salt marshes found on the coast. Outlier detection tests and changes in expression from RANseq data could provide insight onto the genetic mechanisms potentially responsible for difference in salinity preference that has been used to ecologically delimit *N. fasciata* and *N. clarkii*.

CHAPTER 3
EVIDENCE OF SELECTION IN THE NERODIA FASCIATA/CLARKII SPECIES COMPLEX

Introduction

Local adaptation can create pockets of phenotypic divergence even if there have been few genetic changes. One classic example of how impactful a small genetic change can be if a selection pressure is strong is the *Agouti* fur color gene in deer mice of the Nebraska sand hills (Pfeifer et al., 2018). Gene flow between populations on and off the sandhills has been reduced because their fur colors differ. The sandhill mice have a light color matching the sand and cannot survive well on the darker soil off the hills without that camouflage.

Selection pressure can be so strong that it can be deleterious for organisms without the appropriate adaptations to live in that environment. An example of an environmental pressure that has been shown to this intense is salt versus fresh water. Even over short distances it can have a large effect such as in the Common scurvy grass of Norway which have three ecotypes depending on the amount of salt in their environment (Brandrud, Paun, Lorenzo, Nordal, & Brysting, 2017). The danger of dehydration in salt water is a powerful selective force that requires extensive morphological and behavioral adaptions. One group of organisms who have been able to repeatedly evolve the necessary adaptions for survival in salt water is the reptiles (Rasmussen et al., 2011).

Out of more than 10000 species and subspecies of reptiles approximately 100 extensively utilize or live in salt water habitats (Rasmussen et al., 2011). About 80 these are sea snakes and the remainder are sea turtles, the salt water crocodile, and the marine iguana. Each has extensive adaptations to maintain their ionic and osmotic homeostasis, such as the sea turtle salt glands that expel excess salt as a liquid with more salt than sea water (Rasmussen et al., 2011). Not all species go as far as returning to the ocean, but they may make extensive use of brackish coastal environments for activities like forging. An example of this is *Nerodia clarkii*, the salt marsh snake. The preference for brackish water habitats is the key ecological distinction between *N. clarkii* and its sister species *N. fasciata* (Pettus, 1963). Unlike other reptiles, *N. clarkii* does not have salt glands, and continued investigation into its specific adaptations have reached mixed results (Dunson, 1980; H.I. Kochman, 1992; Pettus, 1963). Pettus (1963) demonstrated that if individuals of both species were placed in salt water for an indeterminate amount of time, *N. fasciata* would begin drinking the water and die while *N. clarkii* showed a strong aversion to it. Physiologically when immersed in sea water the freshwater species was more permeable to both water and salts (Dunson, 1980). A recent examination of the kidneys, cloaca, and

colon of the taxa indicate there is little differentiation in osmoregulation (Babonis et al., 2011; Babonis et al., 2012). Although *N. clarkii* was able to maintain its plasma ion balance better than its sister species, there was little difference in morphology, mucus production, or distribution of ion transporters/water channels between them. To date, no one has tested for genetic differences which may play a role in salt tolerance.

To test for genetic evidence of local adaptation, outlier detection methods have become more common (Ahrens et al., 2018). In brief, an outlier is a gene or locus whose level of genetic differentiation (i.e. F_{st}) is outside of the normal range found across the rest of the genome. This differentiation is unlikely to be explained by genetic drift alone, indicating that natural selection may be the explanatory factor causing the locus to be an outlier. Another method includes environmental variables. An environmental association analysis (EAA) identifies genetic variants strongly associated with specific environmental factors. An EAA potentially can uncover evidence of adaptation missed by other outlier detection tests (Rellstab et al., 2015) and provides more information on the potential environmental selection pressure. In this study outlier tests could identify genes unrelated to salt tolerance whereas an EAA can more directly address it. In this way, we can generate a hypothesis as to the function of the outliers. Unlike in the past where we had a few genes or microsatellites, we can now look at thousands of loci from across the genome (Lexer et al., 2013).

The *Nerodia fasciata/clarkii* species complex offers the opportunity to investigate if the different selection regimes of salt and fresh water result in local adaptation. The historically debated difference in adaptation to salt water makes finding osmoregulatory genes of interest. This study could provide the first evidence of selection tied to specific genes within the complex, some of which may be related to salt tolerance. The relative ease and cost of reduced representation libraries, like double digest restriction site-associated sequencing (ddRADseq), have allowed genomic level studies to be implemented on non-model organisms like *N. clarkii*. Using population genomic data this study will be able to examine the *N. fasciata/clarkii* species complex for signs of natural selection to both try and uncover the genetics of salt tolerance.

In summary, my objective is to determine if different populations or the species in the *Nerodia fasciata/clarkii* species complex are experiencing different selection regimes. The regimes may be inferred based on differences in the allele frequencies of loci identified using outlier detection tests and an EAA. Founded on the ecological separation of the two species based on water salinity I developed two hypotheses: 1) if water salinity is acting as a selective pressure then genes associated with osmoregulation will have different frequencies between the coast and inland populations and 2) if water salinity is not acting as a selective pressure then there with either be no osmoregulatory genes found or the allele frequencies will not differ between the inland and coastal populations.

Materials and Methods

Sampling and SNP Generation

Snakes were collected from 2012-2015 either by road cruising or by searching marshes, ponds, wetlands, and other aquatic bodies. The snakes were euthanized or tail clippings and blood was taken in the field. Tissue was preserved in ethanol at -20˚C. From these samples a subset of 139 (Fig 2.2) was selected to generate ddRADseq libraries. Individuals were identified based on morphological characters and which subspecies range they were collected in. In total I identified and selected 26 *N. c. taeniata*, 19 *N. c. clarkii*, 26 *N. c. compressicauda*, 14 *N. f. fasciata*, and 45 *N. f. pictiventris*. The sampling also included seven intergrades, which could not be identified beyond the species level and one *Nerodia* that appeared to be a hybrid of the two species but could not be definitively identified as one species.

DNA was extracted using a standard phenol-chloroform procedure from blood, liver, or scale clippings. The ddRADseq protocol follows Peterson et al. (2012) with some modifications. Briefly, a minimum of 500ng of DNA was digested for a minimum of 8 hours at 37°C using the rare cutting restriction enzyme SbfI-HF (8 bp) and the common cutting Sau3AI (4 bp). The samples were cleaned using AMPure magnetic beads (Beckman Coulter, Inc., Irving, TX, USA) at a 1.1:1 bead-to-DNA ratio and eluted in 42uL of tris-HCl ph8.0 for quantification. Samples were split into groups of 8 to minimize differences in DNA quantity. A barcode was added to each group and ligated at 16°C for 5 hours. After ligation, each sample was cleaned using AMPure beads and DNA quantified. The groups were pooled and cleaned a second time with AMPure magnetic beads and eluted into 30uL of TE ph8.0. Each pooled group was size selected for fragments 300-700bp long on the Blue Pippin Prep (Sage Science) with a 1.5% agarose gel. This size-range was selected to recover the maximum number of potential loci for downstream analyses (Schield et al., 2015). One out of eight indexed primers were added to each group and amplified with PCR. The number of barcode-index combinations was 128. Samples were cleaned twice at a 0.7:1 bead ratio and quantified on the Qubit. Samples were further quantification using a high-sensitivity DNA chip for the BioAnalyzer 2100 (Agilent Genomics) to confirm the appropriately sized fragments were amplified. Sequencing was performed on an Illumina Nextseq 550 using 150 bp paired-end reads.

The reads were *de novo* assembled using the ipyrad pipeline. Reads were filtered if their average phred score offset was less than 33 and if they had 5 or more low quality base calls per read. Bases were trimmed of the end of the remaining reads to a minimum of 100 if they had a Qscore less than 20. These were clustered at 90%. All other parameters were

left at default settings. Any samples with fewer than 1000 loci were removed from further analyses.

Vcftools v0.1.15 (Danecek et al., 2011) was used to further filter SNPs. To avoid sampling bias due to uneven sampling no more than 7 individuals within a 10-mile radius were allowed. Samples were removed haphazardly if that condition was not met. Only biallelic SNPs were retained. Any SNP with a maximum of 50% missing data, a minimum mean depth of 10x, and minor allele frequency greater than or equal to 0.05 was retained. Any individuals with a mean depth below 5x were removed.

Outlier Loci

In brief, the data set contained 103 samples after filtering samples if they had fewer than 1000 loci, more than 7 samples per 10mi radius, and mean depth of coverage at 5x or less. These samples were clustered into 5 populations using k-means clustering and a DAPC (Fig 2.4). The tests for outliers were repeated for two new populations consisting of a mostly *N. fasciata* inland population consisting of clusters 2 and 3 and a mostly *N. clarkii* coastal population consisting of clusters 1, 4, and 5. The tests on the inland and coastal populations may provide information on if the species are separate due to an adaptive difference.

Loci potentially under selection were found using two outlier loci tests. The first was PCAdapt (Luu, Bazin, & Blum, 2017), an R package that uses Principal Component Analysis (PCA) to statistically test for loci outside the neutral distribution. The population genetics VCF file from Chapter 2 was converted to a unique pcadapt format using the read.pcadapt function. To choose the number K principal components (PCs) I ran a test run using K=50. Using the scree plot function the number K of PCs was kept, where the recommended K corresponds to the number before the plateau in the plot. To confirm the scree plot K, score plots were made using 2 PCs at a time until the population clusters could no longer be distinguished. PCAdapt was run using the Mahalanobis distance to generate a p-value for each site. P-values were only generated for SNPs with a minor allele frequency greater than 0.05 causing 49 to be removed.

Outlier detection tests, including PCAdapt, are known for a high rate of false positives. I compared the results of PCAdapt with a more conservative outlier test to try and avoid false positives. Bayescan v2.1 (Fischer, Foll, Excoffier, & Heckel, 2011) finds F_{st} outliers outside the neutral F_{st} model it generates using a Bayesian approach. Bayescan tests two alternative models for each locus (selection vs neutral) and uses a reversible-jump MCMC algorithm to estimate the posterior probability of each model. This probability is used to create Bayes Factors (BFs) which are used to compare models. I used PGDSpider v2.1.1.5 (Lischer & Excoffier, 2011) to convert the SNP data from VCF to Bayescan format with the 5 populations identified from the DAPC (Fig 2.4). The program was run

twice as two independent chains using default parameters: 5000 iterations, thinning interval of 10, 20 pilot runs 5000 iterations long, and a burn-in of 50000. Bayescan also requires the prior odds of the neutral model which defines our skepticism about the chance of each locus being under selection. Following the manual, I increased the odds from 10 to 100. Convergence was checked using three tests implemented in the R package coda (Plummer, Best, Cowles, & Vines, 2006). The Geweke diagnostic, and the Heidelberg and Welch's diagnostics each test the convergence of a single chain whereas the Gelman-Rubin diagnostic compared the two chains. The chain with the better Geweke plot was used for further analysis.

Outlier Selection with MINOTAUR

Outlier loci were selected using the R package MINOTAUR (Verity et al., 2017). MINOTAUR uses multivariate distance measures to compare statistics generated from outlier tests and selects the most likely outliers. The PCAadapt p-value and the Bayescan q-value were used as input for the harmonic mean distance calculation. The top 2.5% were selected for mapping and plotted using the r package qqman. This was performed for the 5 population clusters and between inland and coastal populations.

Outlier loci identified through MINOTAUR are potentially in linkage disequilibrium with genes under positive selection. To find any of these candidate genes I used the *Thamnophis sauritus* genome (McGlothlin et al., 2014) to map them to as it is the most closely related species with a draft genome (Guo et al., 2012; Pyron, Burbrink, & Wiens, 2013). Using Blast+ 2.7.0 (Camacho et al., 2009) the *Thamnophis sauritus* genome was made into a database and outlier loci were blasted to the database using megablast. Only the top alignment for each outlier was kept. Taking the scaffold and position information from each, the windows function of bedtools v2.27.1 (Quinlan & Hall, 2010) was used to search for genes within 20kb and 50kb of each outlier. I was unable to find any studies on linkage disequilibrium in snakes so a conservative and liberal window size was chosen, as several of my populations are likely inbred increasing linkage. To get an idea of putative function, each gene was inputted into the PANTHER classification system (Mi, Muruganujan, Ebert, Huang, & Thomas, 2018; Mi et al., 2019) to look for gene ontology biological processes (GO BP) terms using the *Anolis carolinensis* (green anole lizard) annotated genome. If the genes could not be found in the *Anolis* database they were checked against *Homo sapiens.* Any genes which contained GO terms associated with kidney or renal function and development, solute transport, or osmoregulation were considered candidates that were further investigated using genecards.org (Stelzer et al., 2016). This provides function summaries, mostly based from human studies, of each gene and a list of phenotypes it was associated with in genome wide association studies (GWAS) registered on the GWAS Catalog.

Environmental Association Analysis

Bayenv2 (Günther & Coop, 2013) tests for correlations between allele frequencies and environmental variables. This may allow bayenv2 to detect a difference between *N. clarkii* and *N. fasciata* due to the former's preference for salt water. Bayenv2 accounts for population structure by creating a null model based on the covariance in allele frequencies between populations. For environmental variables I downloaded the 19 bioclimatic layers at 2.5 resolution from WorldClim (Fick & Hijmans, 2017). I did not have a good sampling of salinities so I used altitude from WorldClim as a proxy. Lower altitude water bodies at the coast are more likely to have salt water than water bodies inland. SNP data was converted to Bayenv format from VCF using PGDSpider v2.1.1.5. To calculate the covariance matrix Bayenv2 was run without the environmental factors for 100000 iterations. Standardized environmental factors were generated for annual mean temperature and annual precipitation. These were selected to avoid correlations with the other bioclimatic variables while still being important selective pressures for many species. Altitude was used as a proxy for elevation where water sources at higher elevation are more likely to be fresh water whereas lower elevation, often at the coast, is more likely to contain salt. The value for each variable at all sampling locations was extracted then averaged for each population. Each was standardized by subtracting the mean of all populations and dividing by the standard deviation. Bayenv2 was run for 100000 iterations to calculate the BF for each SNP. Additionally, Spearman's rank correlation coefficient ρ was calculated to make sure the underlying model Bayenv2 used was correct. This was repeated for the coastal versus inland populations.

The top 1% of SNPs were correlated to at least one environmental factor. This was an equivalent of BF>20 and Spearman's $\rho > |0.2|$. Following the same procedure as with the previous outlier selection analysis the loci that contained these SNPs were then blasted to the *Thamnophis sirtalis* and then used bedtools to search for genes. Searching for GO BP terms for osmoregulation is not appropriate for genes cound correlated to temperature or precipitation. Instead I separated the outliers according to environmental variable and genes found at 50kb were then used as input for the PANTHER over-representation test. The test utilizes Fisher's Exact test to examine if the GO terms associated with my list are found more often than are represented in the *Anolis carolinensis* genome. A p-value represents the significance of the over-representation. To further explore these GO terms I used the Reduce and Visualize Gene Ontology (REVIGO) discovery tool (Supek, Bošnjak, Škunca, & Šmuc, 2011). REVIGO summarizes long lists of GO terms by finding a representative subset using semantic similarity measures to reduce redundancy then plots them in semantic space. REVIGO was run with an allowed similarity score of 0.7.

Results

Outlier Analysis

MINOTAUR found 132 outlier loci across the five populations and 136 between the inland-coastal populations. The total number of genes putatively in linkage with the outliers that successfully blasted to the *T. sirtalis* genome were roughly equivalent between the two population schemes however not all genes have been previously characterized (Table 3.1, Table S3, S4). Approximately 50 genes at 20kb and 100 genes at 50kb are named indicating there us some information on their function (Table 3.1). The genes that were predicted or have an unknown function (e.g. LOC106538175) require more study as they could contain gene coding for important adaptive phenotypes. Of the characterized genes some have putative functions (e.g. osmoregulation) that may be related to the diversification of this complex. These are listed in Table 3.2.

Table 3.1: Number of genes found within 20kb and 50 kb of outliers that successfully blasted to the *Thamnophis sirtalis* genome.

Groupings	20kb		50kb	
	Total	Characterized	Total	Characterized
5 clusters	80	52	175	102
Coastal vs Inland	77	53	166	101

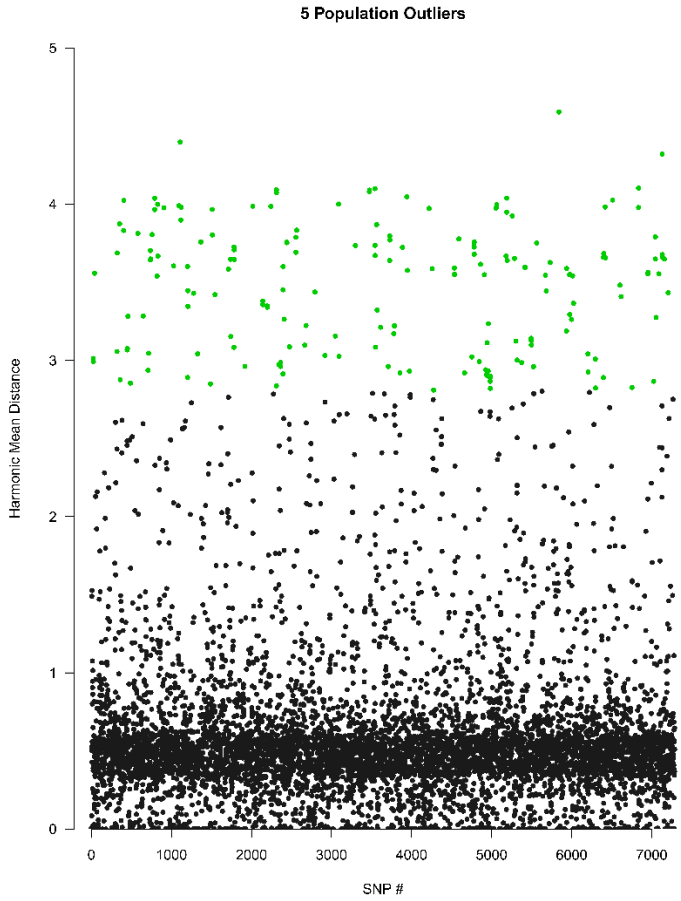

Figure 3.1: 5 population manhattan plot of the harmonic mean distance for each SNP. The x-axis corresponds to a single SNP. Green dots represent the top 2.5% SNPs which are found on 132 loci

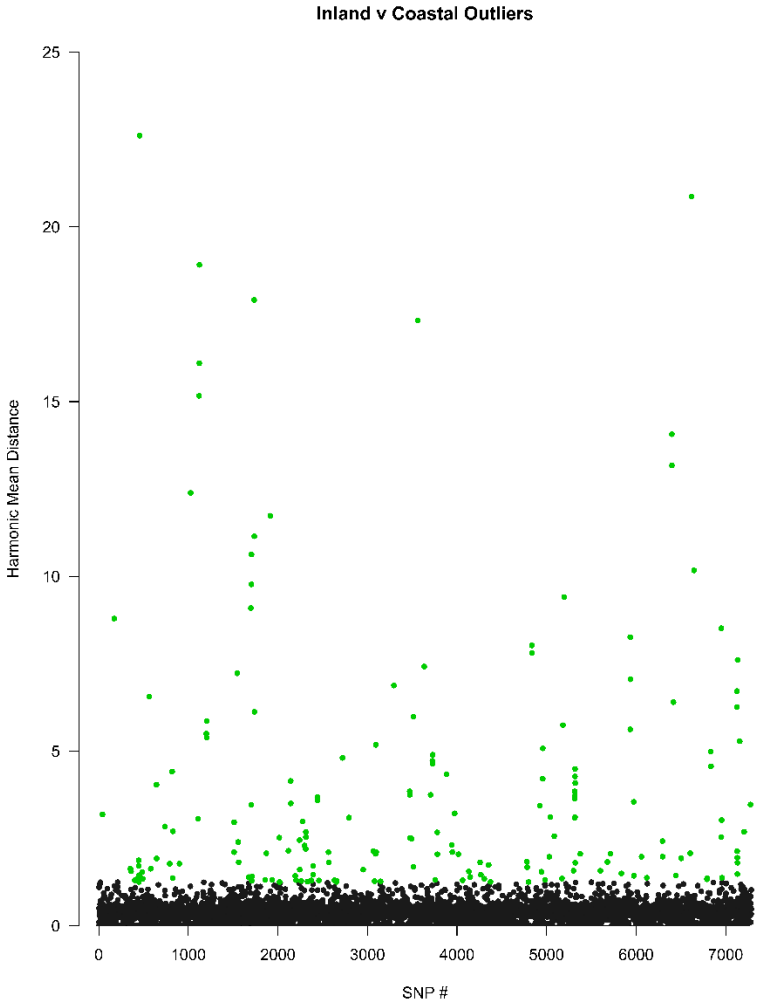

Figure 3.2: Inland v coastal manhattan plot of the harmonic mean distance for each SNP. The x-axis corresponds to a single SNP. Green dots represent the top 2.5% SNPs which are found on 136 loci.

Table 3.2: List of genes near the MINOTAUR outlier loci potentially of importance to the complex's diversification. Important GO BP terms listed. .....Phenotypes that were found associated with the genes, along with the mean association score (i.e. 8=p-value≈$1x10^{-8}$), are included.

	Gene Name	GO BP Terms	GWAS Phenotypes (score)
5 Clusters	*FYN*	Regulation of calcium ion import across plasma membrane	Skin pigmentation (5)
	ITPR3	Ca^{2+} transport, Ca^{2+} transmembrane transport, Ca^{2+} transport into cytosol	Kidney disease (11.7)
	MED1	Angiogenesis, thyroid hormone mediated signaling pathway, thyroid hormone generation	Glomerular filtration rate (18.5)
	NOS3	Angiogenesis, regulation of sodium ion transport, nitric oxide biosynthetic process, response to fluid shear stress, negative regulation of K+ transport	N/A
	SLC13A3	Sodium ion transport	Chronic kidney disease (8.7), glomerular filtration rate (8.7)
	SLC9A3	Cation transport, Na^+ transport, regulation of pH, Na^+ import across plasma membrane	N/A
	SLC9A8	Cation transport, Na^+ transport, regulation of pH, Na^+ import across plasma membrane	N/A
Inland v Coastal	*ADM2*	Angiogenesis, feeding behavior, digestion	N/A
	NDUFS3	Mitochondrial electron transport, NADH ubiquinone	N/A
	SLC13A3	Sodium ion transport	Chronic kidney disease (8.7), glomerular filtration rate (8.7)
	SLC9A8	Cation transport, Na^+ transport, regulation of pH, Na^+ import across plasma membrane	N/A
	WNK1	Ion transport, regulation of sodium ion transport	N/A

The differences in allele frequencies from the Minotaur outliers are found in Table 3.3. The frequencies showed several patterns such as population 2 and 3 always had the same frequency pattern. Populations 4 and 5 shared frequencies for all outliers. Population 1 never had more of the alternate allele and was more likely to be similar to clusters 4 and 5. Only population 1 had a gene where it differed from all other populations.

Table 3.3: Allele frequencies per population for each gene putatively of importance to the complex's divergence. Each was found within 50kb of a locus identified by MINOTAUR as being putatively under selection. Instances where the alternate allele had a frequency >0.5 are bolded.

	window (kb)	*Pop2*	*Pop3*	*Pop1*	*Pop4*	*Pop5*	*Coastal*	*Inland*
FYN	50	**0.25/0.75**	**0.25/0.75**	1/0	**0/1**	**0.42/0.58**	-	-
ITPR3	20	0.97/0.03	0.98/0.02	1/0	**0.07/0.93**	**0.08/0.92**	-	-
MED1	50	**0/1**	**0.07/0.93**	0.67/0.33	1/0	0.85/0.15	-	-
NOS3	20	0.81/0.19	0.81/0.19	0.87/0.13	0.85/0.15	0.77/0.23	-	-
SLC13A3	20	**0.08/0.92**	**0.14/0.86**	0.88/0.12	1/0	1/0	0.96/0.045	**0.12/0.88**
SLC9A3	50	**0/1**	**0.07/0.93**	0.67/0.33	1/0	0.85/0.15	-	-
SLC9A8	20	**0.07/0.93**	**0.02/0.98**	0.78/0.22	1/0	0.95/0.05	0.99/0.01	0.83/0.17
ADM2	50	-	-	-	-	-	0.86/0.14	**0.01/0.98**
NDUFS3	50	-	-	-	-	-	0.78/0.22	**0.02/0.98**
WNK1	50	-	-	-	-	-	**0.42/0.58**	0.88/0.12

Environmental Association Analysis

Across the five populations, bayenv2 found 52 loci to be putatively associated with an environmental variable. Thirty-six were correlated to annual mean temperature, 28 with annual precipitation, and only a single locus was associated with altitude (Table S7). This locus did not have a gene within either sized window. Within 50kb on the *T. sirtalis* genome there were 98 genes and 62 characterized. Within 20kb the numbers were 54 and 36. No loci between the inland and coastal populations reached the BF cutoff thus they did not show a strong association with any environmental variable.

I found 38 genes putatively associated with temperature. PANTHER found 167 GO terms associated with these genes, none of which were over-represented, and REVIGO reduced it down to 81 term clusters (Fig. 3.3). PANTHER found 165 GO terms across 44 genes correlated to precipitation. None of these terms were over-represented. REVIGO reduced this list to 83 (Fig 3.4).

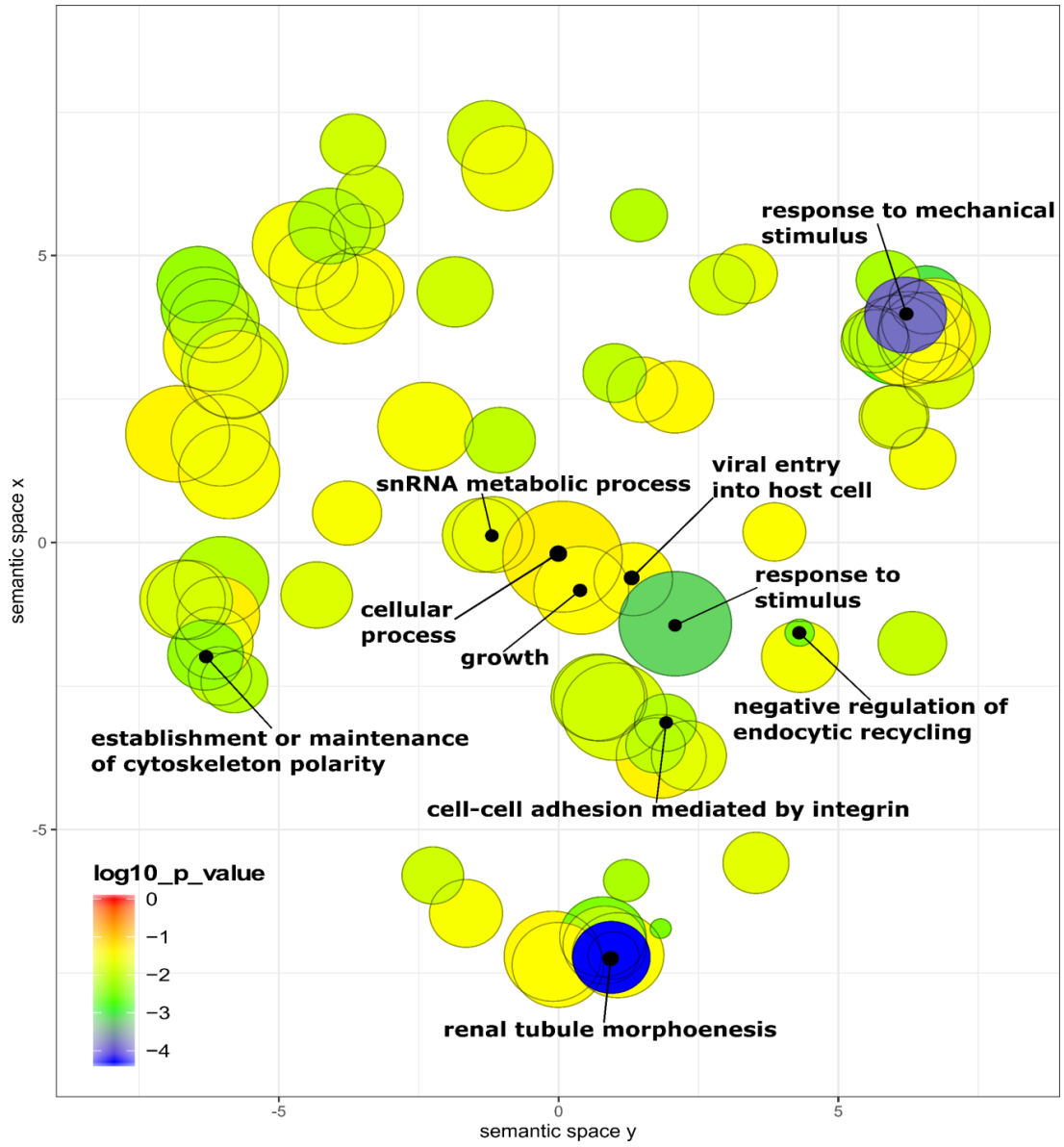

Figure 3.3: REVIGO scatterplot of GO terms from genes found by Bayenv2 to be correlated to annual mean temperature. Each circle represents a cluster of GO terms that are >70% semantically similar. Their size represents how many genes in UniProt are included in that GO term. Circles are placed in 2-dimentional space where the closer they are semantically the closer they are in space. Circles are colored by the log10 of the p-value derived from the PANTHER over-representation test. The more blue the color, the more significant the p-value.

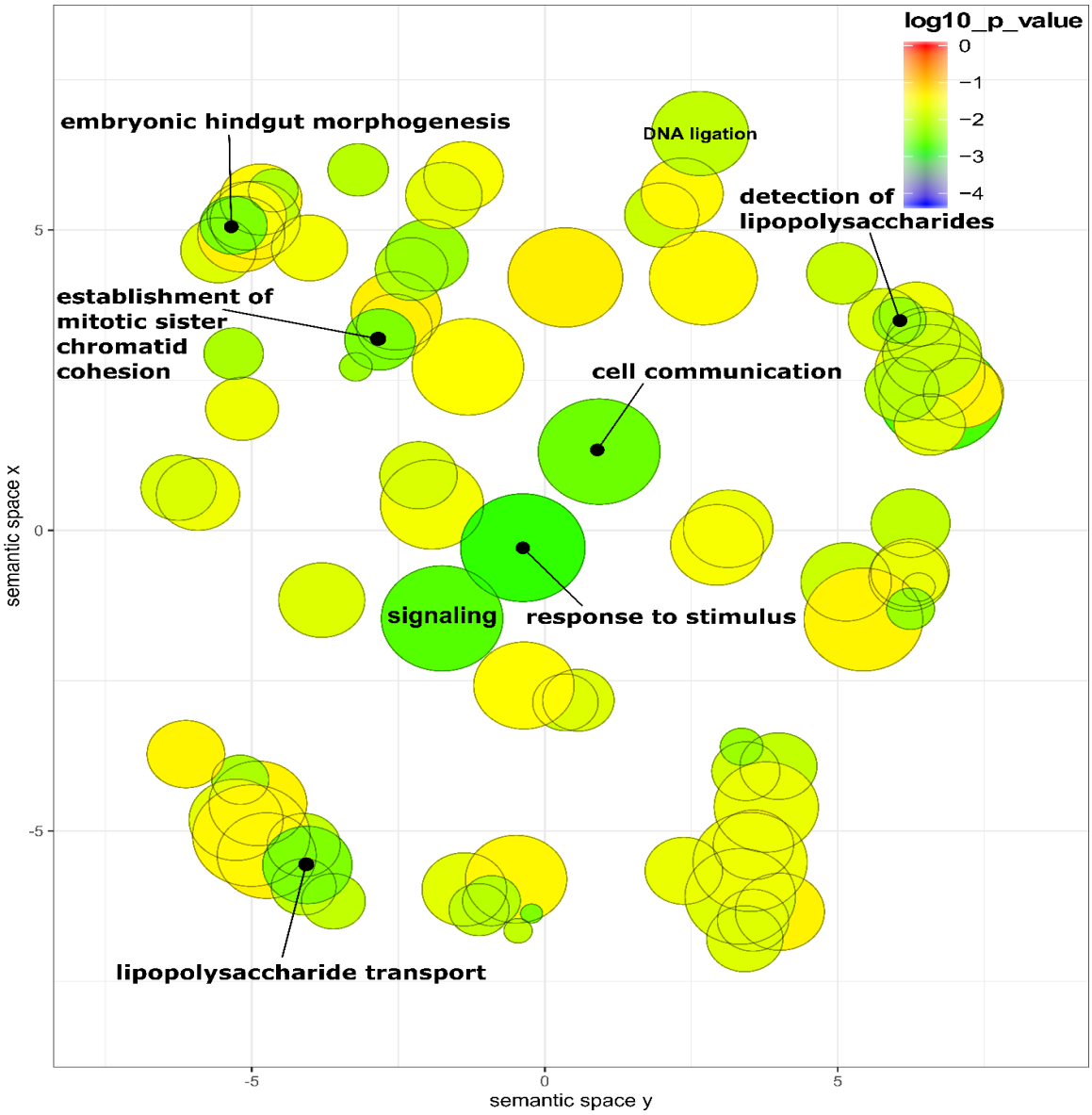

Figure 3.4: REVIGO scatterplot of GO terms from genes found by Bayenv2 to be correlated to annual precipitation. Each circle represents a cluster of GO terms that are >70% semantically similar. Their size represents how many genes in UniProt are included in that GO term. Circles are placed in 2-dimentional space where the closer they are semantically the closer they are in space. Circles are colored by the log10 of the p-value derived from the PANTHER over-representation test. The more blue the color, the more significant the p-value.

Discussion

In this study I aimed to find genes that may play a part in the adaptation of *Nerodia clarkii* to salt water. My results contribute a list of candidate genes that may be under selection which provides some evidence for the validity of separating of *N. clarkii* and *N. fasciata* based on ecology. The outlier detection scan found evidence that there was a difference in allele frequencies of potentially osmoregulatory genes putatively in linkage with the outlier loci (Table 3.3). Five genes were found when specifically testing between inland and coastal populations and seven among the five populations. When comparing the five populations the coastal populations 4 and 5 almost always differed from the inland populations 2 and 3. Population 1 did not always share frequencies with the other coastal populations possibly because of the geographic separation between the east and west coasts reduced gene flow allowing it to diverge. While these allele frequencies do lend some support the ecological importance of salt water the EAA failed to find any gene putatively correlated to altitude. This may be because altitude is a poor proxy for salinity due to the low profile of Florida. The highest points in Florida are found in the Northern Florida Highlands and the Lake Wales Ridge, each less than 350ft high (Upchurch, Scott, Alfieri, Fratesi, & Dobecki, 2019). This may not provide a sufficient cline to find associations. Without the correlation to altitude we do not have evidence more directly identifying salinity as a selective pressure though both temperature and precipitation appear important.

Of the candidate genes found in the outlier scan many have been demonstrated to be important for osmoregulation in other taxa. For example, solute carrier 9 member A3 (SLC9A3) is an acid-based control of Na^+ and has been demonstrated to function in the rainbow trout renal system (Ivanis, Braun, & Perry, 2008). SLC9A8 functions similarly and has been found to be upregulated in threespine stickleback in response to increased salinity (Gibbons, Metzger, Healy, & Schulte, 2017). SLC13A3 is not definitive in its purpose as the previous two solute carriers but it is associated with both chronic kidney disease and glomerular filtration rate (Nanayakkara et al., 2013). Separate from the solute carrier families is WNK1 which a candidate due to its well-documented history of regulating NaCl in fish and mammals (Delpire & Gagnon, 2008; W. S. Marshall, Cozzi, & Spieker, 2017). In addition, it is specifically an outlier between the inland and coastal populations lending support to the ecological separation between the two species. Potentially the other genes found among the five-population test may indicate independently evolved adaptations. For example, the FYN variation is unique to population 1. Many of the allele frequencies are somewhat different in population 1 which may be attributed to drift. Many plants and animals exhibit a Gulf coast-Atlantic coast discontinuity (Soltis et al., 2006). The coastal populations follow this pattern (Fig 2.9) and

there is little evidence that the coastal habitats would be different thus the differences in allele frequencies may be attributed to drift.

Bayenv2 did not provide support for salinity as a selection pressure. It failed to detect any correlation between allele frequencies and environmental variables between the inland a coastal population. Only one locus was found correlated to altitude among the five populations and no genes were found near it. Temperature and precipitation each had correlations to multiple loci though there was no statistically significant over-representation of any GO terms of genes near these loci. In semantic space there were several clusters of importance. For temperature the terms renal tubule morphogenesis (most significant p-value), response to mechanical stimulus, and establishment or maintenance of cytoskeleton polarity had multiple terms clustered near them (Fig 3.3). We know that elevated temperatures during incubation affect the phenotype of *Nerodia fasciata* (Osgood, 1978a) while in chickens it can reduce organ size (Leksrisompong, Romero-Sanchez, Plumstead, Brannan, & Brake, 2007). Feasibly the genes related to organ development may be adapting to the change in temperature across the state. Terms involved with lipopolysaccharides appear correlated to precipitation. I found no studies relating lipopolysaccharides (found in Gram-negative bacteria) to precipitation but they are demonstrated to cause changes in thermoregulation of reptiles (Deen & Hutchison, 2001; Merchant, Fleury, Rutherford, & Paulissen, 2008). Specifically *Thamnophis sirtalis* exhibits hypothermia though *Nerodia sipedon* appears to have no reaction (Burns, Ramos, & Muchlinski, 1996). Similar to the temperature correlations, the development of organ systems forms a cluster, however it appears to mainly focus on the digestive tract (Fig 3.4). Unfortunately, other GO terms do not have evident relationships with temperature or are general enough to be unhelpful in forming hypotheses.

In summary, this study has found evidence for selection within the *Nerodia fasciata/clarkii* species complex. Evidence for salinity being an important environmental selection pressure is not conclusive. Genes putatively under selection with potential osmoregulatory function were consistently found to have different allele frequencies between inland and coastal populations (Tables 3.2&3.3) though drift may also play a strong role in population 1. However, the osmoregulatory genes were only a part of the ~100 genes found in each population comparison. The EAA was unable to find a gene correlated with altitude. This may be due to it being a poor proxy for salinity suggesting a need for salinity data. These data may not conclusively show adaptation to salt water but I believe it is warranted to continue to pursue the hypothesis. Temperature and precipitation, not previously studied as important to the taxa, did appear to be potential selective pressures. In conclusion, these results indicate that there is likely be different selection regimes within this species complex. These selection regimes appear to largely be separated between the coast and the inland populations based on the outlier scan and the

EAA suggest finer scale regimes between the five populations based on temperature and precipitation. Given these data further analyses based on salinity, as well as temperature and precipitation, are warranted. For now, the differences between the inland and coastal samples provide further evidence for the validity of *N. fasciata* and *N. clarkii*, even if it is not clear which specific ecological factor they are adapting to nor what the adaptation is. Independent adaptation in individual populations may also be used as evidence for subspecies.

CHAPTER 4

DISCUSSION

Populations often begin to diverge due to the influences of evolutionary mechanisms, such as gene flow and selection. If the populations do not reunite, they will continue to accrue differences and may ultimately speciate. In this study I examined the evolution of the *Nerodia fasciata/clarkii* species complex in Florida. The complex contains 5 subspecies found in Florida but their validity has been questioned due to unreliable morphological characters and sparse genetic data. To understand the evolutionary relationships within the group I identified potential population structure and estimated their phylogeny. I examined gene flow patterns and searched for evidence of selection within the complex. Based on my results I can come to two conclusions. First, based on multiple lines of evidence there are at least 4 diverging populations (at least 2 per species). They are likely in the grey zone of speciation (De Queiroz 2007) where they may still converge into one taxon again. Second, the use of morphology does not accurately reflect the population structure of the complex. Given the difficulties of quick genetic assessment I suggest that geography be used as an initial subspecies identifier with morphology acting as a supplement.

To investigate the complex's evolutionary relationships, I generated a population genomic level data set of 103 individuals and 7,336 SNPs using double digest restriction site-associated sequencing. Using this dataset, I found evidence that there are multiple well supported populations within the species complex that roughly align with the distributions of the 5 subspecies, though the subspecies identifications (based on morphology) did not reliably match these population clusters. Using mitochondrial and nuclear loci I found a few well supported clades at older nodes but there was little support in young nodes. Notably it did differentiate two populations, one in the panhandle and one in the peninsula. No subspecies were monophyletic nor were the species. With the higher resolution afforded by the population genomic data 4-6 populations, 5 being most likely, were identified and an unrooted coalescent phylogeny was estimated. Only *N. c. clarkii* was monophyletic with all of its members forming a single population and clade. All other subspecies formed least one well supported clade but there was evidence of admixture. Several clades were not well supported and contained multiple subspecies from multiple populations, likely as a result of the close evolutionary relationship the taxa share. Migration among the groups is restricted at the coasts with a corridor of elevated migration from the panhandle to the bottom of the peninsula. This migration corridor consists mostly of *N. fasciata* supporting its separation from *N. clarkii* on the coast. There also appeared to be barriers to gene flow around populations 1, 4, and 5. These results show support for two species, and five

population clusters. However, these genetic clusters do not agree with the morphologically based current taxonomy; except for *N. c. clarkii*.

Using two outlier detection measures I identified candidate genes that may be playing a role in population divergence. Selection on genes related to traits associated with osmoregulation is likely occurring more in individual populations than between species due to fewer candidates being found between the inland v coastal populations which largely represent *N. fasciata* and *N. clarkii*, respectively. Populations 2 and 3 then populations 4 and 5 have the most closely related selection regime according to allele frequencies of candidate genes. The former two populations are separated by the Suwannee River but the migration corridor and shared allele frequencies between them suggests this biogeographic boundary is weaker in this complex than in other taxa. I believe future investigations of the historical biogeographic origins of these populations would be worthwhile. Populations 4 and 5 have the second lowest measures for F_{st}, G_{st}, and D. At K=4 they even form one west coast cluster separate from population 1, which is characteristic of a biogeographic separation of the Gulf and Atlantic coasts. To find potential causes for selection I also performed an environmental association analysis using mean annual temperature, annual precipitation, and altitude. Altitude was to be a proxy for salinity but only one locus correlated to it. I cannot say with this data if this was due to it being a poor proxy or if salinity is actually not important. Based on these results selection appears to be playing a role in the differentiation between and within these two species. Not only was there a difference between the inland and coastal populations but it could be seen when comparing among the five populations that the *N. clarkii* and *N. fasciata* majority populations often had different allele frequencies. Finally, population 1 was somewhat different the other four populations but more similar to 4 and 5, possibly indicating its ecological situation is different from the others.

Taxonomy and Conservation

With available evidence, I suggest there is a case for five subspecies but the use of morphological characters used to delimit *N. f. fasciata*, *N. f. pictiventris, N. c. clarkii, N. c. compressicauda,* and *N. c. taeniata* are insufficient. All were more often grouped with their own subspecies though only *N. c. clarkii* formed a monophyletic group and population cluster above K=4. Those that did not cluster with their own subspecies indicates that there is introgression occurring and/or that the morphology is not accurate at identifying subspecies or differentiation. At the species level *N. fasciata* and *N. clarkii* subspecies each were more likely to cluster within their species than without. The inland and coastal populations comprised mostly of one species each did have putative adaptive divergence further strengthening their species level divergence. I assert that that genetic clusters are

the best representation of the subspecies followed by geography. Given that on the spot genetic assignments can't be made infield identifications should first be based on geography then using morphology as a supplement in areas were two subspecies come in contact, for example, the Tampa bay region.

The federally threatened Atlantic Salt Marsh snake (*N. c. taeniata*) was harder to make conclusions on due to the intense sampling bias requiring the removal of many samples. Instead of a taxonomic change I suggest using an evolutionarily significant unit (ESU). An ESU is a population or group of populations that has a high genetic and ecological distinctiveness (Funk, McKay, Hohenlohe, & Allendorf, 2012). My data indicates that population 1 has multiple lines of evidence suggesting it could warrant that distinction. It is isolated by a barrier to gene flow and has the most unique allele frequencies compared to the relative uniformity of the inland populations and the Gulf coast populations. This population also has the most morphological and subspecific variation among the three subspecies. This includes all the intergrades who could not be distinguished between subspecies using meristic characters. The point of contention may be the lower F_{st} values and its tendency to form clades with other subspecies. I do not see this as a problem because my gene flow data suggests it now isolated with opportunity to diverge. *N. c. taeniata* (found entirely in population 1) is considered to be the most recently evolved subspecies according to Lawson et al. (1991) so the genetic distinctiveness may not be particularly high. I suggest that population 1 be designated as an ESU and the area that encompasses it be protected. This will preserve this unique area and the threatened subspecies for further investigations into the lineage. I believe there may be a case for making all of population 1 *N. c. taeniata* considering it is separated from every other population, it is found approximately in the subspecies' historical range, and the reason the individuals are not considered *N. c. taeniata* now is because of unreliable morphological characters.

Future Directions

I believe this study encourages future studies to follow several avenues of research focusing on evolutionary history and selection. The phylogeographic history of the complex is still debatable. These data provide a starting point for investigations into biogeographic hypotheses on the origin of the species and subspecies in the complex. A southward expansion by *N. f. pictiventris* followed by allopatry may have given rise to *N. f. pictiventris*. A single origin of *N. clarkii*, due to the need for salt tolerance, on the Pliocene islands of Florida as suggested by Carr and Goin (1942) seems likely but dating the divergence is necessary. This also assumes that salinity plays an important role which could not be confirmed here. My outlier scan did find several candidate genes that could have played a role in the divergence based on salinity. However, it could be coincidental

as there has been no functional test of these genes to differentiate them from simple drift following normal Floridian phylogeographic patterns.

Another avenue would be to develop experiments to find the functional differences between the species based on some of the candidate gene found in this study. The first step would be to gather actual salinity data for the EAA to find any additional candidates. Altitude may have been a poor proxy due to the overall low elevation of Florida. From there any genes found correlated with salinity and the candidates from this study can be the basis of gene expression and functional assays. Following Babonis et al. (2011) total RNA could be collected for each species for an expression analysis. Designing probes for the candidate genes to test for a change in expression after exposure to salt water would be a powerful line of evidence concerning salt tolerance. Additionally, the total RNA could include genes that ddRADseq missed as many of these loci come from non-functional regions of the genome.

APPENDICES: Supplementary Tables and Figures

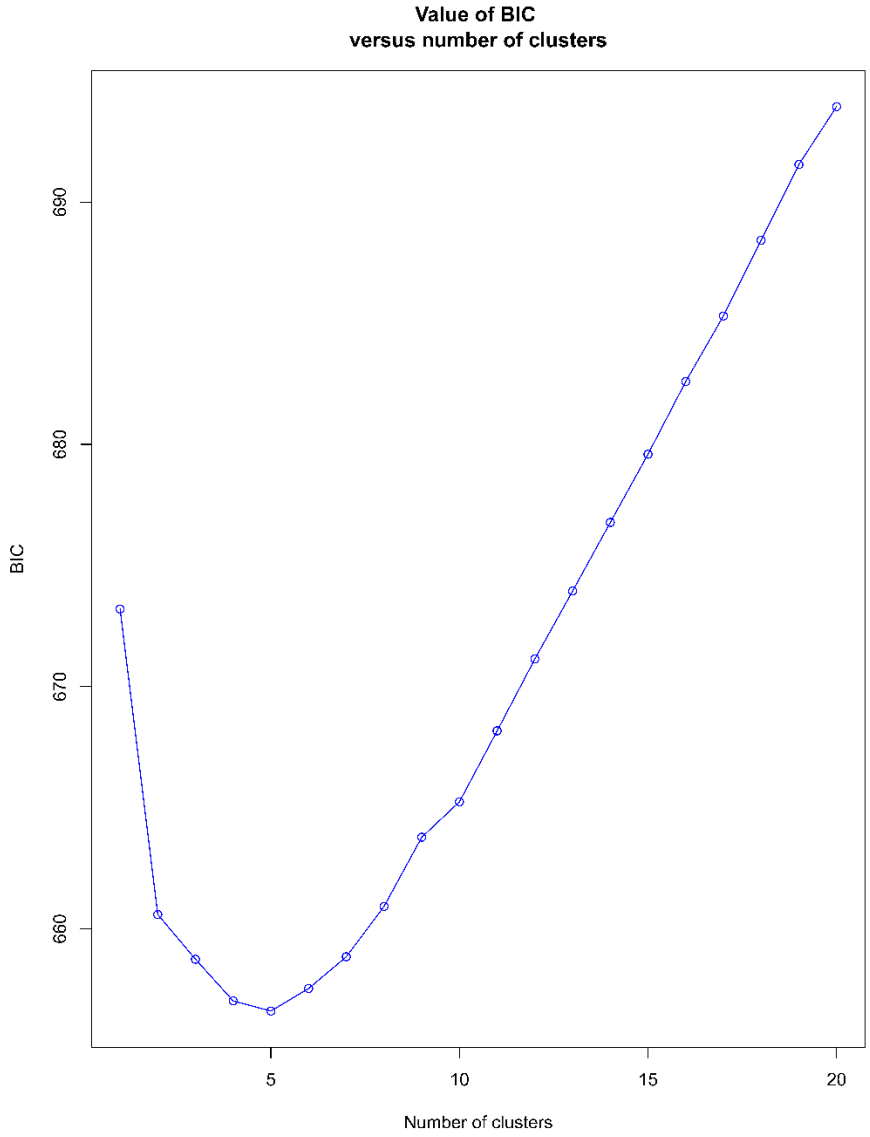

Figure S1: BIC score plot generated by K-means clustering in adegenet. Scores were 657.021, 656.6027, and 657.5342 for clusters 4, 5, and 6, respectively.

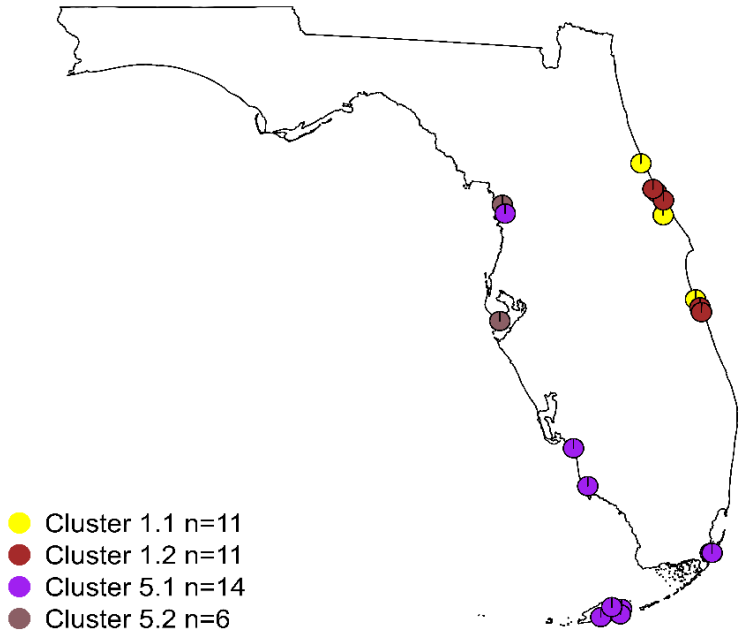

Figure S2: Subclusters of clusters 1 and 5 from the best supported DAPC.

Tables S1: List of all samples with CLP#, subspecific identification, sampling location, and which dataset they were analyzed in.

Sample	Latitude	Longitude	Species	Subspecies	Sanger	ddRADseq
CLP0954	28.55704	-81.2069	N. fasciata	pictiventris		Y
CLP0966	25.58333	-80.55	N. fasciata	pictiventris		Y
CLP0980	30.25403	-82.5128	N. fasciata	pictiventris		Y
CLP0983	30.97475	-84.058	N. fasciata	fasciata		Y
CLP1031	30.38562	-84.2322	N. fasciata	fasciata		Y
CLP1036	29.51715	-82.2227	N. fasciata	pictiventris		Y
CLP1038	29.51715	-82.2227	N. fasciata	pictiventris		Y
CLP1129	28.81816	-80.8601	N. clarkii	integrade		Y
CLP1130	28.81852	-80.86	N. clarkii	integrade		Y
CLP1135	28.81832	-80.86	N. clarkii	integrade		Y
CLP1137	28.81716	-80.8606	N. clarkii	integrade		Y
CLP1139	26.54261	-80.1033	N. fasciata	pictiventris		Y
CLP1140	29.4583	-82.4399	N. fasciata	pictiventris		Y
CLP1141	28.81858	-80.8594	N. clarkii	integrade		Y
CLP1147	28.81871	-80.86	N. clarkii	intergrade	Y	
CLP1148	28.81782	-80.8601	N. fasciata	integrade		Y
CLP1149	28.81782	-80.8601	N. clarkii	integrade		Y
CLP1177	27.71064	-82.6866	N. clarkii	compressicauda		Y
CLP1180	27.7106	-82.6868	N. clarkii	compressicauda	Y	Y
CLP1184	27.71101	-82.6861	N. clarkii	compressicauda	Y	
CLP1189	27.82373	-80.6058	N. fasciata	pictiventris		Y
CLP1190	27.85456	-80.4485	N. clarkii	compressicauda	Y	Y
CLP1191	27.93761	-80.4992	N. clarkii	compressicauda		Y
CLP1192	27.85422	-80.4493	N. clarkii	compressicauda		Y
CLP1194	27.85449	-80.4485	N. clarkii	compressicauda		Y
CLP1215	27.71065	-82.6866	N. fasciata	pictiventris		Y
CLP1217	27.71065	-82.6866	N. clarkii	compressicauda		Y
CLP1227	29.63154	-81.209	N. fasciata	pictiventris		Y
CLP1231	29.5298	-83.3771	N. clarkii	clarkii	Y	Y
CLP1232	30.02405	-84.3678	N. clarkii	clarkii	Y	
CLP1233	30.01765	-83.8711	N. clarkii	clarkii		Y
CLP1234	30.01745	-83.8708	N. clarkii	clarkii	Y	
CLP1235	30.01582	-83.869	N. clarkii	clarkii	Y	
CLP1237	30.01539	-83.8661	N. clarkii	clarkii	Y	
CLP1238	30.01562	-83.8631	N. clarkii	clarkii	Y	
CLP1239	30.01554	-83.8626	N. clarkii	clarkii	Y	
CLP1240	30.01553	-83.8619	N. clarkii	clarkii		Y
CLP1241	30.01514	-83.8607	N. clarkii	clarkii	Y	
CLP1242	30.01433	-83.8542	N. clarkii	clarkii	Y	
CLP1243	30.01553	-83.8619	N. clarkii	clarkii	Y	Y
CLP1244	30.01503	-83.8515	N. clarkii	clarkii	Y	
CLP1246	30.01505	-83.8512	N. clarkii	clarkii	Y	
CLP1248	30.0144	-83.8542	N. clarkii	clarkii	Y	
CLP1250	30.01595	-83.8692	N. clarkii	clarkii	Y	
CLP1251	30.01563	-83.8687	N. clarkii	clarkii	Y	Y
CLP1252	30.0144	-83.857	N. clarkii	clarkii	Y	
CLP1255	29.31543	-81.0871	N. fasciata	pictiventris		Y
CLP1288	28.43314	-80.6955	N. clarkii	compressicauda		Y
CLP1292	28.49332	-80.8887	N. fasciata	pictiventris	Y	
CLP1297	29.36071	-81.1096	N. fasciata	pictiventris		Y
CLP1311	28.92621	-82.6588	N. fasciata	pictiventris		Y
CLP1314	28.92627	-82.659	N. fasciata	pictiventris	Y	
CLP1315	28.92627	-82.659	N. fasciata	pictiventris		Y
CLP1316	28.92722	-82.659	N. fasciata	pictiventris		Y
CLP1318	28.83563	-82.6273	N. fasciata	pictiventris	Y	Y
CLP1323	25.32595	-80.7982	N. fasciata	pictiventris		Y
CLP1368	25.43726	-80.4607	N. fasciata	pictiventris		Y
CLP1371	25.40382	-80.5085	N. fasciata	pictiventris	Y	
CLP1375	28.42713	-81.4498	N. fasciata	pictiventris		Y

CLP1382	28.42713	-81.4498	N. fasciata	pictiventris	Y	
CLP1385	28.21387	-80.9116	N. fasciata	pictiventris		Y
CLP1394	27.93929	-82.407	N. fasciata	pictiventris		Y
CLP1395	28.23258	-82.5484	N. fasciata	pictiventris		Y
CLP1396	28.1893	-82.4267	N. fasciata	pictiventris	Y	
CLP1416	28.89742	-80.8528	N. fasciata	pictiventris		Y
CLP1420	28.92084	-80.8707	N. fasciata	pictiventris		Y
CLP1428	26.85369	-80.2964	N. fasciata	pictiventris	Y	Y
CLP1436	25.28372	-80.333	N. clarkii	compressicauda	Y	Y
CLP1441	30.53459	-82.3283	N. fasciata	pictiventris	Y	Y
CLP1447	28.92084	-80.8707	N. fasciata	pictiventris	Y	
CLP1450	30.29888	-87.4256	N. fasciata	fasciata	Y	Y
CLP1451	30.29888	-87.4256	N. fasciata	fasciata		Y
CLP1452	30.29888	-87.4256	N. fasciata	fasciata	Y	Y
CLP1453	30.9207	-87.0481	N. fasciata	fasciata	Y	Y
CLP1455	30.13719	-85.2015	N. fasciata	fasciata		Y
CLP1456	30.13719	-85.2015	N. fasciata	fasciata	Y	
CLP1457	30.13719	-85.2015	N. fasciata	fasciata		Y
CLP1458	30.01432	-83.8557	N. clarkii	clarkii	Y	
CLP1459	30.01432	-83.8557	N. clarkii	clarkii	Y	
CLP1460	30.01432	-83.8557	N. clarkii	clarkii	Y	
CLP1465	30.01432	-83.8557	N. clarkii	clarkii		Y
CLP1466	30.01432	-83.8557	N. clarkii	clarkii		Y
CLP1467	30.01432	-83.8557	N. clarkii	clarkii		Y
CLP1491	24.67119	-81.404	N. clarkii	compressicauda		Y
CLP1493	24.69779	-81.3262	N. clarkii	compressicauda		Y
CLP1499	25.28428	-80.314	N. clarkii	compressicauda		Y
CLP1500	25.28428	-80.314	N. clarkii	compressicauda		Y
CLP1506	29.64292	-81.6839	N. fasciata	pictiventris		Y
CLP1515	26.7406	-81.829	N. fasciata	pictiventris		Y
CLP1543	30.42347	-81.5642	N. fasciata	pictiventris		Y
CLP1544	30.42347	-81.5642	N. fasciata	pictiventris		Y
CLP1579	25.98289	-81.7039	N. clarkii	compressicauda		Y
CLP1584	27.46241	-82.6543	N. fasciata	pictiventris		Y
CLP1594	26.45463	-80.9799	N. fasciata	pictiventris		Y
CLP1610	30.13719	-85.2015	N. fasciata	fasciata		Y
CLP1672	24.69037	-81.4167	N. clarkii	compressicauda		Y
CLP1674	24.64147	-81.3391	N. clarkii	compressicauda		Y
CLP1700	29.41446	-83.2019	N. clarkii	clarkii		Y
CLP1701	29.40418	-83.2015	N. clarkii	clarkii		Y
CLP1702	29.39887	-83.2054	N. clarkii	clarkii		Y
CLP1704	24.61215	-81.5572	N. clarkii	compressicauda		Y
CLP1708	24.61215	-81.5572	N. clarkii	compressicauda		Y
CLP1709	24.61215	-81.5572	N. clarkii	compressicauda		Y
CLP1722	26.38027	-81.8635	N. clarkii	compressicauda		Y
CLPT02	29.05659	-80.9365	N. clarkii	taeniata	Y	
CLPT021	30.02401	-84.3678	N. clarkii	clarkii		Y
CLPT022	30.05335	-84.4098	N. fasciata	fasciata		Y
CLPT023	30.02401	-84.3678	N. clarkii	clarkii		Y
CLPT024	30.02401	-84.3678	N. clarkii	clarkii		Y
CLPT025	30.02401	-84.3678	N. clarkii	clarkii	Y	Y
CLPT028	30.02401	-84.3678	N. clarkii	clarkii	Y	
CLPT030	30.02401	-84.3678	N. clarkii	clarkii		Y
CLPT040	30.16922	-84.2471	N. fasciata	fasciata		Y
CLPT041	30.12648	-84.2789	N. fasciata	fasciata		Y
CLPT056	29.06038	-80.937	N. clarkii	taeniata	Y	Y
CLPT058	29.06038	-80.937	N. clarkii	taeniata	Y	
CLPT059	28.97529	-80.858	N. clarkii	taeniata		Y
CLPT060	29.06038	-80.937	N. clarkii	taeniata	Y	
CLPT064	24.69791	-81.342	N. clarkii	compressicauda	Y	
CLPT086	29.06038	-80.937	N. clarkii	taeniata	Y	
CLPT088	28.97529	-80.858	N. clarkii	taeniata	Y	
CLPT089	28.97529	-80.858	N. clarkii	taeniata	Y	

CLPT093	28.97529	-80.858	N. clarkii	taeniata	Y	
CLP094	28.97529	-80.858	N. clarkii	taeniata	Y	
CLPT095	28.97529	-80.858	N. clarkii	taeniata	Y	
CLPT096	28.97529	-80.858	N. clarkii	taeniata	Y	
CLPT097	28.97529	-80.858	N. clarkii	taeniata		Y
CLPT098	28.97529	-80.858	N. clarkii	taeniata	Y	
CLPT105	30.86353	-86.9041	N. fasciata	fasciata	Y	Y
CLPT106	30.29888	-87.4256	N. fasciata	fasciata		Y
CLPT108	25.86368	-81.1006	N. fasciata	pictiventris		Y
CLPT109	24.71717	-81.4355	N. clarkii	compressicauda	Y	Y
CLPT116	28.54229	-80.9456	N. fasciata	pictiventris		Y
CLPT117	28.54229	-80.9456	N. fasciata	pictiventris		Y
CLPT123	28.97345	-80.8575	N. clarkii	taeniata		Y
CLPT132	28.97345	-80.8575	N. clarkii	taeniata		Y
CLPT141	27.80393	-80.433	N. clarkii	taeniata		Y
CLPT240	27.80393	-80.433	N. clarkii	taeniata		Y
CLPT242	27.80393	-80.433	N. clarkii	taeniata		Y
CLPT416	29.09295	-80.9744	N. clarkii	taeniata		Y
CLPT417	29.09295	-80.9744	N. clarkii	taeniata		Y
KW0217	29.89547	-84.6426	N. fasciata	fasciata	Y	
KW0382	30.23487	-84.6757	N. fasciata	fasciata	Y	
KW0427	30.61672	-84.0213	N. fasciata	fasciata	Y	
KW0469	30.53654	-84.2796	N. fasciata	fasciata	Y	
KW0470	30.38807	-84.7818	N. fasciata	fasciata	Y	
KW0475	30.04918	-84.3695	N. clarkii	clarkii	Y	
KW1120	30.52455	-85.8438	N. fasciata	fasciata	Y	
CLP1282	28.4642	-80.9533	Thamnophis sauritus		Y	Y
CLP1284	28.50382	-80.9564	Thamnophis sirtalis		Y	
CLP1351	39.87683	-80.7235	Nerodia sipedon		Y	
CLP1352	39.87683	-80.7235	Nerodia sipedon		Y	
CLP1353	39.87683	-80.7235	Nerodia sipedon		Y	

Table S2: List of samples used in BEAST with CLP#, subspecific identification, sampling location, and which genes were successfully (Y) amplified for them.

Sample	Species	Subspecies	cyt-b	ND1	ND4	TATA	PRLR	M	E
CLP1147	N. clarkii	intergrade	Y	Y	Y	Y	Y	Y	Y
CLP1180	N. clarkii	compressicauda	Y	Y	Y	Y	Y	Y	Y
CLP1184	N. clarkii	compressicauda	Y	Y	Y	Y	Y	Y	Y
CLP1190	N. clarkii	compressicauda	Y	Y	Y	Y	Y	Y	Y
CLP1231	N. clarkii	clarkii	Y	Y	Y	Y	Y	Y	Y
CLP1232	N. clarkii	clarkii	Y	Y	Y	Y	Y	Y	Y
CLP1234	N. clarkii	clarkii	Y	Y	Y	Y	Y	Y	Y
CLP1235	N. clarkii	clarkii	Y	Y	Y	Y	Y	Y	Y
CLP1237	N. clarkii	clarkii	Y	Y	Y	Y	Y	Y	Y
CLP1238	N. clarkii	clarkii	Y	Y	Y	Y	Y	Y	Y
CLP1239	N. clarkii	clarkii	Y	Y	Y	Y	Y	Y	Y
CLP1241	N. clarkii	clarkii	Y	Y	Y	Y	Y	Y	Y
CLP1242	N. clarkii	clarkii	Y	Y	Y	Y	Y	Y	Y
CLP1243	N. clarkii	clarkii	Y	Y	Y	Y	Y	Y	Y
CLP1244	N. clarkii	clarkii	Y	Y	Y	Y	Y	Y	Y
CLP1246	N. clarkii	clarkii	Y	Y	Y	Y	Y	Y	Y
CLP1248	N. clarkii	clarkii	Y	Y	Y	Y	Y	Y	Y
CLP1250	N. clarkii	clarkii	Y	Y	Y	Y	Y	Y	Y
CLP1251	N. clarkii	clarkii	Y	Y	Y	Y	Y	Y	Y
CLP1252	N. clarkii	clarkii	Y	Y	Y	Y	Y	Y	Y
CLP1282	Thamnophis sauritus		Y	Y	Y	Y	Y	Y	Y
CLP1284	Thamnophis sirtalis		Y	Y	Y	Y	Y	Y	Y
CLP1292	N. fasciata	pictiventris	Y	Y	Y	Y	Y	Y	Y
CLP1314	N. fasciata	pictiventris	Y	Y	Y	Y	Y	Y	Y
CLP1318	N. fasciata	pictiventris	Y	Y	Y	Y	Y	Y	Y
CLP1351	Nerodia sipedon		Y	Y	Y				
CLP1352	Nerodia sipedon		Y	Y	Y	Y			
CLP1353	Nerodia sipedon		Y	Y	Y	Y	Y	Y	
CLP1371	N. fasciata	pictiventris	Y	Y	Y	Y	Y	Y	Y
CLP1382	N. fasciata	pictiventris	Y	Y	Y	Y	Y	Y	Y
CLP1396	N. fasciata	pictiventris	Y	Y	Y	Y	Y	Y	Y
CLP1428	N. fasciata	pictiventris	Y	Y	Y	Y	Y	Y	Y
CLP1436	N. clarkii	compressicauda	Y	Y	Y	Y	Y	Y	Y
CLP1441	N. fasciata	pictiventris	Y	Y	Y	Y	Y	Y	Y
CLP1447	N. fasciata	pictiventris	Y	Y	Y	Y	Y	Y	Y
CLP1450	N. fasciata	fasciata	Y	Y	Y	Y	Y	Y	Y
CLP1452	N. fasciata	fasciata	Y	Y	Y	Y	Y	Y	Y
CLP1453	N. fasciata	fasciata	Y	Y	Y	Y	Y	Y	Y
CLP1456	N. fasciata	fasciata	Y	Y	Y	Y	Y	Y	Y
CLP1458	N. clarkii	clarkii	Y	Y	Y	Y	Y	Y	Y
CLP1459	N. clarkii	clarkii	Y	Y	Y	Y	Y	Y	Y
CLP1460	N. clarkii	clarkii	Y	Y	Y		Y	Y	Y
CLPT02	N. clarkii	taeniata	Y	Y	Y	Y	Y	Y	Y
CLPT025	N. clarkii	clarkii	Y	Y	Y	Y	Y	Y	Y
CLPT028	N. clarkii	clarkii	Y	Y	Y	Y	Y	Y	Y
CLPT056	N. clarkii	taeniata	Y	Y	Y	Y	Y	Y	Y
CLPT057	N. clarkii	taeniata	Y	Y	Y	Y	Y	Y	Y
CLPT058	N. clarkii	taeniata	Y	Y	Y	Y	Y	Y	Y
CLPT060	N. clarkii	taeniata	Y	Y	Y	Y	Y	Y	Y
CLPT064	N. clarkii	compressicauda	Y	Y	Y	Y		Y	Y
CLPT086	N. clarkii	taeniata	Y	Y	Y	Y	Y	Y	Y
CLPT088	N. clarkii	taeniata	Y	Y	Y		Y	Y	Y
CLPT089	N. clarkii	taeniata	Y	Y	Y	Y	Y	Y	Y
CLPT093	N. clarkii	taeniata	Y	Y	Y	Y	Y	Y	Y
CLPT094	N. clarkii	taeniata	Y	Y	Y	Y	Y	Y	Y
CLPT095	N. clarkii	taeniata	Y	Y	Y	Y	Y	Y	Y
CLPT096	N. clarkii	taeniata	Y	Y	Y	Y	Y	Y	Y
CLPT098	N. clarkii	taeniata	Y	Y	Y	Y	Y	Y	Y
CLPT105	N. fasciata	fasciata	Y	Y	Y	Y	Y	Y	Y

CLPT109	N. clarkii	compressicauda	Y	Y	Y			Y	Y
KW0217	N. fasciata	fasciata	Y	Y	Y	Y	Y	Y	Y
KW0382	N. fasciata	fasciata	Y	Y	Y	Y	Y	Y	Y
KW0427	N. fasciata	fasciata	Y	Y	Y			Y	
KW0469	N. fasciata	fasciata	Y	Y	Y	Y	Y	Y	Y
KW0470	N. fasciata	fasciata	Y	Y	Y	Y	Y	Y	Y
KW0475	N. clarkii	clarkii	Y	Y	Y	Y		Y	Y
KW1120	N. fasciata	fasciata	Y	Y	Y				

Table S3: List of all candidate genes found among 5 populations in 50kb windows.

Locus	Scaffold	Start	Stop	Gene
99749	NW_013658110.1	284808	319852	ACAD9
75077	NW_013658903.1	134137	137752	ACTL7A
71697	NW_013657748.1	1637921	1671121	ADCY7
75077	NW_013658903.1	137837	149714	APTX
42991	NW_013658321.1	469878	548897	ARSB
153743	NW_013660747.1	16567	23081	ATP5G1
161731	NW_013658286.1	303291	468821	BRE
163572	NW_013658815.1	56491	94116	BSG
18127	NW_013659340.1	22604	74061	CACNB1
18127	NW_013659340.1	86668	115622	CACNB3
155424	NW_013658446.1	495713	538155	CACNG5
130944	NW_013658097.1	75322	145910	CARM1
128558	NW_013658423.1	50497	195185	CATSPER4
101327	NW_013657697.1	1405080	1414146	CCDC42
160175	NW_013657697.1	1405080	1414146	CCDC42
160811	NW_013657951.1	891178	911080	CCND3
1595	NW_013657732.1	1635479	1686592	CNPY1
51454	NW_013657928.1	656649	663912	DCTN6
42991	NW_013658321.1	428922	467117	DMGDH
149267	NW_013658063.1	98654	108308	DMRT2
75077	NW_013658903.1	150702	157622	DNAJA1
17166	NW_013658362.1	508536	540670	DOCK1
129077	NW_013658622.1	85346	275257	EEFSEC
104979	NW_013657947.1	395292	435354	EFCAB11
131519	NW_013658936.1	131263	293809	ELAVL4
58505	NW_013657769.1	627814	721867	EML5
137854	NW_013657869.1	677817	748463	ERBB2
37913	NW_013657784.1	773231	1107871	ERC2
14258	NW_013658374.1	17582	89638	FAM120A
133078	NW_013657682.1	2262227	2280582	FBXW9
102417	NW_013658030.1	2031	202973	FHOD3
33811	NW_013658060.1	327963	328523	FLNB
95827	NW_013658391.1	216353	263023	FLVCR2
53953	NW_013657947.1	680772	753224	FYN
28576	NW_013657849.1	571195	613019	GMPS
71697	NW_013657748.1	1582667	1584654	GPR18
71697	NW_013657748.1	1554427	1561912	GPR183
133078	NW_013657682.1	2315138	2317632	IER2
75077	NW_013658903.1	91742	133984	IKBKAP
28587	NW_013657773.1	670737	680160	IL27
115228	NW_013658023.1	362015	380715	IP6K3
162273	NW_013659976.1	58298	75101	IRF3
115228	NW_013658023.1	394802	532109	ITPR3
95827	NW_013658391.1	317112	318118	JDP2
99749	NW_013658110.1	347467	363567	KIAA1257
11736	NW_013657681.1	2075753	2094152	LOC106538175
116293	NW_013657681.1	2075753	2094152	LOC106538175
127585	NW_013657869.1	38145	76183	LOC106538847
127585	NW_013657869.1	16301	35189	LOC106538857
153489	NW_013657893.1	758784	787312	LOC106539243
118495	NW_013657901.1	691800	705282	LOC106539371
155530	NW_013657917.1	571940	572915	LOC106539627
155530	NW_013657917.1	599393	601684	LOC106539628
160811	NW_013657951.1	989864	1008556	LOC106540196
11736	NW_013657681.1	1985803	2036784	LOC106540201
116293	NW_013657681.1	1985803	2036784	LOC106540201
97335	NW_013657983.1	152051	161741	LOC106540669
101327	NW_013657697.1	1356468	1408363	LOC106540689
160175	NW_013657697.1	1356468	1408363	LOC106540689
126796	NW_013658014.1	435475	552577	LOC106541029

115228	NW_013658023.1	389773	393447	LOC106541200
102417	NW_013658030.1	205870	370690	LOC106541291
166362	NW_013658033.1	761661	784599	LOC106541319
166362	NW_013658033.1	736033	754783	LOC106541320
17369	NW_013658038.1	35112	108049	LOC106541369
33811	NW_013658060.1	274166	280994	LOC106541679
99366	NW_013657703.1	825988	826509	LOC106542502
99366	NW_013657703.1	715582	749467	LOC106542660
35666	NW_013658253.1	419811	448989	LOC106543809
42991	NW_013658321.1	411466	427862	LOC106544491
42991	NW_013658321.1	402111	411561	LOC106544493
145198	NW_013657711.1	888568	922518	LOC106544559
145198	NW_013657711.1	856985	876551	LOC106544579
17166	NW_013658362.1	409527	422167	LOC106544792
122078	NW_013658379.1	95483	97643	LOC106544978
122078	NW_013658379.1	98851	117187	LOC106544979
122078	NW_013658379.1	100124	103159	LOC106544982
122078	NW_013658379.1	86771	95323	LOC106544983
54449	NW_013658428.1	425731	432421	LOC106545375
137877	NW_013658504.1	515479	523076	LOC106546030
133078	NW_013657682.1	2207648	2226548	LOC106546100
134435	NW_013658575.1	229181	241877	LOC106546678
9473	NW_013658626.1	334301	454833	LOC106547029
149366	NW_013658629.1	38646	96124	LOC106547055
105067	NW_013658674.1	458169	459777	LOC106547368
58290	NW_013658736.1	2252	33599	LOC106547825
17508	NW_013658744.1	137162	318808	LOC106547899
33393	NW_013658768.1	191940	214542	LOC106548061
75567	NW_013658831.1	288782	375180	LOC106548434
75616	NW_013658838.1	1110	101893	LOC106548477
30004	NW_013658846.1	281991	299052	LOC106548535
30004	NW_013658846.1	318598	359304	LOC106548536
75077	NW_013658903.1	67945	84769	LOC106548871
153479	NW_013658954.1	163449	179926	LOC106549181
153479	NW_013658954.1	236661	281252	LOC106549183
153479	NW_013658954.1	182379	223011	LOC106549185
153479	NW_013658954.1	227683	233724	LOC106549186
153479	NW_013658954.1	110543	160867	LOC106549192
153479	NW_013658954.1	152246	155449	LOC106549193
126425	NW_013659370.1	82992	86558	LOC106551292
57340	NW_013659433.1	44103	48330	LOC106551555
57340	NW_013659433.1	88492	116022	LOC106551561
57340	NW_013659433.1	32377	34990	LOC106551562
154511	NW_013657742.1	867652	903895	LOC106552693
152506	NW_013657742.1	867652	903895	LOC106552693
9573	NW_013657743.1	1599953	1646669	LOC106552982
9573	NW_013657743.1	1541135	1575313	LOC106553011
162273	NW_013659976.1	85087	126240	LOC106553400
29865	NW_013660795.1	448	40884	LOC106555249
53632	NW_013660996.1	53542	57206	LOC106555576
58505	NW_013657769.1	718731	736410	LOC106557020
28587	NW_013657773.1	700004	722203	LOC106557084
28587	NW_013657773.1	733990	774634	LOC106557085
103667	NW_013657775.1	774702	796540	LOC106557137
124673	NW_013657778.1	1001132	1120794	LOC106557182
148947	NW_013657779.1	1144301	1293792	LOC106557198
129757	NW_013657782.1	134916	308221	LOC106557252
95887	NW_013657788.1	1357053	1399281	LOC106557369
63450	NW_013658178.1	300766	407662	LRRC16A
8917	NW_013657774.1	1172038	1174804	LRRN2
127585	NW_013657869.1	88519	145673	MED1
97335	NW_013657983.1	168046	175409	MIS18A
75077	NW_013658903.1	47287	59991	MNT

161731	NW_013658286.1	234183	238456	MRPL33
125485	NW_013659206.1	213960	230166	MRPS33
101327	NW_013657697.1	1490293	1568687	MYH10
160175	NW_013657697.1	1490293	1568687	MYH10
165168	NW_013658480.1	341650	366563	NCMAP
166362	NW_013658033.1	580424	672300	NCOA3
165168	NW_013658480.1	410107	433948	NIPAL3
17166	NW_013658362.1	424104	449726	NOS3
153479	NW_013658954.1	202164	219892	NT5DC2
33393	NW_013658768.1	69739	136424	P4HA2
137854	NW_013657869.1	641639	665590	PGAP3
28576	NW_013657849.1	665523	768047	PLCH1
57340	NW_013659433.1	120679	135147	PLEKHM3
137854	NW_013657869.1	620601	632954	PNMT
134435	NW_013658575.1	247573	313815	PPIP5K2
25322	NW_013657869.1	463951	523717	PPP1R1B
103667	NW_013657775.1	810968	845395	RAB11FIP4
54449	NW_013658428.1	462472	464322	RASSF10
161731	NW_013658286.1	216849	303133	RBKS
51454	NW_013657928.1	711715	776787	RBPMS
165168	NW_013658480.1	375982	393718	RCAN3
163572	NW_013658815.1	144271	204338	RPRD1A
75077	NW_013658903.1	64466	77508	RPS5
129077	NW_013658622.1	59614	80745	RUVBL1
135864	NW_013657853.1	280597	444358	SEPT9
126796	NW_013658014.1	557374	560759	SIVA1
53079	NW_013658634.1	182421	236805	SLC13A3
53079	NW_013658634.1	275903	299125	SLC2A10
28576	NW_013657849.1	633283	648038	SLC33A1
24957	NW_013659743.1	64969	132559	SLC9A3
128484	NW_013658784.1	228981	293535	SLC9A8
153743	NW_013660747.1	54504	63890	SNF8
138705	NW_013657686.1	1177784	1292336	SNX13
165168	NW_013658480.1	316356	335195	SRRM1
157551	NW_013658450.1	467322	536826	ST6GALNAC3
160811	NW_013657951.1	970800	983501	TAF8
33393	NW_013658768.1	140056	185502	TBK1
137854	NW_013657869.1	611214	613925	TCAP
137877	NW_013658504.1	474155	480124	TCF19
118495	NW_013657901.1	615637	691539	TCF7L1
24957	NW_013659743.1	144071	180461	TMEM245
133078	NW_013657682.1	2226722	2260968	TNPO2
53079	NW_013658634.1	259957	261948	TP53RK
103667	NW_013657775.1	803173	803244	TRNAT-CGU
71697	NW_013657748.1	1511148	1600873	UBAC2
153743	NW_013660747.1	25741	50237	UBE2Z
122451	NW_013657703.1	1789480	1852757	UNC79
97335	NW_013657983.1	103270	147761	URB1
133078	NW_013657682.1	2286004	2293732	WDR83
133078	NW_013657682.1	2293955	2303694	WDR83OS
130944	NW_013658097.1	55624	82044	YIPF2
58505	NW_013657769.1	579071	623276	ZC3H14

Table S4: List of all candidate genes found among 5 populations in 20kb windows.

Locus	Scaffold	Start	Stop	Gene
99749	NW_013658110.1	284808	319852	ACAD9
42991	NW_013658321.1	469878	548897	ARSB
18127	NW_013659340.1	22604	74061	CACNB1
155424	NW_013658446.1	495713	538155	CACNG5
130944	NW_013658097.1	75322	145910	CARM1
51454	NW_013657928.1	656649	663912	DCTN6
42991	NW_013658321.1	428922	467117	DMGDH
149267	NW_013658063.1	98654	108308	DMRT2
129077	NW_013658622.1	85346	275257	EEFSEC
131519	NW_013658936.1	131263	293809	ELAVL4
58505	NW_013657769.1	627814	721867	EML5
37913	NW_013657784.1	773231	1107871	ERC2
14258	NW_013658374.1	17582	89638	FAM120A
133078	NW_013657682.1	2262227	2280582	FBXW9
102417	NW_013658030.1	2031	202973	FHOD3
33811	NW_013658060.1	327963	328523	FLNB
53953	NW_013657947.1	680772	753224	FYN
71697	NW_013657748.1	1582667	1584654	GPR18
75077	NW_013658903.1	91742	133984	IKBKAP
115228	NW_013658023.1	394802	532109	ITPR3
99749	NW_013658110.1	347467	363567	KIAA1257
127585	NW_013657869.1	38145	76183	LOC106538847
155530	NW_013657917.1	599393	601684	LOC106539628
11736	NW_013657681.1	1985803	2036784	LOC106540201
116293	NW_013657681.1	1985803	2036784	LOC106540201
97335	NW_013657983.1	152051	161741	LOC106540669
17369	NW_013658038.1	35112	108049	LOC106541369
35666	NW_013658253.1	419811	448989	LOC106543809
145198	NW_013657711.1	856985	876551	LOC106544579
122078	NW_013658379.1	98851	117187	LOC106544979
9473	NW_013658626.1	334301	454833	LOC106547029
149366	NW_013658629.1	38646	96124	LOC106547055
17508	NW_013658744.1	137162	318808	LOC106547899
75567	NW_013658831.1	288782	375180	LOC106548434
75616	NW_013658838.1	1110	101893	LOC106548477
30004	NW_013658846.1	281991	299052	LOC106548535
75077	NW_013658903.1	67945	84769	LOC106548871
153479	NW_013658954.1	163449	179926	LOC106549181
153479	NW_013658954.1	182379	223011	LOC106549185
57340	NW_013659433.1	88492	116022	LOC106551561
9573	NW_013657743.1	1541135	1575313	LOC106553011
162273	NW_013659976.1	85087	126240	LOC106553400
29865	NW_013660795.1	448	40884	LOC106555249
28587	NW_013657773.1	700004	722203	LOC106557084
28587	NW_013657773.1	733990	774634	LOC106557085
103667	NW_013657775.1	774702	796540	LOC106557137
148947	NW_013657779.1	1144301	1293792	LOC106557198
129757	NW_013657782.1	134916	308221	LOC106557252
95887	NW_013657788.1	1357053	1399281	LOC106557369
63450	NW_013658178.1	300766	407662	LRRC16A
8917	NW_013657774.1	1172038	1174804	LRRN2
165168	NW_013658480.1	341650	366563	NCMAP
17166	NW_013658362.1	424104	449726	NOS3
153479	NW_013658954.1	202164	219892	NT5DC2
137854	NW_013657869.1	641639	665590	PGAP3
28576	NW_013657849.1	665523	768047	PLCH1
134435	NW_013658575.1	247573	313815	PPIP5K2
103667	NW_013657775.1	810968	845395	RAB11FIP4
54449	NW_013658428.1	462472	464322	RASSF10
161731	NW_013658286.1	216849	303133	RBKS

165168	NW_013658480.1	375982	393718	RCAN3
163572	NW_013658815.1	144271	204338	RPRD1A
129077	NW_013658622.1	59614	80745	RUVBL1
53079	NW_013658634.1	182421	236805	SLC13A3
28576	NW_013657849.1	633283	648038	SLC33A1
128484	NW_013658784.1	228981	293535	SLC9A8
153743	NW_013660747.1	54504	63890	SNF8
138705	NW_013657686.1	1177784	1292336	SNX13
157551	NW_013658450.1	467322	536826	ST6GALNAC3
33393	NW_013658768.1	140056	185502	TBK1
137877	NW_013658504.1	474155	480124	TCF19
118495	NW_013657901.1	615637	691539	TCF7L1
24957	NW_013659743.1	144071	180461	TMEM245
133078	NW_013657682.1	2226722	2260968	TNPO2
103667	NW_013657775.1	803173	803244	TRNAT-CGU
71697	NW_013657748.1	1511148	1600873	UBAC2
153743	NW_013660747.1	25741	50237	UBE2Z
122451	NW_013657703.1	1789480	1852757	UNC79
97335	NW_013657983.1	103270	147761	URB1
133078	NW_013657682.1	2286004	2293732	WDR83

Table S5: List of all candidate genes found between inland and coastal populations in 50kb windows.

Locus	Scaffold	Start	Stop	Gene
99749	NW_013658110.1	284808	319852	ACAD9
124225	NW_013659323.1	51761	60979	ADM2
117130	NW_013658316.1	165751	229214	ANKRD52
113933	NW_013659655.1	60327	165985	ANKS1A
114300	NW_013659681.1	113984	116331	APOOL
109968	NW_013657847.1	189002	210849	ARF1
66371	NW_013659131.1	176091	186490	ARL14
75637	NW_013658189.1	78302	82212	ARL4C
42991	NW_013658321.1	469878	548897	ARSB
13571	NW_013657827.1	877980	893976	BAAT
138383	NW_013658122.1	257826	263562	BID
144830	NW_013658145.1	59143	63965	BMP10
34812	NW_013657880.1	783828	992525	BTBD11
51321	NW_013658047.1	73317	84037	BTBD18
13007	NW_013657948.1	514293	540383	C1QTNF4
18127	NW_013659340.1	22604	74061	CACNB1
18127	NW_013659340.1	86668	115622	CACNB3
155424	NW_013658446.1	495713	538155	CACNG5
130944	NW_013658097.1	75322	145910	CARM1
107727	NW_013658097.1	493722	541872	CASQ2
75359	NW_013659469.1	67262	74088	CBLN2
101327	NW_013657697.1	1405080	1414146	CCDC42
160175	NW_013657697.1	1405080	1414146	CCDC42
13007	NW_013657948.1	578948	628470	CELF1
107727	NW_013658097.1	447424	480637	CHD1L
58800	NW_013659234.1	11472	57412	CNNM2
34812	NW_013657880.1	707536	730372	CRY1
51454	NW_013657928.1	656649	663912	DCTN6
42991	NW_013658321.1	428922	467117	DMGDH
35194	NW_013657712.1	967788	1002742	DNAJC17
75219	NW_013659036.1	119134	302357	DOCK3
104979	NW_013657947.1	395292	435354	EFCAB11
58505	NW_013657769.1	627814	721867	EML5
37913	NW_013657784.1	773231	1107871	ERC2
14258	NW_013658374.1	17582	89638	FAM120A
13007	NW_013657948.1	541022	546587	FAM180B
90790	NW_013658547.1	324922	461529	FBLN2
170144	NW_013658547.1	324922	461529	FBLN2
149984	NW_013659822.1	55644	66305	FBXL15
33811	NW_013658060.1	327963	328523	FLNB
95827	NW_013658391.1	216353	263023	FLVCR2
109968	NW_013657847.1	276065	327318	GJC2
37859	NW_013658334.1	235313	312545	GLS
28576	NW_013657849.1	571195	613019	GMPS
109968	NW_013657847.1	244106	270199	GUK1
61755	NW_013657873.1	271367	360993	IKZF1
28587	NW_013657773.1	670737	680160	IL27
95827	NW_013658391.1	317112	318118	JDP2
13007	NW_013657948.1	565058	571596	KBTBD4
35073	NW_013657934.1	464371	529209	KIAA0319L
99749	NW_013658110.1	347467	363567	KIAA1257
66371	NW_013659131.1	132535	160005	KPNA4
11736	NW_013657681.1	2075753	2094152	LOC106538175
116293	NW_013657681.1	2075753	2094152	LOC106538175
109968	NW_013657847.1	221759	235281	LOC106538482
109968	NW_013657847.1	218934	221758	LOC106538490
153489	NW_013657893.1	758784	787312	LOC106539243
116327	NW_013657916.1	71038	85101	LOC106539612
155530	NW_013657917.1	571940	572915	LOC106539627

155530	NW_013657917.1	599393	601684	LOC106539628
13007	NW_013657948.1	581012	590532	LOC106540121
11736	NW_013657681.1	1985803	2036784	LOC106540201
116293	NW_013657681.1	1985803	2036784	LOC106540201
160206	NW_013657680.1	3888059	3896249	LOC106540576
101327	NW_013657697.1	1356468	1408363	LOC106540689
160175	NW_013657697.1	1356468	1408363	LOC106540689
126796	NW_013658014.1	435475	552577	LOC106541029
166362	NW_013658033.1	761661	784599	LOC106541319
166362	NW_013658033.1	736033	754783	LOC106541320
17369	NW_013658038.1	35112	108049	LOC106541369
51321	NW_013658047.1	82779	92152	LOC106541468
51321	NW_013658047.1	967	13539	LOC106541481
51321	NW_013658047.1	20562	21485	LOC106541482
51321	NW_013658047.1	33488	34454	LOC106541483
51321	NW_013658047.1	42365	43506	LOC106541484
33811	NW_013658060.1	274166	280994	LOC106541679
99366	NW_013657703.1	825988	826509	LOC106542502
99366	NW_013657703.1	715582	749467	LOC106542660
144830	NW_013658145.1	28064	51776	LOC106542715
54007	NW_013657706.1	511973	768614	LOC106543426
42991	NW_013658321.1	411466	427862	LOC106544491
42991	NW_013658321.1	402111	411561	LOC106544493
116778	NW_013657711.1	319222	370377	LOC106544530
116778	NW_013657711.1	373444	374157	LOC106544538
37859	NW_013658334.1	132075	178534	LOC106544605
53393	NW_013658342.1	64334	98538	LOC106544652
116778	NW_013657711.1	279478	303027	LOC106544659
116778	NW_013657711.1	287313	290838	LOC106544667
54449	NW_013658428.1	425731	432421	LOC106545375
137877	NW_013658504.1	515479	523076	LOC106546030
170144	NW_013658547.1	516439	547898	LOC106546411
9473	NW_013658626.1	334301	454833	LOC106547029
105067	NW_013658674.1	458169	459777	LOC106547368
89634	NW_013658724.1	99110	100303	LOC106547754
75567	NW_013658831.1	288782	375180	LOC106548434
75616	NW_013658838.1	1110	101893	LOC106548477
75219	NW_013659036.1	137031	144577	LOC106549632
12212	NW_013659313.1	158277	160574	LOC106551017
12212	NW_013659313.1	227239	263525	LOC106551018
126425	NW_013659370.1	82992	86558	LOC106551292
114300	NW_013659681.1	128349	138975	LOC106552540
114300	NW_013659681.1	100171	110519	LOC106552541
114300	NW_013659681.1	119730	124902	LOC106552543
114300	NW_013659681.1	90581	90881	LOC106552544
154511	NW_013657742.1	867652	903895	LOC106552693
152506	NW_013657742.1	867652	903895	LOC106552693
9573	NW_013657743.1	1599953	1646669	LOC106552982
9573	NW_013657743.1	1541135	1575313	LOC106553011
62592	NW_013659839.1	99737	105311	LOC106553035
127801	NW_013660529.1	5062	26978	LOC106554777
127801	NW_013660529.1	55367	91113	LOC106554778
58505	NW_013657769.1	718731	736410	LOC106557020
28587	NW_013657773.1	700004	722203	LOC106557084
28587	NW_013657773.1	733990	774634	LOC106557085
103667	NW_013657775.1	774702	796540	LOC106557137
148947	NW_013657779.1	1144301	1293792	LOC106557198
129757	NW_013657782.1	134916	308221	LOC106557252
91274	NW_013657921.1	661362	683384	MATN4
109968	NW_013657847.1	239046	242278	MRPL55
138383	NW_013658122.1	282132	320896	MTR
58800	NW_013659234.1	61338	69545	MYCL
101327	NW_013657697.1	1490293	1568687	MYH10

160175	NW_013657697.1	1490293	1568687	MYH10
35073	NW_013657934.1	446883	458338	NCDN
165168	NW_013658480.1	341650	366563	NCMAP
166362	NW_013658033.1	580424	672300	NCOA3
13007	NW_013657948.1	552213	564892	NDUFS3
165168	NW_013658480.1	410107	433948	NIPAL3
28576	NW_013657849.1	665523	768047	PLCH1
25322	NW_013657869.1	463951	523717	PPP1R1B
13007	NW_013657948.1	571697	578786	PTPMT1
103667	NW_013657775.1	810968	845395	RAB11FIP4
13007	NW_013657948.1	629264	649622	RAPSN
54449	NW_013658428.1	462472	464322	RASSF10
91274	NW_013657921.1	745197	764815	RBPJL
51454	NW_013657928.1	711715	776787	RBPMS
165168	NW_013658480.1	375982	393718	RCAN3
138383	NW_013658122.1	349004	607375	RYR2
75637	NW_013658189.1	4768	54830	SH3BP4
126796	NW_013658014.1	557374	560759	SIVA1
53079	NW_013658634.1	182421	236805	SLC13A3
6520	NW_013657967.1	823266	850000	SLC15A1
53079	NW_013658634.1	275903	299125	SLC2A10
28576	NW_013657849.1	633283	648038	SLC33A1
117130	NW_013658316.1	243460	265528	SLC39A5
128484	NW_013658784.1	228981	293535	SLC9A8
66371	NW_013659131.1	59791	107139	SMC4
35194	NW_013657712.1	1028172	1073048	SPINT1
165168	NW_013658480.1	316356	335195	SRRM1
37859	NW_013658334.1	160940	214566	STK17A
137877	NW_013658504.1	474155	480124	TCF19
128483	NW_013657703.1	331664	469167	TCF7
113933	NW_013659655.1	165461	180308	TCP11
89634	NW_013658724.1	86315	119836	TNFSF10
53079	NW_013658634.1	259957	261948	TP53RK
66371	NW_013659131.1	114006	119116	TRIM59
58800	NW_013659234.1	74291	89801	TRIT1
103667	NW_013657775.1	803173	803244	TRNAT-CGU
122451	NW_013657703.1	1789480	1852757	UNC79
100831	NW_013658526.1	108453	224728	WNK1
130944	NW_013658097.1	55624	82044	YIPF2
53393	NW_013658342.1	111876	132178	ZC3H12B
58505	NW_013657769.1	579071	623276	ZC3H14
117459	NW_013658747.1	12349	85302	ZNF521
114300	NW_013659681.1	146099	168299	ZNF711
13571	NW_013657827.1	897869	919426	ZP1

Table S6: List of all candidate genes found between inland and coastal populations in 20kb windows.

Locus	Scaffold	Start	Stop	Gene
99749	NW_013658110.1	284808	319852	ACAD9
117130	NW_013658316.1	165751	229214	ANKRD52
114300	NW_013659681.1	113984	116331	APOOL
42991	NW_013658321.1	469878	548897	ARSB
13571	NW_013657827.1	877980	893976	BAAT
144830	NW_013658145.1	59143	63965	BMP10
34812	NW_013657880.1	783828	992525	BTBD11
18127	NW_013659340.1	22604	74061	CACNB1
155424	NW_013658446.1	495713	538155	CACNG5
130944	NW_013658097.1	75322	145910	CARM1
107727	NW_013658097.1	493722	541872	CASQ2
13007	NW_013657948.1	578948	628470	CELF1
51454	NW_013657928.1	656649	663912	DCTN6
42991	NW_013658321.1	428922	467117	DMGDH
75219	NW_013659036.1	119134	302357	DOCK3
58505	NW_013657769.1	627814	721867	EML5
37913	NW_013657784.1	773231	1107871	ERC2
14258	NW_013658374.1	17582	89638	FAM120A
170144	NW_013658547.1	324922	461529	FBLN2
33811	NW_013658060.1	327963	328523	FLNB
109968	NW_013657847.1	244106	270199	GUK1
61755	NW_013657873.1	271367	360993	IKZF1
13007	NW_013657948.1	565058	571596	KBTBD4
99749	NW_013658110.1	347467	363567	KIAA1257
66371	NW_013659131.1	132535	160005	KPNA4
109968	NW_013657847.1	221759	235281	LOC106538482
155530	NW_013657917.1	599393	601684	LOC106539628
13007	NW_013657948.1	581012	590532	LOC106540121
11736	NW_013657681.1	1985803	2036784	LOC106540201
116293	NW_013657681.1	1985803	2036784	LOC106540201
17369	NW_013658038.1	35112	108049	LOC106541369
51321	NW_013658047.1	33488	34454	LOC106541483
51321	NW_013658047.1	42365	43506	LOC106541484
144830	NW_013658145.1	28064	51776	LOC106542715
54007	NW_013657706.1	511973	768614	LOC106543426
116778	NW_013657711.1	319222	370377	LOC106544530
53393	NW_013658342.1	64334	98538	LOC106544652
9473	NW_013658626.1	334301	454833	LOC106547029
75567	NW_013658831.1	288782	375180	LOC106548434
75616	NW_013658838.1	1110	101893	LOC106548477
114300	NW_013659681.1	128349	138975	LOC106552540
114300	NW_013659681.1	119730	124902	LOC106552543
9573	NW_013657743.1	1541135	1575313	LOC106553011
127801	NW_013660529.1	5062	26978	LOC106554777
28587	NW_013657773.1	700004	722203	LOC106557084
28587	NW_013657773.1	733990	774634	LOC106557085
103667	NW_013657775.1	774702	796540	LOC106557137
148947	NW_013657779.1	1144301	1293792	LOC106557198
129757	NW_013657782.1	134916	308221	LOC106557252
109968	NW_013657847.1	239046	242278	MRPL55
138383	NW_013658122.1	282132	320896	MTR
58800	NW_013659234.1	61338	69545	MYCL
35073	NW_013657934.1	446883	458338	NCDN
165168	NW_013658480.1	341650	366563	NCMAP
28576	NW_013657849.1	665523	768047	PLCH1
13007	NW_013657948.1	571697	578786	PTPMT1
103667	NW_013657775.1	810968	845395	RAB11FIP4
54449	NW_013658428.1	462472	464322	RASSF10
165168	NW_013658480.1	375982	393718	RCAN3

75637	NW_013658189.1	4768	54830	SH3BP4
53079	NW_013658634.1	182421	236805	SLC13A3
28576	NW_013657849.1	633283	648038	SLC33A1
128484	NW_013658784.1	228981	293535	SLC9A8
35194	NW_013657712.1	1028172	1073048	SPINT1
37859	NW_013658334.1	160940	214566	STK17A
137877	NW_013658504.1	474155	480124	TCF19
128483	NW_013657703.1	331664	469167	TCF7
113933	NW_013659655.1	165461	180308	TCP11
89634	NW_013658724.1	86315	119836	TNFSF10
58800	NW_013659234.1	74291	89801	TRIT1
103667	NW_013657775.1	803173	803244	TRNAT-CGU
122451	NW_013657703.1	1789480	1852757	UNC79
100831	NW_013658526.1	108453	224728	WNK1
53393	NW_013658342.1	111876	132178	ZC3H12B
117459	NW_013658747.1	12349	85302	ZNF521
114300	NW_013659681.1	146099	168299	ZNF711
13571	NW_013657827.1	897869	919426	ZP1

Table S7: List of each locus identified by Bayenv2. If it was correlated to mean annual temperature, annual precipitation, and/or altitude it was marked with a Y. The full list of loci and genes putatively in linkage can be found in Tables S8 and S9.

locus	Mean Temp.	Precipitation	Altitude
2217	Y		
7632	Y	Y	
8702		Y	
10204	Y		
12669	Y	Y	
17979	Y		
18467	Y		
24697		Y	
26349	Y		
36839	Y		
41441	Y		
47488	Y		
52749			Y
54449	Y	Y	
66491	Y		
67814		Y	
72941		Y	
75077		Y	
75396	Y		
77513		Y	
80485	Y	Y	
81645		Y	
96198	Y		
97513		Y	
97755		Y	
100816	Y		
105206	Y		
114346	Y		
122451	Y	Y	
125355	Y		
126394		Y	
126547	Y	Y	
126796	Y	Y	
126957	Y	Y	
127371		Y	
127864	Y		
128483		Y	
132785	Y		
133078	Y	Y	
137754	Y		
148947	Y	Y	
149267	Y		
149315	Y	Y	
149964	Y	Y	
150221	Y		
153438	Y		
157963		Y	
160303	Y	Y	
162460	Y		
165168		Y	
169339	Y	Y	

Table S8: Genes within 50kb of a Bayenv2 locus.

Locus	Scaffold	Start	Stop	Gene
75077	NW_013658903.1	134137	137752	ACTL7A
72941	NW_013657766.1	365047	404279	ACVRL1
72941	NW_013657766.1	314794	316373	ANKRD33
169339	NW_013660869.1	12771	56918	APBA2
75077	NW_013658903.1	137837	149714	APTX
149964	NW_013657740.1	329841	361528	ARHGAP1
149964	NW_013657740.1	300645	327272	ATG13
126394	NW_013658440.1	241507	301895	BCKDK
149964	NW_013657740.1	415220	487442	CKAP5
96198	NW_013659173.1	7	154155	COL4A1
66491	NW_013660427.1	81115	83210	DAO
160303	NW_013658912.1	31099	33374	DIRAS3
149267	NW_013658063.1	98654	108308	DMRT2
75077	NW_013658903.1	150702	157622	DNAJA1
149964	NW_013657740.1	391211	410647	F2
133078	NW_013657682.1	2262227	2280582	FBXW9
160303	NW_013658912.1	110312	114027	GADD45A
160303	NW_013658912.1	46252	106613	GNG12
17979	NW_013657714.1	470012	510611	GSTCD
133078	NW_013657682.1	2315138	2317632	IER2
75077	NW_013658903.1	91742	133984	IKBKAP
17979	NW_013657714.1	456621	469792	INTS12
153438	NW_013658254.1	533271	547416	KCTD7
162460	NW_013657787.1	117432	123128	KIAA0355
81645	NW_013658438.1	161860	192252	KIAA2018
18467	NW_013657808.1	49239	50623	LOC106537756
18467	NW_013657808.1	77494	79045	LOC106537757
18467	NW_013657808.1	81374	81998	LOC106537761
105206	NW_013657914.1	483816	485033	LOC106539590
105206	NW_013657914.1	456817	478035	LOC106539591
126796	NW_013658014.1	435475	552577	LOC106541029
80485	NW_013658150.1	81692	110610	LOC106542748
127864	NW_013658223.1	380939	381811	LOC106543506
54449	NW_013658428.1	425731	432421	LOC106545375
126394	NW_013658440.1	214908	229572	LOC106545491
133078	NW_013657682.1	2207648	2226548	LOC106546100
8702	NW_013658717.1	285506	287878	LOC106547713
8702	NW_013658717.1	283613	285454	LOC106547714
8702	NW_013658717.1	299330	307754	LOC106547715
132785	NW_013658762.1	167237	412159	LOC106548014
75077	NW_013658903.1	67945	84769	LOC106548871
26349	NW_013659520.1	23405	36006	LOC106551904
26349	NW_013659520.1	42120	60026	LOC106551905
26349	NW_013659520.1	2351	18635	LOC106551908
2217	NW_013659612.1	160578	189722	LOC106552294
97513	NW_013660375.1	42630	46668	LOC106554483
97513	NW_013660375.1	49111	52629	LOC106554484
97513	NW_013660375.1	60211	62812	LOC106554486
66491	NW_013660427.1	16730	17590	LOC106554601
66491	NW_013660427.1	24424	53991	LOC106554603
66491	NW_013660427.1	55277	58015	LOC106554604
66491	NW_013660427.1	7903	31595	LOC106554605
125355	NW_013660504.1	32125	65183	LOC106554730
125355	NW_013660504.1	8451	28534	LOC106554732
125355	NW_013660504.1	68889	80466	LOC106554733
148947	NW_013657779.1	1144301	1293792	LOC106557198
162460	NW_013657787.1	89963	91503	LOC106557337
162460	NW_013657787.1	11564	52031	LOC106557344
8702	NW_013658717.1	216783	281964	LRRC75A
150221	NW_013657682.1	573218	652308	LZTS2

137754	NW_013658073.1	911777	917193	MAP10
114346	NW_013658853.1	353139	388416	MAP3K2
75077	NW_013658903.1	47287	59991	MNT
81645	NW_013658438.1	144403	157907	NAA50
165168	NW_013658480.1	341650	366563	NCMAP
165168	NW_013658480.1	410107	433948	NIPAL3
17979	NW_013657714.1	516969	585528	NPNT
137754	NW_013658073.1	966487	977200	NTPCR
66491	NW_013660427.1	94262	99666	NUDT1
36839	NW_013658411.1	387783	504399	NXPH4
150221	NW_013657682.1	652700	695123	PDZD7
77513	NW_013658760.1	297191	365565	PLEKHA2
97513	NW_013660375.1	97054	98230	POP4
153438	NW_013658254.1	552094	595260	RABGEF1
54449	NW_013658428.1	462472	464322	RASSF10
165168	NW_013658480.1	375982	393718	RCAN3
75396	NW_013660924.1	4662	5717	RPRM
75077	NW_013658903.1	64466	77508	RPS5
97513	NW_013660375.1	64839	97028	SCARB1
36839	NW_013658411.1	319881	355598	SHMT2
81645	NW_013658438.1	192352	254472	SIDT1
126796	NW_013658014.1	557374	560759	SIVA1
127864	NW_013658223.1	461862	564849	SLCO3A1
165168	NW_013658480.1	316356	335195	SRRM1
127864	NW_013658223.1	392261	404290	ST8SIA2
157963	NW_013659963.1	33113	51884	STAC3
128483	NW_013657703.1	331664	469167	TCF7
8702	NW_013658717.1	189166	210657	TMEM248
133078	NW_013657682.1	2226722	2260968	TNPO2
153438	NW_013658254.1	488870	526437	TPST1
97513	NW_013660375.1	56914	56985	TRNAA-UGC
97513	NW_013660375.1	57745	57816	TRNAD-GUC
97513	NW_013660375.1	59415	59486	TRNAD-GUC
97513	NW_013660375.1	58422	58494	TRNAF-GAA
122451	NW_013657703.1	1789480	1852757	UNC79
133078	NW_013657682.1	2286004	2293732	WDR83
133078	NW_013657682.1	2293955	2303694	WDR83OS
149964	NW_013657740.1	361615	378238	ZNF408

Table S9: Genes within 20kb of a Bayenv2 locus.

Locus	Scaffold	Start	Stop	Gene
72941	NW_013657766.1	314794	316373	ANKRD33
169339	NW_013660869.1	12771	56918	APBA2
149964	NW_013657740.1	329841	361528	ARHGAP1
126394	NW_013658440.1	241507	301895	BCKDK
96198	NW_013659173.1	7	154155	COL4A1
149267	NW_013658063.1	98654	108308	DMRT2
133078	NW_013657682.1	2262227	2280582	FBXW9
160303	NW_013658912.1	46252	106613	GNG12
17979	NW_013657714.1	470012	510611	GSTCD
75077	NW_013658903.1	91742	133984	IKBKAP
153438	NW_013658254.1	533271	547416	KCTD7
81645	NW_013658438.1	161860	192252	KIAA2018
18467	NW_013657808.1	49239	50623	LOC106537756
105206	NW_013657914.1	456817	478035	LOC106539591
80485	NW_013658150.1	81692	110610	LOC106542748
126394	NW_013658440.1	214908	229572	LOC106545491
132785	NW_013658762.1	167237	412159	LOC106548014
75077	NW_013658903.1	67945	84769	LOC106548871
26349	NW_013659520.1	23405	36006	LOC106551904
26349	NW_013659520.1	2351	18635	LOC106551908
2217	NW_013659612.1	160578	189722	LOC106552294
97513	NW_013660375.1	49111	52629	LOC106554484
97513	NW_013660375.1	60211	62812	LOC106554486
66491	NW_013660427.1	24424	53991	LOC106554603
66491	NW_013660427.1	55277	58015	LOC106554604
125355	NW_013660504.1	32125	65183	LOC106554730
125355	NW_013660504.1	8451	28534	LOC106554732
148947	NW_013657779.1	1144301	1293792	LOC106557198
162460	NW_013657787.1	89963	91503	LOC106557337
162460	NW_013657787.1	11564	52031	LOC106557344
8702	NW_013658717.1	216783	281964	LRRC75A
150221	NW_013657682.1	573218	652308	LZTS2
137754	NW_013658073.1	911777	917193	MAP10
165168	NW_013658480.1	341650	366563	NCMAP
17979	NW_013657714.1	516969	585528	NPNT
150221	NW_013657682.1	652700	695123	PDZD7
77513	NW_013658760.1	297191	365565	PLEKHA2
153438	NW_013658254.1	552094	595260	RABGEF1
54449	NW_013658428.1	462472	464322	RASSF10
165168	NW_013658480.1	375982	393718	RCAN3
97513	NW_013660375.1	64839	97028	SCARB1
36839	NW_013658411.1	319881	355598	SHMT2
81645	NW_013658438.1	192352	254472	SIDT1
157963	NW_013659963.1	33113	51884	STAC3
128483	NW_013657703.1	331664	469167	TCF7
133078	NW_013657682.1	2226722	2260968	TNPO2
153438	NW_013658254.1	488870	526437	TPST1
97513	NW_013660375.1	56914	56985	TRNAA-UGC
97513	NW_013660375.1	57745	57816	TRNAD-GUC
97513	NW_013660375.1	59415	59486	TRNAD-GUC
97513	NW_013660375.1	58422	58494	TRNAF-GAA
122451	NW_013657703.1	1789480	1852757	UNC79
133078	NW_013657682.1	2286004	2293732	WDR83
149964	NW_013657740.1	361615	378238	ZNF408

REFERENCES

Ahrens, C. W., Rymer, P. D., Stow, A., Bragg, J., Dillon, S., Umbers, K. D. L., & Dudaniec, R. Y. (2018). The search for loci under selection: trends, biases and progress. *Molecular Ecology*, *27*(6), 1–15.

Arevalo, E., Davis, S. K., & Sites Jr, J. W. (1994). Mitochondrial DNA sequence divergence and phylogenetic relationships among eight chromosome races of the Sceloporus grammicus complex (Phrynosomatidae) in central Mexico. *Systematic Biology*, *43*(3), 387–418.

Avise, J. C., & Nelson, W. S. (1989). Molecular genetic relationships of the extinct dusky seaside sparrow. *Science*, *243*(4891), 646–648.

Babonis, L. S., Miller, S. N., & Evans, D. H. (2011). Renal responses to salinity change in snakes with and without salt glands. *Journal of Experimental Biology*, *214*(13), 2140–2156.

Babonis, Leslie S., Womack, M. C., & Evans, D. H. (2012). Morphology and putative function of the colon and cloaca of marine and freshwater snakes. *Journal of Morphology*, *273*(1), 88–102.

Baird, S. F., & Girard, C. (1853). *Catalogue of North American reptiles in the Museum of the Smithsonian Institution*. Smithsonian institution.

Balkau, B. J., & Feldman, M. W. (1973). Selection for migration modification. *Genetics*, *74*(1), 171–174.

Bert, T. M. (1986). Speciation in western Atlantic stone crabs (genus *Menippe*): the role of geological processes and climatic events in the formation and distribution of species. *Marine Biology*, *93*(2), 157–170.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Brandrud, M. K., Paun, O., Lorenzo, M. T., Nordal, I., & Brysting, A. K. (2017). RADseq provides evidence for parallel ecotypic divergence in the autotetraploid *Cochlearia officinalis* in Northern Norway. *Scientific Reports*, *7*(1), 5573.

Brooks, W. (2008). *Atlantic salt marsh snake (Nerodia clarkii taeniata) 5-Year Review: Summary and Evaluation*.

Burbrink, F. T. (2002). Phylogeographic analysis of the cornsnake (*Elaphe guttata*) complex as inferred from maximum likelihood and Bayesian analyses. *Molecular Phylogenetics and Evolution*, *25*(3), 465–476.

Burbrink, F. T., Lawson, R., & Slowinski, J. B. (2000). Mitochondrial DNA phylogeography of the polytypic North American rat snake (*Elaphe obsoleta*): a critique of the subspecies concept. *Evolution*, *54*(6), 2107–2118.

Burns, G., Ramos, A., & Muchlinski, A. (1996). Fever response in North American snakes. *Journal of Herpetology*, 133–139.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*(1), 421.

Carr Jr, A. F., & Goin, C. J. (1942). Rehabilitation of *Natrix sipedon taeniata* Cope. In *Proc. New England Zool. Club*, *21*, 47–54).

Carstens, B. C., & Knowles, L. L. (2007). Shifting distributions and speciation: species divergence during rapid climate change. *Molecular Ecology*, *16*(3), 619–627.
Carter, B. D., & McKinney, M. L. (1992). Eocene echinoids, the Suwannee Strait, and biogeographic taphonomy. *Paleobiology*, *18*(3), 299–325.
Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, *30*(23), 3317–3324.
Clay, W. M. (1938). A Synopsis of the North American Water Snakes of the Genus *Natrix*. *Copeia*, *1938*(4), 173–182.
Cliburn, J. W. (1960). The phylogeny and zoogeography of North American *Natrix*. (Doctoral dissertation, University of Alabama).
Conant, R. (1963). Evidence for the specific status of the water snake *Natrix fasciata*. *American Museum Novitates*, *2122*, 1–38.
Conroy, C. J., Papenfuss, T., Parker, J., & Hahn, N. E. (2009). Use of tricaine methanesulfonate (MS222) for euthanasia of reptiles. *Journal of the American Association for Laboratory Animal Science*, *48*(1), 28–32.
Cope, E. D. (1895). On some new North American snakes, *Natrix compressicauda taeniata* subsp. nov. *American Naturalist*, *29*, 676–680.
Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., & Sherry, S. T. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.
Deen, C. M., & Hutchison, V. H. (2001). Effects of lipopolysaccharide and acclimation temperature on induced behavioral fever in juvenile *Iguana iguana*. *Journal of Thermal Biology*, *26*(1), 55–63.
Delpire, E., & Gagnon, K. B. E. (2008). SPAK and OSR1: STE20 kinases involved in the regulation of ion homoeostasis and volume control in mammalian cells. *Biochemical Journal*, *409*(2), 321–331.
Dunson, W. A. (1978). Role of the skin in sodium and water exchange of aquatic snakes placed in seawater. *American Physiological Society*, 151–159.
Dunson, W. A. (1979). Occurrence of partially striped forms of the mangrove snake *Nerodia fasciata compressicauda* kennicott and comments on the status of *N. f. taeniata* Cope. *Florida Scientist*, *42*(2), 102–112.
Dunson, W. A. (1980). The relation of rodium and water balance to survival in sea water of estuarine and freshwater races of the snakes *Nerodia fasciata, N. sipedon and N. valida*. *Copeia*, *1980*(2), 268.
Dupuis, J. R., Sim, S. B., San Jose, M., Leblanc, L., Hoassain, M. A., Rubinoff, D., & Geib, S. M. (2017). Population genomics and comparisons of selective signatures in two invasions of melon fly, *Bactrocera cucurbitae* (Diptera: Tephritidae). *Biological Invasions*, 1–18.
Dutton, A., Carlson, A. E., Long, Aj., Milne, G. A., Clark, P. U., DeConto, R., & Raymo, M. E. (2015). Sea-level rise due to polar ice-sheet mass loss during past warm periods. *Science*, *349*(6244).
Eaton, D. A. R. (2014). PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, *30*(13), 1844–1849.
Engen, S., & Sæther, B. E. (2016). Phenotypic evolution by distance in fluctuating

environments: The contribution of dispersal, selection and random genetic drift. *Theoretical Population Biology*, *109*, 16–27.
Felsenstein, J. (1976). The theoretical population genetics of variable selection and migration. *The Annual Review of Genetics*, *571*, 107–120.
Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315.
Fischer, M. C., Foll, M., Excoffier, L., & Heckel, G. (2011). Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (Microtus arvalis). *Molecular Ecology*, *20*(7), 1450–1462.
Fuller, R. C., McGhee, K. E., & Schrader, M. (2007). Speciation in killifish and the role of salt tolerance. *Journal of Evolutionary Biology*, *20*(5), 1962–1975.
Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology and Evolution*, *27*(9), 489–496.
Gibbons, T. C., Metzger, D. C. H., Healy, T. M., & Schulte, P. M. (2017). Gene expression plasticity in response to salinity acclimation in threespine stickleback ecotypes from different salinity habitats. *Molecular Ecology*, *26*(10), 2711–2725.
Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184–186.
Günther, T., & Coop, G. (2013). Robust Identification of Local Adaptation from Allele Frequencies. *Genetics*, *195*(1), 205 LP – 220.
Guo, P., Liu, Q., Xu, Y., Jiang, K., Hou, M., Ding, L., & Burbrink, F. T. (2012). Out of Asia: Natricine snakes support the Cenozoic Beringian Dispersal Hypothesis. *Molecular Phylogenetics and Evolution*, *63*(3), 825–833.
Harrisson, K. A., Pavlova, A., Telonis-Scott, M., & Sunnucks, P. (2014). Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications*, *7*(9), 1008–1025.
Hebrard, J. J., & Lee, R. C. (1981). A large collection of brackish water snakes from the central atlantic coast of Florida. *Copeia*, (4), 886–889.
Ingvarsson, P. K. (2001). Restoration of genetic variation lost–the genetic rescue hypothesis. *Trends in Ecology & Evolution*, *16*(2), 62–63.
Ivanis, G., Braun, M., & Perry, S. F. (2008). Renal expression and localization of SLC9A3 sodium/hydrogen exchanger and its possible role in acid-base regulation in freshwater rainbow trout (*Oncorhynchus mykiss*). *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *295*(3), R971–R978.
James, F. C. (1983). Environmental component of morphological differentiation in birds. *Science*, *221*(4606), 184–186.
Jansen, K. P. (2001). Ecological genetics of the salt marsh snake *Nerodia clarkii*.
Jansen, K. P., Mushinsky, H. R., & Karl, S. A. (2008). Population genetics of the mangrove salt marsh snake, *Nerodia clarkii compressicauda*, in a linear, fragmented habitat. *Conservation Genetics*, *9*(2), 401–410.
Jiang, Z. J., Castoe, T. A., Austin, C. C., Burbrink, F. T., Herron, M. D., McGuire, J. A., & Pollock, D. D. (2007). Comparative mitochondrial genomics of snakes:

extraordinary substitution rate dynamics and functionality of the duplicate control region. *BMC Evolutionary Biology*, *7*(1), 123.
Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405.
Jost, L. O. U. (2008). GST and its relatives do not measure differentiation. *Molecular Ecology*, *17*(18), 4015–4026.
Kaplan, N. L., Hudson, R. R., & Langley, C. H. (1989). The "hitchhiking effect" revisited. *Genetics*, *123*(4), 887–899.
Kennicott, R. (1860). Descriptions of new species of North American serpents in the Museum of the Smithsonian Institution, Washington. *Proc. Acad. Nat. Sci. Philadelphia*, *12*, 331).
Kingman, J. F. C. (2000). Perspectives anecdotal, historical and critical commentaries on genetics origins of the coalescent: 1974–1982. *Genetics*, *156*, 1461–1463.
Knaus, B. J., & Grünwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44–53.
Kochman, H.I. (1992). Atlantic Salt Marsh Snake. In P. E. Moler (Ed.), *The Rare and Endangered Biota of Florida*, 111–116). Gainesville, FL: University of Florida Press.
Kochman, Howard Irwin. (1977). Differentiation and hybridization in the *Natrix fasciata* complex (Reptilia: Serpentes): a nonmorphological approach. (Doctoral dissertation, University of Florida).
Lanfear, R., Calcott, B., Ho, S. Y. W., & Guindon, S. (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, *29*(6), 1695–1701.
Lawson, R., Meier, A. J., Frank, P. G., & Moler, P. E. (1991). Allozyme variation and systematics of the *Nerodia fasciata-Nerodia clarkii* complex of water snakes (Serpentes : Colubridae). *Copeia*, *1991*(3), 638–659.
Leksrisompong, N., Romero-Sanchez, H., Plumstead, P. W., Brannan, K. E., & Brake, J. (2007). Broiler Incubation. 1. Effect of Elevated Temperature During Late Incubation on Body Weight and Organs of Chicks1. *Poultry Science*, *86*(12), 2685–2691.
Lemmon, A. R., Emme, S. A., & Lemmon, E. M. (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, *61*(5), 727–744.
Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology and Evolution*, *17*(4), 183–189.
Lenormand, T. (2012). From local adaptation to speciation: Specialization and reinforcement. *International Journal of Ecology*, 2012.
Lexer, C., Mangili, S., Bossolini, E., Forest, F., Stölting, K. N., Pearman, P. B., & Salamin, N. (2013). "Next generation" biogeography: Towards understanding the drivers of species diversification and persistence. *Journal of Biogeography*, *40*(6), 1013–1022.
Lischer, H. E. L., & Excoffier, L. (2011). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298–299.

Losos, J. B., Warheitt, K. I., & Schoener, T. W. (1997). Adaptive differentiation following experimental island colonization in Anolis lizards. *Nature*, *387*(6628), 70.
Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67–77.
Margres, M. J., Patton, A., Wray, K. P., Hassinger, A. T. B., Ward, M. J., Lemmon, E. M., & Rokyta, D. R. (2019). Tipping the scales: the migration-selection balance leans toward selection in snake venoms. *Molecular Biology and Evolution*, *36*(2), 271–282.
Marshall, J. C., Kingsbury, B. A., & Minchella, D. J. (2009). Microsatellite variation, population structure, and bottlenecks in the threatened copperbelly water snake. *Conservation Genetics*, *10*(2), 465–476.
Marshall, W. S., Cozzi, R. R. F., & Spieker, M. (2017). WNK1 and p38-MAPK distribution in ionocytes and accessory cells of euryhaline teleost fish implies ionoregulatory function. *Biology Open*, *6*(7), 956–966.
Maynard Smith, J., & Haigh, J. (1973). The hitch-hiking effect of a favourable gene. *Genetics Research*, *89*(5–6), 391–403.
McGlothlin, J. W., Chuckalovcak, J. P., Janes, D. E., Edwards, S. V, Feldman, C. R., Brodie Jr, E. D., & Brodie III, E. D. (2014). Parallel evolution of tetrodotoxin resistance in three voltage-gated sodium channel genes in the garter snake *Thamnophis sirtalis*. *Molecular Biology and Evolution*, *31*(11), 2836–2846.
Mcvay, J. D., Flores-Villela, O., & Carstens, B. (2015). Diversification of North American natricine snakes. *Biological Journal of the Linnean Society*, *116*(1), 1–12.
Merchant, M., Fleury, L., Rutherford, R., & Paulissen, M. (2008). Effects of bacterial lipopolysaccharide on thermoregulation in green anole lizards (*Anolis carolinensis*). *Veterinary Immunology and Immunopathology*, *125*(1–2), 176–181.
Mi, H., Muruganujan, A., Ebert, D., Huang, X., & Thomas, P. D. (2018). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research*, *47*(D1), D419–D426.
Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X., & Thomas, P. D. (2019). Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nature Protocols*, *14*(3), 703–721.
Nanayakkara, S., Stmld, S., Abeysekera, T., Chandrajith, R., Ratnatunga, N., Edl, G., & Komiya, T. (2013). An integrative study of the genetic, social and environmental determinants of chronic kidney disease characterized by tubulointerstitial damages in the North Central Region of Sri Lanka. *Journal of Occupational Health*, 13–172.
Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, *70*(12), 3321–3323.
Osgood, D. W. (1978a). Effects of Temperature on the Development of Meristic Characters in Natrix fasciata. *Copeia*, *1978*(1), 33–47.
Osgood, D. W. (1978b). Effects of Temperature on the Development of Meristic Characters in *Natrix fasciata*. *Copeia*, *1978*(1), 33–47.
Pante, E., Abdelkrim, J., Viricel, A., Gey, D., France, S. C., Boisselier, M. C., & Samadi, S. (2015). Use of RAD sequencing for delimiting species. *Heredity*, *114*(5), 450–

459.
Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, *7*(5).
Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, *48*(1), 94.
Pettus, D. (1958). Water Relationships in *Natrix sipedon*. *Copeia*, *1958*(3), 207–211.
Pettus, D. (1963). Salinity and Subspeciation in *Natrix sipedon*. *Copeia*, *1963*(3), 499–504.
Pfeifer, S. P., Laurent, S., Sousa, V. C., Linnen, C. R., Foll, M., Excoffier, L., & Jensen, J. D. (2018). The evolutionary history of Nebraska deer mice: Local adaptation in the face of strong gene flow. *Molecular Biology and Evolution*, *35*(4), 792–806.
Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R News*, *6*(1), 7–11.
Pyron, R. A., Burbrink, F. T., & Wiens, J. J. (2013). A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evolutionary Biology*, *13*, 93.
Quérouil, S., Silva, M. A., Freitas, L., Prieto, R., Magalhães, S., Dinis, A., & Hammond, P. S. (2007). High gene flow in oceanic bottlenose dolphins (*Tursiops truncatus*) of the North Atlantic. *Conservation Genetics*, *8*(6), 1405.
Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842.
Rambaut, A. (2012). FigTree v1. 4.
Räsänen, K., & Hendry, A. P. (2008). Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecology Letters*, *11*(6), 624–636.
Rasmussen, A. R., Murphy, J. C., Ompi, M., Gibbons, J. W., & Uetz, P. (2011). Marine reptiles. *PLoS ONE*, *6*(11).
Razkin, O., Sonet, G., Breugelmans, K., Madeira, M. J., Gómez-Moliner, B. J., & Backeljau, T. (2016). Species limits, interspecific hybridization and phylogeny in the cryptic land snail complex Pyramidula: The power of RADseq data. *Molecular Phylogenetics and Evolution*, *101*(May), 267–278.
Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, *24*(17), 4348–4370.
Remington, C. L. (1968). Suture-zones of hybrid interaction between recently joined biotas. In *Evolutionary biology* (pp. 321–428). Springer.
Roth, T. C., & Greene, B. D. (2006). Movement patterns and home range use of the northern watersnake (*Nerodia sipedon*). *Copeia*, *2006*(3), 544–551.
Roy, K., Valentine, J. W., Jablonski, D., & Kidwell, S. M. (1996). Scales of climatic variability and time averaging in Pleistocene biotas: implications for ecology and evolution. *Trends in Ecology & Evolution*, *11*(11), 458–463.
Saunders, N. C., Kessler, L. G., & Avise, J. C. (1986). Genetic variation and geographic differentiation in mitochondrial DNA of the horseshoe crab, *Limulus polyphemus*.

Genetics, *112*(3), 613–627.
Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews. Genetics*, *14*(11), 807–820.
Schield, D. R., Adams, R. H., Card, D. C., Corbin, A. B., Jezkova, T., Hales, N. R., & Smith, L. L. (2018). Cryptic genetic diversity, population structure, and gene flow in the Mojave Rattlesnake (*Crotalus scutulatus*). *Molecular Phylogenetics and Evolution*, *127*, 669–681.
Schield, D. R., Card, D. C., Adams, R. H., Jezkova, T., Reyes-Velasco, J., Proctor, F. N., & Castoe, T. A. (2015). Incipient speciation with biased gene flow between two lineages of the Western Diamondback Rattlesnake (*Crotalus atrox*). *Molecular Phylogenetics and Evolution*, *83*, 213–223.
Service, U. S. F. and W. (1977). Listing of the Atlantic salt marsh snake as a threatened species. *Federal Register*, *42*, 60743–60745.
Service, U. S. F. and W. (1993). Atlantic Salt Marsh Snake Recovery Plan. *Atlanta, Georgia*.
Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution*, *68*(1), 1–15.
Soltis, D. E., Morris, A. B., McLachlan, J. S., Manos, P. S., & Soltis, P. S. (2006). Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, *15*(14), 4261–4293.
Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., & Lancet, D. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Current Protocols in Bioinformatics*, *54*(1), 1.30.1-1.30.33.
Sukumaran, J., & Holder, M. T. (2010). DendroPy: a Python library for phylogenetic computing. *Bioinformatics*, *26*(12), 1569–1571.
Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*, *6*(7).
Swofford, D. L. (2003). PAUP*: phylogenetic analysis using parsimony, version 4.0 b10.
Team, Q. D. (2016). QGIS geographic information system. *Open Source Geospatial Foundation Project*.
Team, Rs. (2015). RStudio: integrated development for R. *RStudio, Inc., Boston, MA URL Http://Www. Rstudio. Com*, *42*, 14.
Territo, G. P. (2013). Biogeography and Systematics of the *Nerodia clarkii/Nerodia fasciata* clade in Florida.
Tigano, A., & Friesen, V. L. (2016). Genomics of local adaptation with gene flow. *Molecular Ecology*, *25*(10), 2144–2164.
Tollis, M., & Boissinot, S. (2014). Genetic variation in the green anole lizard (*Anolis carolinensis*) reveals island refugia and a fragmented Florida during the quaternary. *Genetica*, *142*(1), 59–72.
Townsend, T. M., Alegre, R. E., Kelley, S. T., Wiens, J. J., & Reeder, T. W. (2008). Rapid development of multiple nuclear loci for phylogenetic analysis using genomic resources: An example from squamate reptiles. *Molecular Phylogenetics and Evolution*, *47*(1), 129–142.

Tzika, A. C., Koenig, S., Miller, R., Garcia, G., Remy, C., & Milinkovitch, M. C. (2008). Population structure of an endemic vulnerable species, the Jamaican boa (*Epicrates subflavus*). *Molecular Ecology*, *17*(2), 533–544.

Uetz, P. (2010). The original descriptions of reptiles. *Zootaxa*, *68*(2334), 59–68.

Upchurch, S., Scott, T. M., Alfieri, M. C., Fratesi, B., & Dobecki, T. L. (2019). Geologic Framework of Florida. In *The Karst Systems of Florida* (pp. 59–91). Springer.

Vaughan, T., Wu, C., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2 : A Software Platform for Bayesian Evolutionary Analysis, *10*(4), 1–6.

Verity, R., Collins, C., Card, D. C., Schaal, S. M., Wang, L., & Lotterhos, K. E. (2017). minotaur: A platform for the analysis and visualization of multivariate results from genome scans with R Shiny. *Molecular Ecology Resources*, *17*(1), 33–43.

Weir, B. S., & Cockerham, C. C. (1984). Estimating f-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370.

Whitlock, M. C., & McCauley, D. E. (1999). Indirect measures of gene flow and migration: FST not equal to 1/(4Nm + 1). *Heredity*, *82 (Pt 2)*(November 1998), 117–125.

Williams, G. (1966). *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton University Press.

Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, *18*(2), 87–100.

Wood, D. A., Halstead, B. J., Casazza, M. L., Hansen, E. C., Wylie, G. D., & Vandergast, A. G. (2015). Defining population structure and genetic signatures of decline in the giant gartersnake (*Thamnophis gigas*): implications for conserving threatened species within highly altered landscapes. *Conservation Genetics*, *16*(5), 1025–1039.

Wood, D. A., Vandergast, A. G., Lemos Espinal, J. A., Fisher, R. N., & Holycross, A. T. (2011). Refugial isolation and divergence in the Narrowheaded Gartersnake species complex (*Thamnophis rufipunctatus*) as revealed by multilocus DNA sequence data. *Molecular Ecology*, *20*(18), 3856–3878.