2017

MITOCHONDRIAL AND NUCLEAR PATTERNS OF CONFLICT AND CONCORDANCE AT THE GENE, GENOME, AND BEHAVIORAL SCALES IN *DESMOGNATHUS* SALAMANDERS

Justin D. Kratovil
University of Kentucky, jdkr222@gmail.com
Digital Object Identifier: https://doi.org/10.13023/ETD.2017.381

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Justin D. Kratovil, Student

Dr. David W. Weisrock, Major Professor

Dr. David F. Westneat, Director of Graduate Studies

MITOCHONDRIAL AND NUCLEAR PATTERNS OF CONFLICT AND CONCORDANCE AT THE GENE, GENOME, AND BEHAVIORAL SCALES IN *DESMOGNATHUS* SALAMANDERS

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Arts and Sciences
at the University of Kentucky

By
Justin David Kratovil

Lexington, Kentucky

Director: Dr. David Weisrock, Associate Professor of Biology

Lexington, Kentucky

2017

ABSTRACT OF DISSERTATION

MITOCHONDRIAL AND NUCLEAR PATTERNS OF CONFLICT AND CONCORDANCE AT THE GENE, GENOME, AND BEHAVIORAL SCALES IN *DESMOGNATHUS* SALAMANDERS

Advancements in molecular sequencing have revealed unexpected cryptic genetic diversity and contrasting evolutionary histories within genes and between genomes of many organisms; often in disagreement with recognized taxonomy. Incongruent patterns between the mitochondrial and nuclear evolutionary history can have several plausible explanations, but widespread systematic conflict inevitably challenges our conceptions of species boundaries when there is discordance between coevolving and coinherited genomes. It is unknown to what degree mitonuclear conflict drives the process of divergence, or how ubiquitous these patterns are across the tree of life. To understand the evolutionary relevance of intergenomic discordance we must identify the conflicting patterns that exist in natural systems by generating robust estimates of the underlying species history, quantify support for alternative hypotheses of lineage formation, and describe patterns of genetic variation present in robust nuclear genomic datasets. Empirically testing correlations between mitonuclear genomic conflict and reduced gene flow at the organism level will contribute toward a better understanding of lineage boundaries and how intergenomic interactions shape the process of divergence.

Mitochondrial introgression has been inferred in many salamander systems with limited perspective from nuclear sequence data. Within dusky salamanders (*Desmognathus*), these patterns have been observed between morphologically and geographically disparate populations. I sequenced regions throughout the nuclear genome to reconstruct species trees, performed population-level analyses testing concordance between the mitochondrial, nuclear datasets, and nuclear genes with mitochondrial functions with the expectation that coevolutionary interactions among genomes are more likely to manifest in these regions. I also estimated migration rates between populations that may have experienced historical mitochondrial introgression to evaluate phylogeographic patterns.

Using these data we definitively reject species models in which genetic boundaries are based solely on mitochondrial data, favoring geographic models instead. Furthermore,

analyses soundly reject current taxonomic models based on morphological characteristics, suggesting there is greater lineage diversity than is currently recognized.

I also used empirical assays of pre-zygotic reproductive mating behavior within and among populations containing diverse mitochondrial lineages to test metrics of reproductive isolation, and to determine if introgression shapes the evolution of complex traits directly influencing rates of divergence. These results may explain incongruent patterns observed between the mitochondrial and nuclear data as a function of inheritance and population dynamics rather than directly functioning to suppress nuclear gene flow.

This research builds upon recent studies suggesting that speciation is a highly complex and often non-bifurcating process in which introgression can have a profound and lasting signature on the nuclear evolutionary history. Mechanisms responsible for divergence with gene flow challenge evolutionary biologists to reevaluate our notions and definitions of species boundaries to accommodate seemingly conflicted genomic patterns within and between genomes.

KEYWORDS: mitonuclear discordance, divergence with gene flow, lineage boundaries, introgression, *Desmognathus* salamanders

Justin D. Kratovil

July 18, 2017

MITOCHONDRIAL AND NUCLEAR PATTERNS OF CONFLICT AND CONCORDANCE AT THE GENE, GENOME, AND BEHAVIORAL SCALES IN *DESMOGNATHUS* SALAMANDERS

By

Justin David Kratovil

David W. Weisrock, Ph.D.
Director of Dissertation

David F. Westneat, Ph.D.
Director of Graduate Studies

July 18, 2017

To my mother for encouraging my ’satiable curiosity

“Most species do their own evolving, making it up as they go along, which is the way Nature intended. And this is all very natural and organic and in tune with mysterious cycles of the cosmos, which believes that there’s nothing like millions of years of really frustrating trial and error to give a species moral fibre and, in some cases, backbone. This is probably fine from the species’ point of view, but from the perspective of the actual individuals involved it can be a real pig.”

– Terry Pratchett, *Reaper Man*

Acknowledgements

I've had the privilege of having many wonderful instructors who helped kindle my interest in the life sciences at an early age, but it probably wasn't until having courses with Stan Sessions at Hartwick College that I realized scientific research and teaching was an equally viable and exciting career path. He taught me that there's something profound about walking through a rain forest or peering through a microscope and perhaps seeing something for the very first time ever, and the importance of preserving that knowledge. To paraphrase his words, how many other career paths offer an opportunity to catch a rare glimpse “under the curtain” to see how it all works? Stan's exuberance for learning is infectious, and I hope to emulate his devotion and passion for teaching in the many years ahead.

We all stand on the shoulders of giants, and for me that giant would be Steve Tilley. I can't thank Steve enough for introducing me to the world of desmogs, plethodontids, and the natural history of the Blue Ridge Mountains. Steve has contributed so much to our understanding of the life history, ecology, and diversity of salamanders in this region, teaching generations of students, herpetologists, and enthusiasts about these precious and fascinating organisms. It was a privilege to work with him at Smith College, and continue working with him through my Ph.D. He's an amazing naturalist and a good friend.

When I began looking at graduate programs, I was intrigued by the prospect of working with Dave Weisrock. At the time he was a recently hired Assistant Professor at the University of Kentucky seeking to grow his lab and research program. Although he was hesitant to continue doing research on dusky salamanders (an incredibly challenging system), he had confidence in my abilities, and his excitement and motivation secured my

decision to join his lab. Little did he know that, like a stochastic MCMC chain, I'd haphazardly entertain many different research paths and new questions even when it seemed like things were already converging nicely. Without his careful guidance, prodding, and encouragement, who knows what rabbit hole I might have fallen into. Dave has taught me to always think about the broader questions and importantly, work on selling these ideas to others. I'm truly grateful for having Dave as my Ph.D. advisor. He's an inspiring scientist, a genuinely great person, and I look forward to continuing our collaborations into the future.

I also want to thank my other advisory committee members Catherine Linnen, Jim Krupa, Chuck Fox, Nicholas McLetchie, and Steve Price for their helpful guidance, constructive comments, and support over the years. I'd also like to thank all the members of the Linnen-Weisrock "Super Lab" who have been like an extended family, but specifically acknowledge Tim Vatovec (an honorary member of the lab and great friend), Robin Bagley, Kara Jones, Eric O'Neill, Paul Hime, Dana Weisrock, Mason Murphy, Danielle Herrig, and numerous undergraduates: Mary Foley, Jose Bocanegra, Kelly Sovacool, and Mary Virginia Gibbs for making this possible. Other members of the UK community who have supported me over the years include Jacqueline Burke, Cheryl Edwards, Beverly Taulbee, Randal Voss, Jim Shaffer, Brittany Slabach, Rudy Yoshida, Dave Haws, Jennifer Webb, Elisaveta Arnaudova, John Seabolt, Jerzy and Jolanta Jaromczyk, and Chris Schardl. Much of this research would not be possible without the use of the UK Lipscomb HPCC and the advice of Vikram Gazula at the UK Center for Computational Sciences. I am also grateful for the financial support made possible through funding from the National Science Foundation (Grant numbers DEB-0949532, and DEB-

1355000 to D. Weisrock and his collaborators). Additional funding sources are individually acknowledged in each chapter. I also want to thank the staff at South Mountain State Park, specifically Allen Rogers for helping with various projects.

Lastly, I thank my wonderful wife Renee and daughter Theresa. Renee followed me into the Bluegrass, successfully established a legal career, all while enduring the many challenges associated with having a spouse in graduate school (e.g., long hours and work weeks extending into weekends). We now have a family, and they have helped provide me with an important perspective about balancing life in academia, getting sufficient rest, and fulfilling my scholarly commitments. I would not have been able to finish (while retaining my sanity) without Renee's patience, love, and understanding.

Table of Contents

List of Tables

List of Figures

Chapter 1 : Introduction

1.1 – SPECIES BOUNDARIES

Identifying the boundaries which delimit species are fundamental to the study of evolution because defining such characteristics contribute toward our understanding of meaningful patterns and mechanisms that influence the process of divergence and speciation. Species limits are important because they help evolutionary biologists identify cohesive units from which to infer evolutionary relationships among taxa, categorize the grandeur of diversity within the tree of life, make inferences about the mechanisms responsible for generating and maintaining complex traits, and give a sense of the peculiar genetic patterns that we observe at the population level all the way through deeper branches in our shared evolutionary history. A palpable challenge in this process, is that ever since taxonomists first started classifying organisms based on shared and derived characteristics there has been little agreement about the criteria used to define species. This led to contentious arguments and minimal progress beyond generating a multitude of species concepts, which were sometimes valid, but difficult to universally apply despite generally broad consensus in the scientific community that species, as a concept exist. Overviews of species definitions and historical arguments exist elsewhere (Hey 2006; Mallet 2007).

The initial causes of divergence are usually unknown and the criteria historically used to delimit species boundaries, when applicable, were only appropriate for later stages of divergence in many organisms. Often criteria could not be universally applied across diverse taxa and could not explain many exceptions to the rules (e.g., hybrid organisms, asexual reproduction, etc.) (Coyne, JA and Orr 2004). For example, traditional species definitions were primarily based on evidence examining intrinsic reproductive isolation

(Mayr 1942), ecological niche disparity (Van Valen 1976), or reciprocal monophyly (Baum & Shaw 1995). These particular measures may be relevant for the inference of some population boundaries, but no single criterion can be universally applied and it is impossible to guarantee that evolutionary barriers or distinctive characteristics are absolute (Kirkpatrick & Ravigne 2002). There has been a recent shift from defining species based on narrow or universal criteria to using multiple properties of existing populations to estimate divergence between lineages through time. This perspective is known as the general lineage species concept (de Queiroz 2005; de Queiroz 2007).

One consistent theme in our understanding of species limits is the acknowledgement that speciation is usually a long and protracted process. In recent years, aided through the acquisition of unprecedented amounts of genomic sequence data (followed by the computational resources to process and analyze it), there is much greater appreciation that divergence occurs long before rigid species boundaries begin to evolve. In addition, there is mounting evidence that species boundaries are often porous or incomplete, and that divergence can occur despite the homogenizing effect of gene flow. All of this research has contributed to our understanding that speciation is a highly dynamic process, potentially filled with a mosaic of discordant patterns across the genome that can confound efforts to characterize species boundaries and evolutionary relationships among taxa. Ultimately, an underlying species history exists, but there is uncertainty about the degree to which genomic discordance is a reflection of stochastic lineage sorting or a signature of historical or ongoing population level processes during nascent speciation and divergence. Before it is even possible to address this major question, we must first confidently estimate species histories.

1.2 – MOLECULAR SYSTEMATICS

Traditional taxonomic approaches based on morphological characteristics may be a good starting point for proposing and testing hypotheses of species divergence, however these methods pose a significant challenge in systems with low phenotypic or ecological disparity, which obscure high levels of cryptic genetic variation that can exist among populations (Wiens 2004; Adams *et al.* 2009). The burgeoning field of molecular systematics has changed the perception of species boundaries from rigid barriers after reproductive isolation, to a continuum that can operate anywhere genetic variation is present. For example, the onset of divergence can begin when ancestral polymorphisms differentially fix across population-lineages (Zhou *et al.* 2007), or persist in the presence of ongoing gene flow (Pavey *et al.* 2010).

Recovering the species history and discerning factors influencing divergence between metapopulations necessitates an integrated approach that is robust to evolutionary rate heterogeneities and gene tree histories which can deviate from the underlying species tree. Early approaches in molecular systematics were limited to examining the physiochemical migratory properties of allozyme loci, which enabled investigators to uncover profound genetic diversity between morphologically cryptic lineages (Murphy *et al.* 1996; Highton & Peabody 2000). Sanger sequencing initiated another revolution enabling the quantification of nucleotide variation often through the use of highly variable mitochondrial DNA (mtDNA) sequences (as well as conserved ribosomal RNA sequences) effectively bridging the gap between population genetics and systematics (Avise *et al.* 1987; Avise 2009). Other markers such as microsatellites and AFLPs have also been useful

in examining population genetic variation, but as in other molecular tools, this method required significant investment and development for individual study systems.

Within the past decade, next generation sequencing technologies have entered a renaissance facilitated by the rapid and cost effective accumulation of genetic data across entire genomes opening up the possibility of reconciling species histories from individual gene histories (Brito & Edwards 2009). This wealth of sequence data is pertinent to the robust delimitation of species boundaries in order to identify evolutionary patterns of selection and gene flow which can disentangle phylogenetic relationships among lineages. Despite the deluge of data resulting from next generation sequencing, there are many phylogenetic challenges to studying population divergence due to the discordant nature of genome evolution. Gene histories frequently contain conflicting topological relationships, and depending on the rate of divergence, can fail to produce well supported species trees. Single gene-tree estimates are individual histories and inadequately describe the complex and dynamic process of divergence across the genome, population, or species. Given such complex histories, gene lineages are rarely monophyletic or strictly bifurcating, and are subject to processes such as incomplete lineage sorting, duplication, extinction, and branch reticulations caused by introgression and hybridization which create gene-tree species-tree discordances (Maddison & Knowles 2006).

Incomplete lineage sorting is the stochastic fixation of alleles during the initial segregation of diverging populations that can cause gene trees to have different topologies from the actual species tree history (Maddison 1997; Degnan & Rosenberg 2006) . The amount of discordance caused by incomplete lineage sorting, as in genetic drift, is dependent upon allele frequencies and standing genetic variation between diverging

population lineages over time. Shared polymorphisms that persist during divergence appear to resolve much later than the initial splitting of lineages. Likewise, rapid fixation of alleles may suggest earlier divergence or population structure in different loci than might be predicted from recent gene flow or evolutionary rate heterogeneity in other genealogies across the genome. Depending on the amount of standing genetic variation in an ancestral population and the number of loci examined, it may be impossible for polymorphic alleles in divergent species to become fixed for most loci preventing delimitation based on genealogical consensus alone (Hudson & Coyne 2002).

Introgression is another source of discordance between gene and species histories that creates reticulate evolutionary patterns between species that do not share recent common ancestors or genetic backgrounds. Horizontal exchange can be mediated by vectors that transpose genetic material between distantly related lineages, or through asymmetric gene flow during hybridization, an effect which is exacerbated during rapid radiations (Wiens *et al.* 2006; Bryson *et al.* 2010). Generally, incompatibilities between anciently derived descendants prevent widespread introgression across the genome, however there are instances of cytoplasmic capture that have been described between major branches in the tree of life (Dowling & Secor 1997; Twyford & Ennos 2012). Population cohesion, as measured by intraspecific gene flow (recombination), will reduce signatures of introgression in the genome over time, suggesting that horizontally transferred genes have limited utility for delimiting species (Petit & Excoffier 2009). The identification of introgressed regions remains an important aspect of recovering a species history as these regions can be a source of genetic novelty contributing to adaptive selection and divergence.

Recent approaches for estimating species histories while resolving genealogical conflict have focused on the recovery of information from independent multi-locus markers across both nuclear and mitochondrial genomes. Generating sequence data across the genome is still a challenge, especially for organisms with massive genomes, but several methods of generating reduced representation libraries have been developed recently (Davey *et al.* 2011; Lemmon & Lemmon 2013). These innovative techniques have enabled individual laboratories to leverage next generation sequencing to answer questions in systematically challenging organisms without prior genetic resources. Difficulties associated with analyzing the hundreds or thousands of loci needed for confident species tree and parameter rich population genetic estimation are substantial, necessitating efficient algorithms and innovations in high memory parallel processing, storage, and access (Muir *et al.* 2016).

Probabilistic models using the multi-species coalescent (MSC) have emerged as one way of processing massive data sets allowing uncertainty from gene-tree discordance (Rannala & Yang 2003; Fujita *et al.* 2012; Leaché *et al.* 2014). The MSC incorporates the stochastic properties of genetic drift among lineages within branches of a species tree (Liu *et al.* 2009). Coalescent models of divergence are derived from population genetic theory which estimate the probability of lineage convergence from the present to common ancestors in the past (Kingman 1982; Rosenberg & Nordborg 2002). Coalescent models are used to estimate the probability of gene trees given a fixed topography, and can be used to calculate the probability of a gene tree given a species tree through the combined probability of all coalescent branching events. The application of coalescent modeling enables investigators to calculate probabilities under competing hypotheses of lineage

composition directly related to hypothetical species limits (Carstens & Dewey 2010). As gene-tree incongruence is expected to be highest during rapid radiations or events which produce non-bifurcating topologies, coalescent models are particularly well suited for studying gene flow and introgression between populations (Rosenberg & Nordborg 2002; Degnan & Salter 2005). Although the MSC assumes selection or assortative mating is absent, using multiple unlinked markers across the genome ensures species tree estimates are objective and less sensitive to violations of the neutral coalescent (Heled & Drummond 2010; Bryant *et al.* 2012; Edwards *et al.* 2016).

1.3 – INTEGRATIVE APPROACHES TO DEFINING SPECIES LIMITS

Integrative approaches utilizing information from the natural history, ecology, and behavior of organisms can lead to more confident estimates of species boundaries when used in concert with molecular genetic data (Rissler & Apodaca 2007). For all of the genetic data that has been generated, it is often unknown if patterns of population genetic variation and substructure are correlated with the evolution of reproductive barriers which promote the divergence of sister lineages. Reproductive isolation is an important criterion for definitively delineating lineages and therefore has a considerable evolutionary role in facilitating or impeding rates of divergence. From a practical perspective, assays of reproductive isolation can serve to inform species limits by characterizing the intrinsic reproductive barriers that may have evolved long after lineages initially began to diverge.

Fundamentally, studies of reproductive isolation can be used as a diagnostic tool to test the adherence of diverging population-lineages to the Biological Species Concept (Mayr 1942). The absolute contribution of any individual isolating barrier to the process of speciation is extremely difficult to test because multiple factors can prevent the unification

of gametes or the formation of reproductively viable offspring, but also because contemporary barriers are not necessarily the causal factors that initially drove speciation (Coyne & Orr 2004). Tests for pre-zygotic sexual isolation are an ideal proxy for reproductive isolation in comparisons of allopatric and non-allopatric populations because investigators can measure the progression of behaviors leading to isolation along the speciation continuum, and the important stages contributing to reproductive incompatibilities.

At the phylogeographic level, reproductive assays may help explain patterns of genealogical discordance, identify traits related to assortative mating, or support inferences of population structure and admixture. Combined with population genetic and species tree analyses aimed at delimiting lineage boundaries between sister taxa, measures of sexual isolation between populations can be used to test hypotheses of evolutionary divergence and help in the identification of contact zones and important geographical barriers or corridors facilitating or preventing gene exchange (Arnold *et al.* 1993).

Investigators examining sexual isolation and speciation between populations can use studies of pre-zygotic reproductive isolation to begin to make inferences about the factors that limit gene flow, strengthening their overall inferences of whether or not populations are independently evolving. When carefully planned, pre-zygotic reproductive isolation experiments can eliminate many of the confounding variables that influence rates of assortative mating in natural populations, such as mate size preferences or sexual interference (Houck *et al.* 1988). However, other important ecological differences may persist.

1.4 – *DESMOGNATHUS* AS A STUDY SYSTEM

The evolutionary relationships between many lineages of *Desmognathus* salamanders are uncertain despite an abundance of research studying their natural history (Dunn 1926; Petranka 1998) and many molecular systematics studies which have been conducted over the last few decades (Tilley & Schwerdtfeger 1981; Karlin & Guttman 1986; Titus & Larson 1996; Highton & Peabody 2000; Chippindale *et al.* 2004). Investigations of genetic variation within *Desmognathus* have uncovered unexpectedly high genetic diversity among population-lineages despite low ecological or morphological disparities (Titus & Larson 1996; Kozak *et al.* 2005; Tilley *et al.* 2008, 2013; Beamer & Lamb 2008). The detection of gene flow at low levels and incomplete sexual isolation between *Desmognathus* species have complicated delimitation efforts (Verrell 1990a; Petranka 1998; Bonett 2002). Allozyme electrophoretic migration patterns have been useful for distinguishing genetic differences between populations (Tilley & Mahoney 1996; Bonett 2002), however these methods are not as amenable to phylogenetic comparisons and these data are difficult to interpret across studies. Mitochondrial DNA (mtDNA) sequences have been utilized extensively for inferring phylogenetic relationships within *Desmognathus* and among plethodontid genera (Mueller 2006), shedding light on the profound genetic diversity among morphologically similar lineages (Tilley & Mahoney 1996; Kozak *et al.* 2005; Kratovil 2007; Beamer & Lamb 2008; Wooten *et al.* 2010). However, phylogenetic resolution has also been confounded by the lack of informative nuclear markers and over reliance on linked mitochondrial genes for phylogenetic inference (Kratovil 2007). The use of multi-locus next generation sequencing to investigate

species and population-level genetic variation may resolve some of the phylogenetic and taxonomic ambiguity that has persisted in this challenging system.

Desmognathus fuscus are a major component of the diverse plethodontid communities distributed throughout eastern North America. Of particular focus in this dissertation is that genetically diverse mtDNA assemblages have been discovered within isolated mountain ranges of the South and Brushy Mountains east of the Blue Ridge Escarpment in the Piedmont of North Carolina (Fig. 1.1; (Kratovil 2007; Tilley *et al.* 2008). MtDNA gene trees indicate that *D. fuscus* populations are often composed of a single predominant mitochondrial clade (Fig. 1.1), and these patterns frequently disagree with allozyme characteristics (Tilley *et al.* 2013). The high sequence similarity observed between *D. fuscus* and *D. carolinensis* mtDNA lineages suggests there are complex historical patterns of contact and asymmetric mtDNA transfer between these two species, which do not currently have overlapping geographic distributions (Tilley *et al.* 2013). These studies suggest that populations morphologically referable to *D. fuscus* are non-monophyletic, however the stochastic nature of genetic drift, and the propensity for mtDNA to introgress across species boundaries (Weisrock *et al.* 2005), lead to the possibility that the mtDNA gene trees we observe may be discordant from the actual species tree history.

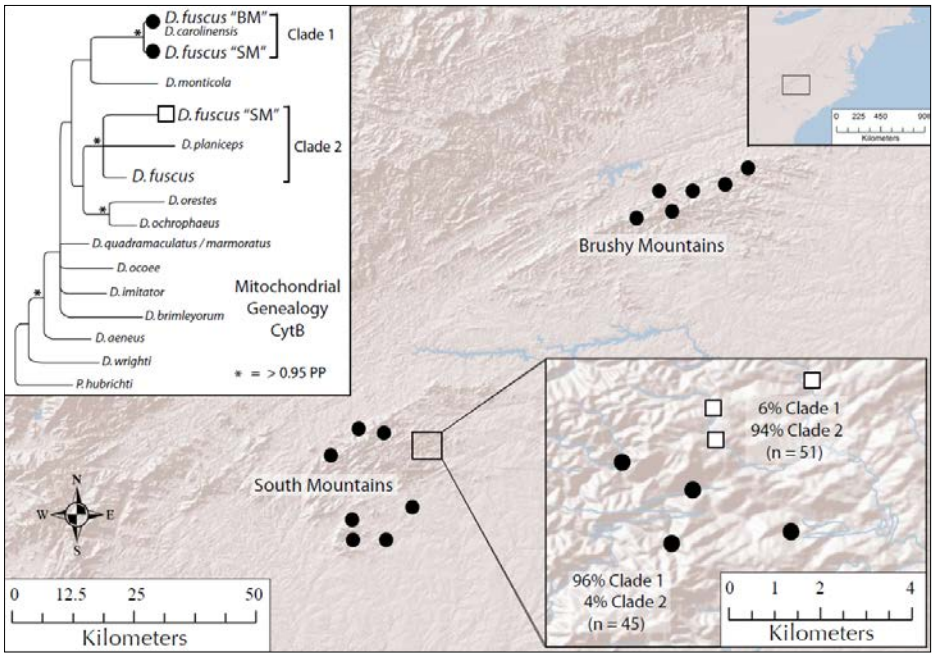


Figure 1.1 – Shaded relief map of *D. fuscus* populations and corresponding mtDNA clades in the western Piedmont of North Carolina

MtDNA clades are indicated by either filled circles (Clade 1) or open squares (Clade 2). (Inset top left) Bayesian gene tree estimated from a 387bp fragment of *cytb* collected from representative Desmognathus taxa. Posterior probabilities greater than 0.95 are indicated by an asterisk.

1.5 – THESIS OBJECTIVES AND STRATEGIES

It is currently unknown whether patterns of mtDNA lineage divergence in *D. fuscus* are: 1) concordant with the underlying evolutionary history of a single cohesive population or multiple reproductively isolated species, 2) the result of biased introgression with *D. carolinensis* (nested within Clade 1; Fig.1.1), or 3) the result of deep coalescence during rapid diversification. To address these basic questions, I have investigated genetic

divergence at the phylogenomic and population-level using multi-locus sequence data collected using two different methods of generating reduced representation genomic libraries. I collected data from both presumptively neutral loci across the genome and separately targeted nuclear loci associated with mitochondrial functions in order to compare evolutionary patterns, and coupled these genetic results with behavioral studies of pre-zygotic sexual isolation between *Desmognathus* populations.

My principle interest in using *Desmognathus* as a study system is that it presents an opportunity to not only resolve persistent questions about species relationships, but to test different models for lineage boundaries, and quantify the support for alternative hypotheses of lineage formation using previously unavailable genome wide sequence data. This research is particular relevant in the study of mtDNA introgression, which has become a commonly observed phenomenon within diverse vertebrate and non-vertebrate taxa. It is generally unknown if and to what degree mtDNA introgression acts as a driving force in nuclear genome evolution and the process of divergence, either at the phylogenetic or population genetic levels. My secondary interest is exploring the effect that mtDNA introgression has on the evolution of complex traits, namely, courtship behavior and pre-zygotic reproductive isolation, which can have an important role in modulating the rate of divergence between lineages.

In the second chapter, I address uncertainty in the phylogenetic relationships between *D. fuscus* populations that contain highly divergent mtDNA lineages. First, I generated a gene tree from longer mtDNA sequence data using a different locus which has more publicly available reference data (*nad2*). Nuclear data was collected using double digest restriction associated DNA sequencing (ddRAD)(Peterson *et al.* 2012), which is one

method of generating short read sequence data from anonymous loci across the genome with minimal investment in library optimization. I used several genetic clustering methods, including algebraic quartet analysis (SVDquartets)(Kubatko 2014) and discriminate analysis of principle components (DAPC)(Jombart *et al.* 2010), to generate hypotheses of lineage formation. I also tested multiple models of lineage formation based on mtDNA clade association and recognized taxonomy using Bayes factor delimitation (BFD)(Leaché *et al.* 2014). Species trees were estimated using the best supported models from BFD. Hypotheses of biased introgression were tested using 4 taxon D-statistics, which measure and compare patterns of shared alleles that deviate from expectations under drift and incomplete lineage sorting. Lastly, we estimated rates of migration between *D. carolinensis* and geographically separated *D. fuscus* populations, and migration rates between populations containing different mtDNA clades.

In the third chapter, I also address uncertainty in the phylogenetic relationships between *D. fuscus* populations that contain highly divergent mtDNA lineages, but instead used a different method of generating nuclear data, used deeper population level sampling, and more representative *Desmognathus* taxa. I used parallel tagged amplicon sequencing (PTAS)(O'Neill *et al.* 2013), that relied on nuclear marker development, individual PCR reactions, and pooled and indexed libraries prior to high throughput sequencing. Again, I used BFD to test multiple models of lineage formation based on mtDNA clade association and recognized taxonomy, and to compare results between different methods. Species trees were estimated using the best supported models from BFD, however, because we sampled representative taxa across the genus, we can more confidently reconstruct relationships among delimited taxa and representative taxa. We compared topological relationships

using SVDquartets and population genetic structure analyses across different nuclear markers; putatively neutral nuclear markers and nuclear encoded mitochondrial genes (NEMGs).

In the fourth chapter, I used experimental enclosures to observe mating behavior during paired crosses and calculate metrics of pre-zygotic reproductive isolation between multiple *D. fuscus* populations, *D. carolinensis,* and another population of uncertain status (*D. sp* Lemon Gap). *D. fuscus* populations were either geographically isolated from one another (South Mountain and Brushy Mountains), or were from the same region (South Mountains) but contained different mtDNA lineages (Clade 1: hereafter the α clade, and Clade 2: hereafter the *D. fuscus* clade). We digitally recorded mating trials to test hypotheses measuring equal durations of mating latency, courtship, "tail straddle walk" (the last observed behavior prior to spermatophore deposition and insemination), or total duration between populations. Behavioral data were used in conjunction with genetic data from previous chapters to address patterns of gene flow, discordance, and the evolution of complex traits.

In the fifth chapter, I summarize results from the previous chapters in the context of the original objectives of this dissertation. Namely, 1) the unresolved phylogenetic relationships among *D. fuscus* populations using nuclear data, 2) concordance and discordance between nuclear and mitochondrial data, 3) the role that mtDNA introgression has on nuclear divergence using patterns revealed by analyzing nuclear genes and NEMGs, and lastly, 4) how this dissertation research contributes to our understanding of species boundaries and the complex process of divergence with gene flow.

Chapter 2 : Model based hypothesis testing biased mtDNA introgression and gene flow using ddRAD sequence data in *Desmognathus*

2.1 – INTRODUCTION

Divergence among species is often a protracted process involving gene flow that can occur long after the initial stages of nascent speciation. This has become particularly evident with increased access to genome-level data, which has led to an abundance of studies identifying reticulate evolutionary histories among species for portions of the genome in a variety of diverse organisms, including birds (Zarza *et al.* 2016), butterflies (Pardo-Diaz *et al.* 2012; The Heliconius Genome Consortium 2012), cicadas (Marshall *et al.* 2011), horses (Jónsson *et al.* 2014), ragworts (Osborne *et al.* 2016), and wild tomatoes (Pease *et al.* 2016). Divergence with gene flow may characterize many of the rapid species radiations within the eukaryotic Tree of Life, and may serve as an important source of localized adaptive genetic variation for populations (Seehausen 2004; Hedrick 2013; Mallet *et al.* 2016).

A commonly identified pattern of post-divergence gene flow is the putatively biased introgression of the mitochondrial genome without corresponding nuclear gene flow (Funk & Omland 2003; Toews & Brelsford 2012). This discordance between genomes has most commonly been seen in phylogeographic studies using a mtDNA marker and a small to moderate number of nuclear markers, where the mtDNA gene tree yields non-monophyletic patterns for species, while nuclear markers yield more-or-less consistent evolutionary patterns (e.g., Sequeira *et al.* 2011; Ruane *et al.* 2014). The mechanisms underlying this biased movement of mitochondrial genomes across species boundaries may

be varied, with hypotheses ranging from relaxed selection on the mitochondrial genome, relative to the nuclear genome (Funk and Omland 2003), to adaptation through capture of a more-fit foreign genome (Sloan *et al.* 2017), to Haldane's rule and the expectation of greater female-facilitated gene flow (and thus, biased mtDNA gene flow) across hybrid zones in XY heterogametic species (Chan & Levin 2005; Patten *et al.* 2015).

Identifying the potential for, and magnitude of, biased mtDNA introgression can be important in systematic or phylogeographic studies given the disproportionate effect it can have on species tree estimation and population parameter estimation, even when using methods developed to accommodate conflicting phylogenetic patterns and independent histories of loci. One solution may be to simply ignore evolutionary information from the mitochondrial genome because of its potential disproportionate contribution to phylogenetic signal (Jockusch *et al.* 2015) or parameter estimation (Ballard & Whitlock 2004). However, excluding or effectively marginalizing mtDNA history can have the undesired consequence of ignoring the evolutionary history shared between coinherited genomes. The propensity for biased mitochondrial introgression may actually provide a signature of historic contact among diverging populations when nuclear markers haven't recorded this dynamic history (Weisrock *et al.* 2005).

While many studies have identified putative histories of biased mtDNA introgression, few have yet to thoroughly test this hypothesis, or assess the magnitude of discordance, using broad sampling of the nuclear genome (see Bernal *et al.* 2016). These patterns may represent true differences in the movement of these genomes across contact zones. Alternatively, they may be an artifact of limited sampling of nuclear markers, which have missed picking up the more general signal of gene flow across species boundaries.

Quantifying differences between competing evolutionary hypotheses using genomic data will contribute to a more complete understanding of the mechanisms driving discordance, the degree to which mitochondrial evolutionary history shapes reticulations in the nuclear genome, and how introgression may facilitate or serve as a product of post divergent gene flow.

Rapid radiations are notoriously tough to phylogenetically disentangle, and *Desmognathus* salamanders, which are a major component of salamander communities distributed throughout eastern North America (Petranka 1998; Rissler & Taylor 2003; Wake 2009), are no exception (Kozak *et al.* 2005; Tilley *et al.* 2013). Rapid diversification within this clade, possibly in response to the reacquisition of biphasic development (Chippindale *et al.* 2004), is associated with the evolution of ecomorphologies partitioned by life history and habitat use that describe most fully aquatic, semi-aquatic, and terrestrial desmognathine taxa (Titus & Larson 1996; Kozak & Wiens 2006; Bruce 2010; Blankers *et al.* 2012). MtDNA based studies have identified multiple non-monophyletic lineages within *D. fuscus,* a medium sized semi-aquatic species inhabiting lower elevations (Kozak *et al.* 2005; Kratovil 2007; Beamer & Lamb 2008; Tilley *et al.* 2013).

Emblematic of the challenges associated with disentangling *Desmognathus* phylogenies, some *D. fuscus* lineages may have exchanged mitochondrial haplotypes with *D. carolinensis*, a smaller bodied montane species found near springs and seeps with a more restricted range distribution and terrestrial ecology (Mead *et al.* 2001). For example, there appears to be minimal divergence between sequenced regions of the mitochondrial *cytb* locus comparing *D. carolinensis* and *D. fuscus* individuals from populations in eastern TN referable as Sinking Creek form (SCF) (0.028 mean nucleotide base differences per

site; range 0 – 0.054). Likewise, comparing sequenced regions of *cytb* between *D. carolinensis* and *D. fuscus* individuals from populations in the Piedmont of North Carolina are also minimally divergent (0.041 mean nucleotide base differences per site; range 0.031 – 0.047). These data suggest a more complex history of mitochondrial introgression between *D. carolinensis* and *D. fuscus* populations on eastern and western sides of the southern Appalachians (Tilley *et al.* 2013).

In stark contrast, while mtDNA patterns are suggestive of introgression between *D. carolinensis* and *D. fuscus* populations, nuclear genetic data in the form of fixed differences in allozyme electrophoretic migration patterns across 22 presumptive loci suggest limited gene flow between taxa. Nei unbiased genetic distances between *D. carolinensis* and (1) *D. fuscus* SCF are rather high, ranging from 0.39 to 0.80; (2) *D. fuscus* individuals in the Piedmont of North Carolina are moderate, ranging from 0.29 to 0.36, and (3) *D. fuscus* individuals from populations in Massachusetts are elevated, ranging from 0.32 to 0.55 (Tilley *et al.* 2013). No nuclear genetic sequence data have been generated to provide a nuclear perspective on these peculiar patterns.

In this study we use genome wide single nucleotide polymorphisms (SNPs) and short read loci generated from double digest restriction site associated DNA (ddRAD) sequencing to test models of lineage formation, population structure, and gene flow between *D. fuscus* and *D. carolinensis*; currently non-sympatric species which we hypothesize have undergone post divergent mitochondrial introgression despite apparent ecomorphological disparities. Our sequencing efforts are focused on *D. fuscus* populations in the isolated South Mountains and Brushy Mountains within the Piedmont physiographic region of North Carolina. Many individuals from these populations share highly similar

mtDNA haplotypes that nest among *D. carolinensis* mtDNA lineages and are referable as the α clade (Kozak *et al.* 2005; Kratovil 2007; sensu Tilley *et al.* 2013). For a more comprehensive assessment of nuclear patterns in *D. fuscus*, we include individuals from populations across the northern and central portions of the species range, referable as either the *D. fuscus* clade or α clade mtDNA lineages, including SCF present in the western edge of the southern Appalachians (Tilley *et al.* 2013).

We are primarily interested in quantifying competing (but not necessarily mutually exclusive) hypotheses of lineage assignment using Bayes factor Delimitation (BFD) (Grummer *et al.* 2014; Leache *et al.* 2014) to determine if the most likely model defining lineage boundaries are consistent with "discovery" based gene clustering methods inferred using SVDquartets (Chifman & Kubatko 2014) and PCA based discriminate analysis (DAPC)(Jombart *et al.* 2010), *a priori* assignments based on recent or ongoing mtDNA introgression, currently recognized taxonomy, or geographic patterns. We also test for historic signatures of biased introgression between *D. carolinensis* and either similar or dissimilar mtDNA lineages of *D. fuscus* using paired D-statistical tests (e.g., ABBA/BABA tests)(Durand *et al.* 2011a) and migration parameter estimates using IMa2p (Sethuraman & Hey 2016). This study provides important quantitative insights into likelihood differences between competing hypotheses of lineage membership using genome wide data and the forces promoting rapid adaptive radiations within salamander communities. This research contributes to a body of recent literature studying the post divergent process of speciation and the reticulate evolutionary patterns shared between mitochondrial and nuclear genomes.

2.2 – MATERIALS AND METHODS

2.2.1 – Sample collection

We sequenced a total of 46 individuals for this study (Fig. 2.2, Table 2.1), with localities chosen based on previous mtDNA sequencing of *cytb* that identified several populations containing *D. fuscus* clade and α clade lineages (Tilley *et al.* 2013). 25 of these individuals were collected from Piedmont *D. fuscus* populations in the Brushy Mountains and South Mountains of North Carolina, with an additional 13 individuals sampled from across the *D. fuscus* range including Massachusetts, Kentucky, Tennessee, and Virginia. To evaluate historical gene flow and introgression between currently non-sympatric Piedmont *D. fuscus* and *D. carolinensis*, we included 8 individuals of *D. carolinensis*, collected 4 km from its type locality (Yancey County, NC). All salamanders used in this study were collected by JK and SGT.

2.2.2 – MtDNA sequence data generation

DNA was extracted using a Qiagen DNeasy Blood and Tissue kit following the manufacturer supplied protocol. To verify and attribute lineage assignments we amplified the mitochondrial *nad2* gene (~1041 bp) for all samples using PCR primers and protocols from Weisrock *et al.* (2001). MtDNA sequence data were generated from individuals using either: (1) Sanger sequencing on an ABI 3730 sequencer, or (2) parallel tagged amplicon sequencing (PTAS; O'Neill *et al.* 2013) using Nextera XT libraries and an Illumina MiSeq with 250 bp paired-end sequencing. Consensus sequences for each individual were assembled using the Muscle aligner in Geneious v6.1.8. We combined *nad2* sequences from 46 individuals from this study with 4 additional *Desmognathus* sequences collected

Table 2.1 – Sample localities and individuals used to generate ddRAD libraries

Locality	Specimen ID	Designation	State	County	Latitude	Longitude
1	JK 1033	*Desmognathus fuscus*	KY	Rockcastle	37.396538	-84.292358
2	JK 1036	*Desmognathus fuscus*	KY	Rockcastle	37.368591	-84.211845
3	SGT 35147	*Desmognathus fuscus*	VA	Roanoke	37.1826	-80.1397
4	SGT 3515	*Desmognathus fuscus*	VA	Giles	37.35085	-80.601
5	SGT 3516	*Desmognathus fuscus*	VA	Rockbridge	37.970913	-79.464053
6	SGT 3520	*Desmognathus fuscus*	VA	Carroll	36.612083	-80.7713
7	JK 12372	*Desmognathus fuscus*	MA	Franklin	42.57225	-72.922033
7	SGT 35636	*Desmognathus fuscus*	MA	Franklin	42.57225	-72.922033
8	SGT 35428	Sinking Creek form	TN	Greene	36.1251	-82.6455
9	SGT 35467	Sinking Creek form	TN	Unicoi	36.1768	-82.281533
9	SGT 35469	Sinking Creek form	TN	Unicoi	36.1768	-82.281533
10	SGT 35619	Sinking Creek form	TN	Carter	36.33345	-82.27547
11	SGT 35675	Sinking Creek form	TN	Washington	36.34665	-82.308983
12	JK 13112	*Desmognathus carolinensis*	NC	Yancey	35.726307	-82.243401
12	JK 13113	*Desmognathus carolinensis*	NC	Yancey	35.726307	-82.243401
13	JK 13117	*Desmognathus carolinensis*	NC	Yancey	35.727383	-82.241486
13	JK 13140	*Desmognathus carolinensis*	NC	Yancey	35.727383	-82.241486
13	JK 13145	*Desmognathus carolinensis*	NC	Yancey	35.727383	-82.241486
13	JK 13147	*Desmognathus carolinensis*	NC	Yancey	35.727383	-82.241486
14	JK 13130	*Desmognathus carolinensis*	NC	Yancey	35.740059	-82.231735
14	JK 13141	*Desmognathus carolinensis*	NC	Yancey	35.740059	-82.231735
15	JK 13010	Piedmont “B”	NC	Burke	35.610933	-81.626383
15	JK 13046	Piedmont “B”	NC	Burke	35.610933	-81.626383
15	JK 13049	Piedmont “B”	NC	Burke	35.610933	-81.626383
15	JK 13065	Piedmont “B”	NC	Burke	35.610933	-81.626383
15	JK 13067	Piedmont “B”	NC	Burke	35.610933	-81.626383
15	JK 13100	Piedmont “B”	NC	Burke	35.610933	-81.626383
15	JK 13101	Piedmont “B”	NC	Burke	35.610933	-81.626383
15	JK 13102	Piedmont “B”	NC	Burke	35.610933	-81.626383
15	JK 13104	Piedmont “B”	NC	Burke	35.610933	-81.626383
16	JK 13082	Piedmont “B”	NC	Burke	35.612199	-81.622226
17	JK 13001	Piedmont “B”	NC	Burke	35.628617	-81.694967
17	JK 13053	Piedmont “A”	NC	Burke	35.628617	-81.694967
17	JK 13054	Piedmont “A”	NC	Burke	35.628617	-81.694967
17	JK 13069	Piedmont “A”	NC	Burke	35.628617	-81.694967
17	JK 13083	Piedmont “A”	NC	Burke	35.628617	-81.694967
18	JK 13017	Piedmont “A”	NC	Burke	35.606711	-81.647785
18	JK 13052	Piedmont “A”	NC	Burke	35.606711	-81.647785
18	JK 13055	Piedmont “A”	NC	Burke	35.606711	-81.647785
19	JK 13012	Brushy Mountain	NC	Wilkes	36.07151	-81.176845
19	JK 13021	Brushy Mountain	NC	Wilkes	36.07151	-81.176845
20	JK 13014	Brushy Mountain	NC	Wilkes	36.116072	-81.128333
20	JK 13015	Piedmont “A”	NC	Wilkes	36.116072	-81.128333
20	JK 13056	Piedmont “A”	NC	Wilkes	36.116072	-81.128333
20	JK 13060	Piedmont “A”	NC	Wilkes	36.116072	-81.128333
21	JK 13036	Brushy Mountain	NC	Wilkes	36.088783	-81.174227

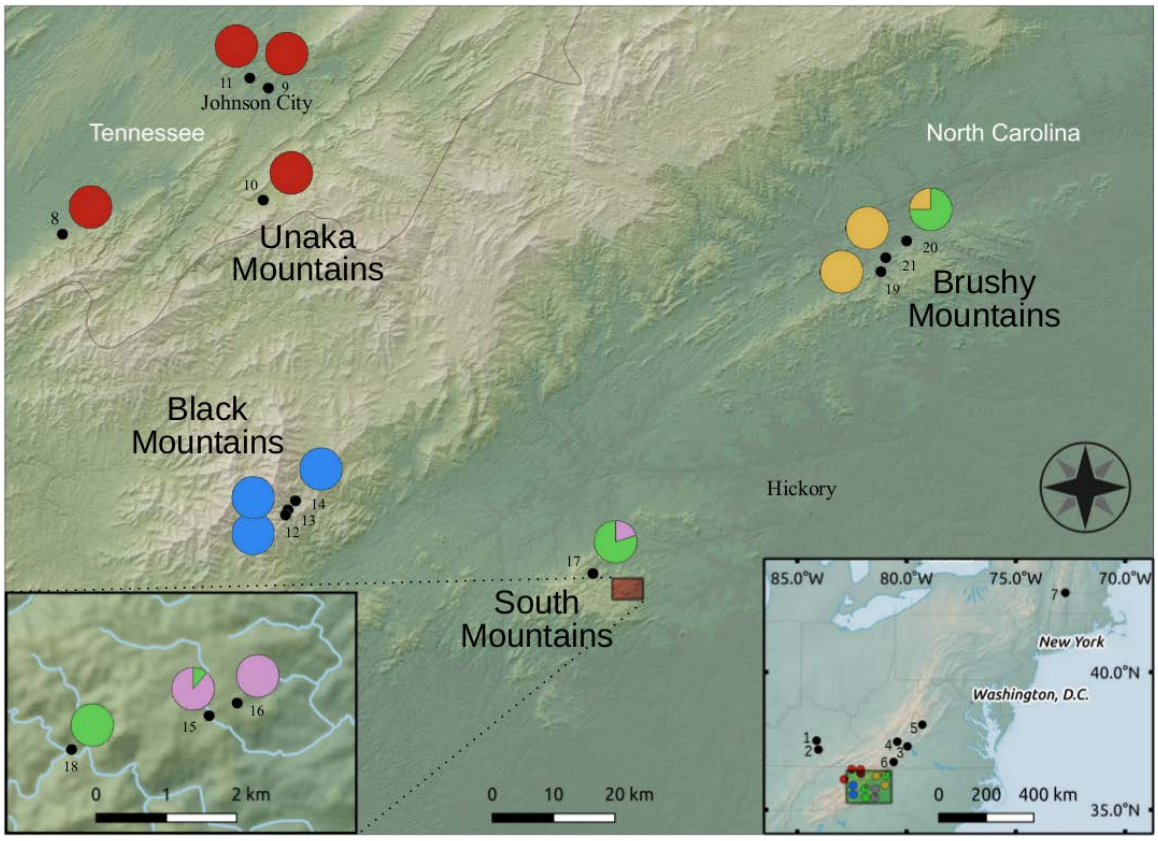


Figure 2.2 – Shaded relief map of *Desmognathus* collection localities and mtDNA *nad2* haplotypes.

Shaded relief map of western North Carolina with numbered collection localities and filled circles indicating proportion of *nad2* mtDNA lineages identified at that location. Inset left are localities in the South Mountains (red highlighted box) illustrating an abrupt boundary between *nad2* lineages. Inset right are numbered localities of *D. fuscus* from the northern and central distribution of the species range in eastern North America. The green highlighted box is the location of the primary map. Filled circles indicate lineage designations based on mtDNA haplotype and taxonomy. Numbers correspond to collection localities in Table 2.1.

by JK and SGT, and 127 sequences from *Desmognathus* downloaded from GenBank (Table S2.1). Uncorrected substitution rates per site after the removal of gaps, missing data, and ambiguous positions (were calculated for 51 haplotype comparisons across *D. fuscus* and *D. carolinensis* mtDNA clades, including within and between pairwise average distances (Table S2) using the program Mega 5.2.2 (Tamura *et al.* 2011).

2.2.3 – MtDNA gene tree estimation

We estimated the best-fitting substitution model for the *nad2* data using PartitionFinder (2.1.1) (Guindon *et al.* 2010; Lanfear *et al.* 2012), which identified a single model and subset for all codon positions. Corrected Akaike Information Criterion (AICc) favored a GTR+I+G model. Gene tree estimation using maximum likelihood was performed using RAxML v8.2.8. ML analyses used the GTR+Γ model without using a parameter for the proportion of invariable sites, as recommended by the program authors. Tree search was performed using 1000 random starting trees. The optimal number of rapid bootstrap searches was determined using the MRE-based Bootstopping criterion option within RAxML, which ran 350 bootstrap replicates. Gene tree estimation was also performed using a Bayesian approach in MrBayes v3.1.2 (Ronquist *et al.* 2012). Analyses were performed using four Markov chains run for 5 million generations and sampled every 1000 generations. A single partition analyses were performed using a GTR+I+G substitution model. Default uniform or flat prior probability distributions were used for most parameters along with random starting trees. Four independent replicates were performed and we assessed convergence through the comparisons of posterior distributions in Tracer v1.6 (Rambaut *et al.* 2014). We discarded the first 25% of samples from each

replicate analysis as burnin and then combined the samples across runs. We then generated a 50% majority-rule consensus tree from the combined posterior distributions (Fig. 2.3).

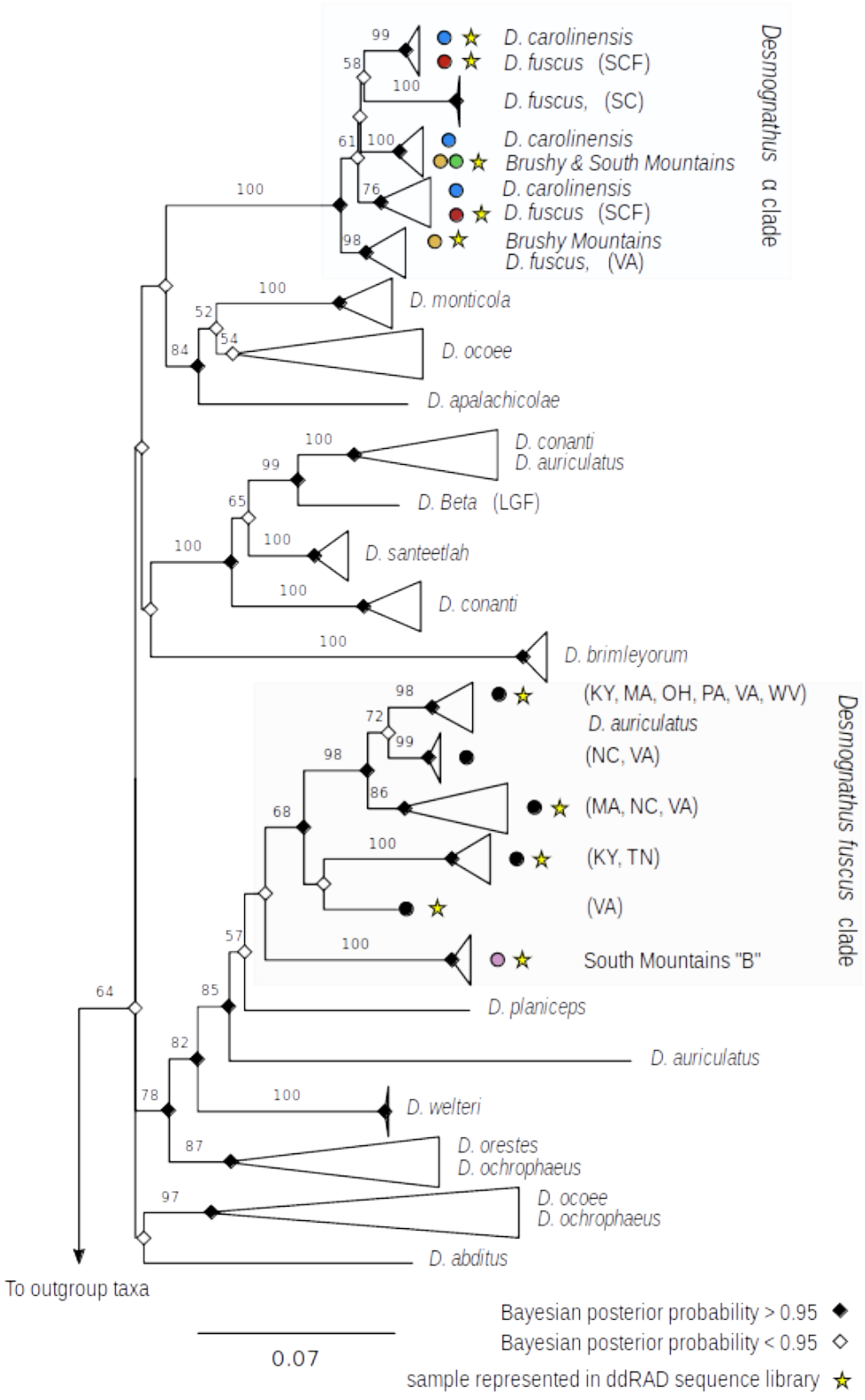


Figure 2.3 – Best ML tree for *nad2* with Bayesian posterior probabilities.

Filled diamonds indicate nodes with ≥0.95% posterior probability support; unfilled diamonds indicate nodes with < 0.95% posterior probability support. The pale blue box indicates *Desmognathus* α clade lineages while the gray box highlights the *Desmognathus fuscus* clade. Filled circles indicate the following *nad2* mtDNA lineages: (blue) *D. carolinensis* α clade, (red) *D. fuscus* Sinking Creek form α clade, (green) *D. fuscus* South Mountain α clade, (orange) *D. fuscus* Brushy Mountain α clade, (black) *D. fuscus* populations from the northern and central distribution of the species range *D. fuscus* clade, (orchid) *D. fuscus* South Mountain *D. fuscus* clade.

2.2.4 – ddRAD library preparation and sequencing

We prepared ddRAD sequencing libraries following a protocol slightly modified from (Peterson *et al.* 2012), which accounted for the substantially larger genome size of *Desmognathus*. This modification involved using a total of 1.5 ug of DNA instead of the original protocol recommendation of 200-500 ng. DNA from each individual was digested using the restriction enzyme combination of EcoRI and SphI in individual 50 uL reaction volumes. Digested DNA was cleaned using AMPure XP beads and individual in-line and indexing Illumina sequencing oligos were then ligated onto the resulting DNA fragments. Libraries were combined into six separate pools and each was size selected for fragments with a mean insert length of 376 ±10% (ranging from 338 to 414bp) using a Pippin Prep (Sage Science). We size selected this region after initial restriction enzyme digestion tests showed multiple potentially repetitive regions and additional genomic fragments in regions >426bp. Size selected fragments were amplified using a high-fidelity DNA polymerase (Bio-Rad) to increase concentration prior to sequencing. We verified in silico (using

Geneious) that the double restriction enzyme digest combination would produce mtDNA fragments outside the selected length range of this study using published whole mitochondrial genome sequence data from *D. fuscus* (GenBank: AY728227 from Mueller *et al.* (2004). All sequencing libraries were combined into a single final pool. Paired-end 150 bp sequencing was performed on two lanes of an Illumina HiSeq 2500 at the Florida State University College of Medicine.

2.2.5 – Bioinformatics

We used the bioinformatic pipeline pyRAD (Eaton & Ree 2013) to generate short read loci and SNP datasets for downstream analyses. PyRAD uses global alignment clustering of sequences (allowing for insertion-deletion variation) using sequence similarity thresholds to cluster sequences within individuals and across distantly related samples. Forward and reverse reads were demultiplexed using unique barcodes to produce 178,527,849 paired-end reads between 262bp and 338bp in length, with an average of 3.88 million reads per individual (range = 705,000 to 7,513,000).

The range of short size selected fragments in our ddRAD library produced a substantial proportion of sequencing overlap between mated pairs, therefore we merged together ~82 million (0.4592) of the paired-end sequences less than 270 bp using the paired-end read merger PEAR (Zhang *et al.* 2014), following the default parameters of the program. Approximately 96 million paired-end sequences were not merged (0.5381). Merged and unmerged paired-end reads were separately quality filtered. Sites with quality scores < 20 were converted to Ns and reads with > 10 Ns were discarded. Approximately 0.5 million reads were discarded due to poor quality (0.0027). Non-overlapping forward

and reverse reads were concatenated for consensus clustering. We used a threshold of 0.85 sequence similarity for within individual clustering. Candidate loci were filtered of potential paralogs and highly repetitive markers by culling loci with excessive heterozygosity and sequencing depth. Putative loci retained during within-individual clustering required a minimum depth of five reads, a maximum depth of five hundred reads, the presence of only diploid alleles, and no more than 10 polymorphic sites per locus, which equates to a polymorphism rate less than 10/242, or 0.0413 in the shortest merged consensus sequence. Each individual had on average 54,096 unique loci (range 23,360 - 67,724) with a mean depth of 29.4 reads per locus. Singletons and unique loci with less than 5x coverage account for approximately 59% of the total data and these reads were removed from downstream analyses. Ultimately, we identified 139,431 unique loci that matched multiple individuals after final filtering.

We used a threshold of 0.85 sequence similarity for clustering ortholog loci together among individuals. Loci were discarded as paralogous if more than 0.15 of individuals shared a heterozygous site across a locus, which may be produced when paralogs cluster together. Given the potentially diverse *Desmognathus* taxa used in this study, we used the discontiguous megablast function in Geneious to compare putative ddRAD loci to the full mitochondrial genome for *D. fuscus* to ensure loci were not of mitochondrial origin.

We did not exhaustively explore threshold parameters or other filtering strategies, but when using a higher threshold of 0.90 sequence similarity for within and among individual clustering, we observed a consistent pattern of additional unique loci being generated per individual with more loci being discarded after final filtering (singletons/low depth). More conservative filtering using a minimum of 10x coverage depth for loci

produced on average 47 thousand reads with a mean depth of 37x across individuals. These strategies did not produce noticeably different results when we used basic population genetic clustering analyses, therefore we only present data for which we used a 5x minimum read depth for loci.

We generated phased sequence alignments for each locus and data sets containing a single randomly sampled SNP from each locus. Two different SNP data sets were assembled that had different levels of missing data and individual sampling. One of these contained a set of 160 SNPs without any missing data across all 46 sampled individuals, hereafter referred as the "46x160 SNP" dataset. The second SNP data set comprised 1380 SNP loci sampled from a subset of 26 individuals that had, at most, one missing individual per locus, hereafter referred as the "26*x1380 SNP" dataset. This dataset included at least one individual from each of the eight OTUs evaluated in downstream analyses and contained less than 0.04 total missing data across individuals.

2.2.6 – Nuclear-based lineage discovery

As a first step in assessing the history of lineage divergence in the nuclear genome, we generated a coalescent-based lineage tree using the program SVDquartets (Chifman & Kubatko 2014) implemented in PAUP* v4.146 (Swofford 2003). This approach treat tips in the tree as the random pairing of gene copies for individuals, and thus can serve as an exploratory method for exploring divergence among clusters of individuals and populations. SVDquartets is tolerant of missing data; therefore, we used a set of 16,173 unlinked SNPs from the full 46 individual data set. Exhaustive sampling was used with the

QFM quartet assembly algorithm. Branch support was evaluated using 1000 bootstrap replicates.

We also used a discriminate analysis of principal components analysis (DAPC)(Jombart *et al.* 2010) implemented in the adgenet R package (Jombart 2008) to assess how our nuclear data were genetically structured without using a prior assignment of individuals to populations or species. DAPC partitions genetic variation by minimizing differences within groups while maximizing differences between groups. K-means clustering is used to identify the number of discrete clusters and then probabilistically assign individuals to them. We performed DAPC using both the 46x160 SNP and 26*x1380 SNP data sets. We used a cross validation method in the poppr R package (Kamvar *et al.* 2014) to identify the maximum number of principle components (PCs) to retain. This used 1000 replicates and a range of PCs from 1 to 35. Individuals were plotted in ordination space using the ade4 R package (Dray & Dufour 2007).

2.2.7 – Nuclear-based model testing

We used Bayes factor delimitation analyses to identify the prevailing history of lineage divergence recorded in the nuclear genome, and determine if the nuclear genome fits a model of divergence similar to, or different from, the mitochondrial genome. To develop the relevant set of models to test, we combined patterns in mtDNA gene trees, exploration of lineage divergence and genetic structure (SVDquartets, DAPC), and general geography to identify the most exclusive set of tips that could be present in an overall species tree for our study system. This resulted in eight hypothetical taxon partitions, or OTUs, that describe the most complex, or parameter rich, model tested (Table 2.2): (1) *D.*

fuscus from KY, (2) *D. fuscus* from MA, (3) *D. fuscus* from VA, (4) *D. fuscus* from TN representing the SCF that contain mtDNA from the α clade, (5) *D. fuscus* from the South Mountains that contain mtDNA from the α clade, (6) *D. fuscus* from the South Mountains that contain mtDNA from the *D. fuscus* clade, 7) *D. fuscus* from the Brushy Mountains that contain mtDNA from the α clade, and 8) *D. carolinensis*, which contain mtDNA from the α clade.

These eight hypothesized tips permitted the testing of 17 models that covered a range of hypotheses regarding divergence and connectivity across our study system. For example, we tested models in which *D. carolinensis* was treated as a lineage separate from other populations in the *D. fuscus* α clade, which would be consistent with a history of biased mtDNA introgression between these groups. Alternatively, models combining *D. carolinensis* with one or more *D. fuscus* α clade population are consistent with histories that lack divergence between these groups and not of mito-nuclear discord in gene flow. Models also explored the combining or splitting of Piedmont *D. fuscus* by mtDNA lineage (α clade vs. *D. fuscus* clade) and by geographical isolate (South Mountain vs. Brushy Mountain). Finally, models tested the distinctiveness of the *D. fuscus* SCF. We do not provide a full enumeration of each model here, and instead specific descriptions of all models can be found in Supplementary Materials (S2.3).

We analyzed our SNP data in a coalescent-based species tree framework and calculated marginal likelihoods for each of the 17 models using the SNAPP module in BEAST v2.3.2. Analyses were performed using both the 46x160 SNP and 26*1380 SNP

Table 2.2 – Bayes factor delimitation models tested.

Color scheme follows Figure 2.3. Shaded boxes indicate combined individuals. SM = South Mountains, BM = Brushy Mountains, SCF = Sinking Creek form. A = α clade, B = *D. fuscus* clade. Individuals per partition are in parentheses (n).

Taxonomy		*D. fuscus*	*D. fuscus*	*D. fuscus*	*D. fuscus*	*D. fuscus*	*D. fuscus*	*D. fuscus*	*D. carolinensis*
Geography		KY	VA	MA	TN: SCF	South Mountains	South Mountains	Brushy Mountains	Black Mountains
mtDNA clade		"*D. fuscus* clade"	"*D. fuscus* clade"	"*D. fuscus* clade"	"α clade"	"α clade"	"*D. fuscus* clade"	"α clade"	"α clade"
Model	**Partitions**	**Assignment Partition (n)**							
1	8	*D. fuscus* KY (2)	*D. fuscus* VA (4)	*D. fuscus* MA (2)	SCF (5)	Piedmont A (8)	Piedmont B (10)	Piedmont BM (7)	*D. carolinensis* (8)
2	7	*D. fuscus* KY (2)	*D. fuscus* VA (4)	*D. fuscus* MA (2)	SCF (5)	Piedmont SM (18)		Piedmont BM (7)	*D. carolinensis* (8)
3	7	*D. fuscus* KY (2)	*D. fuscus* VA (4)	*D. fuscus* MA (2)	SCF (5)	Piedmont B (10)	Piedmont A, BM (15)		*D. carolinensis* (8)
4	6	*D. fuscus* KY (2)	*D. fuscus* VA (4)	*D. fuscus* MA (2)	SCF (5)	Piedmont A, B, BM (25)			*D. carolinensis* (8)
5	6	*D. fuscus* KY, VA, MA (8)			SCF (5)	Piedmont A (8)	Piedmont SMB (10)	Piedmont BM (7)	*D. carolinensis* (8)
6	5	*D. fuscus* KY, VA, MA (8)			SCF (5)	Piedmont SM (18)		Piedmont BM (7)	*D. carolinensis* (8)
7	5	*D. fuscus* KY, VA, MA (8)			SCF (5)	Piedmont B (10)	Piedmont A, BM (15)		*D. carolinensis* (8)
8	4	*D. fuscus* KY, VA, MA (8)			SCF (5)	Piedmont A, B, BM (25)			*D. carolinensis* (8)
9	5	*D. fuscus* KY, VA, MA (8)			SCF (5)	Piedmont A (8)	Piedmont SMB (10)	Piedmont BM, *D. carolinensis* (15)	
10	4	*D. fuscus* KY, VA, MA (8)			SCF (5)	Piedmont B (10)	Piedmont A, BM, *D. carolinensis* (23)		
11	3	*D. fuscus* KY, VA, MA (8)			SCF (5)	Piedmont A, B, BM, *D. carolinensis* (33)			
12	3	*D. fuscus* KY, VA, MA, SCF (13)				Piedmont A, B, BM (25)			*D. carolinensis* (8)
13	3	*D. fuscus* KY, VA, MA (8)			SCF & Piedmont A, B, BM (30)				*D. carolinensis* (8)
14	3	*D. fuscus* KY, VA, MA (8)			Piedmont B (10)	SCF, Piedmont A, Piedmont BM, *D. carolinensis* (28)			
15	2	*D. fuscus* KY, VA, MA, Piedmont B (18)				SCF, Piedmont A, Piedmont BM, *D. carolinensis* (28)			
16	2	*D. fuscus* KY, VA, MA (8)			SCF, Piedmont A, B, BM, *D. carolinensis* (38)				
17	2	*D. fuscus* KY, VA, MA, SCF, Piedmont A, B, BM (38)							*D. carolinensis* (8)

datasets and for each model and set of data we performed two independent replicate analyses. Priors for mutation rates (u and v) were estimated directly from allele frequencies for all SNP loci and were analyzed with a normal distribution using an upper bound of 10 for the 46x160 dataset, an upper bound of 100 for the 26*x1380 dataset, and a lower bound of 0.0 for both analyses (based on initial short runs). Marginal likelihoods were calculated using path sampling with 100 steps and 500,000 MCMC generations per step. Samples were drawn every 1000 generations after a 25% burnin. Convergence of the MCMC on a stable posterior distribution was performed through the comparison of the first path sampling step across replicate analyses using the program Tracer v1.5. Marginal likelihoods from each model were used to calculate Bayes factors for each model and determine the best-fitting model. Interpretations of the strength of support were based on (Kass & Raftery 1995). Finally, we generated a maximum clade credibility tree for the best fitting model using TreeAnnotator and we visualized the posterior distribution of trees using DensiTree v2.2.3. In all cases we used the posterior distribution estimated under step 0 of the path sampling analysis.

2.2.8 – Tests of historical admixture

We tested for signatures of historical admixture events using a four-taxon *D* statistic (i.e., the ABBA BABA test; (Durand *et al.* 2011) implemented in pyRAD. Lineage sorting under drift is expected to fix equal proportions of ancestral alleles between derived lineages (P1 and P2), with deviations in ABBA : BABA patterns resulting from biased gene flow or reticulations between a common lineage (P3) and a derived lineage relative to an outgroup

(P4) (Fig 2.4). This analysis assumes that there is no current gene flow between derived lineages, which would homogenize allelic variation rendering the test non-informative.

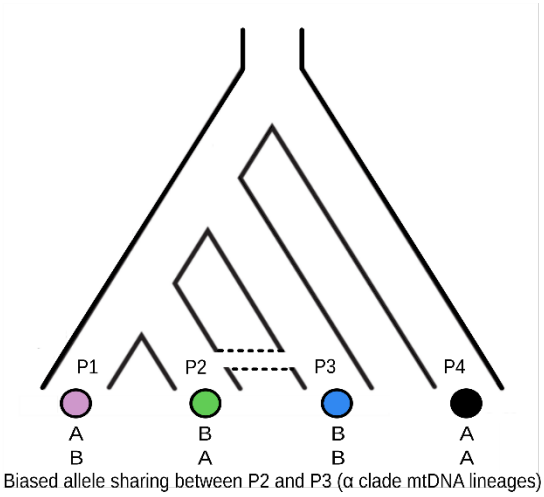


Figure 2.4 – Patterson's Four-taxon D-statistic

A schematic of an ABBA BABA test of introgression between (P1) Piedmont B or (P2) Piedmont A and (P3) *D. carolinensis* which contain similar mtDNA as Piedmont A *D. fuscus*, relative to (P4) outgroup *D. fuscus* from populations in the northern and central species range. Illustrated here is hypothetical historical introgression between P2 and P3, which may be responsible for producing biased ABBA vs. BABA site frequency patterns in nuclear alleles.

We evaluated eight different models for genome-wide signatures of ancestral introgression comparing multiple *D. fuscus* populations east and west of the southern Appalachians with either *D. carolinensis* or *D. fuscus* which contain similar or dissimilar mtDNA (the α clade or *D. fuscus* clade) to test hypotheses of biased nuclear introgression

correlated with sharing a common mtDNA lineage. For example, we tested whether *D. fuscus* individuals from the South Mountains containing (P1) mtDNA from the *D. fuscus* clade or (P2) mtDNA from the α clade, share equal proportions of alleles with (P3) *D. carolinensis* individuals that contain mtDNA from the α clade, relative to (P4) *D. fuscus* from populations in the northern and central distribution of the species range. We do not provide a full enumeration of each model here, and instead specific descriptions of all models and individuals used as test replicates can be found in Supplementary Materials (S2.4).

Instead of testing each D statistic model using a pool of 46 individuals across a limited number of loci, we generated a larger dataset of loci using fewer individuals to balance test replication with more robust hypothesis testing of historic introgression across the genome. We generated 2442 ddRAD loci from 21 individuals, 18 of which were chosen randomly from across 6 OTUs (3 each) assigned to P1, P2, and P3 (treating *D. fuscus* from Massachusetts, Virginia, and Kentucky as a single taxon). We included up to 3 additional *D. fuscus* individuals for the outgroup (P4), which we pooled together to reduce the total number of tests and missing data. This dataset contained less than 0.04 total missing data across individuals. Lastly, we evaluated all four-taxon models derived from combinations of the P1, P2, P3, and P4 groups allowing heterozygous sites, and we used 1000 bootstrap replicates per model to assess significance.

2.2.9 – Population summary statistics

We describe genetic variation among delimited populations using the summary statistic Fst. Fst is the fixation index of within-population genetic variation compared to

the total genetic variance ranging from 1 to 0, where 0 is admixture and 1 is the absence of shared alleles among populations or taxa. We evaluated genetic variation using weighted and mean Weir and Cockerham Pairwise Fst estimates across all population-lineages using 65,307 variable sites from among 2442 ddRAD loci for comparative purposes. We assume these ddRAD loci are distributed genome wide, are unlinked for the purposes of this analysis, are not under selection, and that individuals sampled in this study are not directly related as to influence ascertainment biases in any way.

2.2.10 – Estimates of migration

We used the parallel version of the program IMa2p (Sethuraman & Hey 2016) to estimate all parameters for divergence times (t), migration rates (q), and ancestral population sizes (Ne) between 5 populations that have likely experienced post divergent gene flow or historic introgression: (1) South Mountain *D. fuscus* containing mtDNA from the α clade, (2) South Mountain *D. fuscus* containing mtDNA from the *D. fuscus* clade, (3) Brushy Mountain *D. fuscus* containing mtDNA from the α clade, (4) *D. fuscus* SCF containing mtDNA from the α clade, and (5) *D. carolinensis* containing mtDNA from the α clade. We excluded individual *D. fuscus* from Massachusetts, Virginia, and Kentucky in an effort to decrease the computational time needed for parameter estimates, but otherwise use the species tree generated from our best supported model using BFD as a guide tree for parameter estimation. We used phased output sequences generated by pyRAD for 187 loci across 38 samples without missing data as our input. Loci were formatted using the program Imgc (Woerner *et al.* 2007) and custom scripts to correct allelic violations of the Infinite Sites evolutionary model (which includes recombination within loci). We explored

several parameters in our initial runs to establish sufficient burn-in, adequate mixing of 160 chains, and optimal starting parameters using upper boundaries defined by Watterson's theta estimates calculated across loci in R using the program package pegas (v0.9) (Paradis 2010). We executed 14 separate runs over 24-48 hours each, discarding the first 48 hours of generation step estimates as burnin before sampling genealogies. We evaluated successful convergence of parameter estimates in our marginal likelihood estimates by pooling together 13,267 genealogies and examining linear divergence time plots, with ESS values >1000, high swap rates across chains (>.50-.80), and single peaks for parameter estimates. We present preliminary migration estimates between populations which we will use to guide model tests of migration in future studies when additional genealogies have been generated.

2.3 – RESULTS

2.3.1 – MtDNA gene tree

Gene trees produced by maximum likelihood and Bayesian analyses were largely in agreement in topology for branches with posterior probabilities ≥ 0.95 and bootstrap values ≥ 70. We present the best ML tree with both posterior probabilities and bootstrap values mapped to concordant branches (Fig. 2.1).

The *nad2* gene tree generated here is largely in agreement with analyses of *nad2* data from previous studies (Kozak *et al.* 2005), and has similar topological patterns to those estimated from *cytb* data (Kratovil 2007; Tilley *et al.* 2013). Haplotypes sequenced from *D. fuscus* in this study form two divergent mtDNA lineage clades corresponding to the α clade and the *D. fuscus* clade (*sensu* Tilley *et al.* 2013). Average uncorrected pairwise

sequence divergence between haplotypes in the α clade and *D. fuscus* clade generated from the best ML tree had 0.1103 substitutions per site, using 339 sites. Levels of mean divergence within the α clade and *D. fuscus* clade were 0.0275 and 0.0587 substitutions per site, respectively (Table S2).

The *Desmognathus* α clade contained mtDNA haplotypes from *D. fuscus* sampled from the South Mountain and Brushy Mountain isolates within the Piedmont of North Carolina, as well as from South Carolina and Virginia, and we refer to these as Piedmont "A" haplotypes. Not all individuals sampled from the Piedmont mountain isolates had mtDNA haplotypes from this clade, and we refer to these as Piedmont "B" haplotypes. The α clade also contained all haplotypes sampled from the Sinking Creek form of *D. fuscus* in Tennessee. Finally, all haplotypes sampled from *D. carolinensis* were placed in this clade. *D. carolinensis* haplotypes were identical to, or had very high similarity with haplotypes sampled from either the Sinking Creek form, or to haplotypes sampled from the South and Brushy Mountains (average pairwise uncorrected substitutions per site ranged from 0 to 0.0301)(Table S2).

The mtDNA *D. fuscus* clade primarily comprised haplotypes sampled from *D. fuscus* populations across the northern and central portion of its range. The exception to this was a well-supported clade of haplotypes sampled from the South Mountains, which we refer to as Piedmont "B" haplotypes. We note here that *D. fuscus* populations in the South Mountains contain both Piedmont A and B haplotypes. Average uncorrected pairwise sequence divergence between Piedmont A and B haplotypes in the South Mountains are 0.1268 substitutions per site, using 339 sites. Levels of mean divergence

within Piedmont A and Piedmont B haplotypes were 0.0000 and 0.0029 substitutions per site, respectively (Table S2).

2.3.2 – SVDquartets

Unrooted lineage trees generated by SVDquartets using 16,173 SNPs for 46 samples support the clustering of major lineages with >80% bootstrap support: (1) *D. fuscus* lineages from the northern and central species range (KY, MA, VA), (2) *D. carolinensis*, (3) *D. fuscus* lineages from the South Mountains, (4) *D. fuscus* lineages from the Brushy Mountains, and (5) *D. fuscus* from Tennessee referable as SCF (Fig 2.5). SVDquartet analyses support *D. fuscus* SCF and *D. fuscus* from populations in the Brushy Mountains sharing a more recent common ancestor than other *D. fuscus* Piedmont lineages. *D. fuscus* lineages from populations in the northern and central range distribution share a more recent common ancestor with *D. carolinensis* compared with all other Piedmont and non-Piedmont *D. fuscus* lineages included in this study. We note that in the South Mountains, *D. fuscus* mtDNA Piedmont A and B lineages are nested together with low bootstrap support for nuclear genomic segregation between these mtDNA lineages.

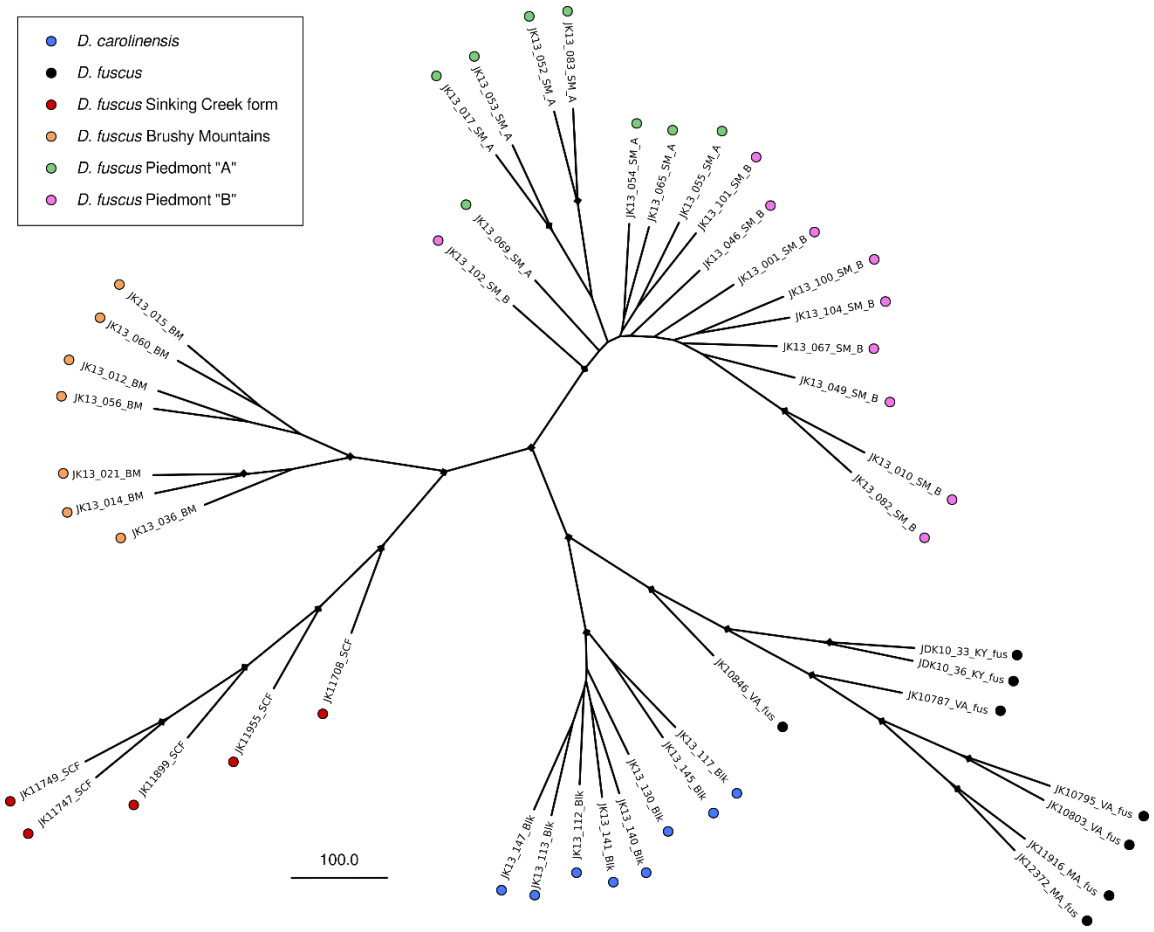


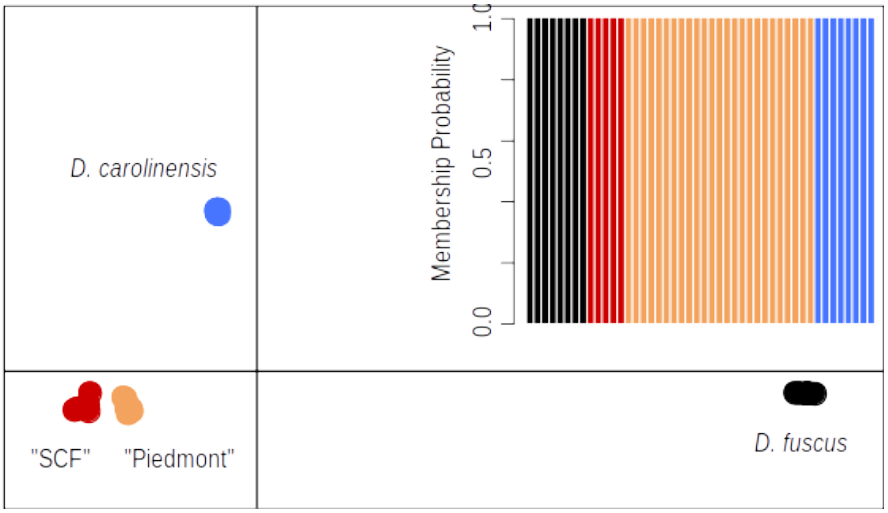
Figure 2.5 – Unrooted lineage tree generated using SVDquartets

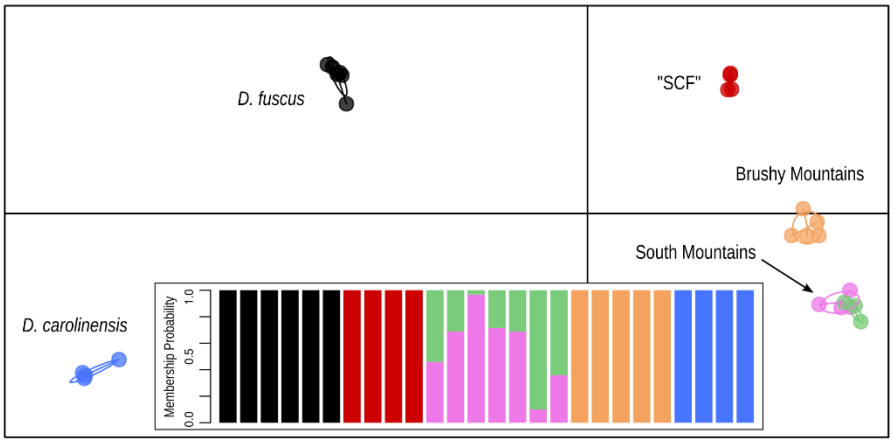
Lineage tree was generated using exhaustive quartet sampling and 1000 bootstrap replicates from 16,173 SNPs across 46 individuals. Lineages are indicated by filled circle (refer to Fig. 2.1 for *nad2* mtDNA clades): (blue) *D. carolinensis*, (red) *D. fuscus* Sinking Creek form, (green) *D. fuscus* South Mountain Piedmont A and (orchid) *D. fuscus* South Mountain Piedmont B, (orange) *D. fuscus* Brushy Mountain, and (black) *D. fuscus* populations from the northern and central distribution of the species range. Piedmont A and B from the South Mountains form a single clade despite comprising different mtDNA lineages. Bootstrap values >80 are indicated with solid diamonds on supported nodes.

2.3.3 – DAPC

We present PCA based discriminate analysis and population genetic clustering based on the entire set of *Desmognathus* individuals for which ddRAD data was collected, followed by more focused analyses of *D. fuscus* individuals from the Piedmont of North Carolina. Using K-means clustering, the lowest Bayes Information Criterion score identified 3-5 clusters which describe genetic variation in the 46x160 SNP dataset and 3 clusters which describe genetic variation in the 26*x1380 SNP dataset. Ten principle components with the lowest mean standard error were retained after K-means cross-validation of our more stringent dataset (46x160SNPs), in which 3 eigenvalues captured 82.7% of the conserved variance. Three major clusters are present in these data representing populations of (1) *D. fuscus* from the northern and central species range (Massachusetts, Kentucky, and Virginia), (2) *D. carolinensis*, and a cluster weakly segregating (3a) *D. fuscus* from Tennessee referable as SCF, and (3b) *D. fuscus* from the Piedmont mountain isolates (South Mountains and Brushy Mountains). K-means clustering based on four discrete clusters assigned all individuals into these membership categories with 100% probability (Fig 2.6).

After K-means cross-validation of our less stringent dataset (26*x1380SNPs), we retained 6 principle components with the lowest mean standard error, in which 5 eigenvalues captured 80.8% of the conserved variance. Several major clusters are present in these data representing populations of (1) *D. fuscus* from the northern and central species range (Massachusetts, Kentucky, and Virginia), (2) *D. carolinensis*, and clusters weakly segregating (3a) *D. fuscus* from Tennessee referable as SCF, (3b) *D. fuscus* from the Brushy Mountains, and (3c) *D. fuscus* from the South Mountains. K-means clustering


D. carolinensis
Membership Probability
0.0
0.5
1.0
"SCF"
"Piedmont"
D. fuscus


D. fuscus
"SCF"
Brushy Mountains
South Mountains
D. carolinensis
Membership Probability
0.0
0.5
1.0

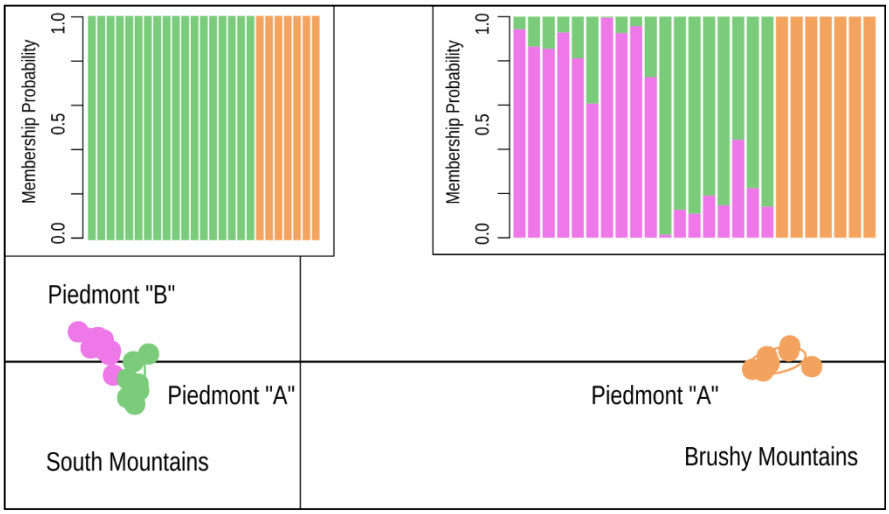

Membership Probability
0.0
0.5
1.0
Membership Probability
0.0
0.5
1.0
Piedmont "B"
Piedmont "A"
South Mountains
Piedmont "A"
Brushy Mountains

Figure 2.6 – DAPC scatterplots and cluster membership probabilities

(top) DAPC scatterplot and cluster membership probabilities of 46 individuals from the 160SNP dataset. Genetic variance can be described as 3 or 4 discriminate clusters illustrated using the following color scheme with dots representing individuals: *D. fuscus* from the northern and central species range (black), *D. carolinensis* (blue), *D. fuscus* from Tennessee "SCF" (red), and *D. fuscus* from the Piedmont mountain isolates (orange). **(middle)** DAPC scatterplot and cluster membership probabilities of 26* individuals from the 1380SNP dataset. Genetic variance can be described as 3 or 4 clusters illustrated using the following color scheme with dots representing individuals and inertial ellipses representing the 95% confidence interval. *D. fuscus* from the northern and central species range (black), *D. carolinensis* (blue), *D. fuscus* from Tennessee "SCF" (red), *D. fuscus* from the Brushy Mountains (orange). Note that *D. fuscus* from the South Mountains Piedmont "A" (green), and *D. fuscus* from the South Mountains Piedmont "B" (orchid) do not form distinct and separate clusters. **(bottom)** DAPC scatterplot and cluster membership probabilities of 25* *D. fuscus* individuals from Piedmont populations using 1280SNP loci after excluding all other individuals in this study. Genetic variance can be described as 2 or 3 clusters illustrated using the following color scheme with dots representing individuals and inertial ellipses representing the 95% confidence interval. *D. fuscus* from the Brushy Mountains (orange) and *D.* fuscus from the South Mountains form separate clusters with high membership probability (inset top left). Note that Piedmont "A" (green) and Piedmont "B" (orchid) mtDNA lineages do not form distinct and separate clusters (inset top right), but begin to segregate in ordination space with the use of more exclusive loci.

based on 5 discrete clusters assigned individuals into 4 of these membership categories with 100% probability, and less confident membership assignment in the last cluster which delineated *D. fuscus* individuals based on their mtDNA lineage (Piedmont A or B) (Fig 2.6).

Given the ambiguous clustering assignments of *D. fuscus* from populations in the Piedmont mountain isolates using the previous dataset (26*x1380SNPs; Fig 2.6), we decided to filter our ddRAD data exclusively for SNPs shared among *D. fuscus* populations in the South Mountains and Brushy Mountains. This exclusive dataset includes more individuals from targeted populations for stronger estimates of genetic variation at the population level. Our original datasets may have reduced variation among these populations if informative loci were excluded for not meeting the minimum thresholds required in other divergent taxa. We obtained 1280 SNP loci representing 2,560 alleles among 25 individuals (allowing at most two missing individuals per locus). Eleven principle components were retained after K-means validation, with 2 eigenvalues capturing 69.5% of the conserved variance in the retained principle components. Two genetic clusters in the Piedmont data are observed with high membership probability, representing each geographic isolate (Fig 2.6, inset top left). When we associated each individual by their respective mtDNA haplotype lineage in PC ordination space, we observed slight segregation between Piedmont A and Piedmont B lineages, however there is low confidence in membership assignments among individuals within these two clusters (Fig 2.6, inset top right).

2.3.4 – Bayes factor delimitation

The computational demands required to generate ML values across 17 competing models for both 46x160SNP and 26*x1380SNP datasets were extensive, therefore we present complete results from two independent runs for each model. Marginal likelihood values of replicates were very similar and we present all models, ranked Bayes factors, and ML differences between replicates (Table 2.3 and Table 2.4). A summary table of BFD results for the 7 best models for both datasets is also presented (Table 2.5).

The model best supported by the 46x160SNP dataset assigned taxa into 8 distinct genetic clusters by geographic divisions as well as mtDNA clades (Table 2.3; model 1). The second best model supported by the data assigns taxa into 6 genetic clusters by combining *D. fuscus* populations from the Piedmont into a single tip. This model has a Bayes factor of +18.7 (Table 2.3; model 4); a difference which corresponds to decisive support (>10) of the best model over all competing alternative models tested (Kass and Raferty 1995). All models combining individuals into tips based on shared mtDNA lineages receive considerably less support compared with strictly geographical models (minimum BF +64.27 to +1,747.25 vs. minimum BF +18.21 to 65.98) and have higher BF rankings (minimum rank 6-16 vs. 2-7). A model using 5 partitions derived from SVDquartets and DAPC analyses (Table 2.3; model 6), has a Bayes factor of +39.99; rank 4, and this model separately combined *D. fuscus* from populations in the northern and central distribution and *D. fuscus* from the South Mountains. Finally, the model based on currently recognized taxonomy which partitions individuals into either *D. fuscus* or *D. carolinensis* has a Bayes factor of +1,270.92; rank 15. We note that models combining

Table 2.3 – BFD results for 46x160SNP analyses.

Marginal likelihood values (ML) and differences between replicate ML scores are indicated for each model (x.1, x.2), along with Bayes factors and sorted rankings. BF are calculated by multiplying x2 the relative difference between the ML of each model and the highest ML value (best-alt.model) x2. General descriptions of models are in Table 2.2. Full descriptions of models are in Supplemental (S2.3).

Model	ML	ML replicate difference	Bayes Factor	Rank
1.2	-1418.8240998616	0.3504643353	0	*
1.1	-1419.1745641969		0.7009286705	
4.1	-1427.9327615477	0.2443136281	18.2173233721	2
4.2	-1428.1770751758		18.7059506283	
2.1	-1434.0808953065	0.0323148371	30.5135908898	3
2.2	-1434.1132101436		30.578220564	
6.1	-1438.8189110605	0.3197702195	39.9896223978	4
6.2	-1439.1386812801		40.6291628369	
5.1	-1442.9516020506	0.2506047237	48.255004378	5
5.2	-1443.2022067743		48.7562138254	
7.2	-1450.7562245041	0.5748575882	63.8642492849	6
3.1	-1450.9584235308	0.2569922811	64.2686473383	6
3.2	-1451.2154158119		64.7826319006	
7.1	-1451.3310820923		65.0139644613	
8.2	-1451.6692310641	0.1487219948	65.6902624049	7
8.1	-1451.8179530588		65.9877063944	
13.2	-1483.9121084685	0.0227948474	130.1760172137	8
13.1	-1483.9349033159		130.2216069085	
12.2	-1700.8949115804	0.0582822819	564.1416234376	9
12.1	-1700.9531938623		564.2581880014	
9.2	-1734.8337062004	0.2959015204	632.0192126776	10
9.1	-1735.1296077208		632.6110157184	
10.1	-1873.4311250572	0.3466714311	909.2140503912	11
10.2	-1873.7777964883		909.9073932534	
14.1	-1936.1017703408	0.0666768872	1034.5553409584	12
14.2	-1936.168447228		1034.6886947328	
11.2	-1959.8372095434	0.0330363421	1082.0262193636	13
11.1	-1959.8702458855		1082.0922920478	
16.2	-2024.425667872	0.1197642601	1211.2031360208	14
16.1	-2024.5454321321		1211.442664541	
17.1	-2054.2461195596	0.0381749557	1270.8440393959	15
17.2	-2054.2842945153		1270.9203893073	
15.1	-2292.4488975298	0.0669178986	1747.2495953363	16
15.2	-2292.5158154284		1747.3834311335	

Table 2.4 – BFD results for 26*x1380SNP analyses.

Marginal likelihood values (ML) and differences between replicate ML scores are indicated for each model (x.1, x.2), along with Bayes factors and sorted rankings. BF are calculated by multiplying x2 the relative difference between the ML of each model and the highest ML value (best-alt.model) x2. General descriptions of models are in Table 2.2. Full descriptions of models are in Supplemental (S2.3).

Model	ML	ML replicate difference	Bayes Factor	Rank
2.2	-23859.4419330731	8.7761229231	0	*
2.1	-23868.2180559962		17.5522458462	
1.1	-23876.1287519689	0.8008619385	33.3736377916	2
1.2	-23876.9296139074		34.9753616686	
4.2	-23977.4432540884	0.5075070733	236.0026420306	3
4.1	-23982.7842196054		246.6845730646	
3.2	-23983.2917266787	0.4483781904	247.6995872112	4
3.1	-23983.7401048691		248.596343592	
6.2	-24425.0640304435	3.5650920658	1131.2441947408	5
6.1	-24428.6291225093		1138.3743788724	
5.1	-24438.8048892902	2.098938914	1158.7259124342	6
5.2	-24440.9038282042		1162.9237902622	
8.1	-24536.0336637523	0.6451986716	1353.1834613584	7
8.2	-24536.6788624239		1354.4738587016	
7.2	-24538.0687508803	1.2476978262	1357.2536356144	8
7.1	-24539.3164487065		1359.7490312668	
13.1	-24761.6686874906	0.1689372341	1804.453508835	9
13.2	-24761.8376247247		1804.7913833032	
9.2	-25791.4875521276	0.0403293806	3864.091238109	10
9.1	-25791.5278815082		3864.1718968702	
10.1	-26212.8660081461	0.7349838516	4706.848150146	11
10.2	-26213.6009919977		4708.3181178492	
11.2	-26371.4627635521	0.4394917776	5024.041660958	12
11.1	-26371.9022553297		5024.9206445132	
12.1	-26430.503135909	0.5602964705	5142.1224056718	13
12.2	-26431.0634323795		5143.2429986128	
14.2	-26636.1002559074	0.7467165358	5553.3166456686	14
14.1	-26636.8469724432		5554.8100787402	
16.1	-26851.9112354571	0.2152426302	5984.938604768	15
16.2	-26852.1264780873		5985.3690900284	
17.2	-27703.3585916207	0.1376986175	7687.8333170952	16
17.1	-27703.4962902382		7688.1087143302	
15.1	-28389.6080115414	0.1535470243	9060.3321569366	17
15.2	-28389.7615585657		9060.6392509852	

Table 2.5 – Summary table of BFD results

Summary table of the best ranked models supported by BFD analyses using the 46x160SNP and 26*x1380SNP datasets. Brief descriptions and number of partitions are followed by marginal likelihoods (ML), Bayes factors (BF), and relative ranks of BFs for each analysis. Full descriptions can be found in Supplemental. Note that BF differences between 46x160SNP model 3 and 7 are negligible.

			46x160SNPs			26*x1380SNPs		
Model	**Description**	**Partitions**	**ML**	**BF**	**Rank**	**ML**	**BF**	**Rank**
1	all separate	8	-1418.82410	–	*	-23876.12875	33.37364	2
2	combine SM	7	-1434.08090	30.51359	3	-23859.44193	–	*
3	combine Piedmont α	7	-1450.95842	64.26865	6	-23983.29173	247.69959	4
4	combine all Piedmont	6	-1427.93276	18.21732	2	-23977.44325	236.00264	3
5	see Table 2.	6	-1442.95160	48.25500	5	-24438.80489	1158.72591	6
6	as model 2 & 5	5	-1438.81891	39.98962	4	-24425.06403	1131.24419	5
7	as model 3 & 5	5	-1450.75622	63.86425	6	-24538.06875	1357.25364	7

SCF with other *D. fuscus* or *D. carolinensis* are all poorly supported (BF +130.176; rank ≥8).

The model best supported by the 26*x1380SNP dataset assigned taxa into 7 distinct genetic clusters by geographic divisions combining *D. fuscus* from the South Mountains containing the mtDNA lineages Piedmont A and Piedmont B (Table 2.4; model 2). The second best model supported by these data assign taxa into 8 genetic clusters and has a Bayes factor of +33.374 (Table 2.4; model 4); a difference which corresponds to decisive support (>10) of the best model over all competing alternative models tested (Kass & Raferty 1995). All models combining individuals into tips based on shared mtDNA lineages receive considerably less support compared with strictly geographical models (minimum BF +247.7 to +9,060.64 vs. minimum BF +0 to 1354.47) and have higher BF rankings (minimum rank 4-17 vs. 1-7). A model using 5 partitions derived from

SVDquartets and DAPC analyses (Table 2.4; model 6), which separately combined *D. fuscus* from populations in the northern and central distribution and *D. fuscus* from the South Mountains, has a Bayes factor of +1,131.24; rank 5. Finally, the model based on currently recognized taxonomy which partitions individuals into either *D. fuscus* or *D. carolinensis* has a Bayes factor of +7,687.83; rank 16. We note that models combining SCF with other *D. fuscus* or *D. carolinensis* are all poorly supported (BF +1,804.45; rank ≥9).

2.3.5 – Species tree estimation

BFD analyses of both SNP datasets support parameter rich models in which all genetic clusters for *D. carolinensis* and *D. fuscus* form distinct tips, with the exception of Piedmont A and B mtDNA lineages, which from a single cluster in the 26*1380SNP dataset. Maximum clade credibility trees were generated from the 1st path sampling step in *BEAST for both datasets, which support the same topological relationships between 8 OTU partitions (>0.95 PP), with the exceptions of SCF and BM (0.48 PP), and *D. fuscus* (VA) and *D. fuscus* (MA) (0.92 PP) in the 160x46 SNP dataset, which are both supported with a 1.0 PP in the 1380x25 SNP dataset. In addition, branch lengths and 95% confidence intervals for highest posterior densities for divergence times are larger in the 160x46 SNP dataset compared with the 1380x25 SNP dataset (Fig 2.7). Both analyses support closer evolutionary relationships between SM samples and a clade containing *D. fuscus* BM and SCF samples. *D. carolinensis* share a more recent common ancestor with *D. fuscus* Piedmont and SCF samples, with these lineages coalescing with *D. fuscus* (KY, MA, VA) lineages deeper in the inferred phylogeny.

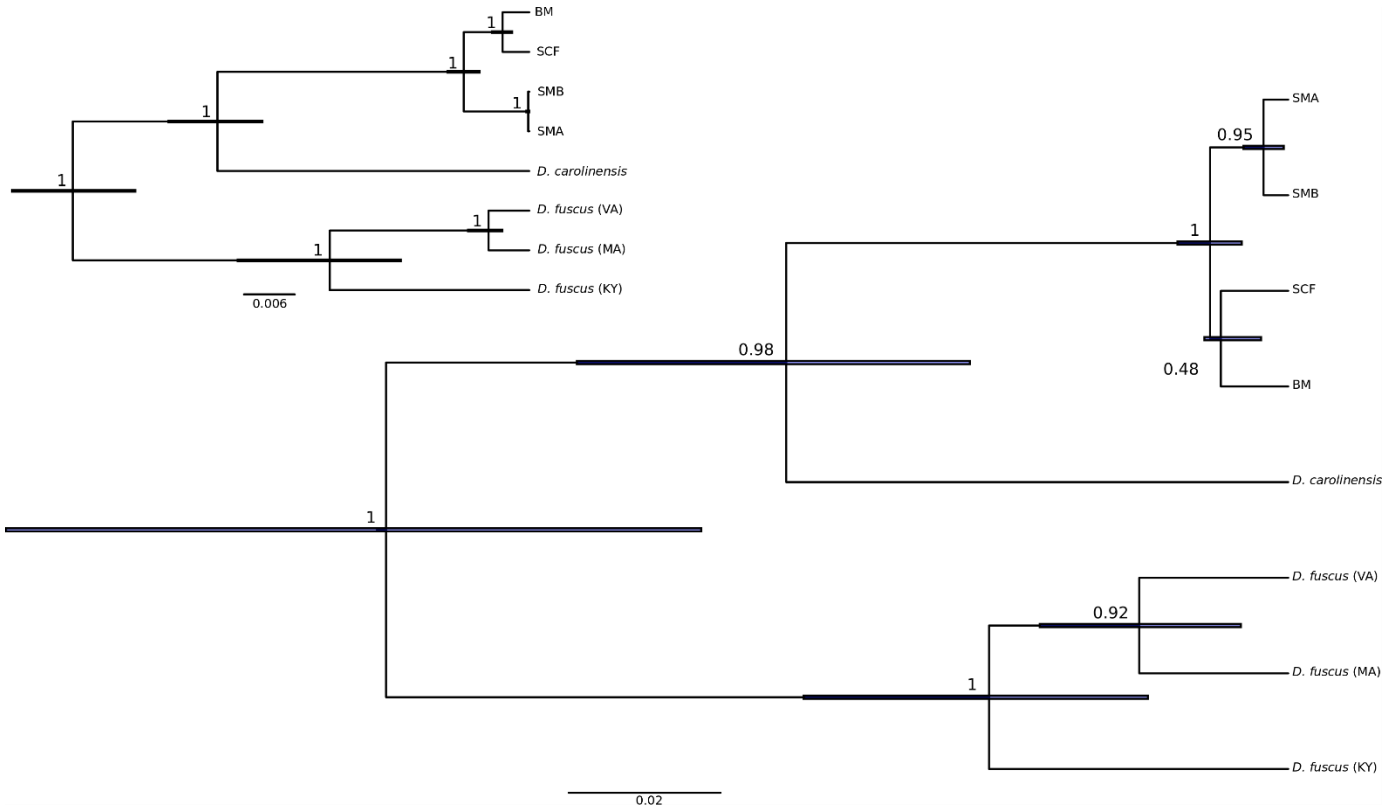


Figure 2.7 – Species trees for BFD lineages

Species trees for 26*1370SNP (top left) and 46x160SNP (bottom right) data sets were generated from *BEAST output files using the full parameter estimation path sampling step during BFD. Posterior probabilities for node support are illustrated, along with the 95% highest posterior density (HPD) in purple.

2.3.6 – Ancestral introgression D-statistics (ABBA BABA)

We used a 4 taxon D-statistic to evaluated 8 different models for genome wide signatures of ancestral introgression between *D. carolinensis* and *D. fuscus* samples from KY, MA, and VA with *D. fuscus* samples from the BM, SM, and SCF populations. We evaluated a dataset of 2442 loci across 21 individuals allowing no more than one missing sample per locus. Bonferroni corrected alpha values were used to account for multiple tests (ranging from 27-108), and z-scores were less than 3 standard deviations (uncorrected p-values equivalent to >0.05) for most individual tests of biased ancestral introgression (Table 2.6). In 1 out of 27 tests examining allele sharing patterns between P1: SMA or P2: SMB and P3: BM, there was increased allele sharing between a P2 sample and a P3 sample (ABBA) over P1 and P3 (BABA), despite conflicting mtDNA lineages (ABBA: 445.38 vs. BABA: 302.88; z-value 3.14 across 2180 loci). Separately pooling in-group P1, P2, and P3 samples for all analyses rendered purported differences among D-statistical values non-significant (z-value range 0.13-2.44; P-values >0.05). There is no statistical support for biased introgression between any populations defined by mtDNA lineage, geographic region, or ecomorphological classes examined in this study.

2.3.7 – Pairwise Fst values between delimited populations

Pairwise Fst values between delimited populations range from 0.0036 to 0.8151 (Table 2.7). *D. fuscus* South Mountain populations containing different mtDNA clades have the lowest pairwise Fst values (0.0036) whereas *D. fuscus* Sinking Creek form and *D. fuscus* populations from Massachusetts have the highest pairwise Fst values (0.8151). Pairwise Fst values between *D. fuscus* and *D. carolinensis* range from 0.2468 to 0.4357

Table 2.6 – Measures of introgression using Patterson's four-taxon D-statistic

Data is presented as a range of Z-scores applied to ddRAD loci across the 8 models tested. Two-tailed P-values are presented with significant results in bold. Individuals and replicates are identified by number codes. P4 include all individuals sampled together [n]. A single test replicate in model 4 identified a single individual (SM101 from the *D. fuscus* clade) with weakly biased pattern of alleles shared with *D. carolinensis* ABBA (445.38) over BABA (302.88) examining 2180 loci.

Model	P1	P2	P3	P4	Z range	P-value	*n*Sig/*n*	N loci range	Pooled sample Z-values
1	SMA 1-3	SMB 1-3	*D. carolinensis* 1-3	*D. fuscus* [4]	(0.1 - 2.2)	0.027807	0/27	(2156 - 2277)	0.32
2	SMA 1-3	SMB 1-3	*D. fuscus* 1-3	*D. carolinensis* [3]	(0.1 - 1.9)	0.057433	0/27	(2291 - 2348)	0.43
3	SMA 1-3	SMB 1-3	BM 1-3	*D. carolinensis* [3]	(0.1 - 2.6)	0.009322	0/27	(2170 - 2337)	1.65
4	SMA 1-3	SMB 1-3	BM 1-3	*D. fuscus* [4]	(0.12 – **3.14***)	0.001689	**1**/27	(2170 - 2337)	1.85
5	SMA 1-3, SMB 1-3	BM 1-3	*D. carolinensis* 1-3	*D. fuscus* [4]	(0.0 - 1.5)	0.133614	0/54	(2087 - 2272)	0.13
6	SMA 1-3, SMB 1-3	SCF 1-3	*D. carolinensis* 1-3	*D. fuscus* [4]	(0.0 - 2.0)	0.0455	0/54	(2163 - 2293)	0.65
7	SMA 1-3, SMB 1-3	BM 1-3, SCF 1-3	*D. carolinensis* 1-3	*D. fuscus* [6]	(0.0 – 1.6)	0.109599	0/108	(2018 – 2293)	0.55
8	SMA 1-3, SMB 1-3	BM 1-3, SCF 1-3	*D. fuscus* 1-3	*D. carolinensis* [3]	(0.04, 3.15**)	0.001633	0/108	(2153-2364)	2.44

>3 Standard deviations equates to an alpha value less than 0.05, uncorrected for multiple tests
* a Z-value >3.113 equates to an alpha value less than 0.05 (Bonferroni adjusted p-value <0.001852 for 27 tests)
** a Z-value >3.5013 equates to an alpha value less than 0.05 (Bonferroni adjusted p-value <0.000463 for 108 tests)

Table 2.7 – Pairwise Fst Estimates among *Desmognathus* populations

Mean pairwise Fst values across loci (left diagonal) and weighted pairwise Fst across loci (right diagonal) estimated from 65,307 sites and 2442 loci using 21 individuals partitioned by the best supported model from BFD (46x160SNPs), indexed as: *D. fuscus* individuals from populations in the (0) South Mountains Piedmont A, (1) South Mountains Piedmont B, (2) SCF, (3) Brushy Mountains, (4) *D. carolinensis*, (5) *D. fuscus* (KY), (6) *D. fuscus* (VA), and (7) *D. fuscus* (MA). Number of individuals per population are indicated (n).

Population (n)	Description	0	1	2	3	4	5	6	7
0 (3)	SM-PA	—	0.0414	0.5231	0.3271	0.5730	0.7616	0.5744	0.7926
1 (3)	SM-PB	0.0036	—	0.5358	0.3331	0.5778	0.7676	0.5814	0.7977
2 (3)	SCF	0.2839	0.2964	—	0.5243	0.6794	0.8861	0.7079	0.9133
3 (3)	BM	0.1581	0.1647	0.3087	—	0.6010	0.7948	0.6055	0.8241
4 (3)	*D. carolinensis*	0.3218	0.3271	0.4357	0.3560	—	0.6464	0.4674	0.6741
5 (2)	*D. fuscus* (KY)	0.5189	0.5293	0.7610	0.5916	0.3896	—	0.4014	0.8862
6 (2)	*D. fuscus* (VA)	0.3420	0.3510	0.5096	0.3870	0.2468	0.2145	—	0.3609
7 (2)	*D. fuscus* (MA)	0.5586	0.5690	0.8151	0.6347	0.4202	0.7722	0.1948	—

with the lowest Fst values being from *D. fuscus* populations in Virginia, and the highest being from *D. fuscus* Sinking Creek form. Fst values between Piedmont *D. fuscus* populations range from 0.1581 to 0.1647. Fst values nearly double when comparing *D. fuscus* from the South Mountains with *D. fuscus* Sinking Creek form, the sister lineage to *D. fuscus* populations in the Brushy Mountains.

2.3.8 – Migration/Gene Flow estimates using IMa2p

We present several preliminary migration estimates for *D. fuscus* and *D. carolinensis* populations using IMa2p. These estimates were obtained from the marginal posterior densities for migration, which are scaled by a factor of 4Neu, or the migration

rate for the effective population size per generation multiplied by the mutation rate across loci. For clarity, we follow the convention of describing migration forwards in time from one source population to another, rather than describing the coalescent backwards in time.

Preliminary IMa2p results suggest there are multiple asymmetric rates of migration between *D. fuscus* population lineages, and between *D. fuscus* and *D. carolinensis*. More specifically, there is a greater probability of proportionally higher gene flow from South Mountain B lineages into South Mountain A lineages, with peak estimates of migration from SMB→SMA at approximately 0.4 migrants per generation (MPG), and peak estimates of migration from SMA→SMB closer to 0 MPG (Figure 2.8). When comparing migration rates between *D. fuscus* Piedmont populations in the South Mountains, there appears to be a greater probability of biased migration from *D. carolinensis* to α clade individuals of similar mtDNA (migration rate of approximately 0.14 MPG from *D. carolinensis*→SMA) over individuals from the *D. fuscus* clade of dissimilar mtDNA (migration rate of approximately 0 MPG from *D. carolinensis*→SMB) (Figure 2.8).

There also appears to be greater probability of historic asymmetric migration from *D. carolinensis* into the common ancestor of *D. fuscus* populations from the Brushy Mountains and SCF, compared with migration rates to the common ancestor of South Mountain *D. fuscus.* Migration rates to the former are approximately 2.33 MPG, vs. approximately 0 MPG to the latter. In addition, the highest probable rates of migration are close to negligible, from the most recent common ancestor of South Mountain *D. fuscus* or Brushy Mountains and SCF *D. fuscus* to *D. carolinensis* (Figure 2.8).

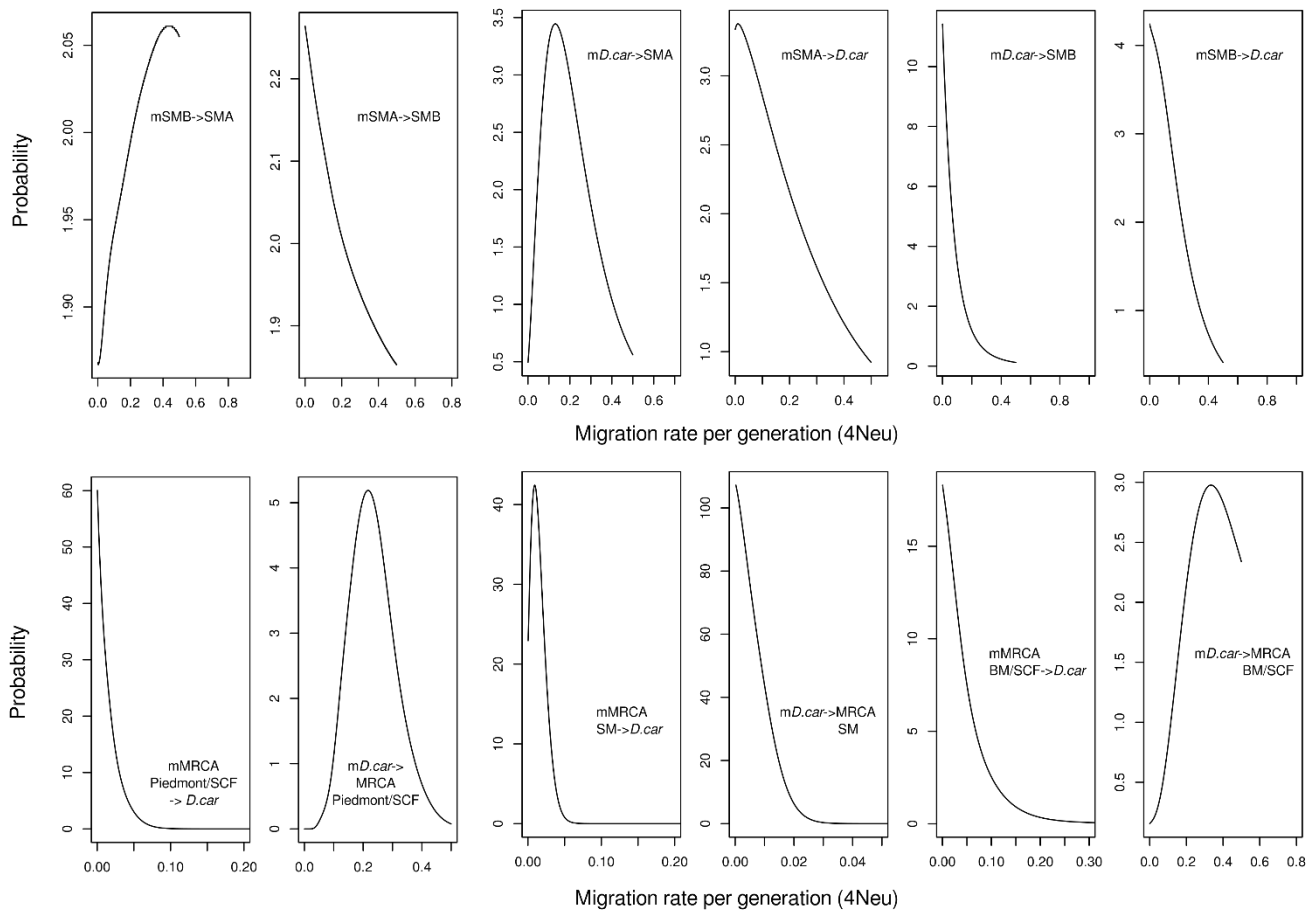


Figure 2.8 – Migration rates between *Desmognathus* lineages

Migration rates per generation (scaled by 4Neu) forward in time from source population → sink population as indicated.

2.4 – DISCUSSION

2.4.1 – MtDNA discordance

Mitochondrial lineages observed in this study are consistent with previous research which identified multiple highly divergent mtDNA clades shared among *D. fuscus* and *D. carolinensis* populations (Kozak *et al.* 2005; Tilley *et al.* 2013). *D. fuscus* in the South Mountains harbor more diverse assemblages than previously recognized; haplotypes which are either genetically similar to *D. carolinensis* lineages (the α clade) or the *D. fuscus* clade. The latter haplotypes, which we refer to as Piedmont B, may represent more ancient mitochondrial lineages within the Piedmont region that have been maintained over time despite widespread historical mtDNA introgression from *D. carolinensis.*

We favor inferences of mtDNA introgression over hypotheses of incomplete lineage sorting due to the high pairwise genotypic similarities observed across a large region of *nad2*, which are identical, or nearly so, among geographically distant populations and across recognized taxa. Less often have we observed multiple divergent haplotypes substantially represented within a region or locality, a possible signature of ILS. The mitochondrial genome also has a higher mutation rate than most coding regions of the nuclear genome, which suggests that for identical haplotypes to be present in non-sympatric *D. fuscus* and *D. carolinensis,* recent mitochondrial exchanges must have occurred, and these loci are probably under strong selection. In this study, we find that patterns of introgression in *Desmognathus* are not limited to the mitochondrial genome.

2.4.2 – Nuclear genetic clustering

Genetic clustering methods used in this research support more complex relationships among 3 to 5 lineages of *D. fuscus* and *D. carolinensis* inferred through the

analysis of both phylogenomic and multidimensional nuclear genetic variation. In stark contrast with recognized taxonomy, nuclear data do not support the monophyly of *D. fuscus.* Lineages inferred from genome wide sequencing data are discordant with genealogical patterns observed in mtDNA loci. Specifically, members of the mtDNA α clade do not form a monophyletic lineage, and *D. carolinensis* forms a separate and distinct lineage in multidimensional space. There is no strong evidence of nuclear divergence among *D. fuscus* mtDNA clades in the South Mountains. While *D. fuscus* from the Brushy Mountains and TN populations referable as SCF are closely related, these populations form distinct but highly supported sister lineages to one another, and are minimally distant in multidimensional space.

2.4.3 – Delimitation of D. fuscus lineages

Model based hypothesis testing through the use of BFD enables the ranking and quantification of genetically distinct clusters to determine which hypothetical models best fit the nuclear data. We find that of all models considered, geographic patterns are the best-fitting, while nuclear data substantially refute models in which lineage tips are based on currently recognized taxonomy or nested combinations of taxa based on similar mtDNA haplotypes. The identification of divergent *D. fuscus* lineages highlight a strong signature of extensive historical nuclear introgression which has shaped evolutionary patterns in *D. fuscus* and *D. carolinensis.* Importantly, nuclear patterns in these data do not reflect the same relationships as inferred from the mitochondrial markers either, which suggests introgression has shaped evolutionary patterns among these genomes in different ways.

An important consideration for resolving evolutionary relationships among taxa using BFD, is that this powerful method delimits population structure rather than strict

species limits (Sukumaran & Knowles 2017). In this study, genetic clustering methods have generated 3-5 hypothetical divisions while the nuclear data support a model with up to 8 primarily geographic divisions. This discrepancy supports the notion that BF limits are driven by isolation by distance, even to the extent that eastern populations of *D. fuscus* in Massachusetts and Virginia are subdivided from Kentucky populations west of the Appalachians. While introgression clearly has influenced evolutionary patterns among *D. fuscus* and *D. carolinensis* lineages, to the extent that models distinguishing them are well supported using BFD, we are left asking at what point should we consider lineages influenced by introgression independent from one another? Furthermore, what additional evidence is needed to support such claims?

2.4.4 – Future directions

An inherent constraint of BFD is the analytical challenges associated with population level sampling and the extent to which broader taxonomic sampling is needed for resolving species or lineage level relationships. Population genetic summary statistics such as Fst values, are potentially informative in determining if lineages represent distinct evolutionary entities by measuring genomic admixture between populations. Reliance on such a metric ignores potentially informative demographic histories of lineages, and the mechanisms responsible for the genetic patterns we observe. Coalescent demographic simulations are an excellent method to infer scenarios likely responsible for producing population genetic variation, however the genetic data required for confidently estimating parameter rich histories across the genome are a limiting factor for many systems and studies.

2.5 – Conclusions

In this study we have taken steps toward estimating migration patterns in *D. fuscus* and *D. carolinensis* to better understand the evolutionary histories and interactions shared between Piedmont and non-Piedmont lineages. Beyond advising against the use of any single marker for inferring species level relationships, these efforts are directed at understanding how migration and divergence with gene flow shape evolutionary patterns across nuclear and mitochondrial genomes, which may help explain patterns of discordance across diverse taxa and the persistence of genetic variation in the face of admixture.

2.5 – Author Contributions

Justin Kratovil and David Weisrock conceived and designed this project. Kelly Sovacool contributed toward data analysis and bioinformatics. Arun Sethuraman consulted and contributed toward migration estimates. Justin Kratovil and Stephen G. Tilley provided tissues. Justin Kratovil and David Weisrock analyzed and interpreted the data and wrote the draft of the paper.

2.6 – Acknowledgements

I would like to thank Robin Bagley and Kara Jones of the Linnen-Weisrock "Super lab" for their advice and expertise during ddRAD library preparation and data analysis for this project. I would also like to thank Paul Hime for figuring out some of the successful practices and restriction enzyme combinations to use for generating ddRAD data in salamanders.

2.7 – Additional Funding Sources

Financial support for library preparation and sequencing was funded through NSF 13-568, DEB-DDIG (Grant No. 1406876) to D. Weisrock and J. Kratovil (co-P.I.s). J. Kratovil was also supported by the University of Kentucky Department of Biology Gertrude Flora Ribble Research Grant.

2.8 – Supplemental information

Table S2.1. – Information for *nad2* sequences collected from NCBI and from other individuals used in this study

GenBank Number	Taxa	Source
AY612340	*Phaeognathus hubrichti*	Kozak et al., 2005
AY612341	*Desmognathus wrighti*	Kozak et al., 2005
AY612342	*Desmognathus aeneus*	Kozak et al., 2005
AY612343	*Desmognathus imitator*	Kozak et al., 2005
AY612344	*Desmognathus marmoratus*	Kozak et al., 2005
AY612345	*Desmognathus marmoratus*	Kozak et al., 2005
AY612346	*Desmognathus marmoratus*	Kozak et al., 2005
AY612347	*Desmognathus quadramaculatus*	Kozak et al., 2005
AY612348	*Desmognathus quadramaculatus*	Kozak et al., 2005
AY612349	*Desmognathus quadramaculatus*	Kozak et al., 2005
AY612350	*Desmognathus quadramaculatus*	Kozak et al., 2005
AY612351	*Desmognathus folkertsi*	Kozak et al., 2005
AY612352	*Desmognathus ocoee*	Kozak et al., 2005
AY612353	*Desmognathus ocoee*	Kozak et al., 2005
AY612354	*Desmognathus ocoee*	Kozak et al., 2005
AY612355	*Desmognathus ocoee*	Kozak et al., 2005
AY612356	*Desmognathus ocoee*	Kozak et al., 2005
AY612357	*Desmognathus ocoee*	Kozak et al., 2005
AY612358	*Desmognathus ocoee*	Kozak et al., 2005
AY612359	*Desmognathus ocoee*	Kozak et al., 2005
AY612360	*Desmognathus ocoee*	Kozak et al., 2005
AY612361	*Desmognathus ocoee*	Kozak et al., 2005
AY612362	*Desmognathus ocoee*	Kozak et al., 2005
AY612363	*Desmognathus orestes*	Kozak et al., 2005
AY612364	*Desmognathus orestes*	Kozak et al., 2005
AY612365	*Desmognathus orestes*	Kozak et al., 2005

AY612366	*Desmognathus ochrophaeus*	Kozak et al., 2005
AY612367	*Desmognathus ochrophaeus*	Kozak et al., 2005
AY612368	*Desmognathus carolinensis*	Kozak et al., 2005
AY612369	*Desmognathus carolinensis*	Kozak et al., 2005
AY612370	*Desmognathus carolinensis*	Kozak et al., 2005
AY612371	*Desmognathus carolinensis*	Kozak et al., 2005
AY612372	*Desmognathus carolinensis*	Kozak et al., 2005
AY612373	*Desmognathus apalachicolae*	Kozak et al., 2005
AY612374	*Desmognathus monticola*	Kozak et al., 2005
AY612375	*Desmognathus monticola*	Kozak et al., 2005
AY612376	*Desmognathus monticola*	Kozak et al., 2005
AY612377	*Desmognathus monticola*	Kozak et al., 2005
AY612378	*Desmognathus monticola*	Kozak et al., 2005
AY612379	*Desmognathus monticola*	Kozak et al., 2005
AY612380	*Desmognathus monticola*	Kozak et al., 2005
AY612381	*Desmognathus conanti*	Kozak et al., 2005
AY612382	*Desmognathus conanti*	Kozak et al., 2005
AY612383	*Desmognathus conanti*	Kozak et al., 2005
AY612384	*Desmognathus conanti*	Kozak et al., 2005
AY612385	*Desmognathus conanti*	Kozak et al., 2005
AY612386	*Desmognathus conanti*	Kozak et al., 2005
AY612387	*Desmognathus conanti*	Kozak et al., 2005
AY612388	*Desmognathus conanti*	Kozak et al., 2005
AY612389	*Desmognathus conanti*	Kozak et al., 2005
AY612390	*Desmognathus conanti*	Kozak et al., 2005
AY612391	*Desmognathus santeetlah*	Kozak et al., 2005
AY612392	*Desmognathus fuscus*	Kozak et al., 2005
AY612393	*Desmognathus fuscus*	Kozak et al., 2005
AY612394	*Desmognathus fuscus*	Kozak et al., 2005
AY612395	*Desmognathus fuscus*	Kozak et al., 2005
AY612396	*Desmognathus fuscus*	Kozak et al., 2005
AY612397	*Desmognathus fuscus*	Kozak et al., 2005
AY612398	*Desmognathus fuscus*	Kozak et al., 2005
AY612399	*Desmognathus fuscus*	Kozak et al., 2005
AY612400	*Desmognathus fuscus*	Kozak et al., 2005
AY612401	*Desmognathus fuscus*	Kozak et al., 2005
AY612402	*Desmognathus fuscus*	Kozak et al., 2005
AY612403	*Desmognathus fuscus*	Kozak et al., 2005
AY612404	*Desmognathus fuscus*	Kozak et al., 2005
AY612405	*Desmognathus fuscus*	Kozak et al., 2005
AY612406	*Desmognathus fuscus*	Kozak et al., 2005
AY612407	*Desmognathus fuscus*	Kozak et al., 2005
AY612408	*Desmognathus fuscus*	Kozak et al., 2005
AY612409	*Desmognathus fuscus*	Kozak et al., 2005
AY612410	*Desmognathus fuscus*	Kozak et al., 2005
AY612411	*Desmognathus fuscus*	Kozak et al., 2005

AY612412	*Desmognathus fuscus*	Kozak et al., 2005
AY612413	*Desmognathus fuscus*	Kozak et al., 2005
AY612414	*Desmognathus auriculatus*	Kozak et al., 2005
AY612415	*Desmognathus auriculatus*	Kozak et al., 2005
AY612416	*Desmognathus welteri*	Kozak et al., 2005
AY612417	*Desmognathus welteri*	Kozak et al., 2005
AY612418	*Desmognathus brimleyorum*	Kozak et al., 2005
AY612419	*Desmognathus brimleyorum*	Kozak et al., 2005
AY612420	*Desmognathus brimleyorum*	Kozak et al., 2005
AY612421	*Desmognathus brimleyorum*	Kozak et al., 2005
AY612422	*Desmognathus brimleyorum*	Kozak et al., 2005
AY612423	*Desmognathus brimleyorum*	Kozak et al., 2005
AY698025	*Desmognathus marmoratus*	Jones et al., 2006
AY698026	*Desmognathus marmoratus*	Jones et al., 2006
AY698027	*Desmognathus marmoratus*	Jones et al., 2006
AY698028	*Desmognathus marmoratus*	Jones et al., 2006
AY698029	*Desmognathus marmoratus*	Jones et al., 2006
AY698030	*Desmognathus marmoratus*	Jones et al., 2006
AY698031	*Desmognathus quadramaculatus*	Jones et al., 2006
AY698032	*Desmognathus quadramaculatus*	Jones et al., 2006
AY698033	*Desmognathus marmoratus*	Jones et al., 2006
AY698034	*Desmognathus marmoratus*	Jones et al., 2006
AY698035	*Desmognathus marmoratus*	Jones et al., 2006
AY698036	*Desmognathus marmoratus*	Jones et al., 2006
AY698037	*Desmognathus marmoratus*	Jones et al., 2006
AY698038	*Desmognathus marmoratus*	Jones et al., 2006
AY698039	*Desmognathus marmoratus*	Jones et al., 2006
AY698040	*Desmognathus marmoratus*	Jones et al., 2006
AY698041	*Desmognathus marmoratus*	Jones et al., 2006
AY698042	*Desmognathus marmoratus*	Jones et al., 2006
AY698043	*Desmognathus marmoratus*	Jones et al., 2006
AY698044	*Desmognathus marmoratus*	Jones et al., 2006
AY698045	*Desmognathus marmoratus*	Jones et al., 2006
AY698046	*Desmognathus marmoratus*	Jones et al., 2006
AY698047	*Desmognathus marmoratus*	Jones et al., 2006
AY698048	*Desmognathus marmoratus*	Jones et al., 2006
AY698049	*Desmognathus marmoratus*	Jones et al., 2006
AY698050	*Desmognathus quadramaculatus*	Jones et al., 2006
AY698051	*Desmognathus quadramaculatus*	Jones et al., 2006
AY698052	*Desmognathus marmoratus*	Jones et al., 2006
AY698053	*Desmognathus marmoratus*	Jones et al., 2006
AY698054	*Desmognathus quadramaculatus*	Jones et al., 2006
AY698055	*Desmognathus monticola*	Jones et al., 2006
AY916020	*Desmognathus ochrophaeus*	Jones et al., 2006
KR732330	*Desmognathus abditus*	Martin et al., 2015
KR732331	*Desmognathus aeneus*	Martin et al., 2015

KR732333	*Desmognathus auriculatus*	Martin et al., 2015
KR732337	*Desmognathus planiceps*	Martin et al., 2015
KR732338	*Desmognathus santeetlah*	Martin et al., 2015
KR732339	*Desmognathus wrighti*	Martin et al., 2015
KR826999	*Desmognathus fuscus*	Martin et al., 2015
KR827000	*Desmognathus marmoratus*	Martin et al., 2015
KR827001	*Desmognathus organi*	Martin et al., 2015
KR827002	*Desmognathus quadramaculatus*	Martin et al., 2015
KR827003	*Desmognathus quadramaculatus*	Martin et al., 2015

Specimen ID or SGT Tag	Taxa	State	County	Latitude	Longitude
JK1032	*Desmognathus fuscus*	KY	Rockcastle	37.396538	-84.292358
JK1033	*Desmognathus fuscus*	KY	Rockcastle	37.396538	-84.292358
JK1036	*Desmognathus fuscus*	KY	Rockcastle	37.368591	-84.211845
JK1037	*Desmognathus fuscus*	KY	Rockcastle	37.368591	-84.211845
JK1038	*Desmognathus fuscus*	KY	Rockcastle	37.368591	-84.211845
SGT35147	*Desmognathus fuscus*	VA	Roanoke	37.1826	-80.1397
SGT35155	*Desmognathus fuscus*	VA	Giles	37.35085	-80.601
SGT35163	*Desmognathus fuscus*	VA	Rockbridge	37.970913	-79.464053
SGT35206	*Desmognathus fuscus*	VA	Carroll	36.612083	-80.7713
SGT35428	Sinking Creek form	TN	Green	36.12510	-82.6455
SGT35467	Sinking Creek form	TN	Unicoi	36.17680	-82.2815333
SGT35469	Sinking Creek form	TN	Unicoi	36.17680	-82.2815333
SGT35589	*Desmognathus orestes*	NC	Caldwell	36.09528	-81.523312
SGT35619	Sinking Creek form	TN	Carter	36.33345	-82.27547
SGT35636	*Desmognathus fuscus*	MA	Franklin	42.57225	-72.9220333
SGT35675	Sinking Creek form	TN	Washington	36.34665	-82.3089833
JK12372	*Desmognathus fuscus*	MA	Franklin	42.57225	-72.922033
JK13001	South Mountain “B”	NC	Burke	35.628617	-81.694967
JK13010	South Mountain “B”	NC	Burke	35.612199	-81.622226
JK13012	Brushy Mountain	NC	Wilkes	36.07151	-81.176845
JK13014	Brushy Mountain	NC	Wilkes	36.116072	-81.128333
JK13015	Brushy Mountain “A”	NC	Wilkes	36.116072	-81.128333
JK13017	South Mountain “A”	NC	Burke	35.606711	-81.647785
JK13021	Brushy Mountain	NC	Wilkes	36.07151	-81.176845
JK13036	Brushy Mountain	NC	Wilkes	36.088783	-81.174227
JK13046	South Mountain “B”	NC	Burke	35.610933	-81.626383
JK13049	South Mountain “B”	NC	Burke	35.612199	-81.622226
JK13052	South Mountain “A”	NC	Burke	35.606711	-81.647785
JK13053	South Mountain “A”	NC	Burke	35.628617	-81.694967
JK13054	South Mountain “A”	NC	Burke	35.628617	-81.694967
JK13056	Brushy Mountain “A”	NC	Wilkes	36.116072	-81.128333
JK13060	Brushy Mountain “A”	NC	Wilkes	36.116072	-81.128333

JK13065	South Mountain “A”	NC	Burke	35.610933	-81.626383
JK13067	South Mountain “B”	NC	Burke	35.610933	-81.626383
JK13069	South Mountain “A”	NC	Burke	35.628617	-81.694967
JK13082	South Mountain “B”	NC	Burke	35.612199	-81.622226
JK13083	South Mountain “A”	NC	Burke	35.628617	-81.694967
JK13100	South Mountain “B”	NC	Burke	35.610933	-81.626383
JK13101	South Mountain “B”	NC	Burke	35.610933	-81.626383
JK13102	South Mountain “B”	NC	Burke	35.610933	-81.626383
JK13104	South Mountain “B”	NC	Burke	35.610933	-81.626383
JK13110	*D. sp.* LGF	TN	Cocke	35.827622	-82.937966
JK13112	*D. carolinensis*	NC	Yancey	35.726307	-82.243401
JK13113	*D. carolinensis*	NC	Yancey	35.726307	-82.243401
JK13117	*D. carolinensis*	NC	Yancey	35.727383	-82.241486
JK13130	*D. carolinensis*	NC	Yancey	35.740059	-82.231735
JK13140	*D. carolinensis*	NC	Yancey	35.727383	-82.241486
JK13141	*D. carolinensis*	NC	Yancey	35.740059	-82.231735
JK13145	*D. carolinensis*	NC	Yancey	35.727383	-82.241486
JK13147	*D. carolinensis*	NC	Yancey	35.727383	-82.241486

Table S2.2. – Summary statistics for *nad2* sequences comparing the α clade and *D. fuscus* clade

Average uncorrected pairwise distances (bottom left)

Average uncorrected pairwise distance (diagonal in bold)

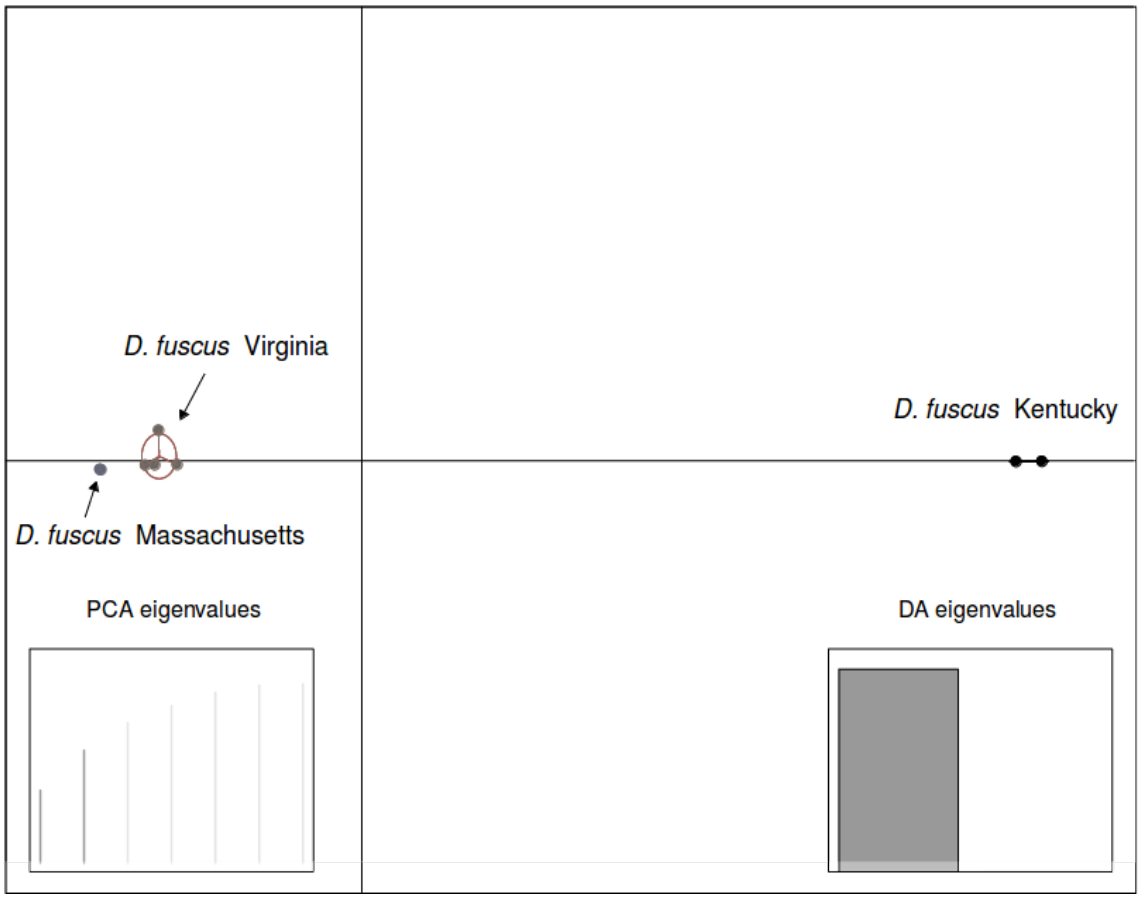
Samples for which average pairwise distances could not be calculated due to low sample size are indicated by **n/c**

	Taxon/Grp	mtDNA clade	1	2	3	4	5	6	7	8	9	10	11	12
1	*D. carolinensis*	α clade	**0.0165**											
2	*D. fuscus BM*	α clade	0.0301	**0.0177**										
3	*D. fuscus NC*	α clade	0.0291	0.0177	**n/c**									
4	*D. fuscus SCF*	α clade	0.0212	0.0324	0.031	**0.0295**								
5	*D. fuscus SMA*	α clade	0.0291	0.0265	0.0324	0.0339	**0.0000**							
6	*D. fuscus VA*	α clade	0.0291	0.0155	0.0118	0.031	0.0324	**n/c**						
7	*D fuscus SC*	α clade	0.0387	0.0383	0.0383	0.0428	0.0354	0.0383	**0.0000**					
8	*D. fuscus*	D. fuscus clade	0.1077	0.114	0.1116	0.1138	0.113	0.1088	0.1069	**0.0457**				
9	*D. fuscus KY*	D. fuscus clade	0.0985	0.1125	0.1128	0.1055	0.1025	0.1128	0.1084	0.0777	**0.0231**			
10	*D. fuscus MA*	D. fuscus clade	0.0988	0.104	0.1047	0.1047	0.1018	0.0988	0.0959	0.0329	0.0715	**0.0029**		
11	*D. fuscus SMB*	D. fuscus clade	0.1176	0.1224	0.118	0.1209	0.1268	0.115	0.1268	0.0885	0.104	0.0782	**0.0029**	
12	*D. fuscus VA*	D. fuscus clade	0.1071	0.1139	0.1143	0.1136	0.1128	0.1084	0.1055	0.0388	0.0767	0.0251	0.087	**0.0320**

S2.3. – Bayes factor delimitation model descriptions

Model 1: 8 total partitions with greatest number of parameters

(1) *D. fuscus* from KY, (2) *D. fuscus* from MA, (3) *D. fuscus* from VA, (4) *D. fuscus* from TN representing the SCF that contain mtDNA from the α clade, (5) *D. fuscus* from the South Mountains that contain mtDNA from the α clade, (6) *D. fuscus* from the South Mountains that contain mtDNA from the *D. fuscus* clade, 7) *D. fuscus* from the Brushy Mountains that contain mtDNA from the α clade, and 8) *D. carolinensis*, which contain mtDNA from the α clade. Preliminary DAPC analysis using 8 *D. fuscus* individuals (11,602 SNPs) from populations in the northern and central species range suggest separate clustering of Kentucky populations, and minimal segregation of Massachusetts and Virginia populations.



Model 2: 7 partitions combining Piedmont SM D. fuscus mtDNA lineages

(1) *D. fuscus* from KY, (2) *D. fuscus* from MA, (3) *D. fuscus* from VA, (4) *D. fuscus* from TN representing the SCF that contain mtDNA from the α clade, (5) *D. fuscus* from the South Mountains that contain mtDNA from the α clade and the *D. fuscus* clade, 6) *D. fuscus* from the Brushy Mountains that contain mtDNA from the α clade, and 7) *D. carolinensis,* which contain mtDNA from the α clade. Both DAPC and SVDquartet analyses suggest admixture of *D. fuscus* in the South Mountains.

Model 3: 7 partitions combining Piedmont SM and BM D. fuscus α clade mtDNA lineages

(1) *D. fuscus* from KY, (2) *D. fuscus* from MA, (3) *D. fuscus* from VA, (4) *D. fuscus* from TN representing the SCF that contain mtDNA from the α clade, (5) *D. fuscus* from South Mountains that contain the *D. fuscus* clade, 6) *D. fuscus* from the South Mountains that contain mtDNA from the α clade and *D. fuscus* from the Brushy Mountains that contain mtDNA from the α clade, and 7) *D. carolinensis*, which contain mtDNA from the α clade. MtDNA haplotypes of *cytb* and *nad2* are similar or identical in the regions sequenced, supporting a model in which Piedmont A lineages from the Piedmont mountain isolates form a single tip exclusive of Piedmont B.

Model 4: 6 partitions combining Piedmont SM and BM D. fuscus mtDNA lineages

(1) *D. fuscus* from KY, (2) *D. fuscus* from MA, (3) *D. fuscus* from VA, (4) *D. fuscus* from TN representing the SCF that contain mtDNA from the α clade, (5) *D. fuscus* from South Mountains and Brushy Mountains that contain mtDNA from the *D. fuscus* clade and α clade

and (6) *D. carolinensis*, which contain mtDNA from the α clade. K-means clustering from DAPC suggest *D. fuscus* Piedmont lineages form a single cluster.

Model 5: 6 partitions combining D. fuscus populations from the northern and central range

(1) *D. fuscus* from KY, MA, and VA, (2) *D. fuscus* from TN representing the SCF that contain mtDNA from the α clade, (3) *D. fuscus* from the South Mountains that contain mtDNA from the α clade, (4) *D. fuscus* from the South Mountains that contain mtDNA from the *D. fuscus* clade, 5) *D. fuscus* from the Brushy Mountains that contain mtDNA from the α clade, and 6) *D. carolinensis*, which contain mtDNA from the α clade. K-means clustering using all 46 *Desmognathus* individuals suggest *D. fuscus* from populations in the northern and central range of the species form a single discrete cluster. DAPC and SVDquartet analyses suggest multiple partitions are viable hypotheses worth testing.

Model 6: 5 partitions combining D. fuscus populations from the northern and central range and separately combining Piedmont SM D. fuscus mtDNA lineages.

(1) *D. fuscus* from KY, MA, and VA, (2) *D. fuscus* from TN representing the SCF that contain mtDNA from the α clade, (3) *D. fuscus* from the South Mountains that contain mtDNA from the α clade and the *D. fuscus* clade, 4) *D. fuscus* from the Brushy Mountains that contain mtDNA from the α clade, and 5) *D. carolinensis*, which contain mtDNA from the α clade. K-means clustering using all 46 *Desmognathus* individuals suggest *D. fuscus* from populations in the northern and central range of the species form a single discrete cluster. DAPC and SVDquartet analyses suggest multiple partitions are viable hypotheses worth testing.

Model 7: 5 partitions combining D. fuscus populations from the northern and central range and separately combining Piedmont SM and BM D. fuscus α clade mtDNA lineages

(1) *D. fuscus* from KY, MA, and VA, (2) *D. fuscus* from TN representing the SCF that contain mtDNA from the α clade, (3) *D. fuscus* from South Mountains that contain the *D. fuscus* clade, (4) *D. fuscus* from the South Mountains that contain mtDNA from the α clade and *D. fuscus* from the Brushy Mountains that contain mtDNA from the α clade, and 5) *D. carolinensis*, which contain mtDNA from the α clade. K-means clustering using all 46 *Desmognathus* individuals suggest *D. fuscus* from populations in the northern and central range of the species form a single discrete cluster. DAPC and SVDquartet analyses suggest multiple partitions are viable hypotheses worth testing.

Model 8: 4 partitions combining D. fuscus populations from the northern and central range and separately Piedmont SM and BM D. fuscus mtDNA lineages

(1) *D. fuscus* from KY, MA, and VA, (2) *D. fuscus* from TN representing the SCF that contain mtDNA from the α clade, (3) *D. fuscus* from South Mountains and Brushy Mountains that contain mtDNA from the *D. fuscus* clade and α clade and (4) *D. carolinensis*, which contain mtDNA from the α clade. K-means clustering using all 46 *Desmognathus* individuals suggest *D. fuscus* from populations in the northern and central range of the species form a single discrete cluster. DAPC and SVDquartet analyses suggest multiple partitions are viable hypotheses worth testing.

Model 9: 5 partitions combining D. fuscus populations from the northern and central range and separately Piedmont Brushy Mountains and D. carolinensis that contain mtDNA from the α clade

(1) *D. fuscus* from KY, MA, and VA, (2) *D. fuscus* from TN representing the SCF that contain mtDNA from the α clade, (3) *D. fuscus* from South Mountains which contain mtDNA from the α clade, (4) *D. fuscus* from the South Mountains which contain mtDNA from the *D. fuscus* clade, (5) *D. fuscus* from the Brushy Mountains and *D. carolinensis* which contain mtDNA from the α clade. This model tests a hypothesis of biased admixture between *D. carolinensis* and Brushy Mountain *D. fuscus* relative to the South Mountain *D. fuscus.*

Model 10: 4 partitions combining D. fuscus populations from the northern and central range and separately D. fuscus populations from the Piedmont South Mountains, Brushy Mountains, and D. carolinensis that contain mtDNA from the α clade.

(1) *D. fuscus* from KY, MA, and VA, (2) *D. fuscus* from TN representing the SCF that contain mtDNA from the α clade, (3) *D. fuscus* from South Mountains which contain mtDNA from the *D. fuscus* clade, (4) *D. fuscus* from the South Mountains, Brushy Mountains, and *D. carolinensis* which contain mtDNA from the α clade. This model tests a hypothesis of segregation between Piedmont *D. fuscus* driven by segregation in mtDNA lineages.

Model 11: 3 partitions combining D. fuscus populations from the northern and central range and separately D. fuscus populations from the Piedmont South Mountains that contain mtDNA from the D. fuscus clade and D. fuscus from the Brushy Mountains and D. carolinensis that contain mtDNA from the α clade.

(1) *D. fuscus* from KY, MA, and VA, (2) *D. fuscus* from TN representing the SCF that contain mtDNA from the α clade, (3) *D. fuscus* from South Mountains which contain mtDNA from the *D. fuscus* clade, *D. fuscus* from the South Mountains, Brushy Mountains, and *D. carolinensis* which contain mtDNA from the α clade. This model tests a hypothesis of segregation between *D. fuscus* on eastern and western populations on opposite sides of the southern Appalachians.

Model 12: 3 partitions combining D. fuscus populations from the northern, central range that contain mtDNA from the D. fuscus clade and TN referable as SCF that contain mtDNA from the α clade, and separately D. fuscus populations from the Piedmont South Mountains that contain mtDNA from the D. fuscus clade and D. fuscus from the South Mountains, Brushy Mountains, and D. carolinensis that contain mtDNA from the α clade.

(1) *D. fuscus* from KY, MA, and VA that contain mtDNA from the *D. fuscus* clade and *D. fuscus* from TN representing the SCF that contain mtDNA from the α clade, (2) *D. fuscus* from South Mountains which contain mtDNA from the *D. fuscus* clade, *D. fuscus* from the South Mountains, Brushy Mountains which contain mtDNA from the α clade, and (3) *D. carolinensis* which contain mtDNA from the α clade. This model can be directly compared with Model 8; treating SCF as an independent lineage from *D. fuscus* populations from the northern and central range.

Model 13: 3 partitions combining D. fuscus populations from the northern and central range and separately D. fuscus from the Piedmont South Mountains that contain mtDNA from the D. fuscus clade, D. fuscus from the South Mountains, TN representing the SCF, and Brushy Mountains that contain mtDNA from the α clade.

(1) *D. fuscus* from KY, MA, and VA, (2) *D. fuscus* from South Mountains which contain mtDNA from the *D. fuscus* clade, *D. fuscus* from the South Mountains, Brushy Mountains, and TN representing the SCF which contain mtDNA from the α clade, and (3) *D. carolinensis* which contain mtDNA from the α clade. Multiple K-means clustering analyses using DAPC suggest genetic variation can be segregated into 3 major clusters. This model explicitly tests this hypothesis against other competing hypotheses.

Model 14: 3 partitions based on mtDNA lineages, splitting D. fuscus from the South Mountains which contain mtDNA from the D. fuscus clade.

(1) *D. fuscus* from KY, MA, and VA, (2) *D. fuscus* from South Mountains which contain mtDNA from the *D. fuscus* clade, (3) *D. fuscus* from the South Mountains, Brushy Mountains, and TN representing the SCF, and *D. carolinensis* which contain mtDNA from the α clade. This model explicitly allows testing the hypothesis that mtDNA lineages referable as Piedmont B from the South Mountains are separate from *D. fuscus* of the *D. fuscus* mtDNA clade.

Model 15: 2 partitions based on mtDNA lineages: the D. fuscus clade and the α clade.

(1) *D. fuscus* from KY, MA, VA, and *D. fuscus* from South Mountains which contain mtDNA from the *D. fuscus* clade, (2) *D. fuscus* from the South Mountains, Brushy Mountains, TN representing the SCF, and *D. carolinensis* which contain mtDNA from the α clade. This model explicitly allows testing the hypothesis that mtDNA lineages referable as Piedmont B from the South Mountains are part of the *D. fuscus* mtDNA clade.

Model 16: 2 partitions combining D. fuscus populations from the northern and central range and separately D. fuscus from the Piedmont South Mountains that contain mtDNA from the D. fuscus clade, D. fuscus from the South Mountains, TN representing the SCF, Brushy Mountains, and D. carolinensis that contain mtDNA from the α clade.

(1) *D. fuscus* from KY, MA, and VA, (2) *D. fuscus* from South Mountains which contain mtDNA from the *D. fuscus* clade, *D. fuscus* from the South Mountains, Brushy Mountains, TN representing the SCF, and *D. carolinensis* which contain mtDNA from the α clade. This model tests a hypothesis of ancient divergence between *D. fuscus* populations from the northern and central species range, and all other *Desmognathus* in this study, or alternatively, that there is a strong signature of nuclear introgression between *D. carolinensis* and formerly sympatric *D. fuscus.*

Model 17: 2 partitions based on recognized taxonomy and ecomorphology

(1) *D. fuscus* from KY, MA, VA, TN representing the SCF, South Mountains, and Brushy Mountains, (2) *D. carolinensis.*

S2.4. – Individuals used for D-statistic tests and model descriptions

	SMA	SMB	BM	SCF	*D. carolinensis*	*D. fuscus*
1	JK13-054	JK13-067	JK13-012	JK11708	JK13-113	JK10787
2	JK13-069	JK13-101	JK13-014	JK11747	JK13-130	JK10846
3	JK13-083	JK13-102	JK13-015	JK11899	JK13-145	JK11916
4						JK12372
5						JK10-36

D-statistic model descriptions

Model 1: testing for biased allele sharing between South Mountain mtDNA α clade lineages and D. carolinensis relative to South Mountain mtDNA D. fuscus clade lineages

(P1) *D. fuscus* from the South Mountains that contain mtDNA from the α clade, (P2) *D. fuscus* from the South Mountains that contain mtDNA from the *D. fuscus* clade, P3) *D. carolinensis*, which contain mtDNA from the α clade, P4) *D. fuscus* from populations in KY, MA, and VA.

Model 2: testing for biased allele sharing between South Mountain mtDNA D. fuscus clade lineages and D. fuscus (KY, MA, VA) relative to South Mountain mtDNA α clade lineages

(P1) *D. fuscus* from the South Mountains that contain mtDNA from the α clade, (P2) *D. fuscus* from the South Mountains that contain mtDNA from the *D. fuscus* clade, P3) *D. fuscus* from populations in KY, MA, and VA, P4) *D. carolinensis*, which contain mtDNA from the α clade.

Model 3: testing for biased allele sharing between South Mountain mtDNA α clade lineages and Brushy Mountain mtDNA α clade lineages relative to South Mountain mtDNA D. fuscus clade lineages using D. carolinensis as the outgroup.

(P1) *D. fuscus* from the South Mountains that contain mtDNA from the α clade, (P2) *D. fuscus* from the South Mountains that contain mtDNA from the *D. fuscus* clade, P3) *D. fuscus* from the Brushy Mountains, which contain mtDNA from the α clade, P4) *D. carolinensis*, which contain mtDNA from the α clade.

Model 4: testing for biased allele sharing between South Mountain mtDNA α clade lineages and Brushy Mountain mtDNA α clade lineages relative to South Mountain mtDNA D. fuscus clade lineages using D. fuscus as the outgroup.

(P1) *D. fuscus* from the South Mountains that contain mtDNA from the α clade, (P2) *D. fuscus* from the South Mountains that contain mtDNA from the *D. fuscus* clade, P3) *D. fuscus* from the Brushy Mountains, which contain mtDNA from the α clade, P4) *D. fuscus*, which contain mtDNA from the *D. fuscus* clade.

Model 5: testing for biased allele sharing between South Mountain lineages and D. carolinensis relative to Brushy Mountain lineages

(P1) *D. fuscus* from the South Mountains, (P2) *D. fuscus* from the Brushy Mountains, which contain mtDNA from the α clade, P3) *D. carolinensis*, which contain mtDNA from the α clade, P4) *D. fuscus*, which contain mtDNA from the *D. fuscus* clade.

Model 6: testing for biased allele sharing between South Mountain lineages and D. carolinensis relative to D. fuscus Sinking Creek form

(P1) *D. fuscus* from the South Mountains, (P2) *D. fuscus* referable as Sinking Creek form, which contain mtDNA from the α clade, P3) *D. carolinensis*, which contain mtDNA from the α clade, P4) *D. fuscus*, which contain mtDNA from the *D. fuscus* clade.

Model 7: testing for biased allele sharing between South Mountain lineages and D. carolinensis relative to D. fuscus Brushy Mountains and D. fuscus Sinking Creek form

(P1) *D. fuscus* from the South Mountains, (P2) *D. fuscus* from the Brushy Mountains and *D. fuscus* referable as Sinking Creek form, both of which contain mtDNA from the α clade, P3) *D. carolinensis*, which contain mtDNA from the α clade, P4) *D. fuscus*, which contain mtDNA from the *D. fuscus* clade.

Model 8: testing for biased allele sharing between South Mountain lineages and D. fuscus (KY, MA, VA) relative to D. fuscus Brushy Mountains and D. fuscus Sinking Creek form

(P1) *D. fuscus* from the South Mountains, (P2) *D. fuscus* from the Brushy Mountains and *D. fuscus* referable as Sinking Creek form, both of which contain mtDNA from the α clade, P3) *D. fuscus*, which contain mtDNA from the *D. fuscus* clade, P4) *D. carolinensis*, which contain mtDNA from the α clade.

Chapter 3 : Phylogeographic analysis of introgressive gene flow among nuclear loci functionally linked to the mitochondrion

3.1 – INTRODUCTION

Species delimitation, the characterization of cohesive genetic lineages, is a fundamental step in studying the process of divergence between population-level lineages. Estimates of the species history lead to a better understanding of the genetic regions governing speciation and help in discerning the mechanisms influencing divergence. For nearly three decades, highly variable mtDNA sequences have been utilized extensively in phylogeography, population genetics, and systematics to study species boundaries (Avise *et al.* 1987; Avise 2009). MtDNA genealogies provide a cost effective and convenient method for surveying genetic variation in animal populations because the mitochondrial of many favorable characteristics (e.g. highly variable genetic resource, abundance within cells, etc.) (Ballard & Whitlock 2004).

These characteristics should describe a genetic marker capable of rapidly tracking population genetic processes; however, several limitations necessitate cautionary use of this molecule for phylogenetic inference and surveys of biodiversity. Any single-locus estimate of the species history should be avoided due to the stochastic nature of lineage sorting and the potential for discordant genealogical histories across the genome (Maddison 1997; Degnan & Rosenberg 2006). The exclusive utilization of mtDNA genealogies presents a larger problem for species delimitation because mtDNA, though often highly diverse in natural populations, is usually a modest fraction of cellular genetic content. In addition, as mtDNA genes are linked along a non-recombining circular chromosome,

lineage histories can be replaced during selective sweeps or horizontal transfer providing imprecise estimates of lineage divergence (Ballard & Whitlock 2004; Rheindt & Edwards 2011).

The inclusion of nuclear sequence data in molecular systematics has provided compelling evidence that there is frequent disagreement between patterns of divergence in nuclear and mitochondrial data. MtDNA introgression, the acquisition of organellar genetic material from one genetic background into another, is far more pervasive than originally thought among diverse animal lineages (Toews & Brelsford 2012). Ancient and ongoing introgression is responsible for the observed patterns of mito-nuclear discordance in sister and non-sister taxa (Good *et al.* 2008; Reid *et al.* 2012), between sympatric populations (Keck & Near 2010), and divergent lineages that hybridize due to similar host affinity (Linnen & Farrell 2007). Determining the mechanisms responsible for discordance is a major challenge, especially when lineages have undergone ancient rapid species radiations (Kozak *et al.* 2006). Recently, coalescent-based analytical methods have been valuable in discerning between incomplete lineage sorting and introgression by modeling genealogical expectations under a history of hybridization (Joly *et al.* 2009) and by incorporating sophisticated models for divergence with gene flow (Pinho & Hey 2010) and ancient admixture (Durand *et al.* 2011).

While there is plenty of evidence supporting nearly ubiquitous patterns of mito-nuclear discordance across animal lineages, relatively little is known about the influence that mtDNA divergence has on the nuclear genome during nascent speciation. Introgression between divergent lineages following speciation may reinforce pre and post zygotic reproductive boundaries through hybrid inviability, or lead to the replacement of

native haplotypes through biased gene flow and selective sweeps. Cytonuclear incompatibilities during this process have resulted in physiological differences that can reduce the fitness of hybrid progeny in yeast (Chou & Leu 2010), copepods (Burton *et al.* 2006), *Drosophila* (Ballard & Whitlock 2004), eels (Gagnaire *et al.* 2012), and birds (Burton *et al.* 2006). Intriguingly, introgression between naturally occurring mtDNA variants in closely related populations can have little or no effect on metabolic function, suggesting that strong functional constraints and coadapted processes may be acting on different genes across the mitochondrial and nuclear genomes (Pichaud *et al.* 2012; Parmakelis *et al.* 2013).

The coordinated interactions between the nuclear and mitochondrial genomes are essential for normal cellular and organismal processes, and these genes may provide important information about the coevolutionary history shared between genomes at the population and species level. Most mitochondrial genes have been transferred to the nucleus over time, with approximately 1,500 mitochondrial proteins encoded in the nucleus (Scharfe *et al.* 2009). Many nuclear encoded mitochondrial genes (NEMGs) are important because they are associated with human mitochondrial disorders, diseases, aging, and obesity (Shen *et al.* 2011; Knoll *et al.* 2013). Efficient oxidative phosphorylation is only possible through the coordinated interactions between nuclear and mitochondrial genome products (Zhang & Broughton 2013), which include cell and tissue specific regulatory proteins, translocases, and intermembrane receptors (Garesse & Vallejo 2001). Highlighting the extent of these interactions, mitochondrial replication and transcription is entirely dependent upon nuclear gene regulation (Zoppoli *et al.* 2011).

To our knowledge, NEMGs have not been examined as prospective genetic markers for the phylogenetic inference of species boundaries or to study population genetic patterns of mito-nuclear gene tree discordance. Coevolution between genomes has been mostly examined in genes involved with oxidative phosphorylation and subunits of the electron transport chain using metrics to identify purifying, neutral, or positive selection based on non-synonymous and synonymous ratios (Parmakelis *et al.* 2013; Zhang & Broughton 2013). Controlled back crosses and mtDNA transfection into similar nuclear backgrounds have been used to examine physiological changes, which may be adaptive (Rand *et al.* 2004), but cannot easily be applied to non-model systems.

If mitochondrial divergence drives a coevolutionary response in the nuclear genome, NEMGs may provide important information about this process and the role that cytonuclear interactions have during speciation. Given the importance of co-regulated interactions between mitochondrial and nuclear genomes, NEMGs may retain signatures of this co-evolutionary history even in the face of recombination elsewhere in the nuclear genome. Comparisons between genealogical estimates using neutral nuclear loci, NEMGs, and mtDNA genes at the population level are essential to thoroughly evaluate these hypotheses.

Reductions in the cost of next-generation sequencing have facilitated the rapid accumulation of multi-locus nuclear data for many organisms increasing the feasibility of reconciling the discordance between species trees and gene trees (Brito & Edwards 2009). These advancements have made it possible to study population genomics at an unprecedented depth within a large array of model and non-model systems (Schuster 2007; Grover *et al.* 2012). While there are many plausible explanations for genealogical

discordance between nuclear and mitochondrial loci, NEMGs might provide important information about the coevolution of these genomes, their contributions to divergence, and speciation.

In this study we compare phylogeographic patterns of divergence between multiple nuclear, NEMG, and mitochondrial loci in populations of *Desmognathus fuscus* and closely related taxa to test hypotheses of lineage boundaries and characterize patterns of introgression and admixture across different classes of markers at the population level. We tested both exclusive and nested hypotheses of lineage boundaries based on currently recognized taxonomy, geographic partitioning of populations, and mtDNA lineage associations using BFD on a subset of individuals sampled across *Desmognathus.* A major objective of this project is to help determine the evolutionary relationships of *D. fuscus* Piedmont lineages with known histories of mtDNA introgression within the context of the broader *Desmognathus* phylogeny. To accomplish these tasks, we generated species trees using algebraic quartet analyses, and explored estimates of population structure using SNPs drawn from putatively unlinked genomic markers. A final objective of this study was to examine concordance among nuclear markers, NEMGs, and mtDNA loci, to test the assumption that nuclear genes associated with mitochondrial function will reflect similar topological histories as mitochondrial genomes due to their coevolutionary interactions.

3.2 – METHODS

3.2.1 – Marker development

The relatively large genome size of *Desmognathus* (13.4-21.5Gb) compared with other vertebrate taxa (Gregory 2017), necessitates a targeted amplicon sequencing

approach to ensure the maximum recovery of nuclear loci sequenced at a sufficient sequencing depth for a large number of population samples. We initially explored the use of anchored hybrid enrichment (Lemmon *et al.* 2012) as a method for targeted sequencing in *Desmognathus*, and while this provided good recovery of targeted loci for one individual (~370 out of 512 loci), the level of multiplexing to yield adequate sequence coverage would be restricted to ~2 individuals per sequence capture reaction and only a few individuals per HiSeq lane, which makes this strategy extremely inefficient for low cost population-level studies.

To generate markers for *Desmognathus*, we generated and mined transcriptome libraries from a diverse set of tissues collected from *D. fuscus* to identify candidate genes and guide marker development and primer design for widespread use. Eight male and female *D. fuscus* were collected from two populations in KY (Table 2.1; locality 1 and 2). RNA was separately extracted from samples of adult brain, eyes, spleen, liver, pancreas, testes, tail muscle, skin (chin, cloaca, back of neck, base of tail) & whole larvae using an RNAeasy Plus extraction kit and QiaShredder (Qiagen) following standard protocols. Tissues were separately pooled so that roughly equal quantities of mRNA were represented across tissues. We used a RiboMinus kit (Invitrogen) to exclude short fragments and reduce highly redundant expressed transcripts in our libraries. Pooled libraries of high quality RNA were processed and sequenced by HudsonAlpha using 454 FLX (Roche) to maximize the length of cDNA sequences. 285,809 transcript reads were generated with an average size of 930nt, N50 of 955. 94.54% of bases were high quality (Q40 or greater). Read data was quality filtered and assembled using the default settings of Newbler software (v2.9) to generate 6,590 contigs. We explored using different parameter values during assembly

ranging from a minimum read overlap of 40bp to 45 bp, and sequence identities between 90-95%.

We used fluctuations in the abundance of gene ontology categories assigned to genes from various transcript assemblies to restrict the pool of candidate markers to loci that were less sensitive to assembly parameter values changes (<1 order of magnitude difference among assemblies). These contigs formed a pool of candidate markers for variant mapping across *D. fuscus* populations and for comparisons among transcriptome libraries concurrently generated by collaborators in a distantly related plethodontid; *Eurycea tynerensis* (Oklahoma salamander). We also compared transcripts with low coverage whole genome shotgun sequence data generated from a single *D. fuscus* sample from KY, sequenced at FSU for use in developing an Anchored Hybrid Enrichment Kit for amphibians (Lemmon *et al.* 2012; Hime *et al.* in prep). Lastly, we utilized annotations from EST libraries and validated reference genomic databases in *Ambystoma* (Putta *et al.* 2004). Candidate markers were filtered using reciprocal tBLASTx searches across the aforementioned databases for mutual best hit searches, forming a pool of over 400 candidate nuclear markers.

Genetic markers for candidate NEMGs were designed in a similar manner as other nuclear loci, but we targeted transcripts associated with mitochondrial function or localization based on gene ontology category assignment identified from Blast2GO and reciprocal tBLASTx searches using a nuclear-encoded mitochondrial protein database (Pagliarini *et al.* 2008; Calvo *et al.* 2016). We specifically targeted transcripts with functional relationships with cellular respiration, autophagy, or proteins localized in the mitochondria, including but not limited to mitochondrial topoisomerase TOP1MT and

proto-oncogene c-myc MYC (Zoppoli *et al.* 2011), uncoupling proteins UPCs (Garesse & Vallejo 2001), mitochondrial "solute carrier family 25" protein encoding genes, and peroxisome proliferator-activated receptor gamma co-activator PGC-1alpha.

We used Primer3 (Untergasser *et al.* 2012) within the program Geneious (v6.1.8) to design forward and reverse primers of standardized length, minimal secondary structure, low pairwise complementarity, and melting temperatures between 54ºC ±2. We prioritized the amplification of coding and non-coding regions ranging from approximately 500-1500 bp in length to generate primers for 96 putatively neutral and unlinked nuclear markers and 30 NEMG candidate markers (IDT). This fragment size range was chosen to help standardize data recovery across loci during library preparation and to obtain uniformly sequence coverage using Illumina MiSeq. We screened and tested amplification conditions by PCR across diverse *Desmognathus* taxa using high density low voltage gel electrophoresis to ensure reliable amplification of a single size fragment without non-specific amplification. This process produced 65 nuclear markers and 12 NEMGs which consistently amplified across test subjects in this study. In addition, we amplified mitochondrial genes *nad2* and *cytb* for comparison with previous studies.

3.2.2 – Sample Collection

We collected tissue samples from 99 *Desmognathus* in this study (Table 1), with localities and taxa chosen based on recognized taxa and previous mtDNA sequencing of *cytb* that identified several *D. fuscus* populations containing mtDNA lineages referable as Piedmont A and B clades (Tilley *et al.* 2013). We expanded the number of *D. fuscus* and *D. carolinensis* individuals sequenced from localities previously examined using ddRAD

sequencing (Kratovil *et al. in preparation*), and included additional populations of uncertain taxonomic status; *D. sp* Lemon Gap. *D. sp* Lemon Gap morphologically resemble *D. carolinensis* but have distinct allozyme electrophoretic mobility characteristics and contain mtDNA more closely related to *D. conanti*. Confounding our understanding of species boundaries in this group, *D. sp* Lemon Gap potentially hybridize with *D. carolinensis* and *D. santeetlah* (Tilley *et al.* 2013; Tilley 2016).

Among North Carolina Piedmont *D. fuscus*, 14 individuals were collected from populations in the Brushy Mountains (BM) and 22 individuals were collected from populations in the South Mountains (SM). An additional 18 individuals were collected across the *D. fuscus* range in Massachusetts, Kentucky, and Virginia. We included 10 *D. fuscus* referable as Sinking Creek form from Tennessee, 11 *D. carolinensis* collected 4 km from the type locality (Yancey County, NC), and 6 *D. sp* near Lemon Gap in Tennessee. To help resolve evolutionary relationships among *D. fuscus* and *D. carolinensis* lineages, we included 18 individuals from the following taxa: *D. auriculatus* (1: MVZ173494), *D. brimleyorum* (1: MVZ145019), *D. conanti* (1), *D. marmoratus* (2), *D. monticola* (3), D. *ocoee* (1), *D. orestes* (3), *D. planiceps* (2), *D. santeetlah* (2), and *D. wrighti* (2). All salamanders used in this study were collected by JK, SGT, and DW, unless otherwise indicated. High quality DNA was extracted using a DNAeasy Blood and Tissue kit (Qiagen) following standard protocols. For all samples we ran 3uL of genomic DNA on an agarose gel to ensure DNA was not degraded.

3.2.3 – PCR amplification

Although we took steps to minimize differences among optimal PCR conditions across target loci, we employed a partial touch down PCR strategy to account for individual differences in annealing temperatures for primers across diverse taxa in this study. Individual PCRs were completed using 96 well reaction plates in 10µL volumes using Taq DNA polymerase (New England Biolabs, Acton MA) and negative controls. PCR cycle conditions were as follows: initial denaturation at 95°C for 120 seconds followed by 12 cycles of denaturation at 95°C for 15 seconds, annealing at 56.5°C for 20 seconds, with each subsequent annealing cycle dropping -0.3°C, followed by extension at 68°C for 105 seconds. Thereafter, denaturation was followed by 23 cycles of annealing at 52.5°C for 20 seconds, and extension at 68°C for 105 seconds. A final 5 minute extension step was used and amplified products were maintained at 4°C. A 3µl sample from each PCR was separated by electrophoresis on a 1.3% agarose gel and digitally photographed to ensure proper amplification of the correctly sized target fragment and to confirm the absence of contamination in the negative control. Additional partial touch down PCR reactions were performed at higher (58°C) or lower (52°C) starting annealing temperatures for specific loci which failed to amplify in the initial plate reaction.

3.2.4 – Next-generation sequencing

PCR amplicons were individually pooled, cleaned using Agencourt AMPure XP beads, and quantitated using a Qubit dsDNA High Sensitivity Assay Kit (Invitrogen). Between 0.5ng and 1ng of each sample was uniquely indexed for PTA sequencing using a Nextera XT kit (Illumina). Briefly, amplicons were randomly fragmented by a *transposase*

which incorporated adapter sequences. Limited-cycle PCR amplified fragments and added unique barcode combinations to each sample for downstream demultiplexing. A subsample of amplicon libraries were quantified using an Agilent Bioanalyzer to ensure we had obtained the expected size fragment distribution in our libraries. All samples were normalized and processed for 250 PE sequencing on one lane of an Illumina MiSeq at the University of Kentucky Medical Center.

3.2.5 – Bioinformatics

Demultiplexed sequences were quality filtered using the program Sickle (v1.33) (Joshi & Fass 2011), which uses a sliding window across sequences to remove low quality (phred scored less than 20) reads from the 3' and 5' ends and remove short sequence fragments less than 20bp. We then used BayesHammer (Nikolenko *et al.* 2013) within the diploid genome assembler program dipSPAdes (v1.0) (Safonova *et al.* 2015), to correct sequencing errors which may result from simple nucleotide repeat regions, chimeras, or other PCR artifacts which can be present when sequence coverage is non-uniform. BayesHammer uses Bayesian subclustering to refine Hamming graphs and k-mer clustering based on read quality. Corrected paired end reads were then merged together using the program PEAR (v.0.9.6) (Zhang *et al.* 2014). Assembled reads ≥50nt were imported into Geneious (6.1.8) for reference sequence based iterative mapping using candidate nuclear, NEMG, and known mitochondrial markers.

All merged-assembled alignments for each locus were individually filtered, requiring at least 10 reads per locus and less than 100 polymorphic sites per Kbp. Individual loci containing more than 2 alleles (biallelic SNPs) in a variable site present in over 0.25%

of sequences at that site were considered either paralogous or the product of excessive PCR duplicate errors and were removed from downstream analyses. Consensus sequences matching 50% of the assembled sequences per locus were individually generated for all loci meeting quality filtering criteria. Bases with quality scores less than 30 were called as Ns. Consensus sequences for each locus were assembled using the Muscle aligner in Geneious. We estimate each locus has a mean coverage of 65x, however coverage varied considerably among markers. We used the command line version of SEQPHASE (Flot 2010) to format nuclear loci for haplotype reconstruction using PHASE (v2.1) (Stephens & Donnelly 2003; Crawford *et al.* 2004) which produced 38 loci which completed processing, 29 which are putatively neutral and 9 of which are NEMGs.

3.2.6 – Nuclear-based model testing

We used Bayes factor delimitation analyses to identify the prevailing history of lineage divergence using nuclear amplicon sequences and compared these patterns with estimates using ddRAD sequencing from a previous study (Kratovil *et al. in prep*). As with the previous study, our primary goal was to determine if the nuclear data fit a model of divergence similar to, or different from, the mitochondrial genome.

For our relevant set of models we tested patterns based on mtDNA lineage, currently recognized taxonomy, and geography to identify the most exclusive set of tips that could be present in an overall species tree for *Desmognathus*. This resulted in 6 hypothetical taxon partitions, or OTUs, that describe the most complex, or parameter rich, model tested (Table 3.8): (1) *D. carolinensis* that contain mtDNA from the α clade, (2) *D. fuscus* from the South Mountains that contain mtDNA from the α clade, (3) *D. fuscus* from

Table 3.8 – Bayes factor delimitation models tested

Shaded boxes indicate combined individuals. SM = South Mountains, BM = Brushy Mountains, A = α clade, B = *D. fuscus* clade.

Model	partitions	*D. orestes*	*D. auriculatus*	Piedmont fuscus (SM.A)	Piedmont fuscus (SM.B)	Piedmont fuscus (BM)	*D. carolinensis*
1	6						
2	5						
3	4						
4	4						
5	5						
6	4						
7	3						
8	2						
9	3						

the South Mountains that contain mtDNA from the *D. fuscus* clade, (4) *D. fuscus* from the Brushy Mountains that contain mtDNA from the α clade, (5) *D. auriculatus* and (6) *D. orestes*, which primarily serve as outgroups to the other taxa (but see below).

As in the chapter 2 study, we did not exhaustively test all possible combinations or divisions of taxa, but generated multiple models that covered a range of hypotheses regarding divergence and connectivity across our study system. For example, we tested models in which *D. carolinensis* and *D. fuscus* from the Piedmont α clade were treated as a single lineage, consistent with nuclear and mtDNA introgression between these groups. Alternatively, we tested models separating *D. carolinensis* from one or more *D. fuscus* α clade populations consistent with nuclear divergence and mito-nuclear discordance. Models also explored the combining or splitting of Piedmont *D. fuscus* according to mtDNA lineage (α clade vs. *D. fuscus* clade) and by geographical isolate (South Mountain vs. Brushy Mountain). Finally, we included models combining *D. fuscus* and *D. auriculatus* to test hypotheses of mtDNA introgression between Piedmont and Atlantic Coastal Plain population lineages, which have been shown to contain well supported clades of *cox1* mtDNA haplotypes (Beamer and Lamb, 2008). We do not provide a full enumeration of each model here, and instead specific descriptions of all models can be found in Supplementary Materials (S3.1).

We analyzed amplicon data using a coalescent-based species tree framework and calculated marginal likelihoods for each of the 9 models using BEAST v1.8.2 (Heled & Drummond 2010). Analyses were performed using an initial set of 25 phased amplicon loci from 10 individuals, ignoring ambiguous sites and allowing at most 1 missing individual per locus as long as all taxa were represented by at least 1 individual. For each model, we

performed 4 independent replicate analyses, generated tree models for each locus using an uncorrelated relax clock and HKY substitution model with flat priors, starting values of 2.0 and a lower bound of 0.0. Marginal likelihoods were calculated using both path sampling and stepping stone estimation over 100 steps taken over 1,000,000 MCMC generations. Samples were drawn every 100,000 generations after a 25% burnin. Convergence of the MCMC on a stable posterior distribution was performed through the comparison marginal distributions and sample size estimates of replicate log files using the program Tracer v1.5. Marginal likelihoods from each model were used to calculate Bayes factors for each model and determine the best-fitting model. Interpretations of the strength of support were based on Kass and Rafferty (1995).

3.2.7 – Species tree reconstruction

Preliminary species trees were estimated using the joint posterior distributions of posterior probabilities from 13 individual phased gene trees in the program *BEAST v1.8.2. For this analysis, we used 18 individuals across 14 representative *Desmognathus* taxa, restricting our analysis to loci that had all 14 taxa present, with at most 1 missing individual. For this analysis, we used a relaxed lognormal molecular clock for each gene tree, allowing the program to estimate the clock rate for each gene. We otherwise used the default recommended priors of the program, estimating parameters over 100,000,000 MCMC generations, sampling trees every 100,000 generations after a 25% burnin. We examined sample size estimates and samples of posterior distributions from log files to find suitable priors and confirm stable MCMC convergence using the program Tracer v1.5. We also present these data as a DensiTree to illustrate uncertainty in the species tree topologies

generated which overlay confidence intervals for parameter estimates (i.e., branch lengths, population sizes, tree topologies) across all input loci.

We also assessed the history of lineage divergence in the nuclear genome by generating coalescent-based lineage trees using the program SVDquartets (Chifman & Kubatko 2014) implemented in PAUP* v4.146 (Swofford 2003). This approach treats tips in the tree as the random pairing of gene copies for individuals, and thus can serve as a method for exploring patterns of divergence among clusters of individuals and populations for different loci. We used a set of 38 concatenated nuclear loci across 44 phased individuals representing samples from across the genus with the highest recovery of sequence data for their taxon or population. This concatenated dataset consisted of 30,455nt of sequence data. For comparisons between nuclear and NEMG datasets, we generated additional lineage trees in SVDquartets for either 29 nuclear loci (23,904nt) or 9 NEMGs (6,551nt) using the same representative individuals. Exhaustive sampling was used with the QFM quartet assembly algorithm with branch support evaluated using 100 bootstrap replicates for all analyses.

3.2.8 – Population genetic approaches

We used the program STRUCTURE v2.3.4 (Pritchard *et al.* 2000; Falush *et al.* 2003) in order to evaluate patterns of underlying population structure and gene flow across nuclear and NEMGs which may respond differently to mtDNA introgression. Specifically, we focus on identifying the proportion of individual nuclear genomic loci that may have resulted from admixture or migration between *D. fuscus* and *D. carolinensis* populations. For this analysis, we examined 29 putatively neutral and unlinked nuclear markers and 9

NEMGs using population level sampling (n = 76 or 75 individuals, respectively). We chose populations and taxa based on previous lineage estimates using BFD (Kratovil *et al. in preparation*), which included *D. fuscus* individuals from KY, VA, MA, and TN (referable as Sinking Creek form), and Piedmont *D. fuscus* from the Brushy Mountains and South Mountains (both Piedmont "A" and "B" mtDNA lineages). For each locus we randomly sampled a single SNP across phased individuals that met a minor allele frequency threshold greater than 0.05. To assess confidence in our initial population structure estimates, we replicated analyses using a different random SNP per locus.

We estimated the best-fitting model for the number of populations (K) by running 4 replicate independent analysis for each assumed value of K ranging from K=1 to K=7, without preassigning individuals to populations. We ran Structure estimates for 1 million generations after a 500,000 replicate burn-in period, using a model which allowed for the possibility of admixture among individuals and allele frequencies correlated among populations. We used the default prior parameters of the program using flat Dirichlet allele frequencies and uniform rates of admixture, and estimated the mean and SD of Fst values for each population. We used the Evanno method as implemented in the program Structure Harvester (Earl & vonHoldt 2012) to determine the best value for K.

3.3 – RESULTS

3.3.1 – Bayes factor delimitation

Marginal likelihood values calculated by using either path sampling or stepping-stone estimation were very similar across models and within replicates, therefore we only present path sampling results for each model across all replicates. The model best supported using

25 phased loci from our preliminary amplicon dataset assigned *Desmognathus* taxa into 5 distinct genetic clusters, which combined Piedmont South Mountain lineages together (Table 3.8; model 2). The second best model supported by the data assigned taxa into 6 genetic clusters which separate *D. fuscus* populations from the South Mountains into two tips partitioned by mtDNA lineage. This model has a Bayes factor of +9.397 (Table 3.9; model 1); a difference which corresponds to strong support (>6) comparing the best model to all other competing alternative models tested (Kass and Raferty 1995).

All models combining individuals into tips based strictly on shared mtDNA lineages (e.g., South Mountain α clade and Brushy Mountain α clade) receive considerably less support compared to the best model (minimum BF +28.934 to +3326.27) and BF rankings range from 3 to 8, out of 9 models tested. Related to patterns of nuclear and mtDNA introgression, in all models tested, there is considerably less support for models combining *D. carolinensis* with any other geographic or taxonomic group (models 4, 6, 8, and 9 which have BF rankings 9, 8, 4, and 6) corresponding to BF values ranging from +192.52 to +3331.42. Two similar models in which *D. auriculatus* are combined with *D. fuscus*, or with *D. fuscus* and *D. carolinensis* (models 7 and 8, respectively), receive considerably less support (BF +3317.87 or +192.52). Lastly, the model based on recognized taxonomy which partitions individuals into tips representing *D. fuscus, D. carolinensis, D. auriculatus*, and *D. orestes* has a Bayes factor of +28.934 rank 3.

Table 3.9 – BFD results for 25 phased loci.

Marginal likelihood values (ML), Bayes factors and sorted rankings. BF are calculated by multiplying x2 the relative difference between the ML of each model and the highest ML value (best-alt.model) x2. General descriptions of models are in Table 3.8.

Model	Description	Replicate	Path Sampling ML	Bayes Factor	Rank
1	6 separate taxa	1	-43271.54729		
	SMA, SMB, BM,	2	-43270.21627	9.39652	2
	D. carolinensis,	3	-43278.01837		
	D. sp outgroups (2)	4	-43277.39442		
2	5 separate taxa	1	**-43265.51801**	*	1
	(Piedmont SM)	2	-43274.00229		
		3	-43280.73072		
		4	-43269.56880		
3	4 separate taxa	1	-43292.14684		
	(Piedmont SM+BM)	2	-43294.48797		
		3	-43279.98519	28.93436	3
		4	-43283.76025		
4	4 separate taxa	1	-44938.40169		
	D. carolinensis	2	-44944.72928		
	+	3	-44931.22694	3331.41788	9
	(SMA + SMB)	4	-44931.24592		
5	5 separate taxa	1	-44890.50091		
	“Piedmont Alpha”	2	-44893.66538		
	(SMA + BMA)	3	-44890.05079	3249.06556	5
		4	-44895.22780		
6	4 separate taxa	1	-44937.73014		
	D. carolinensis	2	-44934.95172		
	+	3	-44929.43808		
	“Piedmont Alpha”	4	-44928.65283	3326.26965	8
7	3 separate taxa	1	-44936.55060		
	all *D. fuscus*	2	-44928.46592		
	+	3	-44924.45247	3317.86894	7
	D. auriculatus	4	-44935.08817		
8	2 separate taxa	1	-43365.62442		
	D. carolinensis	2	-43363.56932		
	+	3	-43361.77674	192.51746	4
	model 7	4	-43366.12672		
9	3 separate taxa	1	-44904.32698	3277.61794	6
	D. carolinensis	2	-44907.46930		
	+	3	-44915.61079		
	Piedmont *D. fuscus*	4	-44910.04662		

3.3.2 – Species tree estimation

Preliminary species trees estimated using 13 phased loci from 18 individuals (allowing at most 1 missing individual per locus) in the program *BEAST v1.8.2 produced node ages with broad and overlapping 95% highest posterior densities (HPD) across recent and ancient bifurcations (Fig 3.9). Three nodes among 14 representative *Desmognathus* taxa are supported by posterior probabilities greater than 0.95. Well supported clades include Piedmont *D. fuscus* lineages from the South and Brushy Mountains, all desmognathine taxa excluding *D. marmoratus* and *D. wrighti*, and lastly, all biphasic desmognathine taxa (excluding *D. wrighti*). Neither *D. carolinensis* nor *D. planiceps* are sister lineages to Piedmont *D. fuscus* containing mtDNA from the α clade or *D. fuscus* clades, although evolutionary relationships among taxa within this clade are generally weak or uncertain ($pp < 0.95$). A DensiTree produced from this 13 locus species tree illustrates the greatest uncertainty in parameter estimates for branch lengths across deeper ancestral nodes and uncertain topological relationships between distal tips of the phylogeny, excluding *D. wrighti* (Figure 3.9).

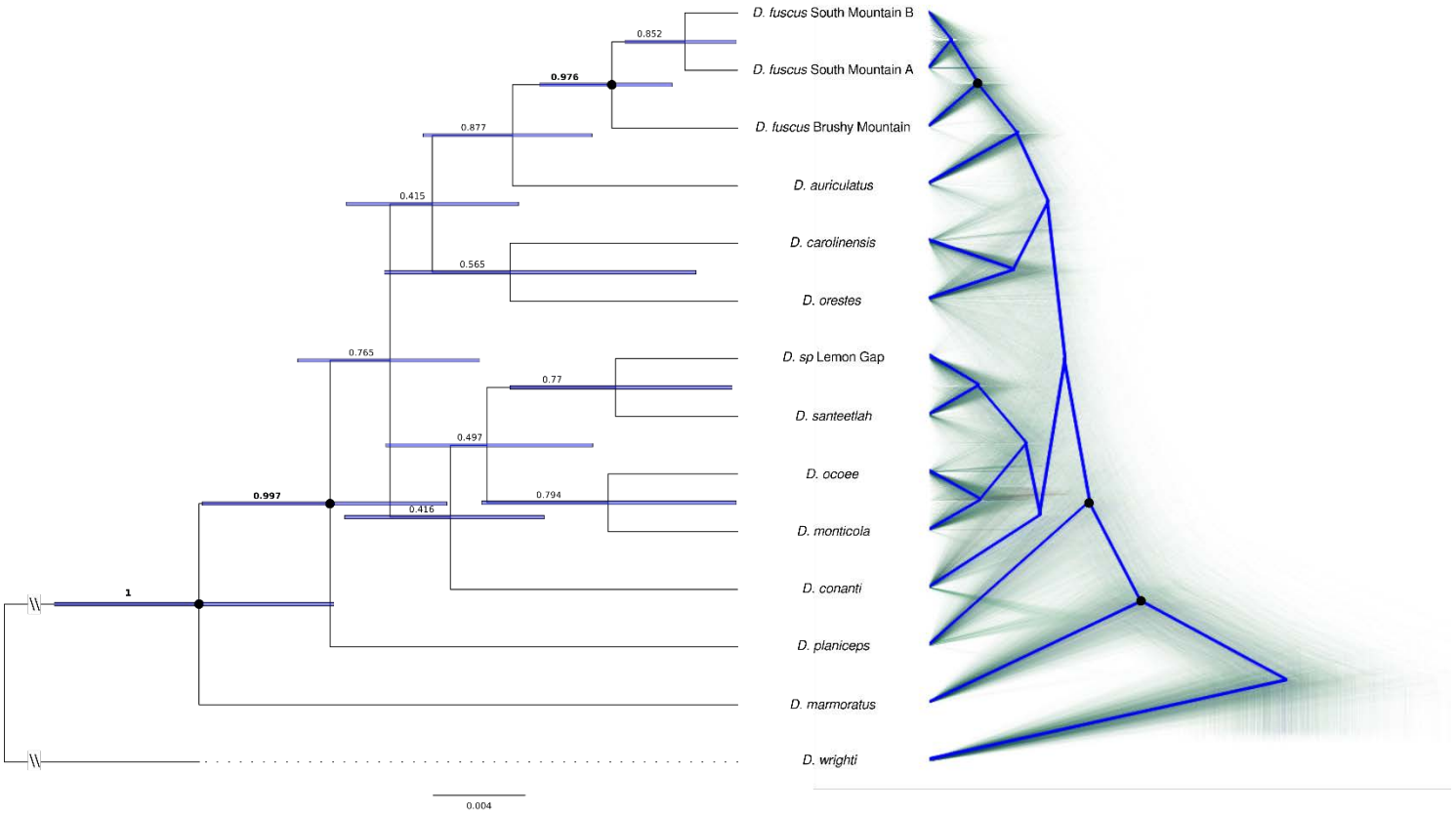


Figure 3.9 – Species tree for *Desmognathus* and parameter uncertainties

Species tree (left) produced by *BEAST with posterior probabilities above branches. Filled circles indicate nodes with >0.95 posterior probabilities. 95% HPD are illustrated as purple bars. DensiTree (right) illustrating parameter uncertainties.

Lineage trees generated using SVDquartets from 38 phased loci across 44 individuals (Figure 3.10) support the clustering of several major lineage divisions with bootstrap support >90: 1) a clade containing Piedmont and Sinking Creek form *D. fuscus* along with *D. carolinensis,* 2) a clade containing *D. sp* Lemon Gap and *D. santeetlah,* 3) *D. orestes*, 4) a clade containing *D. auriculatus, D. planiceps,* and *D. fuscus* from KY, VA, and MA, 5) *D. monticola*, 6) a clade containing *D. ocoee* and *D. conanti*, 7) *D. brimleyorum,* 8), *D. marmoratus,* and 9) *D. wrighti.* Within the clade comprised of *D. fuscus* and *D. carolinensis*, Piedmont *D. fuscus* from the South Mountains form separate well supported lineages which contain either mtDNA associated with the α clade or mtDNA associated with the *D. fuscus* clade (bs > 90). Other *D. fuscus* lineages are geographically partitioned between populations in KY, VA, and MA, however evolutionary relationships between them are unresolved, and these lineages are nested among *D. planiceps* and *D. auriculatus* with strong support (bs >80). All other nominate lineages have bootstrap support greater than 80 toward the distal tips of the trees, and deeper nodes in the tree are equally well supported among major clade divisions.

Lineage trees generated using SVDquartets from 29 phased loci excluding NEMGs (Figure 3.11) are generally similar in topological relationships (with several exceptions), but with weaker branch support compared to the SVDquartet analysis using the full set of 38 loci. 44 individuals cluster into 9 major lineage divisions with bootstrap support ≥90: 1) a clade containing Piedmont and Sinking Creek form *D. fuscus* along with *D. carolinensis,* 2) *D. sp* Lemon Gap weakly supported as a sister lineage to *D. santeetlah,* 3) *D. orestes*, 4) *D. monticola,* 5) a clade containing *D. planiceps, D. auriculatus,* and *D.*

fuscus from KY, VA, and MA, 6) *D. brimleyorum*, 7) *D. marmoratus*, 8) a clade containing *D. ocoee* and *D. conanti*, and 9) *D. wrighti*.

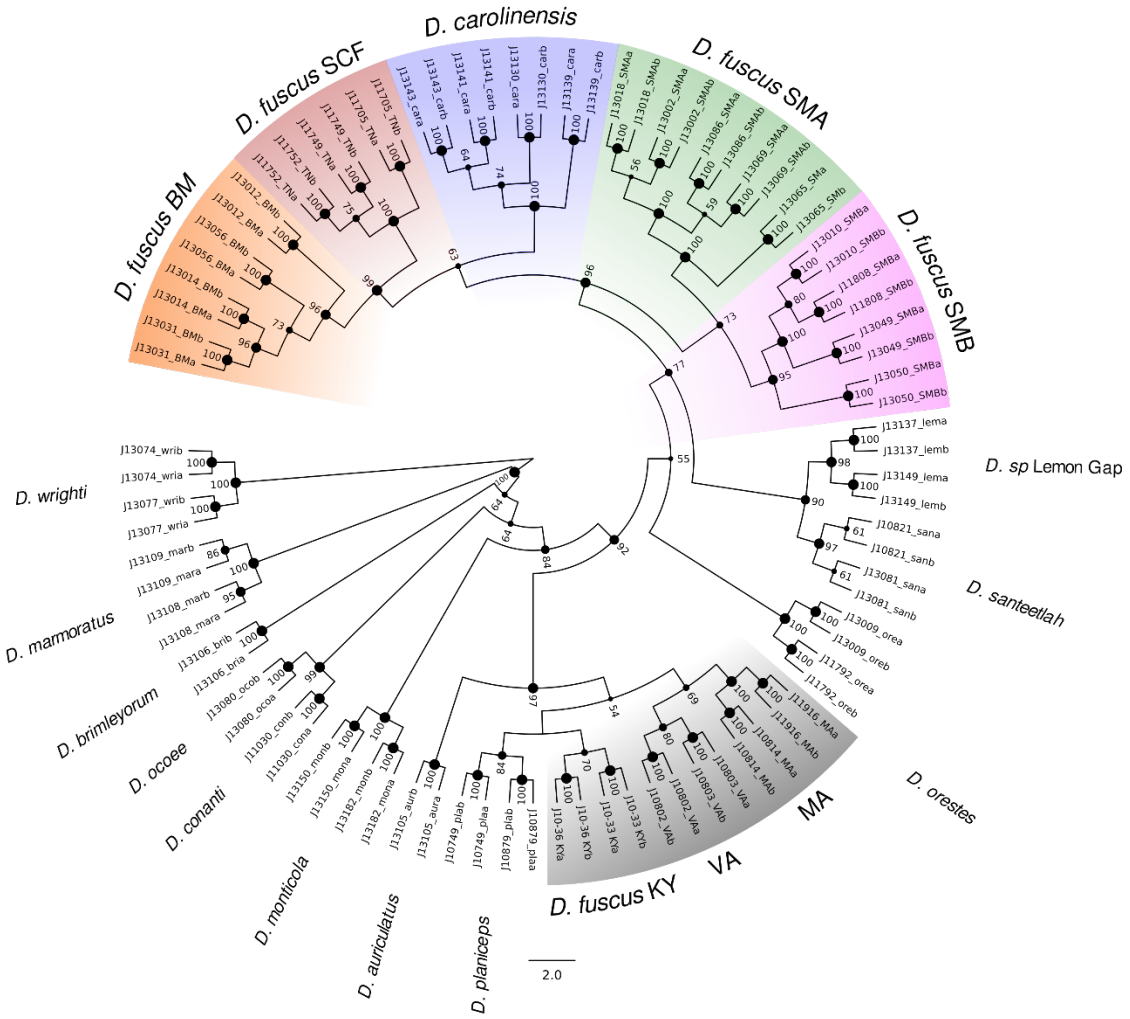


Figure 3.10 – Species tree from all 38 phased loci

Consensus species tree with support values generated from QFM algorithm in SVDquartets using exhaustive sampling and 100 bootstrap replicates. The size of the filled circles are weighted by boot strap support. Well supported focal lineages (>70bs) are shaded following the color scheme from chapter 2.

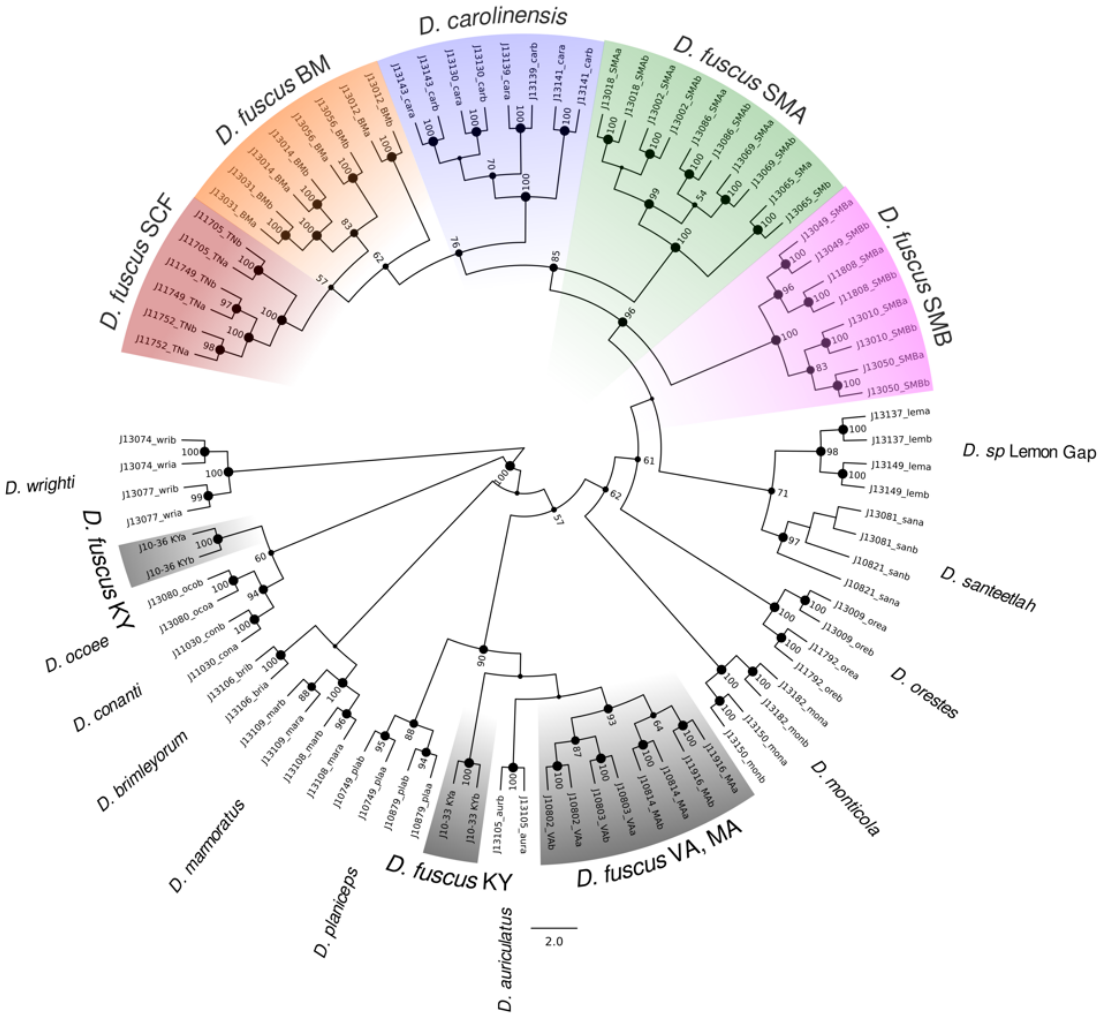


Figure 3.11 – Species tree from 29 phased loci (excluding NEMGs)

Consensus species tree with support values generated from QFM algorithm in SVDquartets using exhaustive sampling and 100 bootstrap replicates. The size of the filled circles are weighted by boot strap support. Well supported focal lineages (>70bs) are shaded following the color scheme from chapter 2.

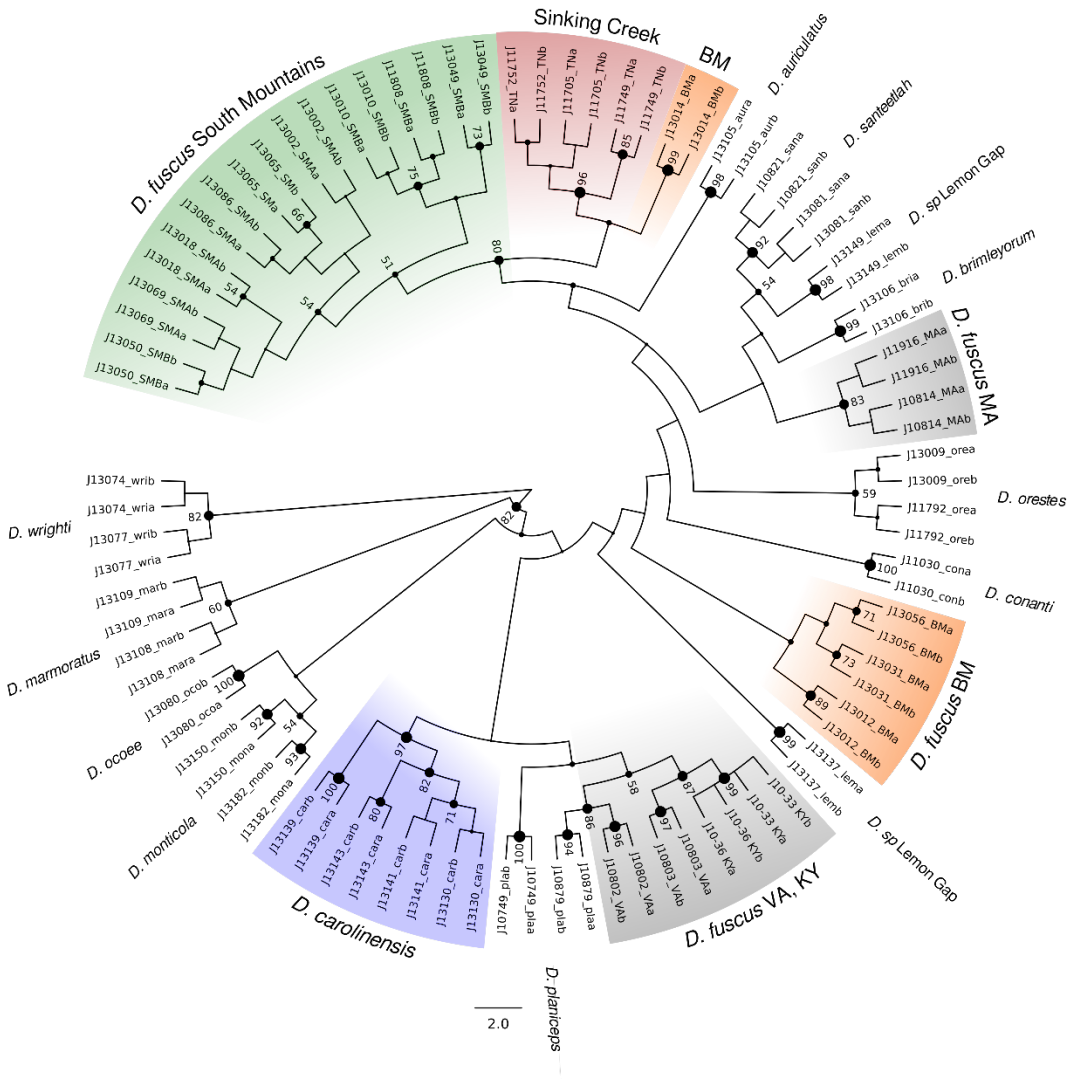


Figure 3.12 – Species tree from 9 phased NEMG loci

Consensus species tree with support values generated from QFM algorithm in SVDquartets using exhaustive sampling and 100 bootstrap replicates. The size of the filled circles are weighted by boot strap support. Well supported focal lineages (>70bs) are shaded following the color scheme from chapter 2.

Within the clade comprised of *D. fuscus* and *D. carolinensis,* Piedmont *D. fuscus* from the South Mountains containing mtDNA associated with the *D. fuscus* clade form a sister lineage to members of the mtDNA α clade (bs > 90). Other *D. fuscus* lineages are still geographically partitioned between populations in KY, VA, and MA, however evolutionary relationships between these tips are unresolved, including a single individual that nests closer to *D. ocoee* and *D. conanti*, but with weak bootstrap support (bs = 60). The majority of the latter *D. fuscus* nested with *D. planiceps* and *D. auriculatus* with strong support (bs ≥90). All other nominate lineages have bootstrap support greater than 80 toward the distal tips of the trees, however deeper nodes in the tree are poorly supported among major clade divisions.

Lineage trees generated using SVDquartets from 9 phased NEMG loci (Figure 3.12) have topological patterns much different from the previous SVDquartet analyses which used a greater number of loci. There is weaker bootstrap support for individuals clustering into the following major lineage divisions ≥80: 1) a clade containing Piedmont and Sinking Creek form *D. fuscus* (which does not include *D. carolinensis*), 2) *D. auriculatus*, 3) *D. santeetlah*, 4) *D. brimleyorum*, 5) *D. fuscus* from MA, 6) *D. conanti*, 7) *D. fuscus* from KY, 8) *D. carolinensis,* 9) *D. ocoee*, and 10) *D. wrighti*. Several other patterns are apparent, such as poorly supported clades for *D. orestes, D. monticola, D. planiceps*, and *D. marmoratus*. Non-monophyletic taxa include *D. sp* Lemon Gap, *D. planiceps*, *D. fuscus* from VA, and Piedmont *D. fuscus* from the Brushy Mountains. Nearly all deep nodes in the phylogeny are poorly supported. There are several prominent topological differences between NEMG and putatively neutral non-NEMG loci in our SVDquartet analyses. Primarily, a clade comprised of nested Piedmont *D. fuscus* and *D.*

carolinensis is not supported. Secondarily, there is no support for separate nuclear lineages in the South Mountains for *D. fuscus* individuals that contain either mtDNA from the α clade or *D. fuscus* clade.

3.3.3 – Patterns of Population Structure

Preliminary population estimates of admixture using the program STRUCTURE indicate similar values of assumed populations (K=2) for either 29 nuclear markers or 9 NEMGs using population level sampling. One replicate analysis using a different random SNP per NEMG locus supported an increased number of assumed populations (K=4) from the data. Analyses using different SNPs from putatively neutral nuclear loci show consistent patterns of population structure between replicates for genetic data collected from *D. fuscus* in the South Mountains, Brushy Mountains, and in Tennessee, referable as Sinking Creek form populations (Fig 3.13). In the first analysis, *D. fuscus* populations in Virginia, Massachusetts, South Mountains, and *D. carolinensis* form a single population, whereas *D. fuscus* populations in KY, TN, and the Brushy Mountains form a second structured population. In the replicate analysis, all individuals including *D. carolinensis* form a single population with the exception of *D. fuscus* individuals from the South Mountains.

When we examine matrices of population structure across NEMGs we observe conflicting patterns between replicates and differences with patterns observed across 29 putatively neutral nuclear markers. In the 1st replicate, *D. fuscus* from Kentucky, Virginia, and Massachusetts form a single structured population with limited admixture with other individuals from populations in Tennessee, the Piedmont, and *D. carolinensis*; these

individuals form a second structured population with a limited number of genomes being of admixed ancestry. The second replicate analysis for K=2 maintains the same structured populations as before, but now includes individuals from populations in Massachusetts and excludes individuals from Virginia. Population structure for either class of marker is not absolute, and there are multiple individuals with low to moderate levels of mixed ancestry in both datasets.

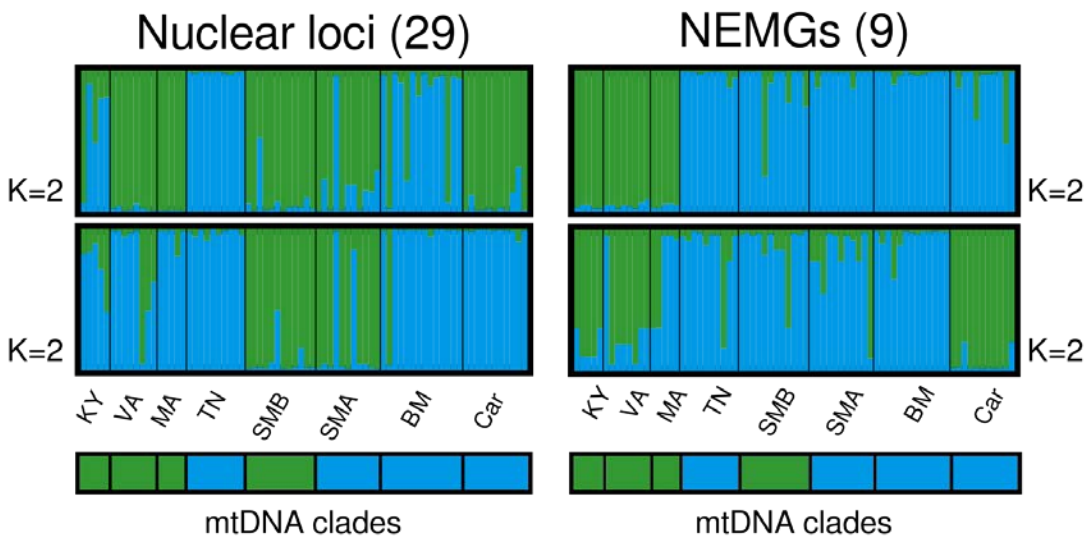


Figure 3.13 – Population structure across *D. fuscus* and *D. carolinensis* comparing nuclear, NEMG loci, and mtDNA clades (K=2)

(Left) Population structure for 29 loci across 76 individuals ordered by population-lineage. (Right) Population structure for 9 NEMG loci across 75 individuals ordered by population-lineage. (Bottom) mtDNA clades for population are arranged in the same order as above. Populations predominantly containing mtDNA from α clade are shaded in blue while populations with mtDNA from the *D. fuscus* clade are shaded in green.

We examined population structure output from replicate analyses with assumed K between 2 and 4 to see other patterns that might exist (Figure 3.14). When K=3, *D. carolinensis* and *D. fuscus* from populations in Kentucky and Virginia form a somewhat structured population, while Piedmont and SCF populations form another admixed population composed of individuals with Brushy Mountain or SCF ancestry. Patterns of structure are less obvious when K=4. There may be a small signature of admixture between *D. carolinensis* and *D. fuscus* from VA, but otherwise there is greater individual variation among regions and a less consistent pattern of geographic structure. Patterns of population genetic structure in NEMGs appear to be more consistent with mtDNA clade lineages, but neither class of markers consistently match patterns from mtDNA clades.

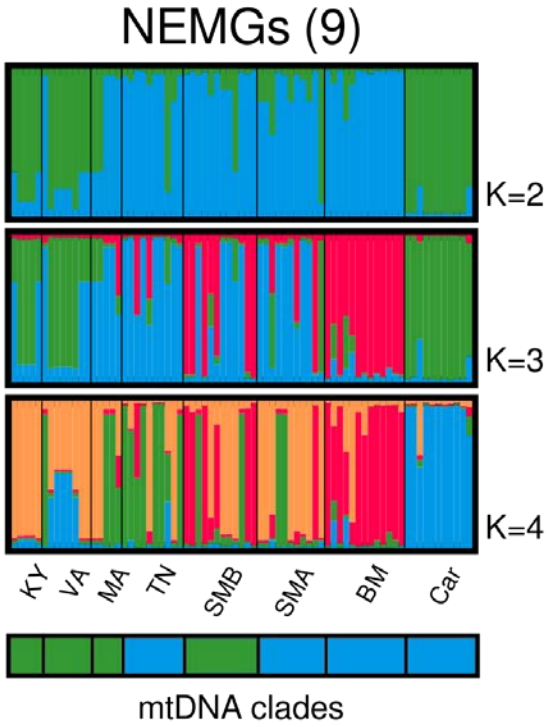


Figure 3.14 – Population structure across a single replicate analysis of *D. fuscus* and *D. carolinensis* comparing NEMG loci and mtDNA clades (K=2 through 4)

(Left) Population structure for 9 NEMG loci across 75 individuals ordered by population-lineage. (Bottom) mtDNA clades for population are arranged in the same order as above. Populations predominantly containing mtDNA from α clade are shaded in blue while populations with mtDNA from the *D. fuscus* clade are shaded in green.

3.4 – Discussion

3.4.1 – PTAS as a sequencing strategy

A major benefit of using PTAS as a method for sequence generation is that it is a viable reduced representation method for organisms with large genomes. This method also balances the initially greater investment in marker development, optimization, and labor with the benefits of obtaining long sequence reads from non-anonymous loci. Due to the

targeted nature of data collection, replicate analyses using PTAS are likely to build upon or improve existing datasets for a particular system and can address specific questions not as well suited for anonymous short read genomic data. A significant cost is that due to primer annealing differences between distantly related taxa, markers may not work outside the system they were developed for leading to marker drop out. Differences at priming sites may also be related to evolutionary rates, therefore markers that more consistently amplify may be found in more conserved genes or regions of the genome.

3.4.2 – Consistency with ddRAD-seq

In this study we successfully obtained long sequence reads for up to 38 targeted loci in representative taxa across *Desmognathus* to test hypotheses of lineage boundaries and examine phylogenetic relationships among taxa with known histories of mitochondrial and nuclear introgression. Initial amplicon data collected from a subset of individuals and taxa was used for BFD analyses, with results being largely in agreement with previous research in *Desmognathus* that used genome wide SNPs across many of the same populations (Kratovil *et al. in prep*). These data support a model separating population-lineages of *D. fuscus* and *D. carolinensis* based on geographic divisions rather than patterns of shared mtDNA haplotype or recognized taxonomy. Models received lower support when *D. fuscus* and *D. carolinensis* were combined based on mtDNA similarities (i.e. whether individuals contain mtDNA from either the α clade or *D. fuscus* clade) or when testing hypotheses based on morphological taxonomy (i.e. a model in which all *D. fuscus* form a single monophyletic lineage). BFD results from this study are in agreement with the analysis of 1380 SNP loci collected using ddRAD, in which nuclear data definitively

support a model where *D. fuscus* lineages within the South Mountains form a single lineage in strong conflict with mtDNA patterns. Together these data support the strong influence of introgression in shaping both the mitochondrial and nuclear genomic histories of these taxa.

3.4.3 – Genus wide patterns

To more definitively place delimited *D. fuscus* and *D. carolinensis* lineages with the genus wide phylogeny, we generated species trees using both Bayesian and quartet analyses. We generated species trees in *BEAST and found that the taxa used for BFD were not distributed evenly throughout the desmognathine phylogeny. If we had included other outgroup lineages for which amplicon data was collected (e.g., *D. conanti*), our BFD results would likely remain the same. Exclusively focusing on BF delimited taxa, the only well supported node (>0.95pp) consists of the ancestor to South Mountain and Brushy Mountain *D. fuscus* Piedmont lineages (Figure 3.9). The other nodes supported by high posterior probabilities are deeper nodes in the phylogeny at the direct ancestor of *D. planiceps*, and at the direct ancestor of *D. marmoratus*. Within the species tree it is apparent that there is greater topological and population parameter uncertainty among deeper nodes and across branches suggesting that the rampant history of gene flow and introgression is not limited to *D. fuscus* and *D. carolinensis* lineages.

3.4.4 – Comparisons between nuclear and NEMG loci

We generated species trees using quartet analyses to compare evolutionary patterns across different classes of loci with the mtDNA genealogy. We expected nuclear genes

with mitochondrial functions to reflect similar patterns as the mitochondrial genome. We were surprised to find that NEMGs did not support the same topological patterns as *nad2,* and that gene trees produced from putatively neutral nuclear markers more often captured the affinity shared between *D. carolinensis* and the introgressed *D. fuscus* lineages east and west of the southern Appalachians. These results seem to suggest that evolutionary patterns in the NEMGs used in this study are decoupled from mtDNA evolutionary patterns. Discordance between different classes of nuclear markers and the underlying species tree is not surprising, and there may be several plausible explanations worth exploring in the future (e.g., incomplete lineage sorting, natural selection, evolutionary rate heterogeneity). Practical considerations, such as the low recovery of candidate NEMG loci and missing data across many taxa may have contributed toward lower confidence in the reconstructed evolutionary histories. Efforts are currently underway to increase representation across *Desmognathus* and improve the capture of sequenced NEMG loci using capture baits for tiled enrichment (RADcapture) (Hoffberg *et al.* 2016).

3.4.5 – Population genetic perspectives

For a population genetic perspective of admixture and gene flow across putatively neutral and NEMG loci, we used SNP data from *D. fuscus* and *D. carolinensis* to generate genotype matrices from 75 or 76 individuals for population structure analyses. These data were compared with mtDNA lineages for populations, which contained either mtDNA from the α clade or *D. fuscus* clade. We initially ran 4 replicates per assumed number of populations (K=1-7), using 1 of 2 random SNPs per locus. We found that evolutionary patterns for neither class of nuclear marker perfectly reflect the dominant mtDNA clade

found in each population, and in only one replicate did NEMG structural patterns more closely match mtDNA clades. Levels of admixture appear much higher in *D. fuscus* populations from the South Mountains regardless of nuclear marker class. Although these results are preliminary, with additional replicate SNP analyses planned in the future, it is apparent that the mtDNA history is a simplistic view that doesn't quite capture the complex demographic history of admixture and gene flow present in the nuclear history of *Desmognathus* populations.

3.4.6 – Future directions

Aside from making improvements to the number of nuclear loci captured and improving representation across *Desmognathus*, we plan to examine heterogeneity and concordance among markers using Bayesian concordance analysis (BCA) in the program BUCKy (Larget *et al.* 2010). Concordance factors are a summary statistic of the proportion of gene trees that contain a given clade, and can be used as a measure of concordance among individual gene trees. BUCKy calculates concordance factors using non-parametric clustering of joint posterior distributions of individual gene trees to inform priors during secondary analyses, without specifying the causes of discordance (Ané *et al.* 2007). Our expectation is that gene trees from NEMGs will have higher concordance factors with each other, in comparison with non-NEMG nuclear loci as a result of evolving under a similar functional constraint. Furthermore, we predict that gene trees reconstructed from NEMGs will have topological patterns that have greater levels of concordance with the mtDNA gene tree, relative to gene trees from non-NEMG markers.

3.5 – Author Contributions

Justin Kratovil and David Weisrock conceived and designed this project. Mary Foley and Jose Bocanegra contributed toward data generation. Justin Kratovil, David Weisrock, Stephen G. Tilley provided tissues. Tissue samples were also provided by MCZ Berkley. Justin Kratovil and David Weisrock analyzed and interpreted the data and wrote the paper.

3.6 – Acknowledgements

I would like to thank Jennifer Webb and Abbe Kesterson for their consultation, advice, and logistical support in preparing DNA libraries for sequencing. I also thank Elisaveta Arnaudova and Jolanta Jaromczyk for their assistance, instruction, and patience in assembling, mining, and managing genomic sequence data.

3.7 – Additional Funding Sources

Financial support for marker development come from the Graduate Research Award through the Society of Systematic Biologists, University of Kentucky Department of Biology Gertrude Flora Ribble Research Grant, NSF EPSCoR Ecological Genomics grant (Grant No. 0814194 and the Commonwealth of Kentucky), NSF 13-568, DEB-DDIG (Grant No. 1406876) to D. Weisrock and J. Kratovil (co-P.I.s). J. Kratovil was also supported by a research assistantship funded through NIH Research Project Grant Program (R01) from the Joint DMS/BIO/NIGMS Math/Bio Program. Grant Number: 1R01GM086888-01. R. Yoshida (P.I)

3.8 – SUPPLEMENTAL INFORMATION

S3.1 – Bayes factor delimitation model descriptions

Model 1: 6 total partitions with greatest number of parameters

(1) *D. fuscus* from the South Mountains that contain mtDNA from the α clade, 2) *D. fuscus* from the South Mountains that contain mtDNA from the *D. fuscus* clade, and 3) *D. fuscus* from the Brushy Mountains that contain mtDNA from the α clade, 4) *D. carolinensis*, which contain mtDNA from the α clade, 5) *D. orestes*, 6) *D. auriculatus.*

Model 2: 5 partitions combining D. fuscus from the South Mountain

(1) *D. fuscus* from the South Mountains that contain mtDNA from the α clade and *D. fuscus* clade, 2) *D. fuscus* from the Brushy Mountains that contain mtDNA from the α clade, 3) *D. carolinensis*, which contain mtDNA from the α clade, 4) *D. orestes*, 5) *D. auriculatus.*

Model 3: 4 partitions combining D. fuscus from the Piedmont (South & Brushy Mountains)

(1) *D. fuscus* from the South Mountains and Brushy Mountains that contain mtDNA from the α clade and D. fuscus clade, 2) *D. carolinensis*, which contain mtDNA from the α clade, 4) *D. orestes*, 5) *D. auriculatus.*

Model 4: 4 partitions combining D. fuscus from the South Mountains and D. carolinensis

(1) *D. fuscus* from the South Mountains and *D. carolinensis* that contain mtDNA from the α clade and *D. fuscus* clade, *2) D. fuscus* from the Brushy Mountains, which contain mtDNA from the α clade, 4) *D. orestes,* 5) *D. auriculatus.*

Model 5: 5 partitions combining D. fuscus from the South Mountains and Brushy Mountains that contain mtDNA from the α clade

(1) *D. fuscus* from the South Mountains and Brushy Mountains which contain mtDNA from the α clade, 2) *D. fuscus* from the South Mountains which contain mtDNA from the *D. fuscus* clade, 3) *D. carolinensis* that contain mtDNA from the α clade, 4) *D. orestes,* 5) *D. auriculatus*.

Model 6: 4 partitions combining α clade lineages

(1) *D. fuscus* from the South Mountains, Brushy Mountains, and *D. carolinensis* that contain mtDNA from the α clade, 2) *D. fuscus* from the South Mountains, which contain mtDNA from the *D. fuscus* clade, 3) *D. orestes,* 4) *D. auriculatus.*

Model 7: 3 partitions combining D. fuscus and *D. auriculatus* lineages

(1) *D. fuscus* from the South Mountains, Brushy Mountains, and *D. auriculatus,* 2) *D. carolinensis,* 3) *D. orestes*

Model 8: 2 partitions combining D. fuscus, D. carolinensis, and D. auriculatus lineages

(1) *D. fuscus* from the South Mountains, Brushy Mountains, *D. carolinensis,* and *D. auriculatus,* 3) *D. orestes*

Model 9: 3 partitions combining α clade lineages and D. fuscus from the South Mountains containing mtDNA from the D. fuscus clade

(1) *D. fuscus* from the Piedmont and *D. carolinensis*, 2) *D. orestes*, 3) *D. auriculatus*

Chapter 4 : Sexual isolation between divergent lineages of *Desmognathus*

4.1 – INTRODUCTION

Reproductive isolation is undoubtedly one of the most important criteria in evolution for distinguishing discrete species boundaries and for studying mechanisms leading to speciation (Coyne & Orr 2004). Metrics of reproductive isolation can be used to diagnose the potential for gene flow, identify traits contributing to isolation or admixture, and in conjunction with molecular genetic analyses can inform species delimitation efforts. Recently, the utility of genomic sequencing has made it apparent that non-bifurcating evolutionary histories and divergence with gene flow are not as exceptional or rare a pattern as once thought (Wen *et al.* 2016; Kumar *et al.* 2017). Introgression can influence the evolutionary topology of lineages and generate discordant patterns at the genomic level, but it is not well understood how quickly this process influences the expression of complex traits and elaborate behaviors which facilitate or inhibit gene flow among taxa (Baack & Rieseberg 2007).

Signatures of introgression provide an intriguing glimpse into the genetic regions potentially capable of influencing adaptive range expansion, creating genetic novelty, or instigating divergence and the process of speciation (Dowling & Secor 1997). Tests for introgression are also informative in helping investigators choose loci appropriate for species delimitation, as the genetic background of individual non-reticulate gene histories are expected to more closely follow the underlying species history (Petit & Excoffier 2009). Introgressed regions can represent a relatively small and anomalous fraction of the genome that hasn't been lost due to selection, high intraspecific gene flow or drift,

signifying an important remnant of past genetic exchange during the evolutionary history shared between species.

Dusky salamanders (genus *Desmognathus*) are a major component of plethodontid communities in the eastern North America (Hairston Sr 1986; Rissler & Taylor 2003). Despite advancements in molecular systematics over the past forty years, the evolutionary relationships within *Desmognathus* have been a significant challenge to resolve. Species exhibit cryptic genetic variation in allozyme electrophoretic migration profiles and in mtDNA sequences in stark contrast with morphological characteristics which are habitat associated and conserved across taxa (Petranka 1998). A limited number of nuclear markers are available for plethodontid salamanders but deeper evolutionary relationships among *Desmognathus* are poorly resolved (Kozak *et al.* 2009).

Molecular studies in this system have heavily relied upon mitochondrial markers, and sequence data have revealed remarkable genetic divergence among lineages distributed across the Ridge and Valley (Tilley *et al.* 2013), Blue Ridge Mountains (Kozak *et al.* 2005), Piedmont (Kratovil 2007; Tilley *et al.* 2008), and Atlantic Coastal Plain (Beamer & Lamb 2008). While isolation by distance may be contributing to some of these patterns, mtDNA introgression may account for high sequence identities shared between populations of *D. carolinensis,* a smaller montane species restricted to the Blue Ridge Mountains, and *D. fuscus,* a medium sized salamander with a larger, but non-overlapping species range with *D. carolinensis* (Tilley *et al.* 2013; Figure 4.15).

It is unknown if and to what degree previous historical contact and introgression between populations shape the evolution of reproductive barriers, and these interactions may contribute toward the accumulation and maintenance of *D. fuscus* lineages without

generating morphological disparities. To address these basic questions concerning and provide further insight into species boundaries, we measured prezygotic reproductive isolation using mating trials between populations of *D. fuscus* and *D. carolinensis*. We focus on populations of *D. fuscus* within the South Mountains which contain divergent mtDNA lineages, *D. fuscus* populations in the Brushy Mountains, and *D. carolinensis* collected near the type locality for the species (Table 2.1; Figure 2.2).

Reproductive isolation studies have been utilized extensively in evolutionary and behavioral studies of plethodontid salamanders to help delimit cryptic species, identify contact zones between diverging lineages, and study the evolution of chemical signaling and delivery (Watts *et al.* 2004). Plethodontid salamanders have elaborate courtship behaviors which provide a means to unambiguously measure a number of discrete sexual behaviors that lead to, or prevent, insemination between conspecific or heterospecific pairs. Plethodontid salamanders readily mate multiple times during their breeding season, even in experimental conditions, and it is relatively easy to measure receptivity by timing courtship progression because each stage leading to insemination depends on the co-receptivity of both male and female participants. Courtship can break down at any point due to mechanical, signaling, or sex specific incompatibilities during pheromone delivery, spermatophore deposition, and insemination, allowing investigators to identify the specific causes and timing of sexual incompatibilities (Arnold *et al.* 1993).

Mating trials have been well documented using *Desmognathus* and have been used to identify contact zones, barriers of gene flow between species, and measures of isolation (Houck *et al.* 1988; Tilley *et al.* 1990; Arnold *et al.* 1993, 1996; Arnold 1993; Mead *et al.* 2001). To our knowledge, controlled mating trials between *D. carolinensis* and *D. fuscus*

have never been published, however *D. carolinensis* and *D. fuscus* are both reproductively isolated from *D. ocoee,* a species morphologically similar to *D. carolinensis*. *D. ocoee* is sympatric with *D. fuscus* and has a limited contact zone with *D. carolinensis* (Verrell 1990b; Mead *et al.* 2001).

Figure 4.15 – Images of *D. fuscus* and *D. carolinensis*

Among other characteristics, *D. fuscus* (left) are notably larger, robust, and characterized as having a slightly keeled tail in cross section, whereas *D. carolinensis* (right) are smaller, gracile, and have a rounded tail in cross section. Photo credits (*D.* fuscus, Kratovil 2007, South Mountains State Park; *D. carolinensis,* Tilley 2006)

4.2 – Materials and Methods

4.2.1 – Specimen collection

D. fuscus from localities identified previously as predominantly containing the mtDNA lineages Piedmont A or B (Kratovil 2007; Chapter 1) were collected by hand in the South Mountains and Brushy Mountains for pre-zygotic isolation trials. In addition, *D. carolinensis* and *D. sp* Lemon Gap, were collected to more extensively assay pre-zygotic

isolation between diverse salamander populations with either similar morphologies (e.g. *D. sp* Lemon Gap and *D. carolinensis*) or differing levels mtDNA introgression (e.g. *D. carolinensis* x *D. fuscus* Piedmont A, or *D. carolinensis* x *D. fuscus* Piedmont B). In total, 127 salamander were collected, 31 from group "SMA", 30 from "SMB", 30 from "BM", 24 from "carolinensis", and 12 from "LG". Salamanders were transported in ice chests, and individually maintained in plastic shoeboxes at Highlands Biological Station or the University of Kentucky. To account for the possibility of low frequency admixture or changes in population structure, I confirmed mitochondrial haplotype/lineage identities for each individual used in this study for group designations. All steps for DNA extraction, sequencing, and analysis of *nad2* fragments are described previously in chapter 2.

4.2.2 – Animal care and initial mating trial description

Laboratory conditions for conducting sexual isolation tests and maintaining animals followed the design of Arnold et al. (1990) and Houck et al. (1988) (L. Houck, personal communication). Animals were individually housed in clear plastic shoe boxes (9 x 17 x 31cm) containing damp crumpled unbleached towels at 15-16°C on a natural photoperiod. Salamanders were initially fed *Drosophila* larvae and adults *ad libitum,* and 1 wax worm moth larvae between mating trials.

The schedule for preliminary mating trials was based on three test groups for males and females, corresponding to salamanders from localities containing mtDNA lineages from the South Mountains identified as Piedmont A or B, and Piedmont A mtDNA lineages in the Brushy Mountains. 405 unique serial mating trials took place over the course of nine evenings, four days apart, in which a single female encountered either a within-lineage

(conspecific) male or between-lineage (heterospecific) male in a clean plastic enclosure containing a flat damp paper towel substrate 30-60 minutes before sunset. During the course of this experiment, each female encountered up to 3 different conspecific males, and 6 different heterospecific males (3 from each locality), evenly partitioned across multiple trials to minimize any effects from sequential matings and individual variation over the course of the study. An example of the paired matrix used is available in supplemental (S4.1). In the morning after each trial, males and females were returned to their individual housing containers.

4.2.3 – Mating behavior and observations

Observations and timing of sexual behavior during trials were scored under red light for a minimum of 5 hours with observations recorded in 15-20 minute intervals to determine behaviors leading to courtship cessation and isolation. Observed sexual behaviors were scored qualitatively using descriptions of mating stages and behaviors following the guidance of (Houck & Arnold 2003) which describe initial contact between salamanders beginning at the “discovery” phase, through “courtship”, and “tail straddle walk” which can lead to spermatophore deposition and possibly insemination.

Briefly, initial contact involves physical and chemical contact between males and females in which individuals nudge or tap each other using their heads, teeth, and chins leading to the “pursuit” phase. Observed or inferred initial contact was required for scoring the duration of mating trials. During pursuit, males follow or intercept females often nudging or making physical contact. The courtship phase began when males were observed making specific behaviors for chemical delivery, such as vaccination of the female's

dorsum or tail using vomerine teeth, pheromone delivery via subdermal mental gland to the female's chin or head, or substrate tapping, strutting/posturing or “butterfly” walk using their forelimbs, or tail undulation by the male. Courtship proceeds to the “tail straddle walk” in which the female straddles the tail of the undulating male as he walks ahead of her. Spermatophore “deposition” is readily observable as the male undulating his tail in a stationary position, followed by the pivoting his tail perpendicular to his body as he walks forward to the point of guiding the female for insemination. This information was used to measure the duration of mating stages between conspecific and heterospecific pairings, in particular the “latency to mate”, or the delay between initial contact and courtship displays, duration of courtship, and duration of the tail straddle walk in fifteen minute time intervals. When advanced mating stages were observed without directly observing the first instance of transitional behaviors, I inferred that these behaviors transpired during the previous observational period.

4.2.4 – Metrics of Isolation

On the morning following each trial (12-13 hrs later), mating encounter boxes were examined for the presence of spermatophore caps, gelatinous spermatophore bases, and insemination; the unambiguous presence of a sperm mass in the female's cloaca, which enabled me to measure the probability of insemination given the deposition of a spermatophores/bases during the trial. Measures of sexual isolation were calculated by evaluating *joint isolation* (*JI*), *isolation asymmetry* (*IA*), and *propensity asymmetry* (*PA*) for all crosses using *D. fuscus* lineages following Arnold et al. (1996). Briefly, *JI* is the sum of homotypic rates of successful mating minus the sum of heterotypic rates of success

and can be expressed by the equation: PAA + PBB - (PAB + PBA). *JI* values range from -2 to 2, but effectively range from 0 to 2 because negative values (higher heterotypic probabilities relative to homotypic probabilities of mating) are not expected.

Isolation asymmetry is a metric of differences between heterotypic mating successes between crosses and can be expressed by the equation |(PAB - PBA)|. Asymmetries result from higher willingness to mate or a lack of pre-zygotic boundaries in one direction relative to the other. *Propensity asymmetry* is a similar metric comparing the differences in the rates of mating between homotypic crosses of test groups. *PA* can be expressed by the equation |(PAA - PBB)|. Theoretical expectations for both *IA* and *PA* values range from 0 to 1, with estimates compared against a hypothetical mean of 0, in which isolating barriers are inferred to be absent between test groups.

All trials involved unique paired crosses between individuals mated multiple times, therefore variance and standard errors for hypothesis testing followed procedures designed for salamander mating trials by McCullagh and Nelder (1989). Metrics of isolation were tested against a null hypothesis that coefficients were equal to zero (no isolation) between test groups using t-tests.

4.2.5 – Mating trials between D. fuscus and D. carolinensis

The schedule for mating trials between *D. fuscus* and *D. carolinensis* were based on four test groups using males and females corresponding to salamanders from localities containing mtDNA lineages referable as Piedmont A from the South Mountains, Piedmont B from the South Mountains, *D. carolinensis* 4 km from the type locality of the species (Mt. Mitchel, NC), and *D. sp* near Lemon Gap (Cocke County, TN). To limit the scope of

trials, we excluded *D. fuscus* from the Brushy Mountains. An example of the paired mating strategy is available in supplemental materials (S4.2).

Unlike the preliminary mating trials, each female from one of the 4 major groups encountered all possible conspecific and heterospecific males over the course of 25 trial nights, four days apart, evenly partitioning encounters between different groups across multiple trials to negate the effects of sequential matings or changes in receptivity over the course of the experiment. In addition, to account for differences in the number of males or females collected from a location, we used "excess" individuals for separate and concurrent paired mating trials. We included 14 additional females and 16 additional males from across test group to supplement mating trials or serve as alternate mates in case any individuals needed to be removed from the study (e.g. injured, sick, or deceased). The schedule of supplemental mating trials were randomized to prevent individuals from encountering each other more than once and males and balance encounters between test groups across the duration of the study (S4.2).

In total, 779 unique serial mating trials took place using the same experimental conditions as previously described to record mating success. After each trial, males and females were returned to their individual housing containers.

Mating behaviors were digitally recorded from 8 mating crosses each night using SwannView security cameras. Two random females from each test group were recorded with infrared digital video recorders positioned above enclosures for the duration of mating trials. Digital recording were saved to a hard drive and we measured the duration of mating stages between pairs that successfully mated. Mating duration was measured from initial contact until "pursuit", courtship until tail straddle walk, and duration of "tail straddle

walk" leading to successful spermatophore deposition. Mating was scored as successful if at least a single spermatophore was deposited in the enclosure, or if a female was inseminated. It is important to note that salamander mating is a lengthy process, linear in progression from stage to stage, but sometimes would repeat stages before advancing further. For purposes of measuring mating stage durations, once a mating stage advanced forward, the duration of that phase was tallied even if the mating pair repeated previous stages multiple times.

4.3 – RESULTS

4.3.1 – Preliminary mating trials

4.3.1.1 – Metrics of reproductive isolation

There were a total of 381 unique trials of which 126 were between homotypic (within-lineage) pairs (Table 4.10) of *D. fuscus*. Twenty four randomized crosses initially planned did not take place due to extreme size differences between large and small salamanders (which may result in cannibalism) and the absence of available substitute partners of equal size from either population during trial set-up. Differences between the number of females and males collected from each population also limited the total number of trials possible. No encounters were excluded from analysis, including non-mating (non-participating) individuals over the course of the experiment.

Across all test groups examined, successful mating was observed between homotypic and heterotypic pairs at varying frequencies. The frequency of successful matings in which at least 1 spermatophore was deposited varied between homotypic pairs, ranging from 0.422 to 0.143 of encounters (mean 0.2482, std dev 0.1240). The

Table 4.10 – Preliminary *D. fuscus* mating trials

Pair	πAA'	πAB'	πBA'	πBB'	JI ± SE	(P) JI	IA ± SE	(P) IA	PI ± SE	(P) PI
SM1 x SM2	0.422 [0.222] (45)	0.128 [0.051] (39)	0.356 [0.244] (45)	0.180 [0.077] (39)	0.118 ± 0.113 [0.003 ± 0.100]	0.3005 [0.9762]	0.227 ± 0.114 [0.193 ± 0.091]	0.0510 [**0.0381**]	0.243 ± 0.125 [0.145 ± 0.092]	0.0569 [0.1203]
SM1 x BM	0.422 [0.222] (45)	0.340 [0.191] (47)	0.442 [0.209] (43)	0.143 [0.048] (42)	-0.217 ± 0.118 [-0.131 ± 0.108]	0.0713 [0.2332]	0.101 ± 0.138 [0.018 ± 0.091]	0.4652 [0.8438]	0.279 ± 0.121 [0.175 ± 0.076]	0.0239 [**0.0246**]
SM2 x BM	0.180 [0.077] (39)	0.239 [0.174] (46)	0.200 [0.057] (35)	0.143 [0.048] (42)	-0.117 ± 0.114 [-0.107 ± 0.102]	0.3088 [0.2852]	0.039 ± 0.111 [0.117 ± 0.094]	0.3088 [0.2181]	0.037 ± 0.098 [0.029 ± 0.068]	0.7112 [0.6713]

Results of pairwise crosses between *D. fuscus* populations. SM1, SM2, and BM refer to South Mountains and Brushy Mountain populations. The frequency of successful mating results (π) defined by spermatophore deposition is listed above insemination rates in brackets. Females are identified first (A or B) and males are identified second (A' or B'). The total number of encounters for each class is listed in parentheses. Significant results are in bold. JI = Joint Isolation, IA = Isolation Asymmetry, PI = Propensity Asymmetry

insemination frequency between homotypic pairs was lower, ranging from 0.222 to 0.048 of encounters (mean 0.116, std dev 0.076). The frequency of spermatophore depositions and inseminations between heterotypic crosses was similar to homotypic rates, ranging from 0.442 to 0.128 (mean 0.242, std dev 0.114) and from 0.244 to 0.051 (mean 0.15, std dev 0.07) of encounters, respectively.

Metrics of isolation are calculated based on the frequency of successful matings between homotypic and heterotypic crosses with standard errors based on variation in mating success between males and females. Joint isolation estimates for spermatophore deposition in all paired crosses range from -0.217 ± 0.118 to 0.118 ± 0.113 while measures of joint isolation using insemination frequency are lower, ranging from -0.131 ± 0.108 to 0.003 ± 0.100. In all pairwise crosses, measures of joint isolation are no different than

expected under admixture or unrestricted mating (P>0.05) against a t-distribution with a hypothetical mean of 0.

Isolation asymmetry scores for spermatophore deposition comparing heterotypic rates of successful mating between crosses range from 0.227 ± 0.114 to 0.039 ± 0.111. Measures of isolation asymmetry using insemination frequency range from 0.193 ± 0.091 to 0.018 ± 0.091. Measures of isolation asymmetry using spermatophore deposition rates between all crosses are no different than expected under admixture (P>0.05) against a t-distribution with a hypothetical mean of 0. Isolation asymmetry between SM Piedmont A and SM Piedmont B using successful deposition rates is marginal (P= 0.0510). The null hypothesis that isolation asymmetry between males and females of SM Piedmont A and SM Piedmont B using insemination rates can be rejected (IA =0.193 ± 0.091; P=0.0381).

Propensity asymmetry scores are measured by successful mating between homotypic crosses. Spermatophore deposition rates in homotypic crosses range from 0.279 ± 0.121 to 0.037 ± 0.098. Measures of propensity asymmetry using insemination frequency range from 0.175 ± 0.076 to 0.029 ± 0.068. Measures of propensity asymmetry in spermatophore deposition between homotypic crosses of SM mtDNA lineages, and crosses between South Mountain Piedmont B and Brushy Mountain mtDNA lineages are no different than expected under admixture (P>0.05) against a t-distribution with a hypothetical mean of 0. It is noteworthy that the propensity asymmetry scores for spermatophore deposition between South Mountain Piedmont A and South Mountains Piedmont B is marginal (P=0.0569), but that insemination rates cannot reject the hull hypothesis that propensity asymmetry is equal to admixture (P=0.1203). The null hypothesis that propensity asymmetries are equal between homotypic pairs of South

Mountain Piedmont A and Brushy Mountain Piedmont A can be rejected for spermatophore deposition rate (PA = 0.0279 ± 0.121; P=0.0239) and insemination rate (PA = 0.175 ± 0.076; P=0.0246).

4.3.1.2 – Mating stage durations

The latency to mate between *D. fuscus* test groups was measured in 15 minute intervals, and recorded durations were transformed using a base 10 logarithm to generate summary statistics. Latency to mate varied widely across trials ranging from 1.176 log minutes (within the first 15 minutes) to 2.6 log minutes (>6 hrs). There were however no differences in the mean durations for latency to mate comparing all 9 test groups using a one-way ANOVA (P>0.05) (Figure 4.16). In addition, there were marginal differences in the durations of courtship between homotypic and heterotypic *D. fuscus* crosses involving South Mountain Piedmont A or B females mating with South Mountains A or B males (mean 1.381; 2.038, 1.642; P=0.07877)(Figure 4.17). We rejected the null hypothesis that there are no barriers of gene flow between South Mountain Piedmont A and South Mountain B mtDNA lineages (*IA* = 0.194; P=0.0361). We fail to reject the null hypothesis that there are no differences in the mean latency to mate or mean courtship durations between South Mountain A females and B males or South Mountain B females and A males (mean 1.176, 1.48, 1.785, 1.475; P>0.05). Finally, we reject the null hypothesis that there were no differences in mean courtship durations between homotypic crosses for South Mountain Piedmont A and South Mountain Piedmont B (means 1.381 and 2.038; P=0.01365).

4.3.2 – Mating trials between Piedmont D. fuscus, D. carolinensis, and D. sp Lemon Gap

4.3.2.1 – Metrics of reproductive isolation

There were a total of 779 unique mating trials of which 221 were between homotypic (within-lineage) pairs of *D. fuscus*, *D. carolinensis*, and *D. sp* Lemon Gap (Table 4.11). In this study, mating trials occurred between all possible crosses of females

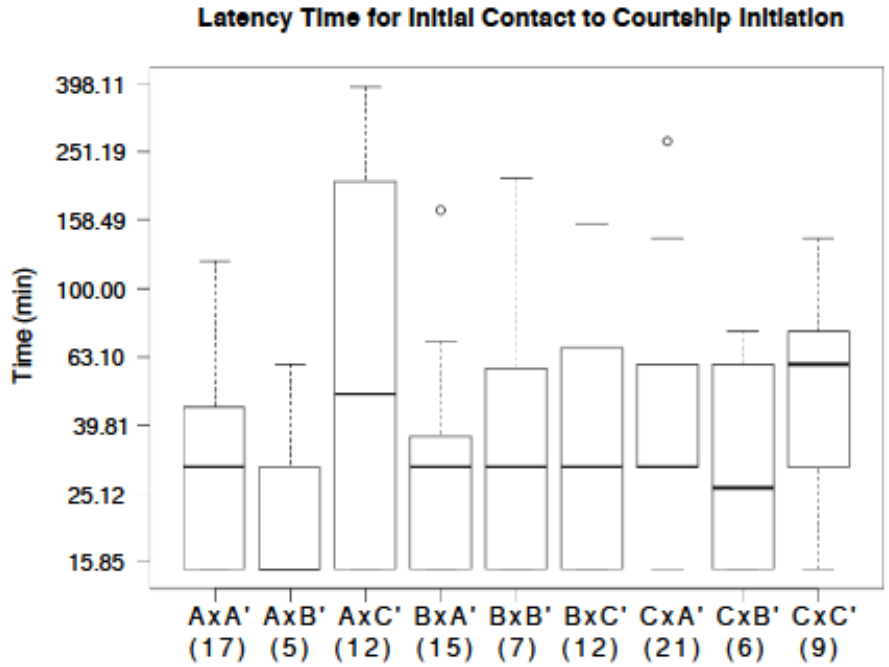


Figure 4.16 – Latency to mate in Piedmont *D. fuscus*

Test crosses are listed as females x males'. “A” refers to South Mountains (Piedmont A), “B” refers to South Mountains (Piedmont B), and “C” refers to Brushy Mountains (Piedmont A). Number of trial observations are listed (n)

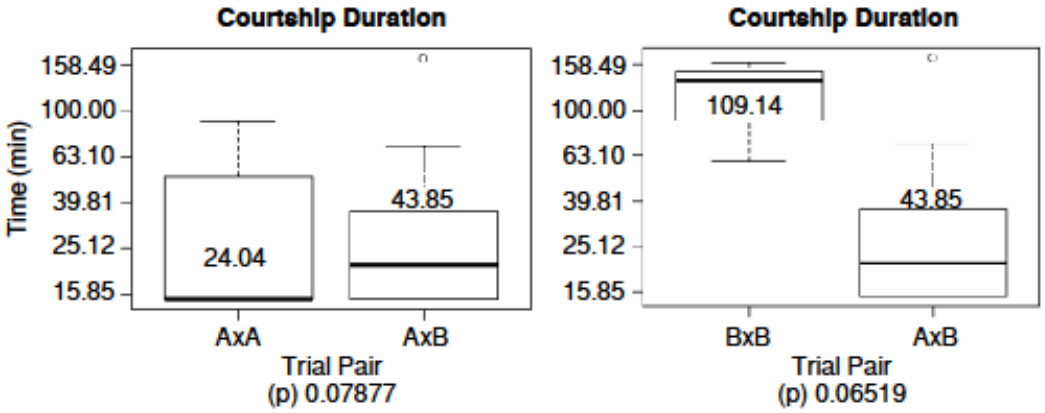


Figure 4.17 – Courtship duration in *Piedmont D. fuscus*

Test crosses are listed as females x males'. “A” refers to South Mountains (Piedmont A), “B” refers to South Mountains (Piedmont B). P-values are listed for each analysis

Table 4.11 – Successful mating trials between *D. fuscus, D. carolinensis,* and *D. sp* Lemon Gap

Results of successful depositions (left) or inseminations (right) between pairwise crosses of *Desmognathus* populations. Females are identified first (A or B) and males are identified second (A' or B'). “LG” = *D. sp* Lemon Gap, “BMC” = Black Mountain *D. carolinensis*, “SMA” = South Mountain *D. fuscus* Piedmont “A”, “SMB” = South Mountain *D. fuscus* Piedmont “B”

Total successful matings (depositions)					Total successful matings (inseminations)				
Pair	AxA'	AxB'	BxA'	BxB'	Pair	AxA'	AxB'	BxA'	BxB'
LG x BMC	18/20	27/39	40/49	71/96	LG x BMC	3/20	8/39	0/49	37/96
LG x SMA	18/20	5/22	18/37	34/60	LG x SMA	3/20	0/22	1/37	0/60
LG x SMB	18/20	2/18	18/41	8/45	LG x SMB	3/20	0/18	0/41	0/45
BMC x SMA	71/96	14/57	44/67	34/60	BMC x SMA	37/96	0/57	3/67	0/60
BMC x SMB	71/96	5/42	36/71	8/45	BMC x SMB	37/96	0/42	2/71	0/45
SMA x SMB	34/60	18/48	29/67	8/45	SMA x SMB	0/60	0/48	0/67	0/45

and males collected from each test population with individuals never encountering each other more than once. As in the preliminary study, encounters between males from each test group were evenly distributed across female test groups over the course of twenty five trial nights, four days apart. No encounters were excluded from analysis, including non-mating (non-participating) individuals during the experiment.

Across all test groups examined, successful mating was observed between homotypic and heterotypic pairs of *D. fuscus*, *D. carolinensis*, and *D. sp* Lemon Gap at varying frequencies (Table 4.12). The frequency of successful mating trials in which at least 1 spermatophore was deposited varied between homotypic pairs, ranging from 0.1778 to 0.9 of encounters (mean 0.5960, std dev 0.2687). The insemination frequency between homotypic pairs was lower, ranging from 0 to 0.3854 of encounters (mean 0.1339, std dev 0.1576). The frequency of spermatophore depositions and inseminations between heterotypic crosses was lower than homotypic rates, ranging from 0.1111 to 0.8163 (mean 0.4257, std dev 0.2150) and from 0 to 0.2051 (mean 0.0254, std dev 0.0562) encounters, respectively.

Metrics of isolation are calculated based on the frequency of successful matings between homotypic and heterotypic crosses with standard errors based on variation in mating success between males and females. Joint isolation estimates for spermatophore deposition in all paired crosses range from -0.0634 ± 0.1795 to 0.7529 ± 0.0684 while measures of joint isolation using insemination frequency are lower, ranging from 0.000 to 0.3572 ± 0.3304. There are no instances of homotypic or heterotypic insemination between Piedmont A and Piedmont B crosses, which prevent us from estimating isolation metrics

Table 4.12 – Reproductive isolation metrics between *D. fuscus, D. carolinensis,* and *D. sp* Lemon Gap

Results of pairwise mating crosses between *Desmognathus* populations. The frequency of successful mating results (π) defined by spermatophore deposition is listed above insemination rates by rows. Females are identified first (A or B) and males are identified second (A' or B'). "LG" = *D. sp* Lemon Gap, "BMC" = Black Mountain *D. carolinensis,* "SMA" = South Mountain *D. fuscus* Piedmont "A", "SMB" = South Mountain *D. fuscus* Piedmont "B". JI = Joint Isolation, IA = Isolation Asymmetry, PI = Propensity Asymmetry

Pair		πAA'	πAB'	πBA'	πBB'	JI	JI SE	(P) JI	IA	IA SE	(P) IA	PI	PI SE	(P) PI
LG x BMC	D	0.9	0.6923	0.8163	0.7396	0.1309	0.2653	0.6231	0.124	0.3483	0.7227	0.1604	0.2481	0.5192
	I	0.15	0.2051	0	0.3854	0.3303	0.0234	**<0.0001**	0.2051	0.1511	0.1782	0.2354	0.217	0.2803
LG x SMA	D	0.9	0.2273	0.4865	0.5667	0.7529	0.0346	**<0.0001**	0.2592	0.2011	0.2025	0.3333	0.1426	**0.022**
	I	0.15	0	0.027	0	0.123	0.1834	0.5036	0.027	0.0384	0.4848	0.15	0.199	0.4532
LG x SMB	D	0.9	0.1111	0.439	0.1778	0.5276	0.0521	**<0.0001**	0.3279	0.2015	0.1091	0.7222	0.1429	**<0.0001**
	I	0.15	0	0	0	0.15	0.1413	0.2905	0	0	0.32	0.15	0.1413	0.2924
BMC x SMA	D	0.7396	0.2456	0.6567	0.5667	0.4039	0.1529	**0.0087**	0.4111	0.1936	**0.0394**	0.1729	0.1941	0.3836
	I	0.3854	0	0.0448	0	0.3406	0.1133	**0.0029**	0.0448	0.0336	0.1849	0.3854	0.1141	**0.0009**
BMC x SMB	D	0.7396	0.119	0.507	0.1778	0.2913	0.1734	0.0942	0.388	0.2207	0.0815	0.5618	0.1935	**0.0043**
	I	0.3854	0	0.0282	0	0.3572	0.1678	**0.0342**	0.0282	0.0329	0.3932	0.3854	0.2028	0.0594
SMA x SMB	D	0.5667	0.375	0.4328	0.1778	-0.0634	0.0911	0.4872	0.0578	0.205	0.7785	0.3889	0.1777	**0.0309**
	I	0	0	0	0	0	NA	NA	0	NA	NA	0	NA	NA

between these test groups. We reject several null hypotheses stating that measures of joint isolation were equal to zero in crosses measuring spermatophore deposition between LG and A (P=0.0001), LG and B (P=0.0001), and *D. carolinensis* and A (P=0.0087). Examining insemination rates, the null hypothesis of no isolating barriers or a mean score of zero is rejected between *D. sp* Lemon Gap and *D. carolinensis* (P=0.0001), *D. carolinensis* and South Mountain Piedmont A (P=0.00145), and *D. carolinensis* and South Mountain Piedmont B (P=0.0171).

Isolation asymmetry scores for spermatophore deposition comparing heterotypic rates of successful mating between crosses range from 0.0578 ± 0.4061 to 0.4111 ± 0.3832. Measures of isolation asymmetry using insemination frequency range from 0 to 0.2051 ± 0.300. Measures of isolation asymmetry using spermatophore deposition rates between *D. carolinensis* and South Mountain Piedmont A are different than expected if there were no barriers to gene flow (P=0.0394). In all other *IA* measures using spermatophore deposition or insemination, the null hypothesis cannot be rejected.

Propensity asymmetry scores for spermatophore deposition comparing homotypic rates of successful mating between crosses range from 0.1604 ± 0.4914 to 0.7222 ± 0.2855. Measures of propensity asymmetry using insemination frequency range from 0 to 0.3854 ± 0.2254. Measures of propensity asymmetry in spermatophore deposition are significantly different from expectations under a null hypothesis of admixture using a mean of 0 between homotypic crosses of *D. sp* Lemon Gap x South Mountain Piedmont A lineages (P=0.0220), *D. sp* Lemon Gap x South Mountain Piedmont B (P=0.0001), *D. carolinensis* x South Mountain Piedmont B (P=0.0043), and crosses between South Mountain Piedmont A and South Mountain Piedmont B lineages (P=0.0309). Regarding *PA* measures using

insemination rates, we only reject the null hypothesis for crosses between *D. carolinensis* and South Mountain A (P=.0009) and crosses between *D. carolinensis* and South Mountain B are marginal (P=0.0594).

4.3.2.2 – Mating durations

Durations of mating stages for 60 successful mating pairs were measured from digital recordings across 24 trial nights. We restricted analyses to only measuring the duration of mating stages between successful mating pairs, measuring the time from initial contact until the first successful spermatophore deposition to facilitate project completion. The stages measured include latency to mate, courtship duration, duration of tail straddle walk, and total mating duration between *D. fuscus* Piedmont A, *D. fuscus* South Mountain Piedmont B, *D. carolinensis,* and *D. sp* Lemon Gap test groups.

Recorded durations were transformed using the natural logarithm prior to statistical analyses using one-way ANOVA and paired t-tests. We used a Tukey HSD Post-hoc test to make pairwise comparisons between groups when mean durations significantly differed after ANOVA statistical tests. For clarity, all comparative analyses are described from the perspective of female salamanders from a single test group mating with males from one of the test groups.

Mean duration of mating stages involving female *D. carolinensis* did not statistically differ among *D. carolinensis*, South Mountains Piedmont A, and *D. sp* Lemon Gap test groups (S4.3; Figure 4.18). We were unsuccessful in digitally recording any successful mating trials between female *D. carolinensis* and South Mountain B males. The

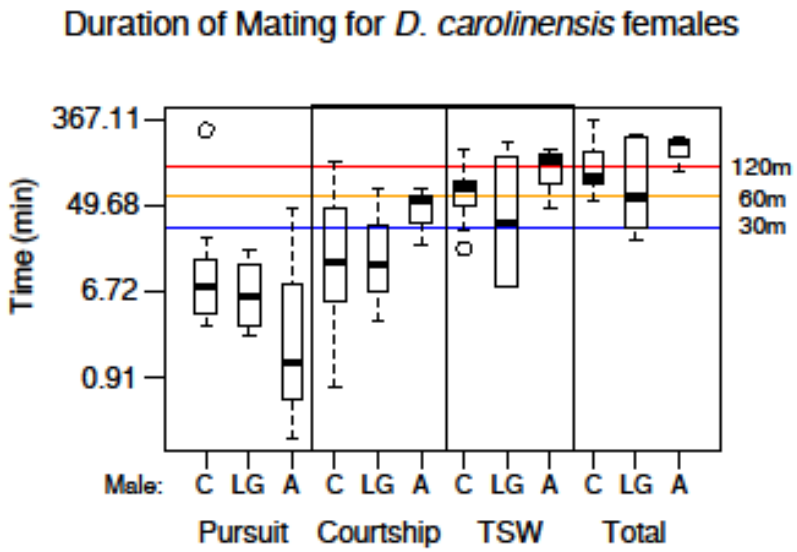


Figure 4.18 – Durations of mating stages for *D. carolinensis* female pairings

Categories of males are listed on the x-axis for each stage. Time is in minutes using a Ln scale. For interpretation, the blue line represents 30 minutes, orange line represents 60 minutes, and red line represents 120 minutes. "C" refers to *D. carolinensis* males, "LG" refers to *D. sp* Lemon Gap males, "A" refers to *D. fuscus* South Mountains (Piedmont A) males.

mean duration of TSW and total duration of mating trials were significantly shorter among South Mountain A females when paired with *D. sp* Lemon Gap males (Table S4.4; Figure 4.19). Mean pairwise Tukey HSD Post-hoc tests revealed longer TSW and total mean durations among all test groups compared with LG males (range 8.342 to 9.068 Ln(s), or approximately 70 to 146 minutes vs. 5.553 Ln(s) or 4.3 minutes; P=0.00001, range 9.39 to 9.51625 Ln(s) or approximately 200 to 226 minutes vs 7.706 Ln(s) or approximately 37 minutes; P=0.00001). Mean duration of mating stages involving female *D. fuscus* South

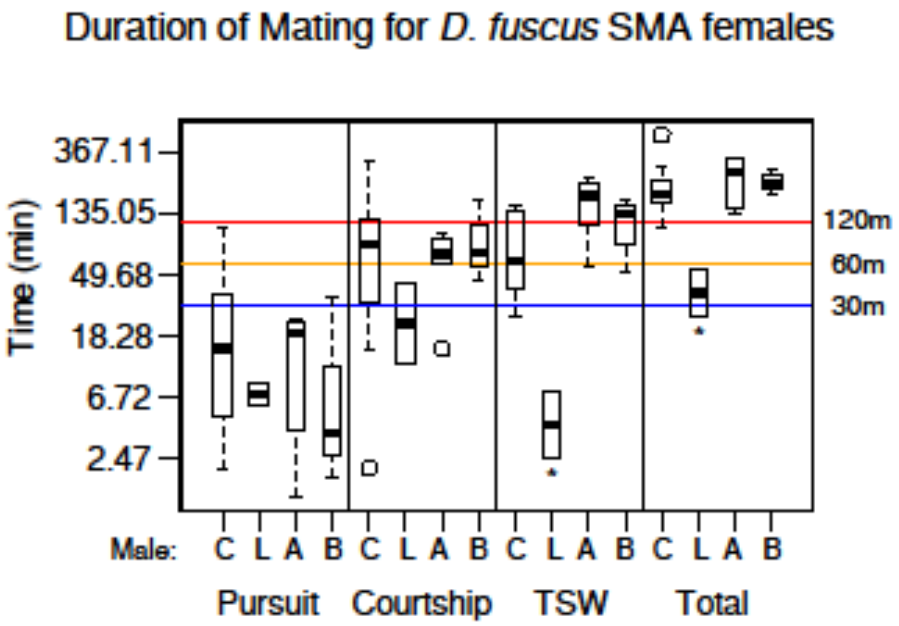


Figure 4.19 – Durations of mating stages for *D. fuscus* South Mountains "Piedmont A" female pairings

Categories of males are listed on the x-axis for each stage. Time is in minutes using a Ln scale. For interpretation, the blue line represents 30 minutes, orange line represents 60 minutes, and red line represents 120 minutes. "C" refers to *D. carolinensis* males, "LG" refers to *D. sp* Lemon Gap males, "A" refers to *D. fuscus* South Mountains (Piedmont A) males, "B" refers to *D. fuscus* South Mountains (Piedmont B). Significant differences are marked with an asterisk (*)

Mountains Piedmont B were limited to comparisons of South Mountain A males and *D. fuscus* males. Mean durations of TSW were significantly shorter in mating crosses between *D. fuscus* South Mountain B females paired with *D. carolinensis* in comparison with longer

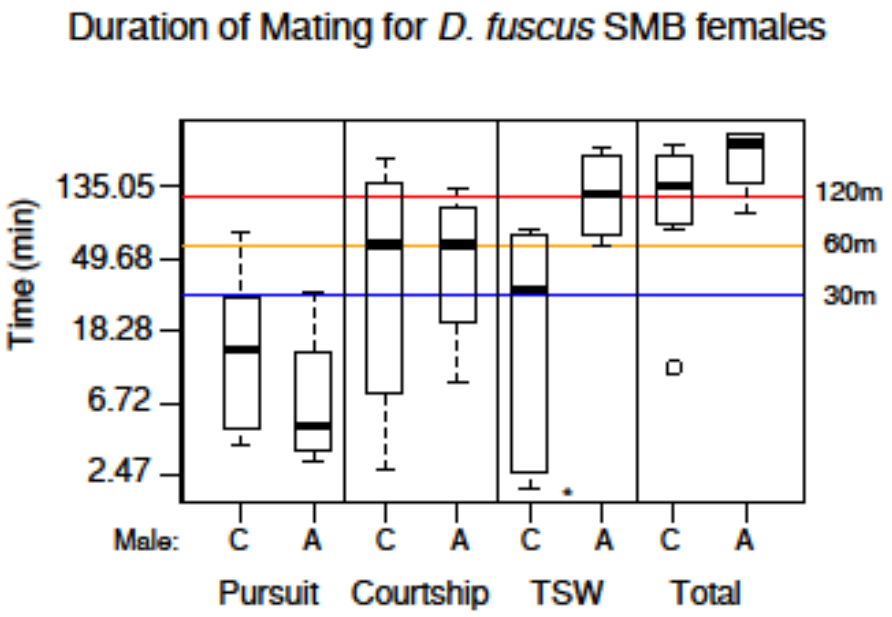


Figure 4.20 – Durations of mating stages for *D. fuscus* South Mountains "Piedmont B" female pairings

Categories of males are listed on the x-axis for each stage. Time is in minutes using a Ln scale. For interpretation, the blue line represents 30 minutes, orange line represents 60 minutes, and red line represents 120 minutes. "C" refers to *D. carolinensis* males, "A" refers to *D. fuscus* South Mountains (Piedmont A) males. Significant differences are marked with an asterisk (*)

mean durations of TSW with South Mountain A *D. fuscus* (6.89 Ln(s) vs. 8.89 Ln(s) or 16.39 minutes vs. 121.19 minutes; P=0.03989). We failed to reject differences in the mean total durations between these test groups (Table S4.5; Figure 4.20). Mean duration of mating stages involving female *D. sp* Lemon Gap did not statistically differ compared with *D. carolinensis* test groups, and only a single observation of successful mating was digitally recorded using South Mountains Piedmont A and South Mountain Piedmont B test groups.

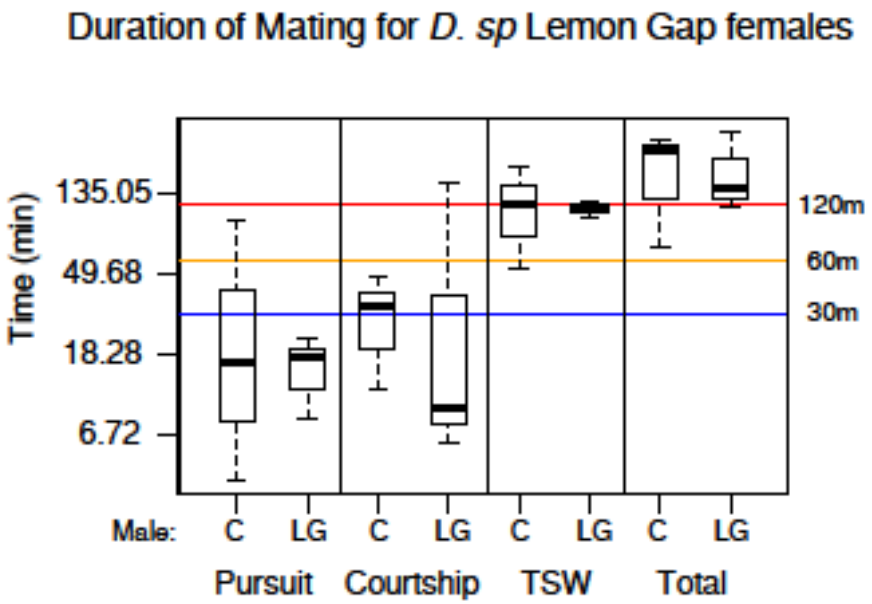


Figure 4.21 – Durations of mating stages for *D. sp* Lemon Gap female pairings

Categories of males are listed on the x-axis for each stage. Time is in minutes using a Ln scale. For interpretation, the blue line represents 30 minutes, orange line represents 60 minutes, and red line represents 120 minutes. "C" refers to *D. carolinensis* males, "LG" refers to *D. sp* Lemon Gap males.

Values for these observations fall within the same range as *D. carolinensis* and *D. sp* Lemon Gap for recorded mating stages (Table S4.6, Figure 4.21).

4.4 – Discussion

4.4.1 – Mating trials between D. fuscus populations

Initial mating trials focused on observed behaviors and the probabilities of mating between *D. fuscus* lineages in the Piedmont of North Carolina. These trials included South Mountain and Brushy Mountain populations previously identified for containing divergent

mtDNA lineages. Despite high uncorrected pairwise distances (0.138) between mtDNA lineages in the South Mountains, we did not observe abrupt pre-zygotic barriers or mating behavior that would disrupt or otherwise directly explain the sharp cline of mtDNA lineages partitioned in this region. Based on the progression of mating behaviors observed, D. fuscus containing mtDNA from the α clade initiated and participated in courtship and mating behaviors with *D. fuscus* containing mtDNA from the D. fuscus clade, and vice versa. Comparing geographically separated populations in the South Mountains and Brushy Mountains, similar behavioral patterns were observed in which animals readily initiated courtship behaviors and mated in captivity, exhibiting similar courtship behaviors regardless of whether individuals contained mtDNA from the α or D. fuscus clade, or if individuals originated from the South Mountains or Brushy Mountains. Consequently, many measures of reproductive isolation suggest reproductive barriers that would potentially limit gene flow are incomplete and fail to restrict gene flow within in the South Mountains (but see below). Likewise, although populations are separated by over 64km, it does not appear that reproductive barriers have evolved allopatrically between individuals of different mtDNA clades in the isolated mountain systems in the Piedmont, but there is a signature of reproductive isolation between individuals with closely related haplotypes within the mtDNA α clade.

Upon closer examination of *D. fuscus* mating trials in the South Mountains we observed weak differences in the spermatophore deposition rates among individuals containing Piedmont A vs. Piedmont B mtDNA. This trend was much stronger when we examined rates of insemination in heterotypic crosses. Specifically, the isolation asymmetry value significantly differs from a null expectation of unrestricted mating

(P=0.0381), which is driven by a lower rate of successful heterotypic inseminations between female *D. fuscus* containing mtDNA from the Piedmont A clade and male *D. fuscus* from the Piedmont B clade. Conversely, there was a much higher rate of successful inseminations between female *D. fuscus* containing mtDNA from the Piedmont B clade and male *D. fuscus* from the Piedmont A clade. If we only consider the outcome of pre-zygotic interactions among existing populations, assuming equal dispersal rates among male and female *D. fuscus*, unbiased sex ratios, the absence of selection, and ignoring post-zygotic isolation; these behaviors would not diminish autosomal nuclear gene flow among populations but could be responsible for sex-biased nuclear patterns or the maternally inherited mtDNA patterns observed in South Mountain *D. fuscus*.

In multiple trials, we observed conspecific rates of successful mating at much lower rates than in heterospecific crosses, as occurs in crosses between mtDNA α clade lineages in the South Mountain and Brushy Mountains. These results imply *D. fuscus* have evolved reproductive barriers to restrict gene flow within their population, rather than between populations, contrary to normal expectations. It is possible that selection and reinforcement have not had an opportunity to act upon the novel encounters used in this experiment, where individuals from these populations may have limited historical contact. Alternatively, inbreeding depression is promoting outcrossing, which may result when population Ne is small or populations have gone through a recent bottleneck. We do not observe any differences in the latency to mate or courtship durations when comparing conspecific or heterotypic crosses in *D. fuscus*. Differences in aspects of mating duration may explain novelty or nuances of mate choice, however these measurements were rather coarse (15-20 minute intervals) and imprecise for recording behaviors at the onset of mating.

4.4.2 – Extended Mating trials

To address additional questions regarding the effect of introgression on the evolution of reproductive barriers, we expanded the number of crosses to include both *D. fuscus*, *D. carolinensis*, and *D. sp* Lemon Gap populations. Observed behaviors are similar to preliminary trials in that all populations readily mate in captivity, but differ in the rates of successful spermatophore depositions and inseminations between crosses. We measured the probability of successful mating trials based on the deposition of at least a single spermatophore. While insemination is a more definitive indication of successful mating, it is contingent upon spermatophore deposition, therefore we included these rates as an informative metric.

In nearly all crosses between population-lineages, joint isolation and propensity isolation metrics significantly differ from a null expectation of unrestricted mating. These results are driven by higher conspecific rates of successful mating relative to heterospecific rates for measures of JI, or much higher conspecific rates of successful mating for one population-lineage relative to the other for measures of PI. These results are split evenly between deposition or insemination rates for JI, but are primarily differences in deposition rates for PI. Only in crosses between members of the mtDNA α clade (*D. carolinensis* and *D. fuscus*) do both deposition and insemination values for JI significantly deviate from unrestricted gene flow. In these crosses, female *D. carolinensis* are less likely to mate with *D. fuscus* males, whereas female *D. fuscus* more readily mate with *D. carolinensis* males. These results hint at introgression having some role in promoting biased rates of gene flow, or at least introgression acting as a signature of historical contact between lineages which may have facilitated the evolution of reproductive barriers between populations through

reinforcement. While most of these populations do not currently have overlapping ranges, weak or asymmetric barriers are in agreement with biased estimates of migration and gene flow between *D. fuscus* and *D. carolinensis,* where a greater number of migrants have been exchanged from *D. carolinensis* to Piedmont lineages (Kratovil *et al.* in prep). These mating results do not explain phylogeographic patterns of mitochondrial introgression, unless we invoke less likely mechanisms of leaky paternal mitochondrial transmittance. Mitochondrial introgression does not explain the evolution of reproductive barriers between *D. carolinensis* or *D. fuscus* and *D. sp* Lemon Gap, which appear to be primarily driven by intrinsic pre-zygotic barriers in allopatry.

Differences in the durations of mating phases may be explained by the novelty of mating crosses that have not evolved intrinsic barriers, nuances in mate choice, or other complex behaviors correlated with barriers of reproductive isolation. While mating in salamanders is generally a prolonged process involving chemical signaling, pheromone delivery to attract and promote female receptivity (Houck & Reagan 1990), and elaborate courtship behaviors, the time and effort of finding a suitable mate must be tempered by the risk of interference by competitors. Introgression could act to facilitate gene flow and diminish intrinsic reproductive barriers between populations or hasten the process of isolation by reinforcing pre-zygotic isolating mechanisms at any stage of the mating process. We therefore used digital recordings of randomized mating trials to explore differences in mating stage durations, and found complex differences in the duration of the tail straddle walk prior to spermatophore deposition, or total duration of successful mating trials between populations.

Closer examination of *D. carolinensis* and *D. sp* Lemon Gap trials revealed no differences in the duration of any mating stages between female *D. carolinensis* or *D. sp* Lemon Gap and any class of paired male. In contrast, substantially shorter durations of tail straddle walk and total mating duration were observed between *D. fuscus* females and *D. sp* Lemon Gap males. In addition, tail straddle walk durations were significantly shorter between *D. fuscus* females and *D. carolinensis* males, verses *D. fuscus* males containing different mtDNA lineages (Piedmont B x α clade). We find that sympatric population-lineages, or populations with historical contact (inferred from mtDNA introgression), have substantially longer mating durations, specifically during the tail-straddle walk phase prior to spermatophore deposition. These results support the idea that historical contact or ongoing interactions among population lineages promote longer mating durations and the evolution of complex mating behaviors or chemical signals. A major caveat of this research is that only a subset of mating trials were digitally recorded, and due to practical limitations, we only analyzed the durations of successful trials. In the future it will be informative to determine if there are differences between the durations of successful and unsuccessful mating stages across populations, and at which stage mating no longer progresses.

In summary, although we find pre-zygotic isolation appears to have evolved between many population-lineages using standard metrics of isolation, and that historical introgression (or ongoing gene exchange) can increase mating durations to the point that we cannot distinguish between conspecific and heterospecific crosses, we are unable to determine why *Desmognathus* populations readily mate with *D. carolinensis* and have presumably replaced their native mtDNA genomes. Incomplete reproductive boundaries may have enabled mtDNA from *D. carolinensis* to be transmitted to other populations.

At best, pre-zygotic mating behaviors are in agreement with patterns of asymmetric biases in gene flow and this may contribute toward the observed phylogeographic patterns in *Desmognathus*. As a diagnostic measure, asserting the independence of lineages based on the evolution of reproductive barriers is challenging. We find that the evolution of pre-zygotic barriers is less often observed between allopatric populations, even when these lineages are genetically highly divergent from one another. In contrast, sympatric lineages, and lineages with complex histories of admixture may promote the evolution of complex behaviors and the dynamic phylogenetic histories within *Desmognathus.*

4.5 – Author Contributions

Justin Kratovil conceived and designed this project. Mary Virginia Gibbs and Mary Foley assisted with data collection and video analysis. Justin Kratovil analyzed, interpreted the data and wrote this chapter.

4.6 – Acknowledgements

I would like to thank Lynne Houck for advice on conducting mating trials and for the use of her equipment at Highlands Biological Station. I would also like to thank Steve Arnold for suggestions on statistical analysis.

4.7 – Additional Funding Sources

This work was supported by the National Science Foundation (grant numbers IOS 0818554 to Lynne D. Houck. and Stevan J. Arnold), the Richard Bruce Family Scholarship in Herpetology through the Highlands Biological Station Grant in Aid of

Research, and a Ribble Grant from the Department of Biology at the University of Kentucky.

4.8 – SUPPLEMENTAL INFORMATION

S4.1 – Scheduled matrix for *D. fuscus* paired mating trials

Group A refers to South Mountain Piedmont "A"; Group B refers to South Mountain Piedmont "B", and Group C refers to Brushy Mountain *D. fuscus.*

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9
Females	Males								
A1	AM1	BM1	CM1	AM6	BM6	CM6	AM11	BM11	CM11
A2	AM2	BM2	CM2	AM7	BM7	CM7	AM12	BM12	CM12
A3	AM3	BM3	CM3	AM8	BM8	CM8	AM13	BM13	CM13
A4	AM4	BM4	CM4	AM9	BM9	CM9	AM14	BM14	CM14
A5	AM5	BM5	CM5	AM10	BM10	CM10	AM15	BM15	CM15
A6	BM1	CM1	AM1	BM6	CM6	AM6	BM11	CM11	AM11
A7	BM2	CM2	AM2	BM7	CM7	AM7	BM12	CM12	AM12
A8	BM3	CM3	AM3	BM8	CM8	AM8	BM13	CM13	AM13
A9	BM4	CM4	AM4	BM9	CM9	AM9	BM14	CM14	AM14
A10	BM5	CM5	AM5	BM10	CM10	AM10	BM15	CM15	AM15
A11	CM1	AM1	BM1	CM6	AM6	BM6	CM11	AM11	BM11
A12	CM2	AM2	BM2	CM7	AM7	BM7	CM12	AM12	BM12
A13	CM3	AM3	BM3	CM8	AM8	BM8	CM13	AM13	BM13
A14	CM4	AM4	BM4	CM9	AM9	BM9	CM14	AM14	BM14
A15	CM5	AM5	BM5	CM10	AM10	BM10	CM15	AM15	BM15
B1	AM6	BM6	CM6	AM11	BM11	CM11	AM1	BM1	CM1
B2	AM7	BM7	CM7	AM12	BM12	CM12	AM2	BM2	CM2
B3	AM8	BM8	CM8	AM13	BM13	CM13	AM3	BM3	CM3
B4	AM9	BM9	CM9	AM14	BM14	CM14	AM4	BM4	CM4
B5	AM10	BM10	CM10	AM15	BM15	CM15	AM5	BM5	CM5
B6	BM6	CM6	AM6	BM11	CM11	AM11	BM1	CM1	AM1
B7	BM7	CM7	AM7	BM12	CM12	AM12	BM2	CM2	AM2
B8	BM8	CM8	AM8	BM13	CM13	AM13	BM3	CM3	AM3
B9	BM9	CM9	AM9	BM14	CM14	AM14	BM4	CM4	AM4
B10	BM10	CM10	AM10	BM15	CM15	AM15	BM5	CM5	AM5
B11	CM6	AM6	BM6	CM11	AM11	BM11	CM1	AM1	BM1

B12	CM7	AM7	BM7	CM12	AM12	BM12	CM2	AM2	BM2
B13	CM8	AM8	BM8	CM13	AM13	BM13	CM3	AM3	BM3
B14	CM9	AM9	BM9	CM14	AM14	BM14	CM4	AM4	BM4
B15	CM10	AM10	BM10	CM15	AM15	BM15	CM5	AM5	BM5
C1	AM11	BM11	CM11	AM1	BM1	CM1	AM6	BM6	CM6
C2	AM12	BM12	CM12	AM2	BM2	CM2	AM7	BM7	CM7
C3	AM13	BM13	CM13	AM3	BM3	CM3	AM8	BM8	CM8
C4	AM14	BM14	CM14	AM4	BM4	CM4	AM9	BM9	CM9
C5	AM15	BM15	CM15	AM5	BM5	CM5	AM10	BM10	CM10
C6	BM11	CM11	AM11	BM1	CM1	AM1	BM6	CM6	AM6
C7	BM12	CM12	AM12	BM2	CM2	AM2	BM7	CM7	AM7
C8	BM13	CM13	AM13	BM3	CM3	AM3	BM8	CM8	AM8
C9	BM14	CM14	AM14	BM4	CM4	AM4	BM9	CM9	AM9
C10	BM15	CM15	AM15	BM5	CM5	AM5	BM10	CM10	AM10
C11	CM11	AM11	BM11	CM1	AM1	BM1	CM6	AM6	BM6
C12	CM12	AM12	BM12	CM2	AM2	BM2	CM7	AM7	BM7
C13	CM13	AM13	BM13	CM3	AM3	BM3	CM8	AM8	BM8
C14	CM14	AM14	BM14	CM4	AM4	BM4	CM9	AM9	BM9
C15	CM15	AM15	BM15	CM5	AM5	BM5	CM10	AM10	BM10

S4.2 – Scheduled matrix for *D. fuscus* and *D. carolinensis* paired mating trials

Scheduled mating trials across different test groups using females (F) and males (M), in which every female encounters every male once (top). Randomized supplemental pairings in which females encounter males only once (bottom). Test crosses included a number of planned "outgroup" crosses which were not included in the final analysis. (LG) refers to Lemon Gap, (BMC1) and (BMC2) refer to different *D. carolinensis* populations near Black Mountain Campground in NC. (A) refers to South Mountain Piedmont "A" *D. fuscus*, and (B) refers to South Mountain Piedmont "B" *D. fuscus*.

Date	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
F	**M**																								
LG F1	LG M1	BM 5	AM 4	BM C1 M3	BM C1 M2	LG M2	BM 1	AM 5	BM C1 M4	BM C1 M3	LG M3	BM 2	AM 1	BM C1 M5	BM C1 M4	LG M4	BM 3	AM 2	BM C1 M1	BM C1 M5	LG M5	BM 4	AM 3	BM C1 M2	BM C1 M1
LG F2	BM C1 M1	LG M5	BM 4	AM 3	BM C2 M2	BM C1 M2	LG M1	BM 5	AM 4	BM C2 M3	BM C1 M3	LG M2	BM 1	AM 5	BM C2 M4	BM C1 M4	LG M3	BM 2	AM 1	BM C2 M5	BM C1 M5	LG M4	BM 3	AM 2	BM C2 M1
LG F3	BM C2 M1	BM C1 M5	LG M4	BM 3	AM 2	BM C2 M2	BM C1 M1	LG M5	BM 4	AM 3	BM C2 M3	BM C1 M2	LG M1	BM 5	AM 4	BM C2 M4	BM C1 M3	LG M2	BM 1	AM 5	BM C2 M5	BM C1 M4	LG M3	BM 2	AM 1
LG F4	AM 1	BM C2 M5	BM C1 M4	LG M3	BM 3	AM 2	BM C2 M1	BM C1 M5	LG M4	BM 3	AM 3	BM C2 M2	BM C1 M1	LG M5	BM 4	AM 4	BM C2 M3	BM C1 M2	LG M1	BM 5	AM 5	BM C2 M4	BM C1 M3	LG M2	BM 1
LG F5	BM 1	AM 5	BM C2 M4	BM C2 M3	LG M2	BM 2	AM 1	BM C2 M5	BM C2 M4	LG M3	BM 3	AM 2	BM C2 M1	BM C2 M5	LG M4	BM 4	AM 3	BM C2 M2	BM C2 M1	LG M5	BM 5	AM 4	BM C2 M3	BM C2 M2	LG M1
BM C1 F1	LG M2	BM 1	AM 5	BM C1 M4	BM C1 M3	LG M3	BM 2	AM 1	BM C1 M5	BM C1 M4	LG M4	BM 3	AM 2	BM C1 M1	BM C1 M5	LG M5	BM 4	AM 3	BM C1 M2	BM C1 M1	LG M1	BM 5	AM 4	BM C1 M3	BM C1 M2

BMC1F2	BMC1M2	LGM1	BM5	AM4	BMC2M3	BMC1M3	LGM2	BM1	AM5	BMC2M4	BMC1M4	LGM3	BM2	AM1	BMC2M5	BMC1M5	LGM4	BM3	AM2	BMC2M1	BMC1M1	LGM5	BM4	AM3	BMC2M2
BMC1F3	BMC2M2	BMC1M1	LGM5	BM4	AM3	BMC2M3	BMC1M2	LGM1	BM5	AM4	BMC2M4	BMC1M3	LGM2	BM1	AM5	BMC2M5	BMC1M4	LGM3	BM2	AM1	BMC2M1	BMC1M5	LGM4	BM3	AM2
BMC1F4	AM2	BMC2M1	BMC1M5	LGM4	BM3	AM3	BMC2M2	BMC1M1	LGM5	BM4	AM4	BMC2M3	BMC1M2	LGM1	BM5	AM5	BMC2M4	BMC1M3	LGM2	BM1	AM1	BMC2M5	BMC1M4	LGM3	BM3
BMC1F5	BM2	AM1	BMC2M5	BMC2M4	LGM3	BM3	AM2	BMC2M1	BMC2M5	LGM4	BM4	AM3	BMC2M2	BMC2M1	LGM5	BM5	AM4	BMC2M3	BMC2M2	LGM1	BM1	AM5	BMC2M4	BMC2M3	LGM2
BMC2F1	LGM3	BM2	AM1	BMC1M5	BMC1M4	LGM4	BM3	AM2	BMC1M1	BMC1M5	LGM5	BM4	AM3	BMC1M2	BMC1M1	LGM1	BM5	AM4	BMC1M3	BMC1M2	LGM2	BM1	AM5	BMC1M4	BMC1M3
BMC2F2	BMC1M3	LGM2	BM1	AM5	BMC2M4	BMC1M4	LGM3	BM2	AM1	BMC2M5	BMC1M5	LGM4	BM3	AM2	BMC2M1	BMC1M1	LGM5	BM4	AM3	BMC2M2	BMC1M2	LGM1	BM5	AM4	BMC2M3
BMC2F3	BMC2M3	BMC1M2	LGM1	BM5	AM4	BMC2M4	BMC1M3	LGM2	BM1	AM5	BMC2M5	BMC1M4	LGM3	BM2	AM1	BMC2M1	BMC1M5	LGM4	BM3	AM2	BMC2M2	BMC1M1	LGM5	BM4	AM3
BMC2F4	AM3	BMC2M2	BMC1M1	LGM5	BM4	AM4	BMC2M3	BMC1M2	LGM1	BM5	AM5	BMC2M4	BMC1M3	LGM2	BM1	AM1	BMC2M5	BMC1M4	LGM3	BM3	AM2	BMC2M1	BMC1M5	LGM4	BM3
BMC2F5	BM3	AM2	BMC2M1	BMC2M5	LGM4	BM4	AM3	BMC2M2	BMC2M1	LGM5	BM5	AM4	BMC2M3	BMC2M2	LGM1	BM1	AM5	BMC2M4	BMC2M3	LGM2	BM2	AM1	BMC2M5	BMC2M4	LGM3
AF1	LGM4	BM3	AM2	BMC1M1	BMC1M5	LGM5	BM4	AM3	BMC1M2	BMC1M1	LGM1	BM5	AM4	BMC1M3	BMC1M2	LGM2	BM1	AM5	BMC1M4	BMC1M3	LGM3	BM2	AM1	BMC1M5	BMC1M4
AF2	BMC1M4	LGM3	BM2	AM1	BMC2M5	BMC1M5	LGM4	BM3	AM2	BMC2M1	BMC1M1	LGM5	BM4	AM3	BMC2M2	BMC1M2	LGM1	BM5	AM4	BMC2M3	BMC1M3	LGM2	BM1	AM5	BMC2M4
AF3	BMC2M4	BMC1M3	LGM2	BM1	AM5	BMC2M5	BMC1M4	LGM3	BM2	AM1	BMC2M1	BMC1M5	LGM4	BM3	AM2	BMC2M2	BMC1M1	LGM5	BM4	AM3	BMC2M3	BMC1M2	LGM1	BM5	AM4
AF4	AM4	BMC2M3	BMC1M2	LGM1	BM5	AM5	BMC2M4	BMC1M3	LGM2	BM1	AM1	BMC2M5	BMC1M4	LGM3	BM3	AM2	BMC2M1	BMC1M5	LGM4	BM3	AM3	BMC2M2	BMC1M1	LGM5	BM4

AF5	BM4	AM3	BMC2M2	BMC2M1	LGM5	BM5	AM4	BMC2M3	BMC2M2	LGM1	BM1	AM5	BMC2M4	BMC2M3	LGM2	BM2	AM1	BMC2M5	BMC2M4	LGM3	BM3	AM2	BMC2M1	BMC2M5	LGM4
BF1	LGM5	BM4	AM3	BMC1M2	BMC1M1	LGM1	BM5	AM4	BMC1M3	BMC1M2	LGM2	BM1	AM5	BMC1M4	BMC1M3	LGM3	BM2	AM1	BMC1M5	BMC1M4	LGM4	BM3	AM2	BMC1M1	BMC1M5
BF2	BMC1M5	LGM4	BM3	AM2	BMC2M1	BMC1M1	LGM5	BM4	AM3	BMC2M2	BMC1M2	LGM1	BM5	AM4	BMC2M3	BMC1M3	LGM2	BM1	AM5	BMC2M4	BMC1M4	LGM3	BM2	AM1	BMC2M5
BF3	BMC2M5	BMC1M4	LGM3	BM2	AM1	BMC2M1	BMC1M5	LGM4	BM3	AM2	BMC2M2	BMC1M1	LGM5	BM4	AM3	BMC2M3	BMC1M2	LGM1	BM5	AM4	BMC2M4	BMC1M3	LGM2	BM1	AM5
BF4	AM5	BMC2M4	BMC1M3	LGM2	BM1	AM1	BMC2M5	BMC1M4	LGM3	BM3	AM2	BMC2M1	BMC1M5	LGM4	BM3	AM3	BMC2M2	BMC1M1	LGM5	BM4	AM4	BMC2M3	BMC1M2	LGM1	BM5
BF5	BM5	AM4	BMC2M3	BMC2M2	LGM1	BM1	AM5	BMC2M4	BMC2M3	LGM2	BM2	AM1	BMC2M5	BMC2M4	LGM3	BM3	AM2	BMC2M1	BMC2M5	LGM4	BM4	AM3	BMC2M2	BMC2M1	LGM5

Supplemental paired mating crosses

F	**M**																			
BMC2F6	AM10	out grp	BMC1M6	BMC2M6	BMC2M7	AM6	AM7	AM8	AM11	out grp	AM9	BM10	out grp	AM12	BM11	BM9	NA	BM12	out grp	LGM6	LGM7
AF6	BM11	BMC1M6	BMC2M6	BMC2M7	AM6	AM7	AM8	BM9	out grp	AM9	BM10	out grp	AM12	NA	out grp	AM11	BM12	out grp	LGM6	LGM7	out grp
BF6	NA	BMC2M7	AM6	AM7	AM8	BM9	BMC2M6	AM9	BM10	out grp	AM12	BM11	out grp	AM11	BM12	out grp	LGM6	LGM7	out grp	out grp	BMC1M6
AF7	NA	BMC2M6	BMC2M7	AM6	AM7	AM8	BM9	BMC1M6	AM9	BM10	out grp	AM10	BM11	out grp	AM11	BM12	out grp	LGM6	LGM7	out grp	out grp
BF7	BMC2M7	AM6	AM7	AM8	BM9	out grp	AM9	BM10	out grp	AM10	BM11	out grp	AM11	BM12	NA	LGM6	LGM7	out grp	out grp	BMC1M6	BMC2M6
AF8	out grp	AM8	BM9	out grp	AM9	BM10	out grp	AM10	BM11	out grp	AM11	BM12	out grp	LGM6	LGM7	out grp	out grp	BMC1M6	BMC2M6	BMC2M7	AM6

BF 8	AM 6	AM 7	AM 8	BM 9	AM 12	AM 9	BM 10	out grp	NA	BM 11	out grp	AM 11	BM 12	out grp	LG M6	LG M7	out grp	out grp	BM C1 M6	BM C2 M6	BM C2 M7
AF 9	AM 8	BM 9	out grp	AM 9	BM 10	out grp	AM 10	BM 11	out grp	BM 3	BM 12	out grp	LG M6	LG M7	out grp	out grp	BM C1 M6	BM C2 M6	BM C2 M7	AM 6	AM 7
BF 9	out grp	BM 10	out grp	AM 10	BM 11	out grp	AM 11	BM 12	out grp	LG M6	LG M7	AM 9	out grp	BM C1 M6	BM C2 M6	BM C2 M7	AM 6	AM 7	AM 8	BM 9	out grp
AF 10	BM 9	out grp	AM 9	BM 10	out grp	AM 10	BM 11	out grp	NA	BM 12	out grp	LG M6	LG M7	out grp	out grp	BM C1 M6	BM C2 M6	BM C2 M7	AM 6	AM 7	AM 8
BF 10	BM 10	out grp	AM 10	BM 11	out grp	AM 11	BM 12	out grp	LG M6	LG M7	out grp	out grp	BM C1 M6	BM C2 M6	BM C2 M7	AM 6	AM 7	AM 8	BM 9	out grp	AM 9
AF 11	out grp	AM 9	BM 10	out grp	AM 10	BM 11	out grp	AM 11	BM 12	out grp	LG M6	LG M7	out grp	out grp	BM C1 M6	BM C2 M6	BM C2 M7	AM 6	AM 7	AM 8	BM 9
BF 11	out grp	AM 10	BM 11	out grp	AM 11	BM 12	out grp	LG M6	LG M7	NA	out grp	BM C1 M6	BM C2 M6	BM C2 M7	AM 6	AM 7	AM 8	BM 9	out grp	AM 9	BM 10
BF 12	AM 9	BM 11	out grp	AM 11	BM 12	BM C2 M7	LG M6	LG M7	out grp	out grp	BM C1 M6	BM C2 M6	AM 6	AM 7	NA	AM 8	BM 9	out grp	NA	BM 10	out grp

S4.3 – Comparisons of successful mating durations between female *D. carolinensis* and available test groups (Car = *D. carolinensis*; A = Piedmont A, LG = *D. sp Lemon Gap*) (top)

			mean	variance	Between group variation				
Group	**Cross**	**Observations (n)**	**Latency**	**Latency**	**SS**	**df**	**MS**	**F**	**P-value**
1	Car x Car	11	6.34466	1.72867	4.71260	2	2.35630	1.12932	0.34634
2	Car x A	3	4.94076	7.41907					
3	Car x LG	6	5.91769	0.66904					
			Courtship	**Courtship**					
1	Car x Car	11	6.69645	2.42323	3.31218	2	1.65609	0.89791	0.42589
2	Car x A	3	7.84856	0.44705					
3	Car x LG	6	6.73437	1.24563					
			TSW	**TSW**					
1	Car x Car	11	8.29138	0.46417	2.78656	2	1.39328	1.26674	0.30704
2	Car x A	3	8.75919	0.54999					
3	Car x LG	6	7.65634	2.59129					
			Total	**Total**					
1	Car x Car	11	8.87213	0.41135	1.87019	2	0.93510	1.49963	0.25130
2	Car x A	3	9.28073	0.17105					
3	Car x LG	6	8.36674	1.22896					

S4.4 – Comparisons of successful mating durations between female *D. fuscus* from the South Mountains (Piedmont A) and available test groups (Car = *D. carolinensis*; B = Piedmont B, LG = *D. sp Lemon Gap*) (top), Tukey HSD Post-hoc test (below)

			mean	variance	Between group variation				
Group	**Cross**	**Observations (n)**	**Latency**	**Latency**	**SS**	**df**	**MS**	**F**	**P-value**
1	A x Car	7	6.73449	2.08902	1.80167	3	0.60056	0.32334	0.80847
2	A x A	5	6.28517	1.69219					
3	A x B	3	5.90489	2.38894					
4	A x LG	2	6.06374	0.06510					
			Courtship	**Courtship**					
1	A x Car	7	8.01830	2.83630	2.08955	3	0.69652	0.42810	0.73626
2	A x A	5	8.11764	0.58266					
3	A x B	3	8.51049	0.45356					
4	A x LG	2	7.20572	0.89519					
			TSW	**TSW**					
1	A x Car	7	8.34216	0.57249	18.69788	3	6.23263	12.95724	**0.00033**
2	A x A	5	9.06848	0.35159					
3	A x B	3	8.76304	0.39679					
4	A x LG	2	5.55323	0.61831					
			Total	**Total**					
1	A x Car	7	9.39495	0.24224	5.48122	3	1.82707	8.97383	**0.00175**
2	A x A	5	9.51625	0.19863					
3	A x B	3	9.51606	0.04457					

4	A x LG	2	7.70605	0.30969

S4.4 (continued) – Tukey HSD Post-hoc test of mating durations between female *D. fuscus* from the South Mountains (Piedmont A) and available test groups (Car = *D. carolinensis*; B = Piedmont B, LG = *D. sp Lemon Gap*) (top). Group descriptions are listed above.

TSW	Diff	95% CI	P-value
Group 1 vs Group 2	0.7263	-0.1208 to 1.5734	0.10370
Group 1 vs Group 3	0.4209	-0.5774 to 1.4192	0.61550
Group 1 vs Group 4	-2.789	-3.9488 to -1.6290	**0.00000**
Group 2 vs Group 3	-0.305	-1.3619 to 0.7510	0.83050
Group 2 vs Group 4	-3.515	-4.7256 to -2.3049	**0.00000**
Group 3 vs Group 4	-3.21	-4.5304 to -1.8892	**0.00000**

Total	Diff	95% CI	P-value
Group 1 vs Group 2	0.12130	-0.2509 to 0.4935	0.77560
Group 1 vs Group 3	0.12110	-0.3176 to 0.5598	0.84850
Group 1 vs Group 4	-1.68890	-2.1986 to -1.1792	**0.00000**
Group 2 vs Group 3	-0.00020	-0.4644 to 0.4640	**0.00000**
Group 2 vs Group 4	-1.81020	-2.3420 to -1.2783	**0.00000**
Group 3 vs Group 4	-1.81000	-2.3903 to -1.2297	**0.00000**

Total	Diff	95% CI	P-value
Group 1 vs Group 2	0.12130	-0.2509 to 0.4935	0.77560
Group 1 vs Group 3	0.12110	-0.3176 to 0.5598	0.84850
Group 1 vs Group 4	-1.6889	-2.1986 to -1.1792	**0.00000**
Group 2 vs Group 3	-0.0002	-0.4644 to 0.4640	**0.00000**
Group 2 vs Group 4	-1.8102	-2.3420 to -1.2783	**0.00000**
Group 3 vs Group 4	-1.8100	-2.3903 to -1.2297	**0.00000**

S4.5 – Comparisons of successful mating durations between female *D. fuscus* from the South Mountains (Piedmont B) and available test groups (Car = *D. carolinensis*; A = Piedmont A)

			mean	variance	Between group variation				
Group	**Cross**	**Observations (n)**	**Latency**	**Latency**	**PCC**	**df**	**mean Diff**	**var Diff**	**P-value**
1	B x Car	8	6.67897	1.18778	0.66182	3	0.10133	0.62099	0.81367
2	B x A	4	6.02579	1.10340					
			Courtship	**Courtship**					
1	B x Car	8	7.66316	2.83864	0.01865	3	0.41901	2.37059	0.62409
2	B x A	4	7.91465	1.35150					
			TSW	**TSW**					
1	B x Car	8	6.89105	2.71155	0.65420	3	-0.94386	0.29329	**0.03989**
2	B x A	4	8.89166	0.43601					
			Total	**Total**					
1	B x Car	8	8.74689	1.00070	-0.65732	3	-0.30465	0.71074	0.52209
2	B x A	4	9.39739	0.27049					

S4.6 – Comparisons of successful mating durations between female *D. sp* Lemon Gap and available test groups (Car = *D. carolinensis*; LG = Lemon Gap)

			mean	variance	Between group variation				
Group	**Cross**	**Observations (n)**	**Latency**	**Latency**	**PCC**	**df**	**mean Diff**	**var Diff**	**P-value**
1	LG x Car	3	7.01777	2.65470	0.93210	3	0.22603	1.33555	0.76705
2	LG x LG	3	6.79175	0.27571					
			Courtship	**Courtship**					
1	LG x Car	3	7.38691	0.53587	0.37127	3	0.25539	2.64612	0.81118
2	LG x LG	3	7.13153	3.06130					
			TSW	**TSW**					
1	LG x Car	3	8.77275	0.39365	0.99740	3	-0.04621	0.26406	0.89052
2	LG x LG	3	8.81896	0.01298					
			Total	**Total**					
1	LG x Car	3	9.20304	0.55179	0.63415	3	-0.03096	0.33061	0.93421
2	LG x LG	3	9.23400	0.24801					

Chapter 5 : Synthesis and conclusions

Delimiting species boundaries provides a foundation for understanding patterns and processes of divergence between lineages. A major challenge in resolving species limits is estimating an underlying evolutionary history that accommodates individual discordances distributed throughout the nuclear genome and does not necessarily reflect the same history as the mitochondrial genome. At the phylogeographic and population levels, discordant patterns may reveal important differences in the retained signature of historical contact and gene flow between populations or lineages, and therefore genetic information from both sources can be relevant in understanding evolutionary patterns. Multiple markers should always be used for more robust phylogenetic and population genetic analyses to account for genealogical differences across markers.

With the advent of genomic sequencing, mtDNA introgression has been an increasingly common observation among diverse organisms, thwarting efforts to resolve phylogenetic questions using a single readily sequenced marker. While introgression may contribute toward topological disagreements between nuclear gene trees, it is less understood how introgression shapes general patterns of divergence among lineages. Does a history of introgression promote segregation, increasing the rate of divergence? Or does introgression facilitate gene flow and break intrinsic barriers of assortative mating? Are these processes mutually exclusive?

In this dissertation, one major objective is to resolve uncertainty in the species boundaries between lineages with a known history of mtDNA introgression. Even a cursory survey of the literature reveals that systematics in *Desmognathus* is fraught with challenges associated with delimiting taxa with conserved morphologies that belie their complex

underlying genetic histories. From a systematic perspective, how do we account for these complex differences both in the nomenclature and in our understanding of species limits? In taking a few steps toward resolving these questions, we leveraged newer genomic sequencing methods to scrutinize loci in the nuclear genome, compared phylogenetic and population level patterns between targeted nuclear loci and nuclear markers with mitochondrial function, and lastly, examined reproductive behaviors that traditionally diagnosed boundaries of biological species.

The combination of model testing, genetic clustering, species tree estimation, and migration estimates in chapter 2 provides compelling support that, characteristically, the evolution of *Desmognathus fuscus* is far more complicated than previously realized. Patterns from nuclear data are highly discordant with mtDNA patterns and recognized taxonomy. MtDNA patterns are also misleading because there is minimal segregation between *D. fuscus* populations containing highly divergent mtDNA lineages in the South Mountains, and a strong signature of segregation between *D. fuscus* populations containing similar mtDNA haplotypes from the α clade. When we examine migration estimates between *D. fuscus* and *D. carolinensis*, curious asymmetries emerge between different mtDNA clades, potentially explaining why mtDNA haplotypes have been retained over time despite nuclear gene flow.

In chapter 3, we confirmed support for the same parameter rich models for lineage boundaries using a different set of nuclear data. Using additional representative taxa across *Desmognathus* we reconstructed relationships among these lineages to find patterns consistent with historical divergence with gene flow and non-bifurcating topologies. When we compare species trees generated from putatively unlinked nuclear loci and those from

NEMGs, we see strongly discordant patterns, neither of which conform to mitochondrial lineages at the population level. A similar pattern is observed in analyses of population structure. We are currently exploring the use of different library preparation methods which are expected to minimize missing data, which may have contributed to lower confidence in our NEMG results.

To assess the evolution of reproductive barriers between *D. fuscus* and *D. carolinensis* lineages, we measured different metrics of reproductive isolation including durations of mating trials using controlled crosses. We find evidence of asymmetric biases, which help in generating hypotheses for some mtDNA patterns, such as the retention of *D. fuscus* clade mtDNA lineages in the South Mountains, however these metrics cannot directly explain differences in durations of mating or prevalence of successful mating between distantly related nuclear lineages.

In conclusion, we find a strong signature of mtDNA introgression and patterns of nuclear divergence indicative of pervasive non-bifurcating topologies, seemingly with limited barriers between lineages. This research contributes to a growing body of literature supporting reticulate evolutionary histories, pervasive mitochondrial introgression, and discordance between nuclear genes, which characterizes divergence and speciation as a complex process that takes place in the present of gene flow.

References

Adams DC, Berns MC, Kozak KH, Wiens JJ (2009) Are rates of species diversification correlated with rates of morphological evolution? *Proc Biol Sci*, **276**, 2729–2738.

Ané C, Larget B, Baum D a, Smith SD, Rokas A (2007) Bayesian estimation of concordance among gene trees. *Molecular biology and evolution*, **24**, 412–26.

Arnold J (1993) Cytonuclear Disequilibria in Hybrid Zones. *Annual Review of Ecology and Systematics*, **24**, 521–554.

Arnold SJ, Reagan NL, Verrell PA (1993) Reproductive Isolation and Speciation in Plethodontid Salamanders. *Herpetologica*, **49**, 216–228.

Arnold SJ, Verrell PA, Tilley SG (1996) The Evolution of Asymmetry in Sexual Isolation: A Model and a Test Case. *Evolution*, **50**, 1024–1033.

Avise JC (2009) Phylogeography: retrospect and prospect. *Journal of Biogeography*, **36**, 3–15.

Avise JC, Arnold J, Ball RM *et al.* (1987) Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics. *Annual Review of Ecology and Systematics*, **18**, 489–522.

Baack EJ, Rieseberg LH (2007) A genomic view of introgression and hybrid speciation. *Current Opinion Genetic Development*, **17**, 513–518.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.

Baum D, Shaw KL (1995) Genealogical perspectives on the species problem. In: *Experimental and Molecular Approaches to Plant Biosystematics* (eds Hoch P, Stephenson A), pp. 289–303. Missouri Botanical Garden, St. Louis.

Beamer D, Lamb T (2008) Dusky salamanders (Desmognathus, Plethodontidae) from the Coastal Plain: multiple independent lineages and their bearing on the molecular phylogeny of the genus. *Molecular phylogenetics and evolution*, **47**, 143–53.

Bernal MA, Gaither MR, Simison WB, Rocha LA (2016) Introgression and selection shaped the evolutionary history of sympatric sister-species of coral reef fishes (genus: Haemulon). *Molecular Ecology*, 639–652.

Blankers T, Adams DC, Wiens JJ (2012) Ecological radiation with limited morphological diversification in salamanders. *Journal of evolutionary biology*, **25**, 634–46.

Bonett R (2002) Analysis of the contact zone between the dusky salamanders Desmognathus fuscus fuscus and Desmognathus fuscus conanti (Caudata: Plethodontidae). *Copeia*, **2002**, 344–355.

Brito PH, Edwards S V (2009) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica*, **135**, 439–55.

Bruce RC (2010) Proximate Contributions to Adult Body Size in Two Species of Dusky Salamanders (Plethodontidae: Desmognathus). *Herpetologica*, **66**, 393–402.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis. *Molecular Biology and Evolution*, **29**, 1917–1932.

Bryson RW, de Oca AN-M, Jaeger JR, Riddle BR (2010) Elucidation of cryptic diversity in a widespread nearctic treefrog reveals episodes of mitochondrial gene capture as frogs diversified across a dynamic landscape. *Evolution; international journal of organic evolution*, **64**, 2315–30.

Burton R, Ellison C, Harrison J (2006) The sorry state of F2 hybrids: consequences of rapid mitochondrial DNA evolution in allopatric populations. *The American Naturalist*, **168**.

Calvo SE, Clauser KR, Mootha VK (2016) MitoCarta2.0: An updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Research*, **44**, D1251–D1257.

Carstens BC, Dewey T a (2010) Species delimitation using a combined coalescent and information-theoretic approach: an example from North American Myotis bats. *Systematic biology*, **59**, 400–14.

Chan KMA, Levin SA (2005) Leaky Prezygotic Isolation and Porous Genomes: Rapid Introgression of Maternally Inherited Dna. *Evolution*, **59**, 720–729.

Chifman J, Kubatko L (2014) Quartet Inference from SNP Data Under the Coalescent Model. *Bioinformatics*, **30**, 3317–3324.

Chippindale PT, Bonett RM, Baldwin AS, Wiens JJ (2004) Phylogenetic evidence for a major reversal of life-history evolution in plethodontid salamanders. *Evolution*, **58**, 2809–2822.

Chou J-Y, Leu J-Y (2010) Speciation through cytonuclear incompatibility: insights from yeast and implications for higher eukaryotes. *BioEssays : news and reviews in*

molecular, cellular and developmental biology, **32**, 401–11.

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

Crawford DC, Bhangale T, Li N *et al.* (2004) Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nature genetics*, **36**, 700–706.

Davey JW, Hohenlohe P a, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature reviews. Genetics*, **12**, 499–510.

Degnan JH, Rosenberg NA (2006) Discordance of Species Trees with Their Most Likely Gene Trees. *PLoS Genetics*, **2**.

Degnan J, Salter L (2005) Gene tree distributions under the coalescent process. *Evolution*, **59**, 24–37.

Dowling TE, Secor CL (1997) The role of hybridization in the evolutionary diversification of animals. *Annual Review of Ecology and Systematics*, **28**, 593–619.

Dray S, Dufour AB (2007) The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, **22**, 1–20.

Dunn EER (1926) *Salamanders of the Family Plethodontidae*. Smith College Fiftieth Anniversary Publications, Northampton, MA.

Durand EY, Patterson N, Reich D, Slatkin M (2011a) Testing for ancient admixture between closely related populations. *Molecular biology and evolution*, **28**, 2239–52.

Durand EY, Patterson N, Reich D, Slatkin M (2011b) Testing for ancient admixture between closely related populations. *Molecular biology and evolution*, **28**, 2239–52.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Eaton D, Ree R (2013) Inferring Phylogeny and Introgression using RADseq Data: An Example from Flowering Plants (Pedicularis: Orobanchaceae). *Systematic biology*, **0**, 1–18.

Edwards S V., Xi Z, Janke A *et al.* (2016) Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution*, **94**, 447–462.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using

multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–87.

Flot JF (2010) Seqphase: A web tool for interconverting phase input/output files and fasta sequence alignments. *Molecular Ecology Resources*, **10**, 162–166.

Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology and Evolution*, **27**, 480–488.

Funk DJ, Omland KE (2003) Species-Level Paraphyly and Polyphyly: Frequency, Causes, and Consequences, with Insights from Animal Mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 397–423.

Gagnaire P-A, Normandeau E, Bernatchez L (2012) Comparative genomics reveals adaptive protein evolution and a possible cytonuclear incompatibility between European and American Eels. *Molecular biology and evolution*, **29**, 2909–19.

Garesse R, Vallejo CG (2001) Animal mitochondrial biogenesis and function: a regulatory cross-talk between two genomes. *Gene*, **263**, 1–16.

Good JM, Hird S, Reid N *et al.* (2008) Ancient hybridization and mitochondrial capture between two species of chipmunks. *Molecular ecology*, **17**, 1313–27.

Gregory T (2017) Animal genome size database.

Grover CE, Salmon A, Wendel JF (2012) Targeted sequence capture as a powerful tool for evolutionary analysis. *American journal of botany*, **99**, 312–9.

Grummer JA, Bryson RW, Reeder TW (2014) Species delimitation using bayes factors: Simulations and application to the sceloporus scalaris species grp (Squamata: Phrynosomatidae). *Systematic Biology*, **63**, 119–133.

Guindon S, Dufayard JF, Lefort V *et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307–321.

Hairston Sr NG (1986) Species packing in Desmognathus salamanders: experimental demonstration of predation and competition. *American Naturalist*, **127**, 266–291.

Hedrick PW (2013) Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, **22**, 4606–4618.

Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular biology and evolution*, **27**, 570–80.

Heliconius T, Consortium G (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, **487**, 94–8.

Hey J (2006) On the failure of modern species concepts. *Trends in Ecology and Evolution*, **21**, 447–450.

Highton R, Peabody R (2000) Geographic protein variation and speciation in salamanders of the Plethodon jordani and Plethodon glutinosus complexes in the southerns Appalachian Mountains with the description of four new species. In: *The Biology of Plethodontid Salamanders* (ed RC Bruce, RG Jaeger, LD Houck E), pp. 31–94. Kluwer Academic/ Plenum Publishers.

Hoffberg SL, Kieran TJ, Catchen JM *et al.* (2016) RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. *Molecular Ecology Resources*, **16**, 1264–1278.

Houck L, Arnold S (2003) *Courtship and mating behavior*.

Houck L, Arnold S, Hickman A (1988) Tests for sexual isolation in plethodontid salamanders (genus Desmognathus). *Journal of Herpetology*, **22**, 186–191.

Houck L, Reagan NL (1990) Male courtship pheromones increase female receptivity in a plethodontid salamander. *Animal Behaviour*, **39**, 729-734.

Hudson RR, Coyne JA (2002) Mathematical consequences of the genealogical species concept. *Evolution*, **56**, 1557–1565.

Jockusch EL, Martínez-Solano I, Timpe EK (2015) The effects of inference method, population sampling, and gene sampling on species tree inferences: an empirical study in slender salamanders (Plethodontidae: Batrachoseps). *Systematic biology*, **64**, 66–83.

Joly S, McLenachan P, Lockhart PJ (2009) A statistical approach for distinguishing hybridization and incomplete lineage sorting. *The American naturalist*, **174**, E54-70.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart T, Devillard S (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics*, **11**, 94.

Jónsson H, Schubert M, Seguin-Orlando A *et al.* (2014) Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 18655–60.

Joshi N, Fass J (2011) Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (version 1.33).

Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, **2**, e281.

Karlin AA, Guttman SI (1986) Systematics and geographic isozyme variation in the plethodontid salamander Desmognathus fuscus (Rafinesque). *Herpetologica*, **42**, 283–301.

Kass R, Raftery A (1995) Bayes Factors. *Journal of the American Statistical Association*, **90**, 773–795.

Keck BP, Near TJ (2010) Geographic and temporal aspects of mitochondrial replacement in Nothonotus darters (Teleostei: Percidae: Etheostomatinae). *Evolution; international journal of organic evolution*, **64**, 1410–28.

Kingman J (1982) On the genealogy of large populations. *Journal of Applied Probability*, **19**, 27–43.

Kirkpatrick M, Ravigne V (2002) Speciation by natural and sexual selection: models and experiments. *American Naturalist*, **159**, S22–S35.

Knoll N, Jarick I, Volckmar A-L *et al.* (2013) Gene set of nuclear-encoded mitochondrial regulators is enriched for common inherited variation in obesity. *PloS one*, **8**, e55884.

Kozak KHKH, Larson A, Bonett RMRM, Harmon LJLJ (2005) Phylogenetic analysis of ecomorphological divergence, community structure, and diversification rates in dusky salamanders (Plethodontidae: Desmognathus). *Evolution*, **59**, 2000–2016.

Kozak KH, Mendyk RW, Wiens JJ (2009) Can parallel diversification occur in sympatry? Repeated patterns of body-size evolution in coexisting clades of North American salamanders. *Evolution; international journal of organic evolution*, **63**,

1769–84.

Kozak KH, Weisrock DW, Larson A (2006) Rapid lineage accumulation in a non-adaptive radiation: phylogenetic analysis of diversification rates in eastern North American woodland salamanders (Plethodontidae: Plethodon). *Proceedings. Biological sciences / The Royal Society*, **273**, 539–46.

Kozak KH, Wiens JJ (2006) Does niche conservatism promote speciation? A case study in North American salamanders. *Evolution; international journal of organic evolution*, **60**, 2604–21.

Kratovil J (2007) Molecular Systematics of Dusky Salamanders (Desmognathus) from Two Isolated Mountain Ranges in the North Carolina Piedmont. Smith College.

Kubatko L (2014) Species Tree Inference using SVDquartets. , 1–10.

Kumar V, Lammers F, Bidon T *et al.* (2017) The evolutionary history of bears is characterized by gene flow across species. *Scientific Reports*, **7**, 46487.

Lanfear R, Calcott B, Ho SYW, Guindon S (2012) Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular biology and evolution*, **29**, 1695–701.

Larget BR, Kotha SK, Dewey CN (2010) BUCKy : Gene Tree / Species Tree Reconciliation with Bayesian Concordance Analysis. , 2–3.

Leaché AD, Fujita MK, Minin VN, Bouckaert RR (2014a) Species delimitation using genome-wide SNP Data. *Systematic Biology*, **63**, 534–542.

Leaché AD, Harris RB, Rannala B, Yang Z (2014b) The influence of gene flow on species tree estimation: A simulation study. *Systematic Biology*, **63**, 17–30.

Leaché AD, Harris RB, Rannala B, Yang Z (2014) The influence of gene flow on species tree estimation: A simulation study. *Systematic Biology*, **63**, 17–30.

Leigh Van Valen (1976) Ecological Species, Multispecies, and Oaks on JSTOR. *Taxon Vol. 25, No. 2/3*, **25**, 233–239.

Lemmon AR, Lemmon EM (2012) Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic biology*, **61**, 727–44.

Lemmon EM, Lemmon AR (2013) High-Throughput Genomic Data in Systematics and Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 99–121.

Linnen CR, Farrell BD (2007) Mitonuclear discordance is caused by rampant

mitochondrial introgression in Neodiprion (Hymenoptera: Diprionidae) sawflies. *Evolution; international journal of organic evolution*, **61**, 1417–38.

Liu K, Raghavan S, Nelesen S, Linder CR, Warnow T (2009) Rapid and accurate large-scale coestimation of sequence alignments and phylogenetic trees. *Science (New York, N.Y.)*, **324**, 1561–4.

Maddison WP Systematic Biol. 1997 Maddison.pdf.

Maddison WP (1997) Gene trees in species trees. *Systematic biology*, **46**, 523–536.

Maddison WP, Knowles LL (2006) Inferring Phylogeny Despite Incomplete Lineage Sorting. *Systematic biology*, **55**, 21–30.

Mallet J (2007) Concepts of Species. *Encyclopedia of Life*, **5**, 1–15.

Mallet J, Besansky N, Hahn MW (2016) How reticulated are species? *BioEssays*, **38**, 140–149.

Marshall DC, Hill KBR, Cooley JR, Simon C (2011) Hybridization, mitochondrial DNA phylogeography, and prediction of the early stages of reproductive isolation: lessons from New Zealand cicadas (genus Kikihia). *Systematic biology*, **60**, 482–502.

Mayr E (1942) *Systematics and the Origin of Species*. Columbia Univ Press, New York.

McCullagh P, Nelder JA (1989) *Generalized Linear Models, 2nd ed.* Chapman and Hall, London.

Mead LS, Tilley SG, Katz L a (2001) Genetic structure of the Blue Ridge dusky salamander (Desmognathus orestes): inferences from allozymes, mitochondrial DNA, and behavior. *Evolution*, **55**, 2287–2302.

Mueller RL (2006) Evolutionary Rates, Divergence Dates, and the Performance of Mitochondrial Genes in Bayesian Phylogenetic Analysis. *Systematic biology*, **55**, 289–300.

Mueller RL, Macey JR, Jaekel M, Wake DB, Boore JL (2004) Morphological homoplasy, life history evolution, and historical biogeography of plethodontid salamanders inferred from complete mitochondrial genomes. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 13820–5.

Muir P, Li S, Lou S *et al.* (2016) The real cost of sequencing: scaling computation to keep pace with data generation. *Genome biology*, **17**, 53.

Murphy RW, Sites JW, Buth DG, Haufler CH (1996) Proteins : Isozyme Electrophoresis.

In: *Molecular Systematics* (eds Hillis DM, Moritz C, Mable BK), pp. 51–120. Sinauer Associates, Inc, Sunderland, MA.

Nikolenko SI, Korobeynikov AI, Alekseyev MA (2013) BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics*, **14**, S7.

O'Neill EM, Schwartz R, Bullock CT *et al.* (2013) Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (Ambystoma tigrinum) species complex. *Molecular ecology*, **22**, 111–29.

Osborne OG, Chapman MA, Nevado B, Filatov DA (2016) Maintenance of Species Boundaries Despite Ongoing Gene Flow in Ragworts. *Genome Biology and Evolution*, **8**, 1038–1047.

Pagliarini DJ, Calvo SE, Chang B *et al.* (2008) A Mitochondrial Protein Compendium Elucidates Complex I Disease Biology. *Cell*, **134**, 112–123.

Paradis E (2010) Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, **26**, 419–420.

Pardo-Diaz C, Salazar C, Baxter SW *et al.* (2012) Adaptive introgression across species boundaries in Heliconius butterflies. *PLoS genetics*, **8**, e1002752.

Parmakelis A, Kotsakiozi P, Rand D (2013) Animal mitochondria, positive selection and cyto-nuclear coevolution: insights from pulmonates. *PloS one*, **8**, e61970.

Patten MM, Carioscia SA, Linnen CR (2015) Biased introgression of mitochondrial and nuclear genes: A comparison of diploid and haplodiploid systems. *Molecular Ecology*, **24**, 5200–5210.

Pavey S a, Nielsen JL, Hamon TR (2010) Recent ecological divergence despite migration in sockeye salmon (Oncorhynchus nerka). *Evolution; international journal of organic evolution*, **64**, 1773–83.

Pease JB, Haak DC, Hahn MW, Moyle LC (2016) Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. *PLoS Biology*, **14**, 1–24.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one*, **7**, e37135.

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *Trends in ecology &*

evolution, **24**, 386–93.

Petranka J (1998) *Salamanders of the United States and Canada*. Smithsonian Institution Press., Washington.

Pichaud N, Ballard JWO, Tanguay RM, Blier PU (2012) Naturally occurring mitochondrial DNA haplotypes exhibit metabolic differences: insight into functional properties of mitochondria. *Evolution; international journal of organic evolution*, **66**, 3189–97.

Pinho C, Hey J (2010) Divergence with gene flow: models and data. *Annual Review of Ecology, Evolution, and …*, 215–232.

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

Putta S, Smith JJ, Walker J a *et al.* (2004) From biomedicine to natural history research: EST resources for ambystomatid salamanders. *BMC genomics*, **5**, 54.

de Queiroz K (2005) Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences of the United States of America*, **102 Suppl**, 6600–7.

de Queiroz K (2007) Species concepts and species delimitation. *Systematic biology*, **56**, 879–86.

Rambaut A, Suchard MA, Xie D, Drummond AJ (2014) Tracer v 1.6, Available from http://beast.bio.ed.ac.uk/Tracer.

Rand DM, Haney RA, Fry AJ (2004) Cytonuclear coevolution: The genomics of cooperation. *Trends in Ecology and Evolution*, **19**, 645–653.

Rannala B, Yang Z (2003) Using DNA Sequences From Multiple Loci. *Genetics*, **1656**, 1645–1656.

Reid N, Demboski JR, Sullivan J (2012) Phylogeny estimation of the radiation of western North American chipmunks (Tamias) in the face of introgression using reproductive protein genes. *Systematic biology*, **61**, 44–62.

Rheindt FE, Edwards S V (2011) Genetic introgression : an integral but neglected component of speciation in birds. *Perspectives in Ornithology*, **128**, 620–632.

Rissler LJ, Apodaca JJ (2007) Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (Aneides flavipunctatus). *Systematic biology*, **56**, 924–42.

Rissler LJ, Taylor DR (2003) The phylogenetics of Desmognathine salamander populations across the southern Appalachians. *Molecular Phylogenetics and Evolution*, **27**, 197–211.

Ronquist F, Teslenko M, Van Der Mark P *et al.* (2012) Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.

Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory, and the analysis of genetic polymorphisms. *Nature Reviews Genetics*, **3**, 380–390.

Ruane S, Bryson RW, Pyron RA, Burbrink FT (2014) Coalescent Species Delimitation in Milksnakes (Genus Lampropeltis) and Impacts on Phylogenetic Comparative Analyses. *Systematic Biology*, **63**, 231–250.

Safonova Y, Bankevich A, Pevzner PA (2015) dipSPAdes: Assembler for Highly Polymorphic Diploid Genomes. *Journal of Computational Biology*, **22**, 528–545.

Scharfe C, Lu HH-S, Neuenburg JK *et al.* (2009) Mapping gene associations in human mitochondria using clinical disease phenotypes. *PLoS computational biology*, **5**, e1000374.

Schuster S (2007) Next-generation sequencing transforms today's biology. *Nature*, **5**, 16–18.

Seehausen O (2004) Hybridization and adaptive radiation. *Trends in ecology & evolution*, **19**, 198–207.

Sequeira F, Sodré D, Ferrand N *et al.* (2011) Hybridization and massive mtDNA unidirectional introgression between the closely related Neotropical toads Rhinella marina and R. schneideri inferred from mtDNA and nuclear markers. *BMC evolutionary biology*, **11**, 264.

Sethuraman A, Hey J (2016) IMa2p - parallel MCMC and inference of ancient demography under the Isolation with migration (IM) model. *Molecular Ecology Resources*, **16**, 206–215.

Shen P, Wang W, Krishnakumar S *et al.* (2011) High-quality DNA sequence capture of 524 disease candidate genes. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 6549–54.

Sloan DB, Havird JC, Sharbrough J (2017) The on-again, off-again relationship between

mitochondrial genomes and species boundaries. *Molecular Ecology*, **26**, 2212–2236.

Stephens M, Donnelly P (2003) Report A Comparison of Bayesian Methods for Haplotype Reconstruction from Population Genotype Data. *Am. J. Hum. Genet*, **73**, 1162–1169.

Sukumaran J, Knowles LL (2017) Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, 1607-1612.

Swofford DL (2003) Phylogenetic Analysis Using Parsimony (*and Other Methods).

Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Tilley SG, Bernardo J, Katz LA *et al.* (2013) Failed species, innominate forms, and the vain search for species limits: Cryptic diversity in dusky salamanders (desmognathus) of eastern tennessee. *Ecology and Evolution*, **3**. 2547–2567.

Tilley SG, Eriksen RL, Katz L a. (2008) Systematics of dusky salamanders, Desmognathus (Caudata: Plethodontidae), in the mountain and Piedmont regions of Virginia and North Carolina, USA. *Zoological Journal of the Linnean Society*, **152**, 115–130.

Tilley SG, Mahoney MJ (1996) Patterns of genetic differentiation in salamanders of the Desmognathus ochrophaeus complex (Amphibia: Plethodontidae). *Herpetological Monographs*, **10**, 1–42.

Tilley SG, Schwerdtfeger PM (1981) Electrophoretic variation in Appalachian populations of the Desmognathus fuscus complex (Amphibia: Plethodontidae). *Copeia*, **1981**, 109–119.

Tilley SG, Verrell P a, Arnold SJ (1990) Correspondence between sexual isolation and allozyme differentiation: a test in the salamander Desmognathus ochrophaeus. *Proceedings of the National Academy of Sciences of the United States of America*, **87**, 2715–9.

Titus TA, Larson A (1996) Molecular phylogenetics of desmognathine salamanders (Caudata: Plethodontidae): A reevaluation of evolution in ecology, life history, and morphology. *Systematic Biology*, **45**, 451–472.

Tllley SG (1997) Patterns of Genetic Differentiation in Appalachian Desmognathine Salamanders. *Journal of Heredity*, **88**, 305–315.

Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular ecology*, 3907–3930.

Twyford AD, Ennos R a (2012) Next-generation hybridization and introgression. *Heredity*, **108**, 179–189.

Untergasser A, Cutcutache I, Koressaar T *et al.* (2012) Primer3-new capabilities and interfaces. *Nucleic Acids Research*, **40**, 1–12.

Verrell PA (1990a) Tests for sexual isolation among sympatric salamanders of the genus Desmognathus. *Amphibia-Reptilia*, **11**, 147–153.

Verrell PA (1990b) Sexual Compatibility among Plethodontid Salamanders: Tests between Desmognathus apalachicolae, and D. ochrophaeus and D. fuscus. *Herpetologica*, **46**, 415–422.

Wake DB (2009) What Salamanders Have Taught Us About Evolution. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 333–352.

Watts R a, Palmer C a, Feldhoff RC *et al.* (2004) Stabilizing selection on behavior and morphology masks positive selection on the signal in a salamander pheromone signaling complex. *Molecular biology and evolution*, **21**, 1032–41.

Weisrock DW, Kozak KH, Larson a (2005) Phylogeographic analysis of mitochondrial gene flow and introgression in the salamander, Plethodon shermani. *Molecular ecology*, **14**, 1457–72.

Weisrock DW, Macey JR, Ugurtas IH, Larson A, Papenfuss TJ (2001) Molecular Phylogenetics and Historical Biogeography among Salamandrids of the " True " Salamander Clade: Rapid Branching of Numerous Highly Divergent Lineages in Mertensiella luschani Associated with the Rise of Anatolia. *Molecular Phylogenetics and Evolution*, **18**, 434–448.

Wen D, Yu Y, Hahn MW, Nakhleh L (2016) Reticulate evolutionary history and extensive introgression in mosquito species revealed by phylogenetic network analysis. *Molecular Ecology*, **25**, 2361–2372.

Wiens JJ (2004) Notes and Comments What Is Speciation and How Should We Study It ? *American Society of Naturalists*, **163**, 914–923.

Wiens JJ, Engstrom TN, Chippindale PT (2006) Rapid diversification, incomplete isolation, and the "speciation clock" in North American salamanders (Genus Plethodon): testing the hybrid swarm hypothesis of rapid radiation. *Evolution; international journal of organic evolution*, **60**, 2585–603.

Woerner AE, Cox M., Hammer MF (2007) Recombination-Filtered Genomic Datasets by Information Maximization. *Bioinformatics*, **23**, 1851–1853.

Wooten J, Camp C, Rissler L (2010) Genetic diversity in a narrowly endemic, recently described dusky salamander, Desmognathus folkertsi, from the southern Appalachian Mountains. *Conservation Genetics*, **11**, 835–854.

Zarza E, Faircloth BC, Tsai WLE *et al.* (2016) Hidden histories of gene flow in highland birds revealed with genomic markers. *Molecular Ecology*, **25**, 5144–5157.

Zhang F, Broughton RE (2013) Mitochondrial-Nuclear Interactions: Compensatory Evolution or Variable Functional Constraint among Vertebrate Oxidative Phosphorylation Genes? *Genome biology and evolution*, **5**, 1781–1791.

Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, **30**, 614–620.

Zhou R, Zeng K, Wu W *et al.* (2007) Population genetics of speciation in nonmodel organisms: I. Ancestral polymorphism in mangroves. *Molecular biology and evolution*, **24**, 2746–54.

Zoppoli G, Douarre C, Dalla Rosa I *et al.* (2011) Coordinated regulation of mitochondrial topoisomerase IB with mitochondrial nuclear encoded genes and MYC. *Nucleic acids research*, **39**, 6620–32.

Vita

Birthplace
Feeding Hills, MA

Education
University of Kentucky, Lexington, KY
Ph.D. Candidate, Department of Biology (anticipated August 2017)
Advisor: David Weisrock
Dissertation title: *Mitochondrial and nuclear patterns of conflict and concordance at the gene, genome, and behavioral scales in Desmognathus salamanders*
Smith College, Northampton, MA
M.S. Biology (2007)

Hartwick College, Oneonta, NY
B.A. Biology (2003)

Professional Experience
University of Kentucky, Lexington, KY
Teaching and Research Assistant, Weisrock laboratory
August 2009-Present
Bio 555 "Vertebrate Zoology"; Guest Lecturer in 2015
Bio 303 "Introduction to Evolution"; Guest Lecturer in 2014
Bio 155 "Introductory Biology Laboratory"
Bio 153 "Principles of Integrative Biology"
Bio 111 "General Biology" for non-majors

Ameripath, Shelton, CT
Clinical Laboratory Technologist, Specialist in Cytogenetics
2008-2009

Smith College, Northampton, MA
Teaching Assistant, Department of Biological Sciences
2005-2007
Bio 267 "Plant Systematics"
Bio 201 "Animal Physiology"
Bio 111 "Molecules, Cells, and Systems"

Baystate Medical Center, Springfield, MA
Clinical Genetics Laboratory Technologist
2005-2006
Clinical Genetics Laboratory Technician
2003-2004

Hartwick College, Oneonta, NY
Undergraduate Teaching Assistant
2003
Bio 318 “Evolution”

Professional Honors, Awards, and Distinctions

2017 UK College of Arts and Sciences Certificate for Outstanding Teaching
2014 Gertrude Flora Ribble Graduate Fellowship, University of Kentucky ($11,000)
2014 NSF Doctoral Dissertation Improvement Grant ($19,485)
2013 Gertrude Flora Ribble Research Grant, University of Kentucky ($550)
2012 Bruce Family Scholarship in Herpetology, Highlands Biological Station, NC Grant-in-Aid of Research ($2100)
2012 Gertrude Flora Ribble Research Grant, University of Kentucky ($500)
2011 Gertrude Flora Ribble Research Grant, University of Kentucky ($500)
2010 Graduate Student Research Award, Society of Systematic Biologists ($1800)
2007 Nominated Associate Member of Sigma Xi, Scientific Research Society
2007 Elizabeth Horner Research Grant, Smith College ($3600)
2006 Elizabeth Horner Research Grant, Smith College ($1443)
2003 TriBeta Membership Award, Department of Biology, Hartwick College
2003 Chapter President of TriBeta, Biology Honor Society, Hartwick College
1998 Eagle Scout, BSA, Pioneer Valley Council, Chicopee MA

Publications

Tilley, S.G, J. Bernardo, L.A. Katz, L. López, J.D. Roll, R.L. Eriksen**, J. Kratovil**, N.K.J. Bittner, and K.A. Crandall. 2013. Failed species, innominate forms, and the vain search for species limits: Cryptic diversity in dusky salamanders (*Desmognathus*) of eastern Tennessee. Ecology and Evolution 3:2547-2567.

Kratovil, J.D. “Molecular systematics of dusky salamanders (*Desmognathus*) from two isolated mountain ranges in the North Carolina Piedmont.” Masters Thesis, Smith College 2007.

Kratovil, J.D., and A.R. Crooker. 2003. Leaf litter arthropod communities at Pine Lake with notes on pseudoscorpion ecology and morphology. Journal of Biological Research: Senior Research Theses Volume 4. Hartwick College Department of Biology, Oneonta NY, 13820

Presentations

Kratovil, J.D., Tilley, S.G., and Weisrock, D.W. 2016. Phylogenomics and the species limits of Piedmont and Non-Piedmont dusky salamanders (*Desmognathus fuscus*). Special Highlands Conference on Plethodontid Salamander Biology, Highlands, NC 2016. *Oral Presentation*

Kratovil, J.D., and Weisrock, D.W. 2015. Delimiting species boundaries in dusky

salamanders (*Desmognathus*), Society for the Study of Reptiles and Amphibians annual conference, Lawrence, KS 2015. *Oral Presentation*

Kratovil, J.D., and Weisrock, D.W. 2014. Reproductive isolation between divergent mitochondrial lineages of *Desmognathus fuscus* in the Piedmont of North Carolina. Evolution annual conference, Raleigh, NC. *Oral Presentation*

Kratovil, J.D., and Weisrock, D.W. 2014. Reproductive isolation between divergent mitochondrial lineages of *Desmognathus fuscus* in the Piedmont of North Carolina. Conference on the Biology of Plethodontid Salamanders, Tulsa, OK. *Poster Presentation*

Kratovil, J.D., and Weisrock, D.W. 2013. Divergent mtDNA lineages and reproductive isolation in southern Appalachian salamanders. American Genetic Association Symposium, Cornell University, Ithaca, NY. *Poster Presentation*

Kratovil, J.D., Bonett, R.M., and Weisrock, D.W. 2010. A framework for using comparative genomics to guide marker development and gene discovery in plethodontid salamanders. Evolution annual conference, Portland, OR. *Poster Presentation*

Kratovil, J.D., 2007. Life-histories of New England salamanders. Smith College hosting Northampton Public Schools (grades 5-8), *Oral Presentation*