2010

Using Landscape Genetics To Assess Population Connectivity In A Habitat Generalist

Tyler Duncan Hether
University of Central Florida

USING LANDSCAPE GENETICS TO ASSESS POPULATION CONNECTIVITY IN A HABITAT GENERALIST

by

TYLER DUNCAN HETHER
B.S. University of Central Florida, 2006

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

Summer Term
2010

Major Professor: Eric A. Hoffman

ABSTRACT

Understanding the nature of genetic variation in natural populations is an underlying theme of population genetics. In recent years population genetics has benefited from the incorporation of landscape and environmental data into pre-existing models of isolation by distance (IBD) to elucidate features influencing spatial genetic variation. Many of these landscape genetics studies have focused on populations separated by discrete barriers (e.g., mountain ridges) or species with specific habitat requirements (i.e., habitat specialists). One difficulty in using a landscape genetics approach for taxa with less stringent habitat requirements (i.e., generalists) is the lack of obvious barriers to gene flow and preference for specific habitats. My study attempts to fill this information gap to understand mechanisms underlying population subdivision in generalists, using the squirrel treefrog (*Hyla squirella*) and a system for classifying 'terrestrial ecological systems' (i.e. habitat types). I evaluate this dataset with microsatellite markers and a recently introduced method based on ensemble learning (Random Forest) to identify whether spatial distance, habitat types, or both have influenced genetic connectivity among 20 *H. squirella* populations. Next, I hierarchically subset the populations included in the analysis based on (1) genetic assignment tests and (2) Mantel correlograms to determine the relative role of spatial distance in shaping landscape genetic patterns. Assignment tests show evidence of two genetic clusters that separate populations in Florida's panhandle (Western cluster) from those in peninsular Florida and southern Georgia (Eastern cluster). Mantel correlograms suggest a patch size of approximately 150 km. Landscape genetic analyses at all three spatial scales yielded improved model fit relative to isolation by distance when including habitat types. A hierarchical effect was identified whereby the importance of spatial distance (*km*) was the strongest predictor of patterns of genetic differentiation above the scale of the genetic patch. Below the genetic patch, spatial distance was still an explanatory variable but was only

approximately 30% as relevant as mesic flatwoods or upland oak hammocks. Thus, it appears that habitat types largely influence patterns of population genetic connectivity at local scales but the signal of IBD becomes the dominant driver of regional connectivity. My results highlight some habitats as highly relevant to increased genetic connectivity at all spatial scales (e.g., upland oak hammocks) while others show no association (e.g., silviculture) or scale specific associations (e.g., pastures only at global scales). Given these results it appears that treating habitat as a binary metric (suitable/non-suitable) may be overly simplistic for generalist species in which gene flow probably occurs in a spectrum of habitat suitability. The overall pattern of spatial genetic and landscape genetic structure identified here provides insight into the evolutionary history and patterns of population connectivity for *H. squirella* and improves our understanding of the role of matrix composition for habitat generalists.

ACKNOWLEDGMENTS

There are several individuals that helped make this endeavor possible. First and foremost, I am in debt to my advisor Eric Hoffman for his guidance and tutelage. I would like to also thank my committee members, Chris Parkinson and David Jenkins, for their support both inside and outside of the classroom. Rosanna Tursi, Sarah May, Juan Daza, Allyson Fenwick, Haakon Kalkvik, Genevieve Metzger, Greg Territo, and Ocean Cohen provided insightful comments and encouragement throughout this process. Todd Earnhardt from the Biodiversity and Spatial Information Center at North Carolina State University helped with providing anthropogenic habitat classification descriptions. Liza McCauley and Kim Medley provided GIS support and theoretical considerations. Genevieve Metzger, Christian Metzger, Allyson Fenwick, Will Fenwick, Greg Territo, and Jessica Hightower helped with sample collections throughout the summer 2009 field season. I would like to thank the following individuals for their assistance in statistical analyses. I am thankful for Pedro Quintana-Ascencio and James Angelo for their inspiration to appreciate statistics. Andrew Storfer and Melanie Murphy from the department of Zoology at Washington State University helped with landscape genetics analyses. Richard Cutler from the department of Mathematics and Statistics at Utah State University aided in partial dependence plots and various scripts. This thesis was supported by Sigma Xi Grants-in-Aid research. Finally, I would like to take this opportunity to acknowledge Genevieve Metzger. She helped virtually in every aspect of this thesis. Without her friendship I would have not pursued this research project. I therefore owe her my deepest gratitude.

TABLE OF CONTENTS

LIST OF FIGURES viii

LIST OF TABLES xii

LIST OF ACRONYMS/ABBREVIATIONS xiv

CHAPTER 1: INTRODUCTION 1

CHAPTER 2: METHODS 1

Study Species 1

Sample Collections 5

Molecular Analysis 6

Spatial Genetic Structure 8

Genetic Clusters 8

Isolation by Distance 10

Habitat Permeability 11

Combining Landscape and Genetic Datasets 11

Running the Full RF Model and Calculating MIR Values 12

Model Selection Algorithm 12

Final RF and Inference 13

Hierarchical Landscape Genetic Structure 13

CHAPTER 3: RESULTS 14

Molecular Analysis 14

Spatial Genetic Structure 15
Genetic Clustering 15
Isolation by Distance 24
Genetic Patch Size 24
Landscape Genetics at the Global Scale 28
Hierarchical Landscape Genetic Structure 33
Landscape Genetic Structure at the Intermediate Scale 33
Landscape Genetic Structure at the Local Scale 33
CHAPTER 4: DISCUSSION 34
Landscape Genetics and Habitat Generalists 34
Evolution and Population Connectivity in the Squirrel Treefrog 39
CHAPTER 5: CONCLUSION 43
APPENDIX A: PREDICTOS USED TO ASSESS HABITAT PERMEABILITY 44
APPENDIX B: SINGLE-SCALE RF MODELS 53
APPENDIX C: CORRELATION AMONG VECTOR BUFFER WIDTHS 56
LIST OF REFERENCES 58

LIST OF FIGURES

Figure 1: Number of population genetic clusters (*K*) identified using two assignment algorithms at the global scale (n = 20 populations). Left panel gives the likelihood of the data for a given *K* (black line; right hand Y-axis) plus (+) and minus (-) one-half of the SD across 10 independent runs. Output of the Evanno criterion (see text) is also provided on the left panel (gray line; left hand Y-axis. Right panel gives the relative density of the number of populations along the MCMC chain following burnin using the Geneland algorithm. 16

Figure 2: Individual assignments at the global scale based on the STRUCTURE algorithm for *K*=2 genetic clusters. Each individual is shown as a column. The proportion of each individual's genome that originated in *K* clusters is shown. Population abbreviations are provided and cross-reference to Table 1. Different colors represent different clusters (*K*), and the length of columns represents the proportion of each individual's genome that originated from the color-coded *K*. For instance, PEN (far left) contains individuals whose genomes are mostly derived from the black (Western) cluster. 17

Figure 3: Geneland output of Posterior Probability (see legend) of Eastern cluster membership at the global scale (n = 20 populations). Population abbreviations cross-reference with Table 1. Posterior probability of the Western cluster is defined as 1 minus the Posterior Probability of the Eastern Cluster. 18

Figure 4: Second-order STRUCTURE analysis showing the number of population genetic clusters (*K*) identified using two assignment algorithms in the Western cluster (n = 4 populations). Left panel gives the likelihood of the data for a given *K* (black line; right hand Y-axis) plus (+) and minus (-) one-half of the SD across 10 independent runs. Output of the Evanno

criterion (see text) is also provided on the left panel (gray line; left hand Y-axis. Right panel gives the relative density of the number of populations along the MCMC chain following burnin using the Geneland algorithm. 20

Figure 5: Third-order STRUCTURE analysis showing the number of population genetic clusters (*K*) identified using two assignment algorithms in the Western cluster but excluding PEN. Left panel gives the likelihood of the data for a given *K* (black line; right hand Y-axis) plus (+) and minus (-) one-half of the SD across 10 independent runs; Output of the Evanno criterion (see text) is also provided on the left panel (gray line; left hand Y-axis. Right panel gives the relative density of the number of populations along the MCMC chain following burnin using the Geneland algorithm. 21

Figure 6: Individual assignments within Western cluster based on the STRUCTURE algorithm. Each individual is shown as a column. The proportion of each individual's genome that originated in *K* clusters is shown. Population abbreviations are provided and cross-reference to Table 1. A: all individuals within the Western cluster (*K* = 2; n = 152 individuals); B: Same as (A) but excluding PEN (*K* = 3; n = 115 individuals). 22

Figure 7: Number of population genetic clusters (*K*) identified using two assignment algorithms within the Eastern Cluster. Left panel gives the likelihood of the data for a given *K* (black line; right hand Y-axis) plus (+) and minus (-) one-half of the SD across 10 independent runs; Output of the Evanno criterion (see text) is also provided on the left panel (gray line; left hand Y-axis. Right panel gives the relative density of the number of populations along the MCMC chain following burnin using the Geneland algorithm. 23

Figure 8: Genetic distance as a function of geographic distance at the global scale (left plots) and intermediate scale (i.e., within the Eastern cluster; right plots) for two response types: genetic distance based on reduction of heterozygosity (top plots) and based on allele frequency distribution (bottom plots). Significance is based on Mantel tests with 9,999 randomizations using the program IBDWS. 26

Figure 9: Mantel correlogram within the Eastern cluster for log-transformed R_{ST} (A) and D_{ps} values (B). Mantel r is similar to Pearson's product-moment coefficient and ranges from -1 to 1. For each distance class (i.e., 50 km interval) the Mantel *r* + 95% confidence interval is shown. Filled circles denote significance at the Bonferroni corrected level $\alpha' = 0.05/8 = 0.00625$ 27

Figure 10: Partial Plots for R_{ST}-based RF models. A) global scale, B) intermediate scale. Values above figures denote Model Improvement Ratios (MIRs) for retained predictors; that is, the importance of a given predictor *relative* to the most important predictor (far left; MIR=1.00). These plots show the predictive function of (log-transformed) R_{ST} on a given predictor while accounting for the average effects of other predictors. For example, R_{ST} has a nonmonotonic partial dependence on *urban*; it decreases nearly linearly throughout the main body of the data (denoted by rug) until approximately 10% cover before it increases. 31

Figure 11: Partial Plots for D_{ps}-based RF models. A) global scale, B) intermediate scale, C) local scale. Values above figures denote Model Improvement Ratios (MIRs) for retained predictors; that is, the importance of a given predictor *relative* to the most important predictor (far left; MIR=1.00). These plots show the predictive function of (log-

transformed) D_{ps} on a given predictor while accounting for the average effects of other predictors. 32

Figure 12: Bivariate partial dependence plot for *Hyla squirella* genetic differentiation in the Eastern cluster (D_{ps}-based model). Here the partial dependence is the effect of two predictor variables, upland oak hammocks (*oak*) and urban land cover (*urban*), on the model after accounting for the effect of spatial distance (*km*). Decreasing partial dependence can be inferred as increasing genetic connectivity. 42

LIST OF TABLES

Table 1: Southeast regional GAP (SEGAP) dataset-derived variables used to assess habitat permeability in this study. Habitats are categorized by their type (anthropogenic or (semi)-natural). For each habitat the name, abbreviation, brief description, ecological justification, general genetic response, and references are given. A detailed list of these habitats can be found in Appendix A1 3

Table 2: Spatial information for 20 sampling localities (i.e., populations) used in this study. For each locality, abbreviation, latitude, longitude, and number (n) of genotyped *Hyla squirella* samples are given. 5

Table 3: Summary Statistics for Mantel tests at the global and intermediate scales for log-transformed R_{ST} values on log-transformed spatial distance (km). P-value based on 9,999 randomizations. 25

Table 4: Summary Statistics for Mantel tests at the global and intermediate scales for log-transformed D_{ps} values on log-transformed spatial distance (km). P-value based on 9,999 randomizations. 25

Table 5. Features associated with genetic connectivity among *Hyla squirella* populations using Random Forest. Presented here are the chosen predictors for following model selection for the 500 m vector buffer width (detailed summary statistics for all vector buffer widths can be found in Appendix B). Models are grouped first by hierarchical scale (see text), next by response (R_{ST} or D_{ps}), and finally by type. $p R^2$ is a pseudo R^2; MSE denotes mean squared error. Summary statistics, based on constructing 30 forests for each sub-model (see Methods), include median and 95% confidence intervals (95% CI) of $p R^2$ and median MSE.

P-value of the chosen sub-model is provided (see text). Model denotes the chosen variables (names cross-reference with Table 1) following model selection. These variables are ordered starting with the most important variable (in terms of MIR values) to the least important. Font style denotes general trend identified using partial dependence plots: standard font, negative association with genetic differentiation among populations; Boldface, positive association with genetic differentiation among populations; italic, complex association (i.e., nonlinear association whereby genetic differentiation has a minimum at some value of percent cover); and underline; weak main effect. 30

LIST OF ACRONYMS/ABBREVIATIONS

AR	Allelic Richness
DOP-PCR	Degenerate Oligonucleotide-Primed Polymerase Chain Reaction
DPS	Genetic distance based on the proportion of shared alleles
FST	Genetic distance based on allelic state
HE	Expected Heterozygosity
HWE	Hardy-Weinberg Equilibrium
HWLE	Hardy-Weinberg and Linkage Equilibria
IBD	Isolation by Distance
K	Number of distinct genetic clusters
MCMC	Markov Chain Monte Carlo
MIR	Model Improvement Ratio
NE	Effective population size
PCR	Polymerase Chain Reaction
RF	Random Forest
RMA	Reduced Major Axis
RST	Genetic distance based on allele size
SEGAP	Southeast regional Gap Analysis Project
SMM	Stepwise Mutation Model

CHAPTER 1: INTRODUCTION

Understanding how genetic variation is partitioned among populations is of fundamental importance in evolutionary biology. Population genetic studies often infer dispersal among populations by correlating pairwise genetic distance (e.g., F_{ST}) with straight-line spatial distance (i.e., isolation by distance; IBD; Wright 1943) to determine the relationship between genetic and geographic distance. Notwithstanding the ubiquity with which IBD is used in studies of population structure, the correlation of spatial distance often only weakly explains genetic distance among populations (Jenkins *et al.* 2010), prompting researchers to identify factors other than Euclidean distance that may explain patterns of gene flow. Within the past decade there has been a surge of research effort aimed at quantifying how extrinsic factors, such as landscape and environmental features, facilitate or inhibit genetic connectivity among natural populations. The field of Landscape Genetics (Manel *et al.* 2003; Storfer *et al.* 2007; Holderegger & Wagner 2008) stems from the realization that the classical models of IBD are overly simplistic in their assumption that the inter-population landscape matrix is homogeneous and does not influence gene flow (Kozak *et al.* 2008). Indeed, previous landscape genetic research has revealed strong correlations between genetic distance and ecologically relevant features including habitat gaps (Pierson *et al.* 2010), cover type and river crossings (Spear *et al.* 2005), species-specific corridors (Banks *et al.* 2005; Spear & Storfer 2010), salinity (Bekkevold *et al.* 2005), slope (Lowe *et al.* 2006), anthropogenic versus natural forest cover (Pavlacky *et al.* 2009), conservation-relevant habitats (Emaresi *et al.* 2009), and spatial scale (Chan *et al.* 2009).

One pattern that has emerged is that much of landscape genetic research to date has focused on the genetic consequences of landscape features for habitat specialist (i.e., species that have

specific habitat requirements)(e.g., Stevens *et al.* 2006; Chan *et al.* 2009). In contrast, there is a general paucity of studies that use a landscape genetics approach for species with less stringent habitat requirements (i.e., generalist species). This disparity may be due to publication bias, as studies conducted for generalist species may lack adequate genetic structure at the scale under study. Two recent studies that have compared genetic structure among related habitat specialist and generalist species determined that generalist species tend to have higher genetic connectivity than specialist species (Brouat *et al.* 2003; Vandergast *et al.* 2004). These studies suggest that habitat specialists display a lower propensity to disperse through unsuitable habitat (i.e., higher landscape resistance; McRae 2006) owing to two mechanisms, population fragmentation and decreased ability to diffuse through the intervening landscape. Subsequently, the magnitude of measurable genetic differentiation for habitat specialists is expected to be larger than for generalists (Geffen *et al.* 2004). Nevertheless, a growing number of studies suggest that landscape features can be paramount in shaping patterns of gene flow and genetic structure in habitat generalists, even though generalist species are often regarded as existing as panmictic populations. As in habitat specialists, discrete barriers (e.g., large rivers or highly trafficked roadways) can serve as barriers to gene flow for habitat generalists (Frantz *et al.* 2010). However, constrained movement across the landscape matrix can be more difficult to analyze for generalists than specialists. For extreme specialists the landscape can be treated as a simple binary matrix: suitable and non-suitable (Chan *et al.* 2009); forests and non-forest (Vandergast *et al.* 2004); scrub and non-scrub (Hokit *et al.* 2010). On the other hand, generalists can inhabit a range of habitats and it has been suggested that the spectrum of habitat optimality can be larger for generalists as compared to habitat specialist (Stewart *et al.* 2010).

Amphibians are model organisms for studying how the landscape alters patterns of gene flow and genetic structure. Many amphibians are philopatric organisms with low dispersal rates

which makes them ideal for detecting fine-scale genetic structure (Duellman & Trueb 1986; Blaustein *et al.* 1994; Blaustein *et al.* 2003; Funk *et al.* 2005b; Manier & Arnold 2006; Giordano *et al.* 2007). Most amphibians have a biphasic life history and thus experience both aquatic and terrestrial habitat degradation and environmental stressors (Duellman & Trueb 1986; Blaustein *et al.* 2003; Steele *et al.* 2009; Storfer *et al.* 2009). Moreover, anurans are particularly useful in landscape genetics due to the ease of sampling, predictable breeding habitats, and relatively short generation time. Not surprisingly, a large portion of the available landscape genetic literature has used this group to investigate features that correlate with genetic distance among populations (e.g., Funk *et al.* 2005a; Spear *et al.* 2005; Stevens *et al.* 2006; Spear & Storfer 2008; Allentoft *et al.* 2009; Angelone & Holderegger 2009; Chan & Zamudio 2009; Lee-Yaw *et al.* 2009; Richards-Zawacki 2009; Wang 2009; Zhao *et al.* 2009; Murphy *et al.* 2010; Spear & Storfer 2010). From these and other studies of anuran ecology and population genetics, it is apparent that salient and discrete *a priori* defined landscape features (e.g., mountain ridges or major rivers separating populations of poor dispersing individuals) tend to correspond with gene flow patterns up to and in excess of the effects of IBD. For example, Funk *et al.* (2005a) suggested that mountain ridges and elevation were associated with greater genetic differentiation among populations of Columbia spotted frogs (*Rana luteiventris*) based on results of simple and partial Mantel tests.

Several analytic tools have been used to assess the influence of landscape features on gene flow (for review see Balkenhol *et al.* 2009b). Among these, the partial Mantel test (Smouse *et al.* 1986) has been used extensively as an extension to the Mantel test (Mantel 1967) whereby three or more distance matrices (e.g., genetic, spatial, and environmental) can be used to determine partial correlation coefficients for all matrices under investigation. The partial Mantel test is an attractive approach in landscape genetics because IBD is often assumed and researchers are interested in the

effects of some other variable while holding spatial distance constant. However, the validity of the partial Mantel test has been called into question (Raufaste & Rousset 2001; Castellano & Balletto 2002; Rousset 2002) and simulations have shown this test can yield a high Type-I error rate (Balkenhol *et al.* 2009b). Moreover, assessing all pairwise combinations among populations may not be biologically meaningful if, for example, a pairwise vector traverses an absolute barrier to dispersal (i.e., ocean separating populations of amphibians). An alternative method to identify the importance of variables uses the nonlinear classification and regression tree-like (CART-like) analysis Random Forest (hereafter RF; Breiman 2001). RF has been used in other disciplines including bioinformatics (Cutler & Stevens 2006), chemoinformatics (Svetnik *et al.* 2003; Svetnik *et al.* 2005), ecology (Cutler *et al.* 2007), landscape ecology (Evans & Cushman 2009), and has been recently introduced to landscape genetics (Murphy *et al.* 2010). Briefly, RF is similar to Bagging, or bootstrap aggregation (Hastie *et al.* 2009), whereby an ensemble of classification or regression trees (regression in this study) are grown, each on a bootstrap sample of the training data (Svetnik *et al.* 2005). The predictions of each bootstrap tree are then averaged (for regression) to give a final prediction (Svetnik *et al.* 2003). Trees notoriously have high variance; small changes in the data can result in different series of splits down the tree. The advantage of averaging the predictions of many bootstrapped trees smoothes out this variance (Hastie *et al.* 2009). RF adds another layer of randomness to the Bagging procedure to further reduce the variance by reducing the correlation between trees; this reduction is accomplished by randomly selecting only a subset of the total predictors (m) as candidates for node splitting (Hastie *et al.* 2009). RF is an appealing technique in landscape genetics as it has the ability to handle wide datasets (i.e., relatively large number of predictors, p, compared to number of observations, n), handle redundant and/or irrelevant predictors, and provide a type of cross-validation in parallel with the training step, and because

variable importance can be used in conjunction with partial dependence plots to aid in biological interpretation (Svetnik *et al.* 2003; Cutler *et al.* 2007; Hastie *et al.* 2009; Siroky 2009).

The overarching goal of this study was to determine the utility of landscape genetics in assessing genetic connectivity for an abundant habitat generalist, the squirrel treefrog (*Hyla squirella*). The landscape genetics approach typically involves two steps to investigate landscape and environmental influences on patterns of gene flow: (1) identify patterns of genetic discontinuity (e.g., clusters, clines and genetic differentiation) and (2) correlate these patterns with landscape and environmental features (Manel *et al.* 2003; Guillot *et al.* 2005a; Holderegger *et al.* 2006). My first aim was to identify whether incorporating landscape variables into models of IBD increased the explanatory power of population connectivity. Thus, IBD served as a null model. If landscape information is important, I asked whether all habitat types that were predicted to increase genetic connectivity uniformly contributed to model fit or if some habitats were more important than others. The former may suggest that habitat can be treated as a binary predictor (suitable/non-suitable) as with many landscape genetics studies for specialist species; the latter would be expected in systems characterized by a spectrum of habitat suitability. I used RF and the general methodology of Murphy *et al.* (2010) to determine (i) if including habitats explained more of the variation in genetic distance among populations than spatial distance alone and (ii) variable importance for each habitat. Second, I tested whether there was a hierarchical effect and I predicted that the contribution of spatial distance would decrease with decreasing spatial scale. These spatial scales were analyzed systematically: the 'global scale' was defined by my sampling scheme; the 'intermediate scale' was defined genetically based upon individual assignment tests; and the 'local scale' was identified by estimating genetic patch sizes within clusters. Typically, landscape genetics studies focus on range restricted species or species with specific habitat requirements. This study uniquely investigates the

extent that landscape genetic approaches can determine the influence of ecological features for a habitat generalist and how the spatial scale under investigation can affect inferences of population connectivity. These data are discussed with regard to how my *a priori* expectations are met given the data and how this study further advances the growing field of landscape genetics.

CHAPTER 2: METHODS

Study Species

Hyla squirella is one of the most abundant treefrogs found along the Atlantic and Gulf Coastal Plains of the United States – occurring from Virginia to eastern Texas and south to the Florida Keys (Lannoo 2005). There are two reasons why *H. squirella* provides an ideal model system for this study. First, as a terrestrial anuran, it likely exhibits many of the beneficial attributes of amphibians in population genetics study (see above); second, it is a habitat generalist. Indeed, Carr (1940) described this species as showing little discrimination in terms of suitable habitats. *Hyla squirella* occur in a wide range of habitats including: fields and urbanized areas (Deckert 1915; Wright 2002); swamps (Lannoo 2005); pine and oak groves (Wright & Wright 1995); and almost anywhere adjacent to food, moisture, and shelter (Conant & Collins 1998). Breeding habitats include grassy, ephemeral pools free of predatory fish such as road side ditches (Wright & Wright 1995; Babbitt & Tanner 1997; Jensen *et al.* 2008). There are reports of preference for open wooded areas (Carr 1940; Wright & Wright 1995) and oviposition usually occurs in open canopy ponds (Binckley & Resetarits 2007). From the available anuran landscape genetics and *H. squirella* literature I predicted that 7 habitat types derived from the 2001 Southeast Gap Analysis Project (SEGAP; Comer & Schulz 2007) would influence gene flow among *H. squirella* populations (Table 1). SEGAP consists of ecological systems (natural or semi-natural), human-modified land (e.g., pastures and urbanized regions), and non-terrestrial land cover type (e.g., lakes). Ecological systems are US National Vegetation Classification (US-NVC) plant community associations that tend to co-occur in areas

with similar ecological dynamics (e.g., flooding, fire regime) and environmental settings and gradients (Comer & Schulz 2007).

Table 1: Southeast regional GAP (SEGAP) dataset-derived variables used to assess habitat permeability in this study. Habitats are categorized by their type (anthropogenic or (semi)-natural). For each habitat the name, abbreviation, brief description, ecological justification, general genetic response, and references are given. A detailed list of these habitats can be found in Appendix A1

Name	Abbreviation	Brief Description	Ecological Justification	Genetic Prediction	Reference
Type - Anthropogenic†					
Urbanization land cover	*urban*	Developed urbanized land of varying intensity.	Houses and buildings provide various degrees of shelter.	Complex with more connectivity at intermediate percent cover	(Wright 2002; Jensen *et al.* 2008)
Silviculture	*sil*	Forests established by planting and/or seeding in; can include dense forest canopy cover.	Many pond-breeding amphibians require upland forested habitats for foraging and overwintering.	Increase connectivity	(Semlitsch 1998; Babbitt *et al.* 2006)
Pastures and Crop land	*pas*	Agricultural land for livestock grazing or the production of seed or hay crops.	Pasture land often comprises a mosaic of ephemeral, open-canopy ponds suitable for breeding *H. squirella*.	Increase connectivity	(Babbitt & Tanner 1997; Babbitt & Tanner 2000; Babbitt *et al.* 2006; Binckley & Resetarits 2007)

Name	Abbreviation	Brief Description	Ecological Justification	Genetic Prediction	Reference
Type - (semi)-natural					
Mesic Flatwoods	*flat*	Forested systems characterized by *Pinus spp.* with frequent, low-intensity fires and subject to seasonally high water tables.	Fire regime in this system concomitant with hydroperiod allows for relatively high occurrence of suitable breeding habitats.	Increase connectivity	(Binckley & Resetarits 2007)
Swamp	*swamp*	Hardwood/deciduous canopy dominants and hydrology dominated by rainfall and sheetflow.	Preferred habitat for *H. squirella*	Increase connectivity	(Jensen *et al.* 2008)
Water and Floodplain Forest	*rff*	Open water and forested systems associated with lotic environments.	Flooding (from nearby rivers) is a key ecological factor in this system which can increase the density of ponds containing predatory fish	Decrease connectivity	(Babbitt & Tanner 2000)
Upland oak hammock	*oak*	Upland oak dominated habitat with infrequent fire frequency	Many pond-breeding amphibians require upland forested habitats for foraging and overwintering.	Increase connectivity	(Semlitsch 1998; Babbitt *et al.* 2006)

† Descriptions of non-ecological systems provide by T. Earnhardt from the Biodiversity and Spatial Information Center at North Carolina State University.

Sample Collections

Sampling localities were chosen with the main goal of facilitating a continuum of pairwise comparisons needed to correlate landscape composition and genetic differentiation at various spatial scales (Guillot *et al.* 2009). First, I bisected the study domain into 75km^2 strata. Second, I randomly sampled two points within each stratum. Third, I buffered these points using 5 km radii using ArcView 9.2 (ESRI, Inc.). Suitable breeding habitats within these buffered zones were surveyed on nights following moderate to heavy rainfall during the summer 2009 breeding season. From these surveys I collected 675 tissue samples (toe clips) from 20 georeferenced breeding sites (Table 2).

Table 2: Spatial information for 20 sampling localities (i.e., populations) used in this study. For each locality, abbreviation, latitude, longitude, and number (n) of genotyped *Hyla squirella* samples are given.

Abbreviation	Latitude	Longitude	n
AST	29.1605	-81.5535	32
CHAR	26.9317	-81.7607	32
CUT	29.5505	-83.1829	30
DISS	29.2771	-81.3343	32
EAPP	30.0282	-84.9879	32
GRAS	29.0147	-82.3232	31
GULF	28.5390	-82.6171	37
HIKE	30.3461	-83.3394	33
LAZY	28.6266	-81.8882	63
OCK	29.5376	-81.7780	36
OST	28.8461	-81.0936	22
PALM	27.9213	-80.5515	19
PEN	30.3196	-87.2634	37
PINE	30.0503	-81.3978	30
SAND	30.2744	-82.2845	32
SPAR	29.3811	-82.0420	46
SR2	30.3849	-86.3761	41
STAR	29.9711	-82.2559	32
WAPP	30.1358	-85.3702	42
WAY	31.2089	-82.4494	16

Molecular Analysis

For DNA extraction I used the standard phenol-chloroform method (Sambrook & Russel 2001). Individuals were genotyped at nine microsatellite loci that were specifically developed for *H. squirella*. Abdoullaye *et al.* (2010) provides a full description of the primer development protocol and accession numbers. Briefly, total genomic DNA was fragmented using a degenerate oligonucleotide-primed polymerase chain reaction (DOP-PCR) and amplicons were hybridized with 5'-biotinylated, 3'-amino modified $(CA)_{15}$ or $(GATA)_8$ repeat motifs bound to streptavidin-coated particles (Ardren *et al.* 2002; Hoffman *et al.* 2003). Hybridization conditions followed Ardren *et al.* (2002) with slight modifications: 1) hybridization temperature profile was 95°C for 5 min, then 52°C for 25 min (ramp speed 0.1°C/sec.) and 2) the final two washes were carried out at 72°C. Enriched product underwent a second DOP-PCR and cloned using a TOPO TA cloning kit (Invitrogen, Carlsbad, CA). Clones containing the repeat motifs were identified using the T3/T7 screening procedure of Cabe & Marshall (2001). Primer pairs were designed from positive clones with adequate flanking region and checked for polymorphism and deviations from Hardy-Weinberg and linkage equilibria (HWLE).

Following PCR, amplicons from these nine loci were visualized on a 2% agarose gel to verify amplification and genotypes were scored on a Beckman CEQ8000 (Beckman-Coulter, Fullerton, CA) following the manufacture's protocol. Genotypes were initially checked for high null allele frequencies (>0.09), allelic dropout, and scoring errors with MICRO-CHECKER *v* 2.23 (Van Oosterhout *et al.* 2004). I tested for significant deviations from HWLE (Fisher's exact test) using GENEPOP *v.* 4.0.7 (Raymond & Rousset 1995; Rousset 2008). Markov chain parameters for all tests

included a dememorization of 10,000 and 500 batches (10,000 iterations per batch). I accounted for multiple comparisons by applying a sequential Bonferroni correction (Rice 1989). Recently colonized populations may show a noticeable decrease in expected heterozygosity or allelic richness. Such recently founded populations may have a low heterozygosity (H_E) and allelic richness (A_R) compared to longer established populations and the permeability of the landscape may not have had sufficient time for a detection of a genetic signature. Populations were screened for evidence of recent colonization by comparing population specific H_E and A_R in FSTAT v 1.2 (Goudet 1995).

Heterozygosity-based estimates of genetic distance between pairwise populations were assessed using SPAGEDI 1.3 (Hardy & Vekemans 2002). Genetic distances based on allele size (i.e., R_{ST}) are expected to be larger than those based on allelic state (i.e., F_{ST}) when loci are at least partially stepwise mutation model-like (SSM-like) and have a high mutation rate (e.g., microsatellites) compared to the effect of drift or migration (Hardy *et al.* 2003). Because I sampled at a broad (i.e., regional) spatial coverage whereby migration rates between populations may be comparably low and/or divergence time long, I tested the null hypothesis that allele sizes do not contribute to the observed genetic differentiation (i.e., $F_{ST} = R_{ST}$). This hypothesis was tested by randomly permuting allele sizes among allelic states and generating a null distribution of R_{ST} values. I considered R_{ST} to be significantly larger than expected under the null hypothesis if the observed R_{ST} values fell within the 5% most extreme of the randomized R_{ST} values (one-tailed test) (Hardy *et al.* 2003).

Estimates of genetic distance that are based on allele frequency distributions are expected to detect more recent population-level differences due to landscape features than compared to genetic distances based on heterozygosity (Murphy *et al.* 2008). Moreover, in systems containing high effective population size (N_e) and measured with highly polymorphic SMM-like loci (e.g., microsatellites) genetic distances based on allele frequency distributions are expected to show more

pronounced differentiation than those based on reduction of heterozygosity. Therefore, in addition to estimating genetic distance based on reduction of heterozygosity (F_{ST} or R_{ST}) I obtained estimates based on the proportion of shared alleles (Dps', herafter Dps; Bowcock *et al.* 1994) in the program MICROSAT v 1.5b (Minch *et al.* 1996).

Spatial Genetic Structure

Genetic Clusters

I used two Bayesian model-based approaches to estimate the number of genetic clusters, *K*, and to assign individuals to these clusters. First, I used the program STRUCTURE. The algorithm in STRUCTURE probabilistically assigns individuals to groups ('clusters') that maximizes within cluster HWE and minimizes among-cluster HWE. It is possible that further substructure can be identified by subsequent STRUCTURE analyses within clusters which may provide insight into the degree of admixture within larger clusters. My STRUCTURE analysis method was similar to Degner *et al.* (2010). Within each cluster identified in STRUCTURE I repeated the algorithm until no further substructure was supported. At the largest level (all 675 individuals representing 20 collecting localities) I performed a short pilot run in STRUCTURE v 2.3.1 (Pritchard *et al.* 2000) for each K = 1-20. Likelihood values for each *K* increased to a point then decreased noticeably after around K = 10 (data not shown). Therefore, I performed 10 independent runs for each K = 1-10 using the admixture model with correlated allele frequencies among subpopulations and allowed the degree of admixture, α, to be inferred from the data. I collected data for 5 x 10^5 iterations (allowing the first 2 x 10^5 iteration to be discarded as burnin). All other parameters were set to their default values. I

inferred the number of true clusters using the Δ*K* criterion (Evanno *et al.* 2005). Because subsequent STRUCTURE runs had less genetic content (i.e., fewer individuals) I included location information to the model. This modification placed a higher prior weight on clustering outcomes when correlated with locality information while still being robust to false detection of genetic structure were none exist (Hubisz *et al.* 2009). All remaining parameters for the higher order STRUCTURE runs were the same as full dataset.

Second, I used the R package Geneland v 3.1.4 (Guillot *et al.* 2005a; Guillot *et al.* 2005b; Guillot 2008; Guillot *et al.* 2008) to corroborate the STRUCTURE results and to obtain estimates of population membership in a geographic context. Geneland is useful in identifying general areas of high landscape resistance or discrete boundaries (e.g., major rivers) where gene flow is reduced. As in the STRUCTURE runs, I hierarchically analyzed the genetic data in Geneland to identify multiple levels of genetic partitioning. In Geneland, for the largest level of hierarchy, I used the spatial model and assumed uncorrelated allele frequencies between subpopulations to estimate genetic clusters. I allowed the number of HWLE populations to be an unknown parameter and allowed for joint updates of population membership and allele frequencies (Guillot 2008). As above, I considered the minimum and maximum number of clusters, *K*, to range from 1 to 10. The maximum rate of the Poisson process was set to the number of individuals (n = 676); the maximum number of nuclei in the free Voronoi tessellation was set to three times the number of individuals as recommended by the program's user manual. The number of MCMC iterations was 3×10^5 (recording every 50 iterations; post process burnin = 2000 saved iterations). All subsequent Geneland runs were performed with similar parameters, adjusting the number of individuals and maximum number of clusters for each level accordingly. Unlike other genetic assignment tests to date, Geneland is unique in that it can explicitly account for the presence of null alleles. Therefore, for my Geneland analyses

I included all nine microsatellite loci; for any locus-population combination that showed evidence of high null allele frequencies (see Results; null allele frequency > 0.09) I attempted to filter out null alleles (i.e., set *filter.NA*=TRUE). For each hierarchical level, I preformed 10 independent runs using the above parameters and used the mean posterior density to choose the best run given the data.

Isolation by Distance

To test for evidence of global IBD in my dataset I performed a Mantel test (Mantel 1967) with 9,999 randomizations and used a reduced major axis (RMA) regression to estimate slope and intercept of IBD. These two analyses were performed in the program IBDWS *v* 1.3 (Jensen *et al.* 2005). A significant Mantel test determines whether there is a statistical dependence between (linearized) genetic distance and (log-transformed) geographic distance. A significant Mantel test may be indicative of IBD, of genetic clustering resulting from some dispersal barrier in otherwise panmictic subpopulations (Fontaine *et al.* 2007; Guillot *et al.* 2009), or both.

To test whether any pattern of IBD is merely a by-product of genetic clustering I chose to perform a series of Mantel tests based on the results of genetic clustering algorithms. That is, within genetic clusters identified by STRUCTURE and Geneland I performed subsequent Mantel tests for genetic (heterozygosity-based and D_{ps}) versus geographic distance.

I predicted that below the scale of detectable IBD, population connectivity would be driven largely by landscape composition. Therefore, in addition to performing Mantel test within genetic clusters I performed a spatial autocorrelation analysis using Mantel's *r* at different distance classes to determine the scale at which populations are genetically more similar to one another than at random (Sokal & Oden 1978a; Sokal & Oden 1978b; Soares *et al.* 2008). I used a Mantel correlogram to estimate the 'local' scale (i.e., genetic patch size; Soares *et al.* 2008). The analysis correlates a matrix

of genetic distance with a binary matrix representing pairwise observations of genetic distance at a given distance class. For this analysis I only considered genetic distances in the Eastern cluster (see below). At each distance class (n = 8; 50 km increments) I tested the null hypothesis of absence of spatial pattern. I corrected for multiple comparisons using a Bonferroni corrected P-value ($\alpha' = 0.05/8 = 0.00625$). The size of the genetic patch was estimated by the intercept of the line connecting Mantel's statistic for each distance class and the expectation under the null hypothesis (Soares *et al.* 2008). The above analyses were carried out for both types of genetic distance (heterozygosity-based and D_{ps}) using 9,999 permutations in the R package ecodist *v* 1.2.2 (Goslee & Urban 2007).

Habitat Permeability

I took a landscape genetics approach to determine the relative importance of spatial distance and to identify which habitat types may have contributed to genetic structuring in this system. Here my methodology was adapted from Murphy *et al.* (2010). This approach can be broken down into four general steps: (1) combine genetic and landscape data, (2) run full RF model and calculate model improvement ratio (MIR) for each variable, (3) perform model selection algorithm, (4) run final RF for chosen sub-model to obtain final variable importance, predicted response, and overall model significance.

Combining Landscape and Genetic Datasets

First, I constructed a network of pairwise combinations among the 20 populations. Because I assumed that the Gulf of Mexico is an absolute barrier to dispersal, I removed all vectors that

overlapped this area. The remaining 128 vectors in the network served as the basis for inference. Next I combined pairwise estimates of genetic differentiation for these 128 vectors with their corresponding pairwise spatial distance (km).

Landscape data, derived from the SEGAP dataset, consisted of 30 m^2 raster cells of habitat types. The number of cells for each pairwise combination in the network depends on (1) the spatial distance between two populations and (2) the buffer width of the inter-population vector. To help alleviate confounding measures of habitat type with these two factors I converted the number of raster cells for a measured habitat type to a percentage of the total number of cells. Because reliable estimates of dispersal distances for *H. squirella* are lacking I performed the RF methodology described below for three pairwise buffer widths (diameter = 500 m, 2 km, and 10 km). Separate RF analyses at these vector buffer widths allowed me to determine the best set of predictors for each width separating populations.

Running the Full RF Model and Calculating MIR Values

I used the R package randomForest *v* 4.5-28 (Liaw & Wiener 2002) to run RF with all predictors (i.e., the 'full' model) in regression mode with 5000 trees. Measures of the importance *I* of each predictor *p* (I_p) are generated automatically in RF and were converted to model improvement ratios (MIRs) by dividing each I_p by the maximum I_p (MIR = I_p / I_{max}).

Model Selection Algorithm

My model selection criteria differed slightly from Murphy *et al.* (2010) in two ways. First, I created sub-models via iteratively removing each predictor starting with the lowest MIR, until only the predictor with the largest MIR was retained. Second, because model fit (in terms of pseudo R-squared; hereafter pR^2) differed somewhat for each RF run for a given sub-model I ran 30 forests for

each of these sub-models and obtained 95% confidence intervals around their means. I considered two sub-models to have significantly different $p\mathrm{R}^2$ values if they had non-overlapping confidence intervals. I selected the sub-model with the fewest retained predictors whose mean was not significantly different from that of the best fitting sub-model.

Final RF and Inference

For the chosen sub-model, I determined overall direction of each predictor while averaging out the effects of other predictors using partial dependence plots (Cutler *et al.* 2007; Hastie *et al.* 2009). Finally, significance was estimated by randomizing the response of the chosen sub-model (i.e., genetic distance) 9,999 times, obtaining model-fit ($p\mathrm{R}^2$) for each iteration, and estimating the tail probability of the Monte Carlo null distribution ($\alpha = 0.05$) as in Murphy *et al.* (2010).

Hierarchical Landscape Genetic Structure

The above landscape genetics analysis examined which habitats correlated with genetic distance at the largest hierarchical scale (i.e. for all 20 populations). Populations within distinct genetic clusters should be more similar to each other than populations between clusters. Consequently, genetic distances should be higher for population pairs that occur between genetic clusters. Not surprisingly when landscape features co-vary with regions characterized by this level of genetic structure they may be associated with a high variable importance. To test whether features identified at the highest level were also important among populations (1) at the intermediate scale (i.e., within genetic clusters) and (2) at the local scale (i.e., within a genetic patch), I also performed the above RF analyses among populations within these genetic subsets.

CHAPTER 3: RESULTS

Molecular Analysis

I genotyped 675 *H. squirella* from 20 localities at all nine loci (Table 1). MICRO-CHECKER identified two loci, *hsq131* and *hsq136*, that consistently showed evidence of high frequency null alleles. I removed these two loci from all subsequent analyses unless their presence could be accounted for (e.g., in Geneland). After applying a sequential Bonferroni correction for the remaining seven loci for all populations, eight comparisons remained significantly out of HWE. No populations had more than one locus out of HWE with the exception of WAPP, for which two loci (*hsq103* and *hsq107*) showed deviation from expectations. Additionally, no loci showed evidence of linkage at the 5% nominal level. Therefore, the remaining seven loci were used in all remaining analyses, although some loci in some populations may have contained low-frequency null alleles. My dataset consisted of a large range of expected heterozygosities (0.00 – 0.95; mean = 0.64 $\pm$ 0.33 standard deviation [SD]). Allelic richness per locus, rarified to 14 diploid individuals, ranged from 1 to 15.2. Despite the large differences in these extreme values, the mean allelic richness (average $\pm$SD across loci) for each population was similar throughout the study domain (7.2 $\pm$ 0.8). The allele size permutation test indicated that the observed global R_{ST} value was significantly larger (P-value < 0.001) than permuted R_{ST} (pR_{ST}) based on allelic state, suggesting that mutations have contributed to the observed genetic differentiation. Therefore, I report (log-transformed) R_{ST} for heterozygosity-based estimates of genetic differentiation hereafter. Global genetic differentiation (R_{ST} = 0.055) was

significantly greater than zero (95% CI = 0.042 – 0.067) and suggested moderate genetic structuring across all populations.

Spatial Genetic Structure

Genetic Clustering

At the highest hierarchical level (i.e., the 'global' scale) both STRUCTURE and Geneland identified two genetic clusters (Figure 1, 2, and 3). Overall, population assignments at this level were congruent between these programs and displayed 'Eastern' and 'Western' genetic clusters. Isoclines in Figure 4 show the posterior probabilities of genetic cluster membership for the Eastern cluster. Qualitatively, the area of inflection between these clusters occurs in Florida's panhandle north of the Apalachee Bay.

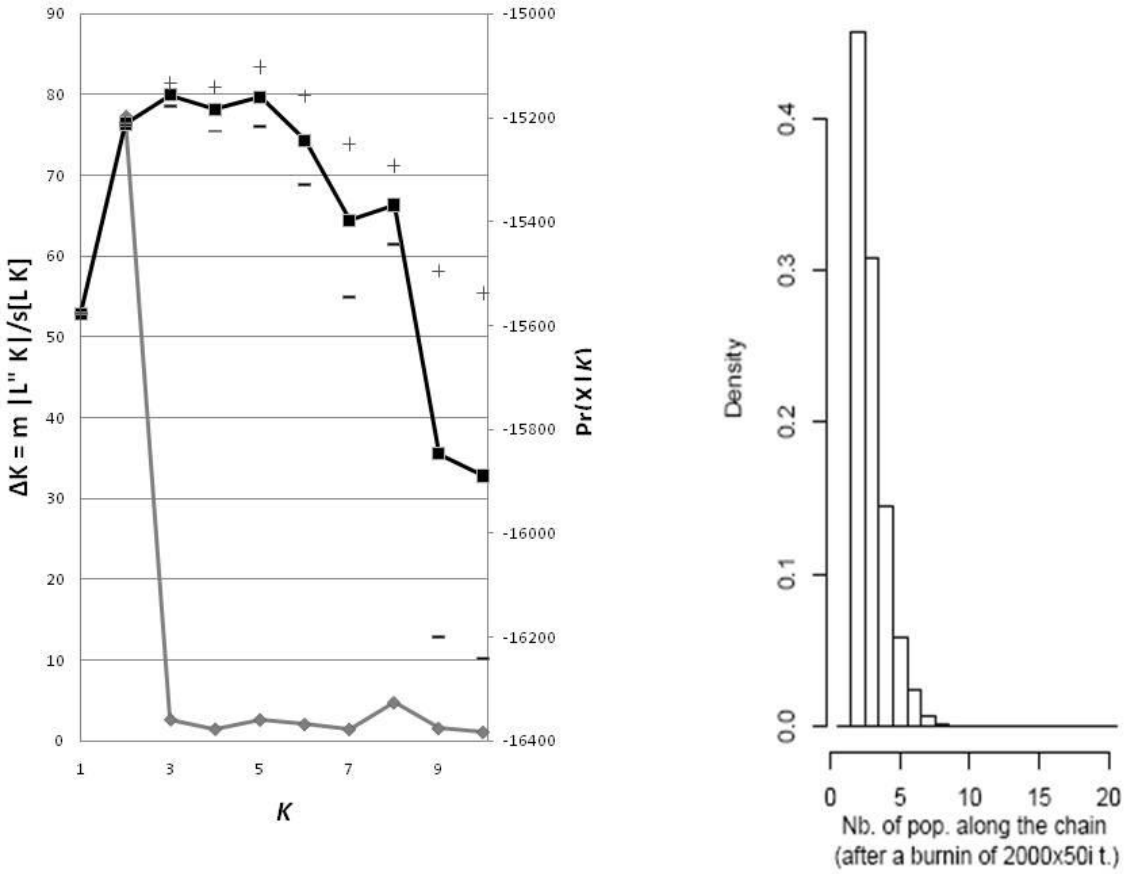

Figure 1: Number of population genetic clusters (*K*) identified using two assignment algorithms at the global scale (n = 20 populations). Left panel gives the likelihood of the data for a given *K* (black line; right hand Y-axis) plus (+) and minus (-) one-half of the SD across 10 independent runs. Output of the Evanno criterion (see text) is also provided on the left panel (gray line; left hand Y-axis. Right panel gives the relative density of the number of populations along the MCMC chain following burnin using the Geneland algorithm.

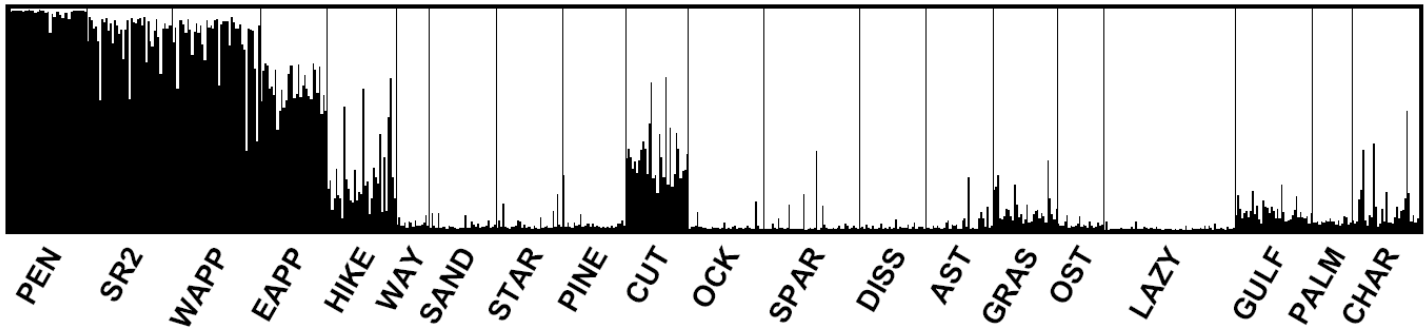

Figure 2: Individual assignments at the global scale based on the STRUCTURE algorithm for *K*=2 genetic clusters. Each individual is shown as a column. The proportion of each individual's genome that originated in *K* clusters is shown. Population abbreviations are provided and cross-reference to Table 1. Different colors represent different clusters (*K*), and the length of columns represents the proportion of each individual's genome that originated from the color-coded *K*. For instance, PEN (far left) contains individuals whose genomes are mostly derived from the black (Western) cluster.

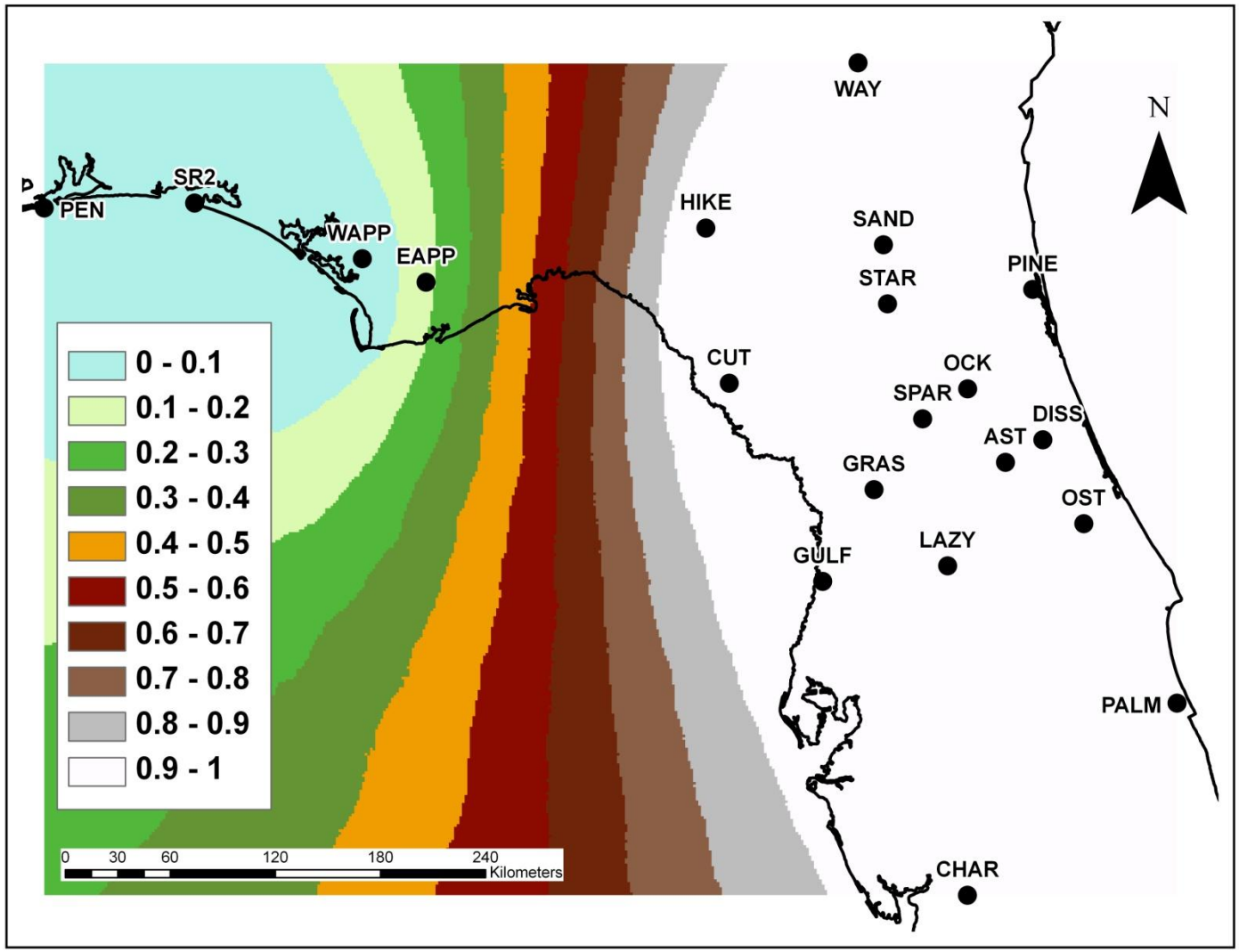

Figure 3: Geneland output of Posterior Probability (see legend) of Eastern cluster membership at the global scale (n = 20 populations). Population abbreviations cross-reference with Table 1. Posterior probability of the Western cluster is defined as 1 minus the Posterior Probability of the Eastern Cluster.

The Western cluster consisted of a relatively small number of sampling localities (n=4). Subsequent hierarchical STRUCTURE analyses considering only these four localities showed evidence of further substructure in this region. Here, STRUCTURE discriminated among all four localities (Figure 4, 5, and 6) whereas Geneland only supported a split between the westernmost locality (PEN) and the remaining three localities (SR2, WAPP, and EAPP; Figure 4 and 5). While assignment tests were able to identify genetic structure within the Western cluster, the relatively low number of localities made it impractical to obtain reliable estimates of landscape genetic patterns. Therefore, the Western region was not utilized in any following landscape genetic structure analyses (see below). In contrast, the Eastern cluster consisted of 16 collecting localities ranging throughout Florida's peninsula to southern Georgia (Figure 3). Despite a considerable spatial distance between the furthest spanning populations (*ca.* 480 km), no further substructure within the Eastern cluster was supported by either assignment test algorithm (Figure 7).

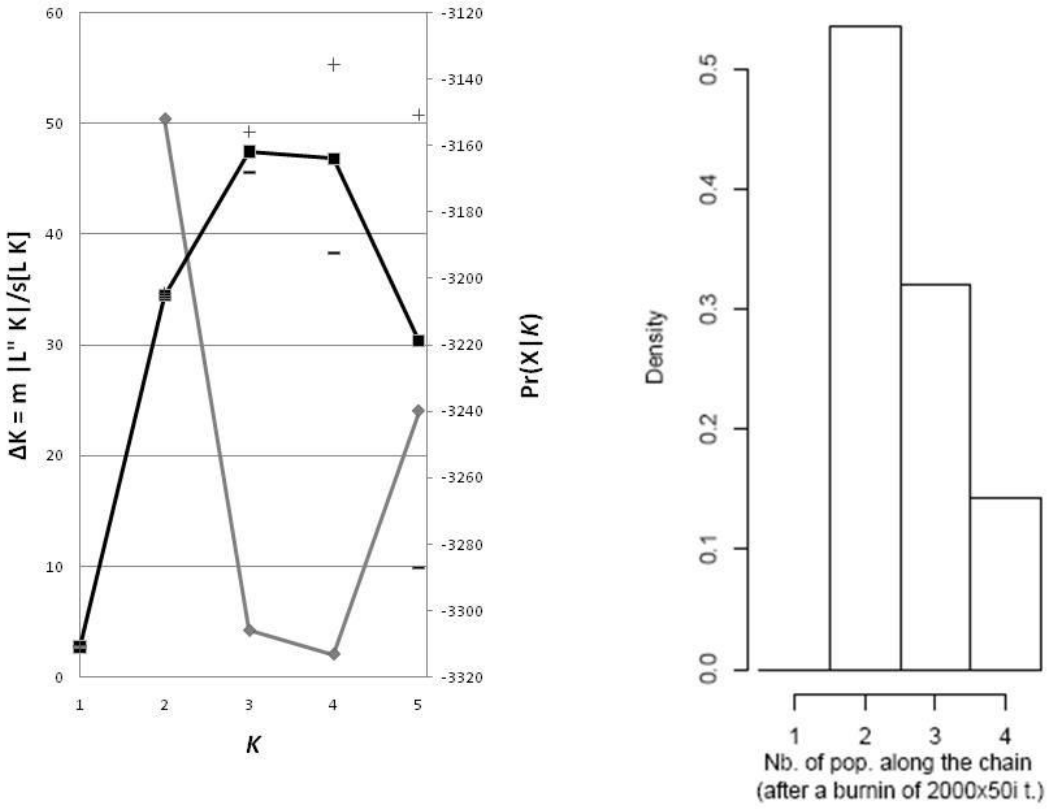

Figure 4: Second-order STRUCTURE analysis showing the number of population genetic clusters (*K*) identified using two assignment algorithms in the Western cluster (n = 4 populations). Left panel gives the likelihood of the data for a given *K* (black line; right hand Y-axis) plus (+) and minus (-) one-half of the SD across 10 independent runs. Output of the Evanno criterion (see text) is also provided on the left panel (gray line; left hand Y-axis. Right panel gives the relative density of the number of populations along the MCMC chain following burnin using the Geneland algorithm.

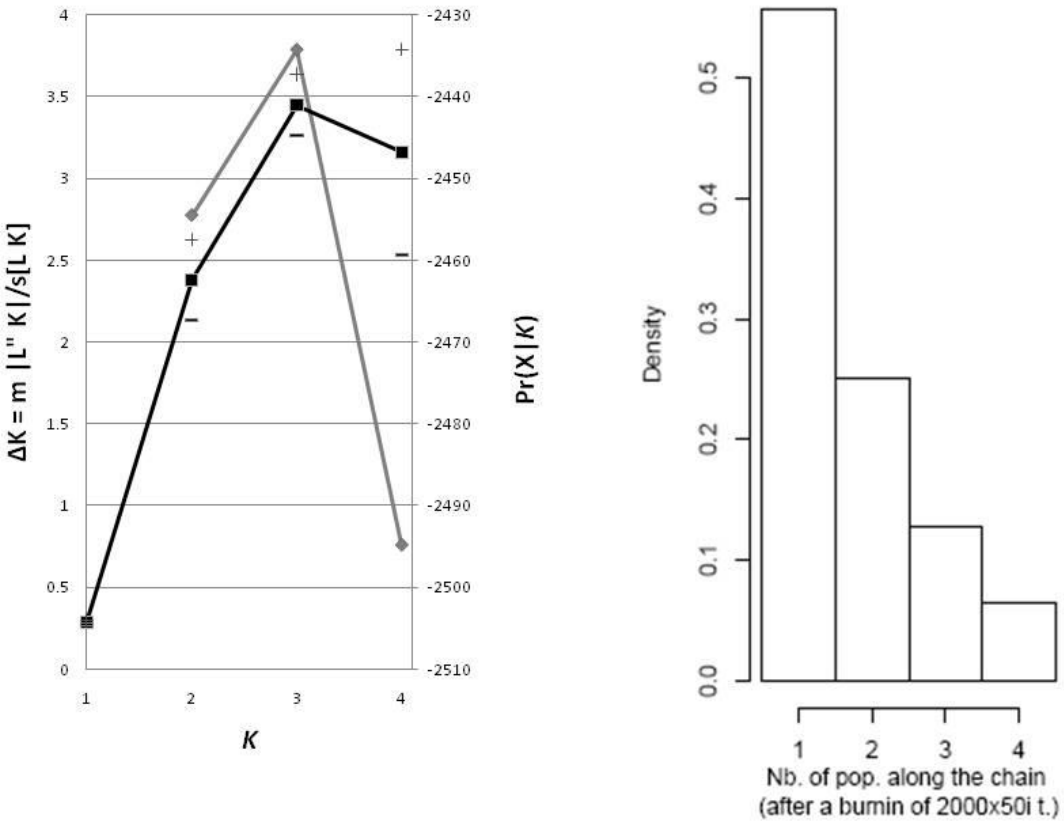

Figure 5: Third-order STRUCTURE analysis showing the number of population genetic clusters (*K*) identified using two assignment algorithms in the Western cluster but excluding PEN. Left panel gives the likelihood of the data for a given *K* (black line; right hand Y-axis) plus (+) and minus (-) one-half of the SD across 10 independent runs; Output of the Evanno criterion (see text) is also provided on the left panel (gray line; left hand Y-axis. Right panel gives the relative density of the number of populations along the MCMC chain following burnin using the Geneland algorithm.

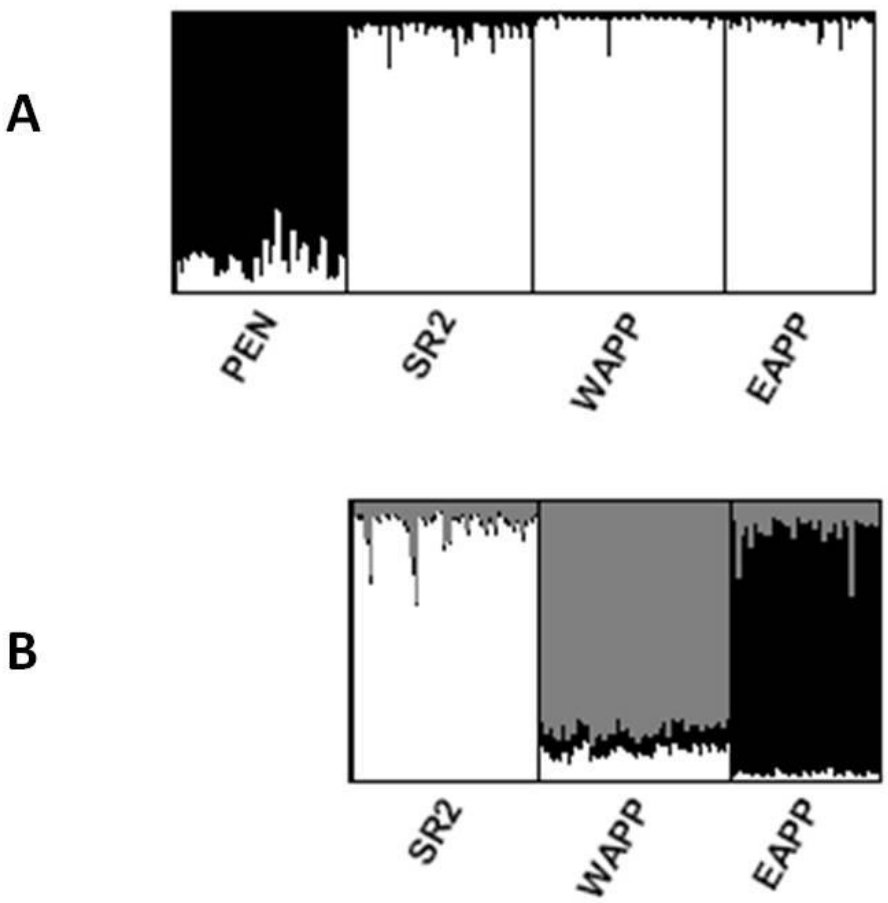

Figure 6: Individual assignments within Western cluster based on the STRUCTURE algorithm. Each individual is shown as a column. The proportion of each individual's genome that originated in *K* clusters is shown. Population abbreviations are provided and cross-reference to Table 1. A: all individuals within the Western cluster ($K = 2$; n = 152 individuals); B: Same as (A) but excluding PEN ($K = 3$; n = 115 individuals).

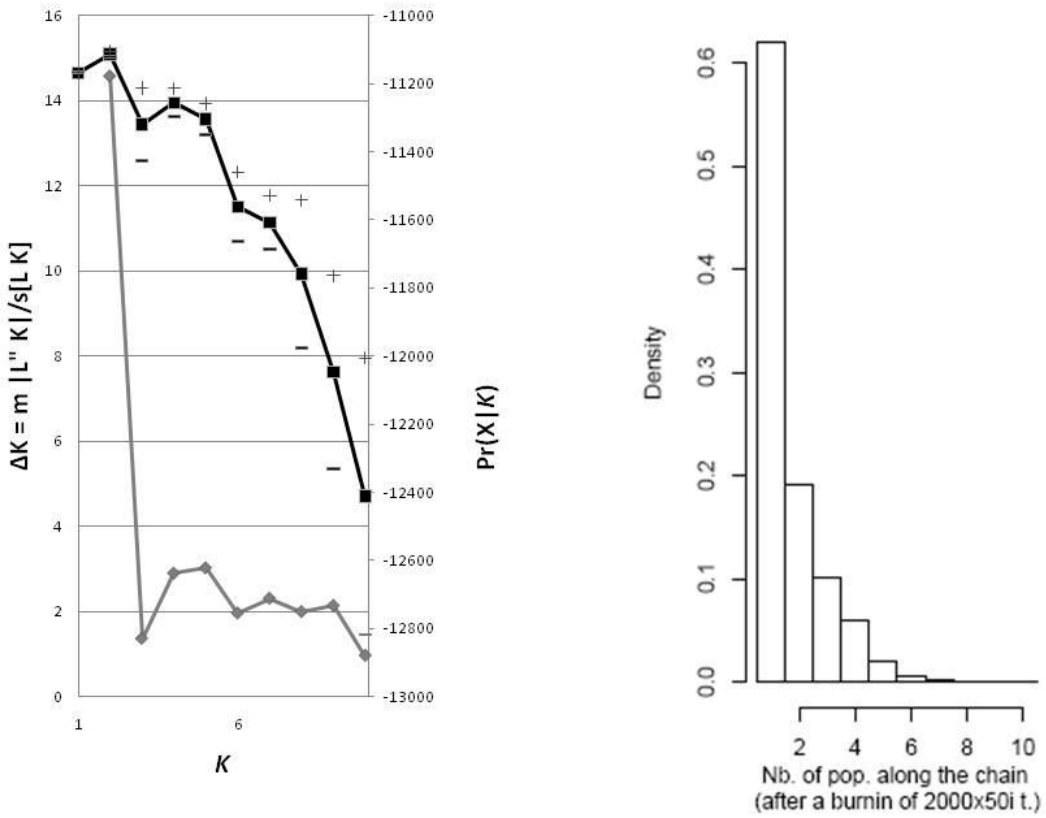

Figure 7: Number of population genetic clusters (*K*) identified using two assignment algorithms within the Eastern Cluster. Left panel gives the likelihood of the data for a given *K* (black line; right hand Y-axis) plus (+) and minus (-) one-half of the SD across 10 independent runs; Output of the Evanno criterion (see text) is also provided on the left panel (gray line; left hand Y-axis. Right panel gives the relative density of the number of populations along the MCMC chain following burnin using the Geneland algorithm.

Isolation by Distance

I tested for IBD in my dataset hierarchically based on the genetic clusters identified in STRUCTURE and Geneland. First, I tested for IBD at the 'global' scale which consisted of all pairwise combinations among 20 populations. Second, I tested for IBD at the 'intermediate' scale; this consisted of all pairwise combinations within the Eastern cluster. For both the global and the intermediate scales the Mantel tests showed evidence of IBD. These tests where significant for both log transformed R_{ST} (Table 3; Figure 8) and log transformed D_{ps} values (Table 4; Figure 8). However, genetic distance based on D_{ps} values produced overall better model fit compared to R_{ST} values.

Genetic Patch Size

Figure 9 shows the Mantel Correlograms. For R_{ST}-based estimates of genetic distance (Figure 9a), no comparison was significantly different from zero for any distance class after Bonferroni correction. Therefore, no estimate of local scale could be estimated. However, for D_{ps}-based estimates (Figure 9b) a spatial gradient is observed with significantly positive autocorrelation at shorter distance classes and significantly negative a larger distance classes. The x-intercept occurs near 150-200 km. However, for distance class 3 (i.e., populations 100 - 150 km apart) the 95% confidence intervals overlap with the Mantel expectation under the null hypothesis. I therefore chose all populations in the Eastern cluster below 150 km apart as my conservative estimate of the genetic patch size.

Table 3: Summary Statistics for Mantel tests at the global and intermediate scales for log-transformed R_{ST} values on log-transformed spatial distance (km). P-value based on 9,999 randomizations.

Scale	R^2	Slope	95% CI	P-value
Global	31.3	0.2041	0.1797—0.2284	< 0.0001
Intermediate	9.3	0.0986	0.0814—0.1157	0.0153

Table 4: Summary Statistics for Mantel tests at the global and intermediate scales for log-transformed D_{ps} values on log-transformed spatial distance (km). P-value based on 9,999 randomizations.

Scale	R^2	Slope	95% CI	P-value
Global	60.2	0.4099	0.3727—0.4471	< 0.0001
Intermediate	40.2	0.3652	0.2969—0.4335	< 0.0001

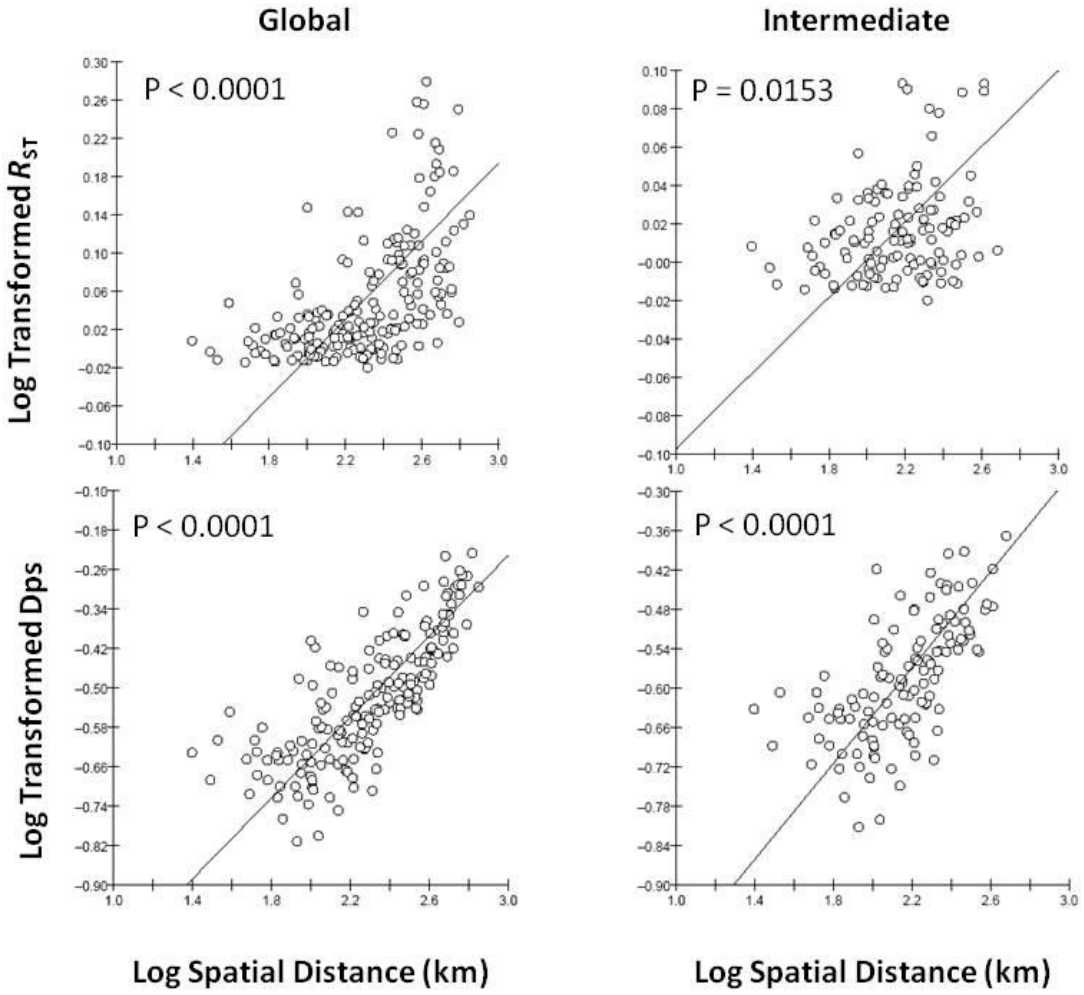

Figure 8: Genetic distance as a function of geographic distance at the global scale (left plots) and intermediate scale (i.e., within the Eastern cluster; right plots) for two response types: genetic distance based on reduction of heterozygosity (top plots) and based on allele frequency distribution (bottom plots). Significance is based on Mantel tests with 9,999 randomizations using the program IBDWS.

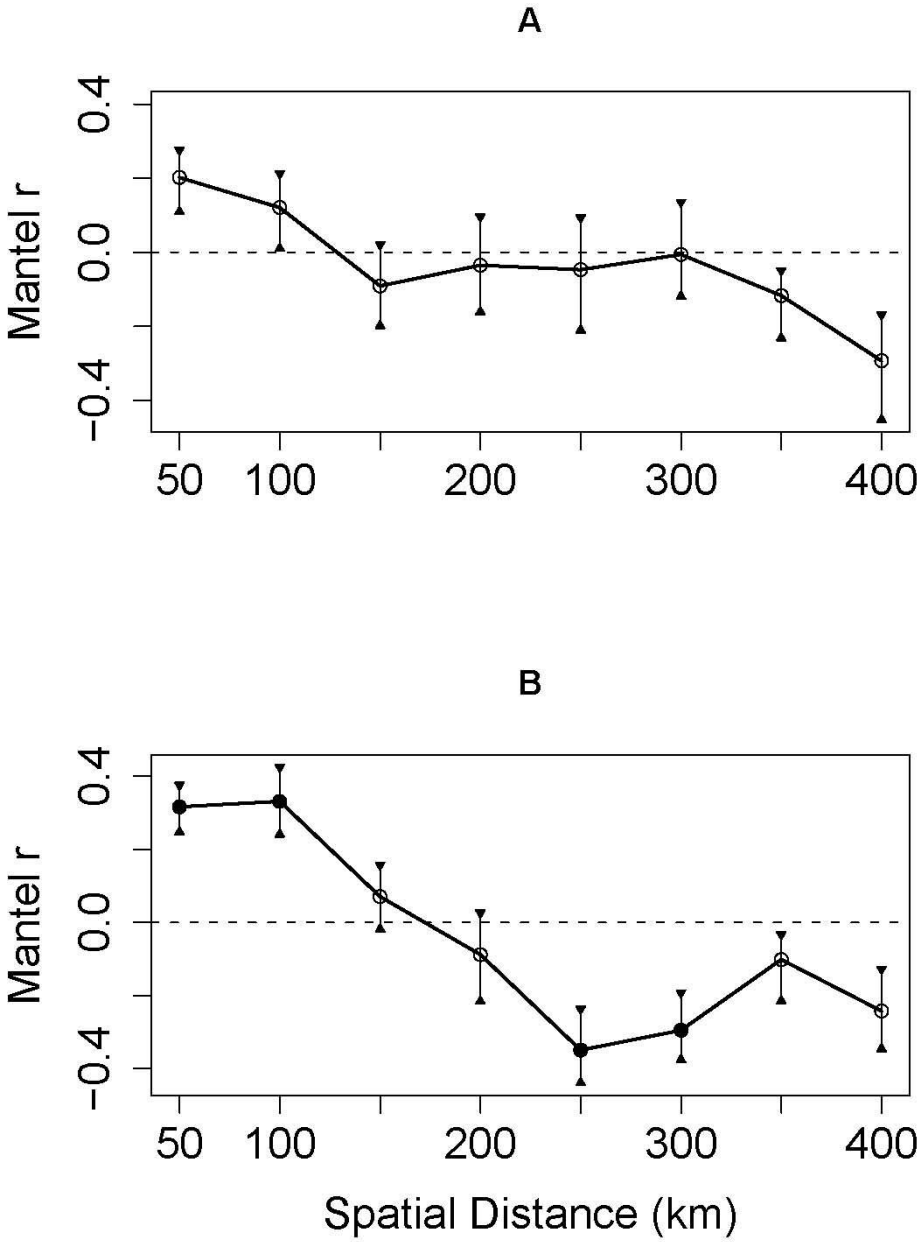

Figure 9: Mantel correlogram within the Eastern cluster for log-transformed R_{ST} (A) and D_{ps} values (B). Mantel r is similar to Pearson's product-moment coefficient and ranges from -1 to 1. For each distance class (i.e., 50 km interval) the Mantel $r \pm$ 95% confidence interval is shown. Filled circles denote significance at the Bonferroni corrected level $\alpha' = 0.05/8 = 0.00625$.

Landscape Genetics at the Global Scale

For a given response type (D_{ps} or R_{ST}), models with different vector buffer widths had similar prediction error (MSE), overlapping $p\mathrm{R}^2$ confidence intervals, and similar predictors retained. Thus, I report only the smallest vector buffer width (see Appendix B for detailed summary of 2 km and 10 km models).

Landscape genetic analyses revealed a strong pattern of model improvement when habitat variables were incorporated into models of IBD. Table 5 presents the median $p\mathrm{R}^2$ ($\pm$ 95% confidence intervals), mean squared error (MSE), and model significance for the chosen models. Models including habitat types consistently explained more of the variation in genetic distance ($p\mathrm{R}^2$) than spatial distance alone (Table 5). As with the Mantel tests, genetic distance based on allele frequency distribution (D_{ps}) yielded improved model fit ($p\mathrm{R}^2$) compared to models with genetic distance based on reduction of heterozygosity (R_{ST}).

At the global scale, RF uncovered an increase in genetic distance with increasing spatial distance for both R_{ST}-based and D_{ps}-based measures of genetic distance. Figures 10 and 11 show the partial dependence plots of the individual variables included in the models for R_{ST} and D_{ps}. These plots show the effect of a given predictor on the model after accounting for the average effects of the other retained predictors. From these plots and MIR values it is apparent that the effect of spatial distance (*km*) is large compared to the next most important predictor at the global scale, regardless of response type. For instance, percent oak hammock was the most important habitat type for the D_{ps}-based model but was only about 54% as relevant as spatial distance. Both models show a decrease in genetic distance with increasing percent oak hammock among localities, and a

complex association for *urban* (Table 5, Figure 10 and 11). However, not all predictor variables were similar for the two genetic distance models. R_{ST}-based RF analyses show a strong predictive function for R_{ST} and percent cover of mesic flatwoods (*flat*). In contrast, *flat* is not a variable that significantly explains variation of D_{ps}. Similarly, *pas* appear to have different effects depending on which genetic distance measure is used (Table 5).

Table 5. Features associated with genetic connectivity among *Hyla squirella* populations using Random Forest. Presented here are the chosen predictors for following model selection for the 500 m vector buffer width (detailed summary statistics for all vector buffer widths can be found in Appendix B). Models are grouped first by hierarchical scale (see text), next by response (R_{ST} or D_{ps}), and finally by type. pR^2 is a pseudo R^2; MSE denotes mean squared error. Summary statistics, based on constructing 30 forests for each sub-model (see Methods), include median and 95% confidence intervals (95% CI) of pR^2 and median MSE. P-value of the chosen sub-model is provided (see text). Model denotes the chosen variables (names cross-reference with Table 1) following model selection. These variables are ordered starting with the most important variable (in terms of MIR values) to the least important. Font style denotes general trend identified using partial dependence plots: standard font, negative association with genetic differentiation among populations; Boldface, positive association with genetic differentiation among populations; italic, complex association (i.e., nonlinear association whereby genetic differentiation has a minimum at some value of percent cover); and underline; weak main effect.

Scale **Response** **Type**	**Median pR^2**	**95% CI**	**Median MSE**	**P-value**	**Model**
Global					
R_{ST}					
km only	0.753	0.333—1.174	2.22E-01	<0.001	**km**
landscape genetics	40.48	40.20—40.94	1.42E-03	<0.001	**km**, **flat**, oak, *urban*, *swamp*
D_{ps}					
km only	32.29	32.03—32.53	1.89E-01	<0.001	**km**
landscape genetics	60.38	60.03—61.46	1.11E-03	<0.001	**km**, oak, *urban*, pas
Intermediate					
R_{ST}					
km only	-38.03	-38.48— - 37.59	5.02E-02	0.5289	km
landscape genetics	14.13	13.75—14.38	4.00E-04	<0.001	**pas**, **km**, *urban*, *swamp*, **flat**
D_{ps}					
km only	23.81	23.57—24.06	1.33E-01	<0.001	**km**
landscape genetics	47.78	47.53—48.16	9.12E-04	<0.001	**km**, oak, *urban*
Local					
D_{ps}					
km only	-31.77	-32.43 — -31.11	1.35E-01	0.3724	km
landscape genetics	30.24	29.43—30.50	7.13E-04	<0.001	<u>flat</u> oak, **km**

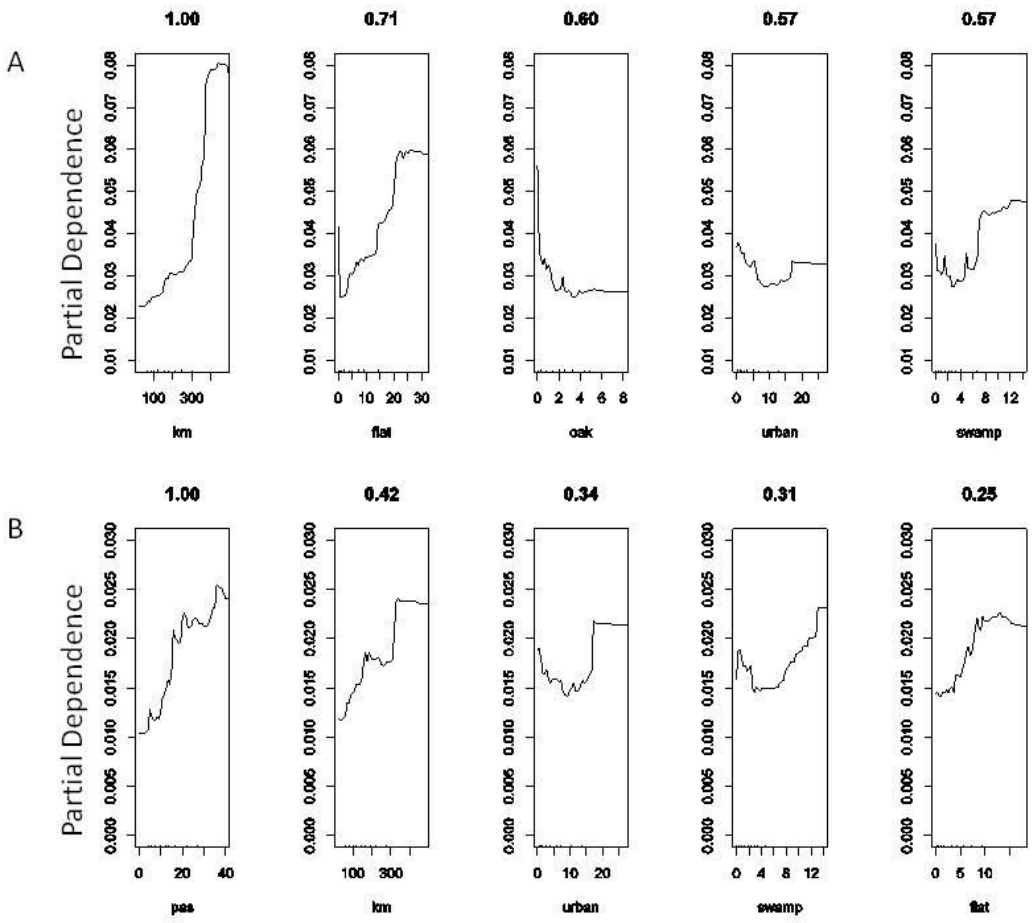

Figure 10: Partial Plots for R_{ST}-based RF models. A) global scale, B) intermediate scale. Values above figures denote Model Improvement Ratios (MIRs) for retained predictors; that is, the importance of a given predictor *relative* to the most important predictor (far left; MIR=1.00). These plots show the predictive function of (log-transformed) R_{ST} on a given predictor while accounting for the average effects of other predictors. For example, R_{ST} has a nonmonotonic partial dependence on *urban*; it decreases nearly linearly throughout the main body of the data (denoted by rug) until approximately 10% cover before it increases.

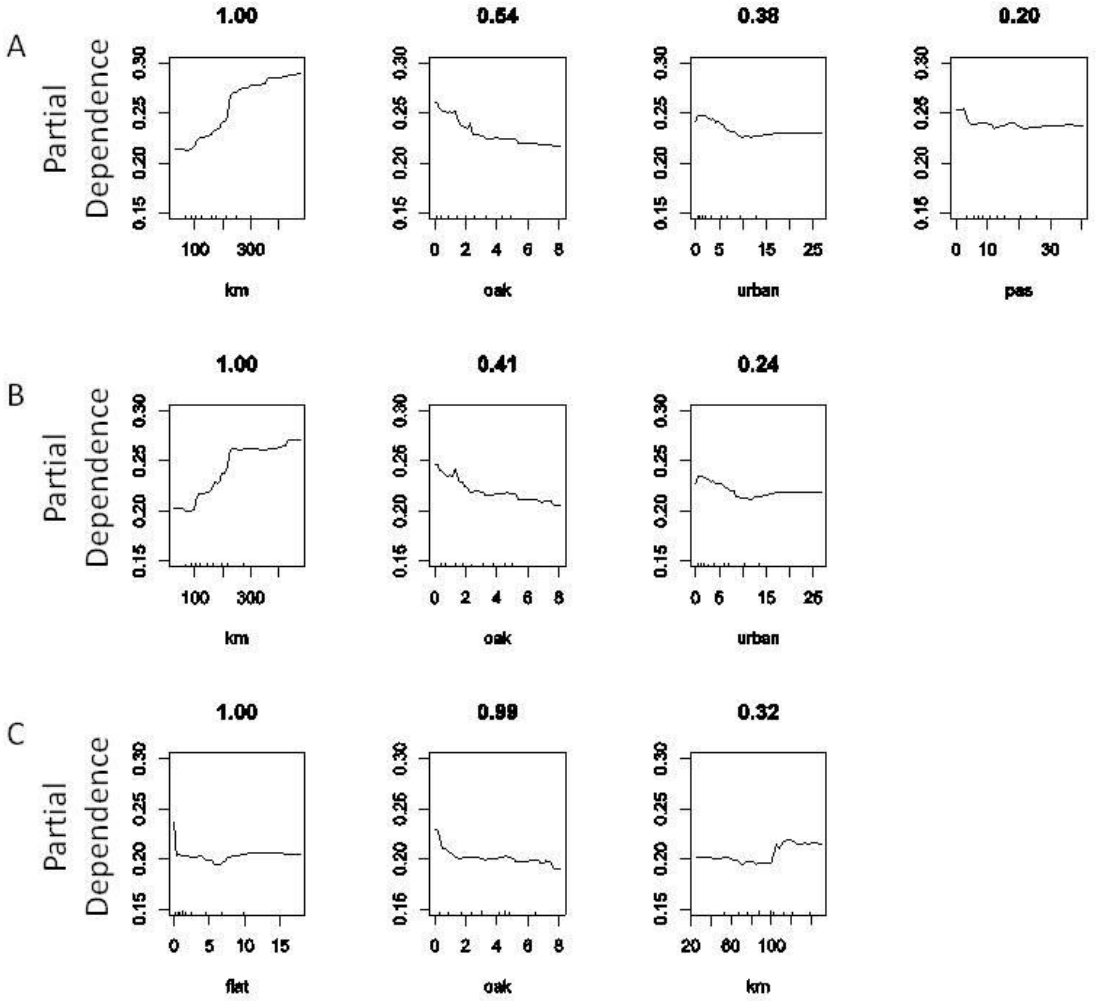

Figure 11: Partial Plots for D_{ps}-based RF models. A) global scale, B) intermediate scale, C) local scale. Values above figures denote Model Improvement Ratios (MIRs) for retained predictors; that is, the importance of a given predictor *relative* to the most important predictor (far left; MIR=1.00). These plots show the predictive function of (log-transformed) D_{ps} on a given predictor while accounting for the average effects of other predictors.

Hierarchical Landscape Genetic Structure

Landscape Genetic Structure at the Intermediate Scale

For this analysis, I included only samples in the Eastern cluster and re-ran the RF algorithm. The effect of spatial distance (*km*) was relatively unimportant for R_{ST}-based analyses. In this model, pasture (*pas*) was the best predictor and showed a positive association with R_{ST}; *urban*, *flat* and *swamp* were also retained. The RF analyses based on D_{ps} showed a large improvement over R_{ST}-based analyses (Table 5). For example, the model fit for D_{ps} is over 3 times that of R_{ST} (Table 5). This difference is largely due to the strong effect of IBD still present for D_{ps}-based analyses (Figure 10; Appendix B). Habitats retained in the Dps-based model at this scale were *oak* and *urban* and generally showed a positive association with genetic connectivity (Figure 11b).

Landscape Genetic Structure at the Local Scale

Spatial distance alone was not significantly associated with D_{ps} at the local scale (RF P-value = 0.372; Table 2). However, the inclusion of habitat types explained approximately 30% of the variation in pairwise D_{ps} (RF P-value <0.001; Table 2). Here RF identified *oak* and *flat* as the two most important variables associated with genetic distance. Both of these variables had nearly identical MIR values (1.000 and 0.989, respectively). Upland oak hammock had a negative association with genetic distance and *flat* displayed a weak main effect (Figure 11).

CHAPTER 4: DISCUSSION

Landscape Genetics and Habitat Generalists

The degree of a species' habitat specialization can have profound effects on its landscape connectivity (With & Crist 1995; With *et al.* 1997) and population dynamics (Krauss *et al.* 2003). Landscape genetics studies that treat habitat suitability as a binary variable, suitable or non-suitable, may be appropriate for habitat specialists but overly simplistic for generalist species that use habitats with variable suitability (Stewart *et al.* 2010). My data support the hypothesis that rates of gene flow for a generalist species can depend on a spectrum of habitat suitability. For example, my landscape genetics models suggest that some habitats are highly relevant (e.g., *oak*) while others (e.g., *sil*) appeared not to influence, or have not had adequate time to influence, genetic connectivity at the considered scale.

The conclusions of my study support those of recent studies that investigated how landscapes influence genetic structure for generalist species. Vandergast *et al.* (2004) compared genetic structure for three spider species of the genus *Tetragnatha*, two forest specialists and one generalist, within and among fragmented forests of the Island of Hawaii. The matrix separating remnant forests consisted mainly of a recent (< 200 year old) lava flow. Data based on mtDNA and allozymes showed that restricted habitat has resulted in genetic structure for the two habitat specialists whereas no evidence of global genetic differentiation or IBD was present in the generalist. Similarly, Brouat *et al.* (2003) assessed levels of IBD at fine spatial scales (up to 13.6 km) for two carabid species (again, one was a forest specialist and one was a generalist). Evidence of IBD based

on microsatellite data was apparent for the specialist species but the generalist species only displayed a weak pattern of IBD. My study is similar to these studies in that the generalist *Hyla squirella* can be found in a wide range of habitats throughout its range (and has a pattern of IBD as in the carabid study). However, I quantified the role of the landscape composition. First, I determined whether habitat types are correlated with an increase in gene flow (e.g. upland oak hammocks) or have no measured effect (e.g., silviculture or river floodplain forests). Second, my landscape genetics models discerned among the effects of superficially similar habitats by partitioning the different effects of different forest types (e.g., upland oak hammocks, mesic flatwoods, floodplain forests, and silviculture). Therefore, the landscape genetics approach utilized here identified evolutionary patterns of dispersal and identified ecological habitats relevant to organism dispersal.

This study further expands our knowledge of how spatial scale influences genetic patterns. While others landscape genetics studies (e.g., Murphy *et al.* 2010) found that variation in genetic distance was better explained at fine scales (i.e., within genetic clusters) my data suggest that variation in genetic distance was better explained at broader scales (e.g., D_{ps} models pR^2 = 60.38, 47.78, 30.24 for global, intermediate, and local, respectively). This discrepancy may be due to the large effect of spatial distance at broad scales in the current study. Perhaps not surprisingly in systems characterized by IBD, a reduced spatial extent (i.e., distance) can decrease explanatory effect of spatial distance as well as overall model fit. Regardless of the attenuating explanatory power of spatial distance with reduced scale, landscape genetic models identified here show that the inclusion of habitat types better explained the variation in genetic distance than spatial distance alone (Table 5). This was most clear at the local scale (i.e., within a genetic patch), where habitat types explained substantially more variation in D_{ps} than spatial distance, for which no effect of IBD was present. This result suggests that landscape processes govern genetic structure within genetic patches

whereas IBD—either directly by the taxon's intrinsic dispersal ability or indirect as acting as a surrogate for some unmeasured landscape feature—shapes patterns of genetic structure at larger spatial scales.

This study found that habitats are used differentially and predictably by *H. squirella* and that spatial scale influenced landscape genetic patterns, but this study also revealed three surprising patterns. Specifically: (1) changing the vector buffer width among population pairs did not meaningfully change which habitats were important; (2) models based on D_{ps} outperformed those based on R_{ST}; and (3) *flat* was positively associated with R_{ST}. First, previous studies suggest that the significance of a particular landscape feature on genetic connectivity can be influenced by the vector buffer width analyzed (Emaresi *et al.* 2009; Murphy *et al.* 2010). The current study found that changing the vector buffer width did not meaningfully change (i) which habitats were important and (ii) the order of importance of these habitats (Appendix B). A possible explanation for the lack of difference among models is the strong correlation among vector buffer widths for a given habitat (Appendix C). The two broader vector buffer widths (2 km and 10 km) always correlated strongly with 500 km vector buffer width for any habitat type, suggesting that a narrow (i.e., 500 km) transect adequately sampled the surrounding landscape composition. Second, RF models had a better model fit and lower prediction error when using genetic distances based on allele frequency distribution than when using heterozygosity-based estimates. However, it is important that both frequency-based and heterozygosity-based estimates be modeled because they may differ in strength and pattern (Fig. 10, 11; Table 5). For example, I found that *pas* showed a positive association with R_{ST} but a negative association with D_{ps}. In this case, D_{ps} result is more likely accurate because it is (1) not subject to equilibrium assumptions (Bowcock *et al.* 1994) and (2) simulations suggest that D_{ps} can better detect more recent landscape genetic signature in systems characterized by high effective population size

(Murphy *et al.* 2008) as may be expected for the abundant *H. squirella* (Babbitt & Tanner 2000; Babbitt *et al.* 2006). One general conclusion is that when both estimates converge on the same pattern, then the confidence in that pattern is strengthened.

Third, *flat* was associated with increased genetic differentiation in R_{ST}-models. Two ecological processes that dominate mesic flatwoods are fire and flooding. The high fire return interval (*ca.* 1 to 4 y) in this system maintains an open canopy while the relatively low slope and poor drainage is ideal for generating a mosaic of ephemeral ponds during seasonal rains (NatureServe 2006). Both of these processes should be ideal for breeding site selection. Indeed, Binckley and Resetarits (2007) report *H. squirella* prefer to oviposit in ponds with open canopy cover. Further, like many amphibians, *H. squirella* effectively avoid breeding in ponds with permanent water and/or predatory fish (Babbitt & Tanner 2000). Two hypotheses may explain the positive association between R_{ST} and mesic flatwoods. First, this species may behaviorally change its degree of philopatry with a change in surrounding habitat. Such behaviorally driven mechanisms are not new in the literature. For example, Sacks *et al.* (2008) examined population genetic structure in a wide ranging generalist *Canis latrans*, a species that displays a behavioral preference for dispersing through areas which are similar to their natal habitat. Using microsatellite data Sacks *et al.* (2008) found genetic structure in *C. latrans* concordant with general habitat subdivisions of the heterogeneous California Floristic Province. As mesic flatwoods are abundant in the zone of inflection between Eastern and Western genetic clusters, this hypothesis predicts that populations of *H. squirella* in the region increase site fidelity and philopatry with increasing open canopy forest with temporary ponds. Further, this hypothesis suggests that contemporary habitat may be the causal mechanism for the East-West pattern seen in the assignment tests (Figure 1 and 2) and that regional genetic structure of this generalist species is flexible through time. Alternatively, this pattern may be due to a historical

footprint (Dionne *et al.* 2008) in which potential phylogenetic-level breaks in this region could have caused the observed pattern, and contemporary habitat may simply be a statistical artifact. Thus, the discrepancy shown for mesic flatwoods in this study presents a challenge in landscape genetics; to reconcile the effects of ecological pattern (i.e., contemporary versus historic landscapes) with evolutionary pattern (i.e., recent genetic structure or older divergence). Phylogeographic-level analyses in this species, including populations in more westerly states (e.g., Texas and Louisiana), may shed light on the importance of contemporary habitats for gene flow patterns at broader spatial scales. Further, the fact that D_{ps}-based landscape genetics models did not uncover the same importance and trend of *flat* at the global scale implies that mesic flatwoods may be less important in influencing relatively recent genetic connectivity.

A landscape genetics approach can be used to test our ecological expectations. For example, semi-aquatic amphibians spend some part of their life-history in upland habitats surrounding breeding ponds; however, the use of these habitats by amphibians remains poorly understood as reliable sampling in upland terrestrial environments can be difficult (Dodd & Cade 1998; Semlitsch 1998; Bulger *et al.* 2003; Semlitsch & Bodie 2003; Trenham & Shaffer 2005). Using a landscape genetics approach, I found evidence that upland oak hammocks were strongly correlated with genetic connectivity among *H. squirella* populations. Moreover, this relationship was evident regardless of the spatial scale under consideration (Table 5). This result is consistent with a recent mark-recapture study (Windes 2010) that examined landscape features correlated with *H. squirella* survival and recapture rates. Windes (2010) found that *H. squirella* display strong site fidelity but recapture rates generally decreased with increased size of surrounding upland woodlot area. In addition, woodlot area was positively associated with *H. squirella* survival. Taken together, the mark-

recapture (Windes 2010) and landscape genetics research (this study) underscores the necessity of upland terrestrial environments for many amphibians.

Evolution and Population Connectivity in the Squirrel Treefrog

The overall patterns of spatial genetic structure may provide insight into the evolutionary history and patterns of population connectivity for *H. squirella*. Within the Western cluster the two assignment test algorithms showed evidence of further substructuring (Figure 4, 5, and 6). Both assignment test suggest PEN is genetically differentiated from the other three population (SR2, WAPP, and EAPP); STRUCTURE further suggested differentiation among SR2, WAPP, and EAPP (Figure 6b). Although the relatively few pairwise observations within the Western cluster made it impractical to rigorously quantify the degree to which the landscape may have contributed to the genetic structure, these results suggest two areas of further research. First, samples collected in PEN occurred approximately 100 km away from the next closest sampled population (SR2; Figure 3). The genetic distance between these two populations was 0.147 and 0.393 for R_{ST} and D_{ps}, respectively. Because H_E and A_R were similar throughout the study it is unlikely that the genetic diversity within PEN or SR2 was a result of a recent founding event. One possible explanation for this relatively large pairwise genetic differentiation is the high resistance for *H. squirella* to move through saline habitats (i.e., the Gulf of Mexico or the Santa Rosa Sound). PEN occurs on Santa Rosa Island in the Gulf of Mexico; this island, oriented east-west, is generally less than 1 km wide and surrounded by salt and brackish water on either side. Assuming populations in this region are in equilibrium—such that the loss of alleles due to genetic drift is countered by the gain in alleles due to gene flow— the

physical geography creates a one-dimensional stepping stone model that amplifies divergence for a given distance relative to other models of gene flow (Wright 1943). Phylogenetic studies in this region have identified a similar pattern of genetic differentiation in beach mice (*Peromyscus polionotus leucocephalus* and *P. p. allophrys*; Van Zant & Wooten 2007); however, the seemingly congruent pattern with *P. polionotus* and *H. squirella* should be interpreted with caution. Phylogenetic-level analyses in this region for *H. squirella* will aid in determining if this pattern is a result of shared biogeographic history or pseudo-congruence (Steele & Storfer 2007). Second, WAPP and EAPP were sampled *ca.* 24 km apart and separated by the Apalachicola River. This river has been invoked to explain genetic discontinuities in several terrestrial taxa (reviewed in Soltis *et al.* 2006). Evidence based on assignment tests (Figure 4 and 6b) showed that WAPP and EAPP are genetically differentiated. While my landscape genetics analyses did not identify floodplain forest as a general barrier to *H. squirella* gene flow, the Apalachicola River may serve as a partial barrier to gene flow. Large, wide rivers (*ca.* 50 m) were barriers to gene flow for the generalist European Badger (*Meles meles*) while small rivers showed no effect on patterns of population connectivity (Frantz *et al.* 2010). Additional sampling effort west and east of this river will shed light on whether the Apalachicola River is a barrier between *H. squirella* populations.

Within the eastern cluster, specific habitats used by *H. squirella* can be better evaluated, including the effects of these habitats on population genetic patterns. Landscape genetic patterns found here also may be relevant to other generalist amphibians in the southeastern Coastal Plain. For example, my landscape genetic data suggest that pastures may increase gene flow among populations (Table 5; Figure 11). This correlation is supported in the literature. Many pastures in the Coastal Plain consist of a mosaic of grassy wetlands. Tadpoles of *H. squirella* occur in wetlands surround by pastures up to 1.2 km away from the closest woodland (Babbitt & Tanner 2000)

suggesting movement through pasture matrix is likely. In addition *H. squirella*, and the green treefrog (*H. cinerea*), may be expanding their ranges north into the Piedmont presumably due to the proliferation of farm ponds in Georgia (Jensen *et al.* 2008).

I predicted that floodplain forests (*rff*) would inhibit gene flow in *H. squirella.* Floodplain forests are periodically inundated and can deliver predatory fish to potential breeding ponds which reduces *H. squirella* abundance (Babbitt & Tanner 2000). Moreover, low dispersing species, such as *H. squirella* (Windes 2010), should be affected by a riverine barrier (Zhao *et al.* 2009). However, my landscape genetic study suggests that (small) rivers and floodplain forests are generally uninformative in assessing population connectivity in *H. squirella*. On one hand, it may be possible that the linear geometry of river floodplain forest habitat, as opposed more patchily distributed habitat, may have precluded identification of the importance of rivers and floodplain forest in this system. On the other hand, floodplain forests may not be a dispersal barrier for *H. squirella* if a suitable number of fish-less breeding ponds are available within these habitats.

Finally, amphibian populations and communities can be reduced by ultraviolet radiation and toxic chemicals (Alford & Richards 1999; Blaustein *et al.* 2003), but habitat loss and alteration are probably the most serious causes of the global amphibian decline (Dodd & Smith 2003). The anuran assemblage in urban-rich regions is largely represented by generalist species such as *B. terrestris*, *Osteopilus septentrionalis*, *H. cinerea*, and *H. squirella* (personal observation). For the latter species I found that urbanized habitats may not hinder gene flow; rather this habitat may act as a facilitator of gene flow (Figure 9 and 10), possibly due to increase in urban-associated temporary ponds (e.g., roadside ditches, retention ponds). Further, it appears that *H. squirella* gene flow is enhanced with more upland oak habitat in inter-population matrices. For example, urban land cover within the Eastern cluster is associated with increased genetic connectivity and this relationship appears more

pronounced with the increased percent upland oak hammocks (Figure 12). Thus, while global trends in amphibian declines may be shaped by the conversion of natural to anthropogenic land cover, natural selection may favor populations and species, such as *H. squirella*, that can sufficiently exploit their surrounding habitats, some of which are novel in the species history.

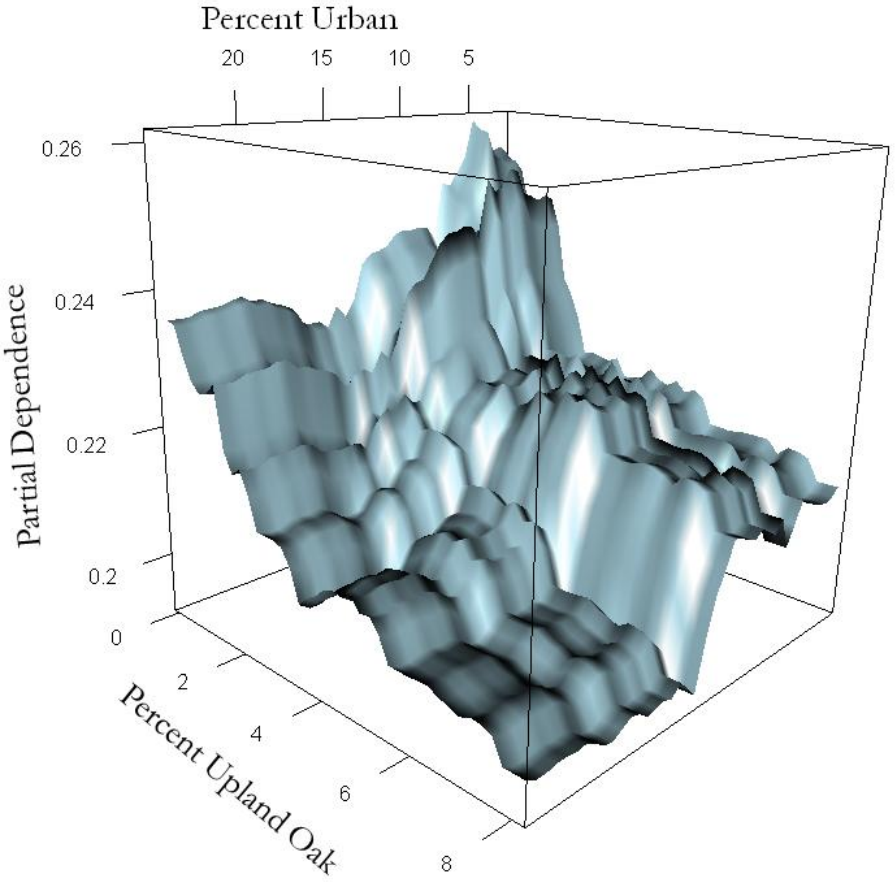

Figure 12: Bivariate partial dependence plot for *Hyla squirella* genetic differentiation in the Eastern cluster (D_{ps}-based model). Here the partial dependence is the effect of two predictor variables, upland oak hammocks (*oak*) and urban land cover (*urban*), on the model after accounting for the effect of spatial distance (*km*). Decreasing partial dependence can be inferred as increasing genetic connectivity.

CHAPTER 5: CONCLUSION

Overall, this study underscores the broad utility of landscape genetics to better infer factors potentially responsible for gene flow and genetic structure among natural populations, including generalist species. Furthermore, it contributes to the growing body of literature that suggests landscape features strongly influence rates of gene flow among populations (reviewed in Storfer *et al.* 2007; Holderegger & Wagner 2008; Balkenhol *et al.* 2009a); that they are important determinants of population genetic connectivity for generalist species (Sacks *et al.* 2008; Frantz *et al.* 2010); and that different landscape types have specific effects that can be identified as to the magnitude and direction of how they influence connectivity. The general methods applied here are useful to investigate evolutionary processes, identify potential dispersal pathways for invasive species, complement niche-based modeling to identify previously unknown populations, model landscape resistance across multiple spatial scales, and aid in corridor design for conservation biology. Many population genetics analyses have considered habitat as simply suitable or non-suitable to explain genetic differentiation. I suggest this technique should be avoided in the future as superficially similar habitats may yield contrasting patterns associated with genetic connectivity; combining all "suitable" habitats may yield a low signal to noise ratio. Moreover, this study demonstrated that IBD and landscape composition have different explanatory power at different spatial scales for the evolutionary patterns in *H. squirella*. Finally, this study indicated landscape features that have potentially influenced patterns of genetic structure in a widespread, generalist species.

APPENDIX A: PREDICTOS USED TO ASSESS HABITAT PERMEABILITY

Appendix A. Detailed list of predictors used to assess habitat permeability in this study. Abbreviation, name, and SEGAP dataset codes are provided. Variables are sorted by whether they are "non-ecological" systems (i.e., anthropogenic land cover types) or semi-natural or natural ecological systems. For each SEGAP dataset code the description (non-ecological systems) or a list of their classifiers (ecological systems) is provided. A complete description of each ecological system can be found at http://www.basic.ncsu.edu/segap/.

Abbreviation	Name	GAP dataset code	Description	
Non-ecological system habitat types				
urban	Urbanization	SEGAP211	Developed open areas such as golf courses and road sides	
		SEGAP220	Low density urbanization; impervious surfaces account for 20-49% of total cover	
		SEGAP230	Medium density urbanization; impervious surfaces account for 50-79% of total cover	
		SEGAP 240	High density urbanization; impervious surfaces account for 80-100% of total cover	
sil	Silviculture	SEGAP410	Plantations dominated by deciduous species.	
		SEGAP420	Plantations dominated by evergreen species.	
pas	Pastures	SEGAP810	Agricultural land cover where pasture/hay vegetation accounts for greater than 20% of the total vegetation.	
		SEGAP820	Agricultural land cover used for the production of annual crops and woody crops such as orchards and vineyards. Crop vegetation accounts for greater than 20 percent of total vegetation.	
(Semi)-natural ecological systems				
			GAP data Classifiers:	
flat	Mesic Flatwoods	CES203.536	Land Cover Class…….	Woody Wetland
			Spatial Scale & Pattern……	Matrix
			Required Classifiers…….	Natural/Semi-

		natural; Vegetated (>10% vasc.); Wetland
	Diagnostic Classifiers……...	None
	Non-Diagnostic Classifiers……..	Woody-Herbaceous; Extensive Wet Flat
	FGDC Crosswalk…….	Vegetated, Tree-dominated, Open tree canopy, Evergreen open tree canopy
CES203.382	Land Cover Class…….	Mixed Upland and Wetland
	Spatial Scale & Pattern……	Matrix
	Required Classifiers…….	Natural/Semi-natural; Vegetated (>10% vasc.); Upland; Wetland
	Diagnostic Classifiers……...	Forest and Woodland (Treed); Woody-Herbaceous; Short Disturbance Interval; Needle-Leaved Tree
	Non-Diagnostic Classifiers……..	None
	FGDC Crosswalk…….	Vegetated, Tree-dominated, Open tree canopy, Evergreen open

		tree canopy
CES203.375a-c	Land Cover Class…….	Vegetated, Tree-dominated, Open tree canopy, Evergreen open tree canopy
	Spatial Scale & Pattern……	Matrix
	Required Classifiers…….	Natural/Semi-natural; Vegetated (>10% vasc.); Upland; Wetland
	Diagnostic Classifiers……...	Forest and Woodland (Treed); Extensive Wet Flat; Short Disturbance Interval; Needle-Leaved Tree
	Non-Diagnostic Classifiers……..	None
	FGDC Crosswalk…….	Vegetated, Tree-dominated, Open tree canopy, Evergreen open tree canopy
CES411.381	Land Cover Class…….	Mixed Upland and Wetland
	Spatial Scale & Pattern……	Matrix
	Required Classifiers…….	Natural/Semi-natural; Vegetated (>10% vasc.); Upland; Wetland
	Diagnostic Classifiers……...	Needle-Leaved

			Non-Diagnostic Classifiers……..	Tree
				Woody-Herbaceous; Extensive Wet Flat
			FGDC Crosswalk…….	Vegetated, Tree-dominated, Open tree canopy, Evergreen open tree canopy
swamp	Swamp	CES203.304a,b	Land Cover Class…….	Woody Wetland
			Spatial Scale & Pattern……	Large patch
			Required Classifiers…….	Natural/Semi-natural; Vegetated (>10% vasc.); Wetland
			Diagnostic Classifiers……...	Forest and Woodland (Treed); Extensive Wet Flat; Needle-Leaved Tree; Broad-Leaved Tree
			Non-Diagnostic Classifiers……..	Organic Peat (>40 cm); Mineral: W/ A-Horizon >10 cm
			FGDC Crosswalk…….	Vegetated, Tree-dominated
		CES203.251	Land Cover Class…….	Woody Wetland
			Spatial Scale & Pattern……	Small patch

			Required Classifiers…….	Natural/Semi-natural; Vegetated (>10% vasc.); Wetland
			Diagnostic Classifiers……...	Forest and Woodland (Treed); Depressional; Needle-Leaved Tree
			Non-Diagnostic Classifiers……..	Isolated Wetland [Partially Isolated]
			FGDC Crosswalk…….	Vegetated, Tree-dominated, Open tree canopy, Evergreen open tree canopy
rff	River Floodplain Forest	CES203.247a	Land Cover Class…….	Woody Wetland
			Spatial Scale & Pattern……	Linear
			Required Classifiers…….	Natural/Semi-natural; Vegetated (>10% vasc.); Wetland
			Diagnostic Classifiers……...	Riverine / Alluvial [Blackwater]
			Non-Diagnostic Classifiers……..	Forest and Woodland (Treed)
			FGDC Crosswalk…….	None
		CES203.249	Land Cover Class…….	Woody Wetland
			Spatial Scale & Pattern……	Linear
			Required Classifiers…….	Natural/Semi-

		natural; Vegetated (>10% vasc.); Wetland
	Diagnostic Classifiers……...	Riverine / Alluvial [Blackwater]
	Non-Diagnostic Classifiers……..	Forest and Woodland (Treed)
	FGDC Crosswalk…….	None
CES203.489a	Land Cover Class…….	Woody Wetland
	Spatial Scale & Pattern……	Linear
	Required Classifiers…….	Natural/Semi-natural; Vegetated (>10% vasc.); Wetland
	Diagnostic Classifiers……...	Forest and Woodland (Treed); Riverine / Alluvial [Brownwater]
	Non-Diagnostic Classifiers……..	None
	FGDC Crosswalk…….	None
CES203.493	Land Cover Class…….	Woody Wetland
	Spatial Scale & Pattern……	Linear
	Required Classifiers…….	Natural/Semi-natural; Vegetated (>10% vasc.); Wetland
	Diagnostic Classifiers……...	Riverine / Alluvial [Blackwater]
	Non-Diagnostic Classifiers……..	Forest and Woodland (Treed)
	FGDC Crosswalk…….	None

oak	Upland oak forest and hammock	CES203.560	Land Cover Class…….	Forest and Woodland
			Spatial Scale & Pattern……	Large patch
			Required Classifiers…….	Natural/Semi-natural; Vegetated (>10% vasc.); Upland
			Diagnostic Classifiers……...	Forest and Woodland (Treed); Broad-Leaved Deciduous Tree
			Non-Diagnostic Classifiers……..	None
			FGDC Crosswalk…….	Vegetated, Tree-dominated, Closed tree canopy, Deciduous closed tree canopy
		CES203.494	Land Cover Class…….	Forest and Woodland
			Spatial Scale & Pattern……	Small patch
			Required Classifiers…….	Natural/Semi-natural; Vegetated (>10% vasc.); Upland
			Diagnostic Classifiers……...	Forest and Woodland

	(Treed); Long Disturbance Interval; Broad-Leaved Evergreen Tree
Non-Diagnostic Classifiers……..	None
FGDC Crosswalk…….	None

APPENDIX B: SINGLE-SCALE RF MODELS

Appendix B. Summary of features associated with genetic connectivity among *Hyla squirella* populations using Random Forest. Models are grouped first by response variable (I or II); then based on distinct genetic clusters identified by Bayesian cluster techniques (1 or 2); hierarchically within the Eastern cluster by (a) all pairwise observations (maximum spatial distance approximately 480 km) and (b) within genetic patch (D_{ps} only; see text). Three single-scale models are reported based on inter-population Vector Buffer Width (VBW) (i) 500 m; (ii) 2 km; and (iii) 10 km. pR^2 is a pseudo R^2; MSE denotes mean squared error. Summary statistics, based on constructing 30 forests for each sub-model (see Methods), include median and 95% confidence intervals (95% CI) of pR^2 and median MSE. P-value of the chosen sub-model was assessed by randomizing the response variable (number of iterations = 9,999) and calculating the tail probability of the empirical (median) pR^2. Model denotes the chosen variables (names cross-reference with Table 1) following model selection. These variables are ordered starting with the most important variable (in terms of MIR values) to the least important. Font style denotes general trend identified using partial dependence plots: standard font, negative association with genetic differentiation among populations; Boldface, positive association with genetic differentiation among populations; italic, complex association; and underline; weak main effect.

Type	Median pR^2	pR^2 95% CI	Median MSE	P-value	Model
I. Genetic distance based on heterozygosity					
I. 1 *Eastern and Western Clusters*					
I. 1.i VBW = 500 m	40.48	40.20—40.94	1.42E-03	<0.001	**km**, **flat**, oak, urban, **swamp**
I. 1.iiVBW = 2 km	40.99	40.32—41.09	1.41E-03	<0.001	**km**, **flat**, oak, *urban*, **swamp**
I. 1.iiiVBW = 10 km	40.22	39.73—40.50	1.43E-03	<0.001	**km**, oak, **flat**, *urban*, **swamp**
I. 2 *Eastern Cluster only*					
I. 2.a All East (480 km)					
I. 2.a.i VBW = 500 m	14.13	13.75—14.38	4.00E-04	<0.001	**pas**, **km**, urban, *swamp*, **flat**
I. 2.a.iiVBW = 2 km	2.59	2.01—2.99	4.54E-04	0.035	**pas, flat,** *swamp***, km**
I. 2.a.iiiVBW = 10 km	13.37	12.90—13.58	4.03E-04	0.001	**pas**, **swamp**, **flat**
II. Genetic distance based on allele frequency distributions					
II.1 *Eastern and Western Clusters*					
II. 1.i VBW = 500 m	60.38	60.03—61.46	1.11E-03	<0.001	**km**, oak, urban, pas
II. 1.iiVBW = 2 km	61.30	61.09—61.41	1.08E-03	<0.001	**km**, oak, urban, pas, sil, **flat**
II. 1.iiiVBW = 10 km	61.27	60.95—62.62	1.08E-03	<0.001	**km**, oak, urban, pas

II.2 *Eastern Cluster only*					
II.2.a All East (480 km)					
II. 2.a.i VBW = 500 m	47.78	47.53—48.16	9.12E-04	<0.001	**km**, oak, urban
II. 2.a.iiVBW = 2 km	51.19	50.77—51.46	8.53E-04	<0.001	**km**, oak, urban, **sil**, flat, rff
II. 2.a.iiiVBW = 10 km	50.23	49.92—50.56	8.70E-04	<0.001	**km**, oak, flat
II.2.b *Subset 150 km*					
II. 2.b.i VBW = 500 m	30.24	29.43—30.50	7.13E-04	<0.001	oak, flat, **km**
II. 2.b.iiVBW = 2 km	34.24	33.63—34.51	6.72E-04	<0.001	flat, oak, **km**, swamp
II. 2.b.iiiVBW = 10 km	27.92	27.33—28.28	7.36E-04	<0.001	oak, flat, **swamp**, **km**, *rff*

APPENDIX C: CORRELATION AMONG VECTOR BUFFER WIDTHS

Appendix C: Linear correlation coefficients among habitat types and vector buffer widths at the global scale (20 populations). The values after "." denote vector buffer width. Boldface indicates r > 0.65.

	urban.500m	sil.500m	pas.500m	rff.500m	flat.500m	oak.500m	swamp.500m
urban.2km	**0.959**	-0.510	0.186	-0.078	-0.112	-0.009	-0.267
urban.10km	**0.933**	-0.525	0.172	-0.048	-0.119	-0.013	-0.259
sil.2km	-0.498	**0.980**	-0.323	-0.024	-0.253	0.151	0.082
sil.10km	-0.539	**0.947**	-0.302	0.017	-0.284	0.133	0.109
pas.2km	0.191	-0.338	**0.967**	-0.244	-0.079	-0.032	-0.075
pas.10km	0.178	-0.364	**0.968**	-0.216	-0.057	-0.022	-0.066
rff.2km	-0.042	-0.053	-0.220	**0.946**	0.074	0.069	-0.141
rff.10km	-0.026	0.019	-0.234	**0.918**	0.049	0.068	-0.203
flat.2km	-0.113	-0.244	-0.064	0.078	**0.997**	-0.477	0.608
flat.10km	-0.098	-0.236	-0.041	0.067	**0.985**	-0.481	0.602
oak.2km	-0.085	0.234	-0.071	0.036	-0.489	**0.959**	-0.542
oak.10km	-0.020	0.206	-0.065	0.035	-0.505	**0.965**	-0.565
swamp.2km	-0.261	0.069	-0.043	-0.159	0.602	-0.551	**0.987**
swamp.10km	-0.267	0.085	-0.039	-0.159	0.606	-0.557	**0.963**

LIST OF REFERENCES

Abdoullaye D, Acevedo I, Adebayo AA, *et al.* (2010) Permanent Genetic Resources added to Molecular Ecology Resources Database 1 August 2009-30 September 2009. *Molecular Ecology Resources* **10**, 232-236.

Alford RA, Richards SJ (1999) Global amphibian declines: A problem in applied ecology. *Annual Review of Ecology and Systematics* **30**, 133-165.

Allentoft ME, Siegismund HR, Briggs L, Andersen LW (2009) Microsatellite analysis of the natterjack toad (*Bufo calamita*) in Denmark: populations are islands in a fragmented landscape. *Conservation Genetics* **10**, 15-28.

Angelone S, Holderegger R (2009) Population genetics suggests effectiveness of habitat connectivity measures for the European tree frog in Switzerland. *Journal of Applied Ecology* **46**, 879-887.

Ardren WR, Miller LM, Kime JA, Kvitrud MA (2002) Microsatellite loci for fathead minnow (*Pimephales promelas*). *Molecular Ecology Notes* **2**, 226-227.

Babbitt KJ, Baber MJ, Brandt LA (2006) The effect of woodland proximity and wetland characteristics on larval anuran assemblages in an agricultural landscape. *Canadian Journal of Zoology-Revue Canadienne De Zoologie* **84**, 510-519.

Babbitt KJ, Tanner GW (1997) Effects of cover and predator identity on predation of Hyla squirella tadpoles. *Journal of Herpetology* **31**, 128-130.

Babbitt KJ, Tanner GW (2000) Use of temporary wetlands by anurans in a hydrologically modified landscape. *Wetlands* **20**, 313-322.

Balkenhol N, Gugerli F, Cushman SA, *et al.* (2009a) Identifying future research needs in landscape genetics: where to from here?, 455-463.

Balkenhol N, Waits LP, Dezzani RJ (2009b) Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography* **32**, 818-830.

Banks SC, Lindenmayer D.B., Ward S.J., Taylor A.C. (2005) The effects of habitat fragmentaion via forestry plantation establishment on spatial genotypic structure in the small marsupial carnivor, *Antechinus agilis*. *Molecular Ecology* **14**, 1667-1680.

Bekkevold D, Andere C., Dahlgren T.G., *et al.* (2005) Environmental correlates of population differentiation in Atlantic Herring. *Evolution* **59**, 2656-2668.

Binckley CA, Resetarits WJ (2007) Effects of forest canopy on habitat selection in treefrogs and aquatic insects: implications for communities and metacommunities. *Oecologia* **153**, 951-958.

Blaustein AR, Romansic JM, Kiesecker JM, Hatch AC (2003) Ultraviolet radiation, toxic chemicals and amphibian population declines. *Diversity and Distributions* **9**, 123-140.

Blaustein AR, Wake D.B., W.P. S (1994) Amphibian Declines: Judging Stability, Persistence, and Susceptibility of Populations to Local and Global Extinctions. *Conservation Biology* **8**, 60-71.

Bowcock AM, Ruizlinares A, Tomfohrde J, *et al.* (1994) High-Resolution of Human Evolutionary Trees with Polymorphic Microsatellites. *Nature* **368**, 455-457.

Breiman L (2001) Random forests. *Machine Learning* **45**, 5-32.

Brouat C, Sennedot F, Audiot P, Leblois R, Rasplus JY (2003) Fine-scale genetic structure of two carabid species with contrasted levels of habitat specialization. *Molecular Ecology* **12**, 1731-1745.

Bulger JB, Scott NJ, Seymour RB (2003) Terrestrial activity and conservation of adult California red-legged frogs *Rana aurora draytonii* in coastal forests and grasslands. *Biological Conservation* **110**, 85-95.

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

Carr AF (1940) Dates of Frog Choruses in Florida. *Copeia*, 55.

Castellano S, Balletto E (2002) Is the partial Mantel test inadequate? *Evolution* **56**, 1871-1873.

Chan LM, Fitzgerald LA, Zamudio KR (2009) The scale of genetic differentiation in the Dunes Sagebrush-Lizard (*Sceloporus arenicolus*), an endemic habitat specialist. *Conservation Genetics* **10**, 131-142.

Chan LM, Zamudio KR (2009) Population differentiation of temperate amphibians in unpredictable environments. *Molecular Ecology* **18**, 3185-3200.

Comer PJ, Schulz KA (2007) Standardized ecological classification for mesoscale mapping in the southwestern United States. *Rangeland Ecology & Management* **60**, 324-335.

Conant R, Collins JT (1998) *A Field Guide to Reptiles and Amphibians of Eastern and Central North America*, Third Edition edn. Houghton Mifflin Company, New York.

Cutler A, Stevens JR (2006) Random forests for microarrays. In: *DNA Microarrays, Part B: Databases and Statistics*, pp. 422-+. Elsevier Academic Press Inc, San Diego.

Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT (2007) Random forests for classification in ecology. *Ecology* **88**, 2783-2792.

Deckert RF (1915) Further Notes on the Salientia of Jacksonville, FLA. *Copeia* **No. 18**, pp. 3-5.

Degner JF, Silva DM, Hether TD, Daza JM, Hoffman EA (2010) Fat frogs, mobile genes: unexpected phylogeographic patterns for the ornate chorus frog (*Pseudacris ornata*). *Molecular Ecology* **19**, 2501-2515.

Dionne M, Caron F, Dodson JJ, Bernatchez L (2008) Landscape genetics and hierarchical genetic structure in Atlantic salmon: the interaction of gene flow and local adaption. *Molecular Ecology* **17**, 2382-2396.

Dodd CK, Cade BS (1998) Movement patterns and the conservation of amphibians breeding in small, temporary wetlands. *Conservation Biology* **12**, 331-339.

Dodd CK, Jr., Smith LL (2003) Habitat destruction and alteration: historical trends and future prospects for amphibians. In: *Amphibian Conservation* (ed. Semlitsch RD), pp. 94-112, Smithsonian Institution, Washingtion, D.C.

Duellman WE, Trueb L (1986) *Biology of Amphibians* The Johns Hopkins University Press, Baltimore.

Emaresi G, Pellet J, Dubey S, Hirzel AH, Fumagalli L (2009) Landscape genetics of the Alpine newt (*Mesotriton alpestris*) inferred from a strip-based approach. *Conservation Genetics. DOI: 10.1007/s10592-009-9985-y.*

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

Evans JS, Cushman SA (2009) Gradient modeling of conifer species using random forests. *Landscape Ecology* **24**, 673-683.

Fontaine MC, Baird SJE, Piry S, *et al.* (2007) Rise of oceanographic barriers in continuous populations of a cetacean: the genetic structure of harbour porpoises in Old World waters. *Bmc Biology* **5**, 16.

Frantz AC, Pope LC, Etherington TR, Wilson GJ, Burke T (2010) Using isolation-by-distance-based approaches to assess the barrier effect of linear landscape elements on badger (*Meles meles*) dispersal. *Molecular Ecology* **19**, 1663-1674.

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

Funk WC, Greene AE, Corn PS, Allendorf FW (2005b) High dispersal in a frog species suggests that it is vulnerable to habitat fragmentation. *Biology Letters* **1**, 13-16.

Geffen E, Anderson MJ, Wayne RK (2004) Climate and habitat barriers to dispersal in the highly mobile grey wolf. *Molecular Ecology* **13**, 2481-2490.

Giordano AR, Ridenhour B.J., Storfer A. (2007) The influence of altitude and topography on genetic structure in the long-toed salamander (*Abystoma macrodactulym*). *Molecular Ecology* **16**, 1625-1637.

Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* **22**, 1-19.

Goudet J (1995) Fstat version 1.2: a computer program to calculate F-statistics. *Journal of Heredity* **86**, 485-486.

Guillot G (2008) Inference of structure in subdivided populations at low levels of genetic differentiation-the correlated allele frequencies model revisited. *Bioinformatics* **24**, 2222-2228.

Guillot G, Estoup A, Mortier F, Cosson JF (2005a) A spatial statistical model for landscape genetics. *Genetics* **170**, 1261-1280.

Guillot G, Leblois R, Coulon A, Frantz AC (2009) Statistical methods in spatial genetics. *Molecular Ecology* **18**, 4734-4756.

Guillot G, Mortier F., Estuop A. (2005b) Geneland: a computer package for landscape genetics. *Molecular Ecology* **5**, 712-715.

Guillot G, Santos F, Estoup A (2008) Analysing georeferenced population genetics data with Geneland: a new algorithm to deal with null alleles and a friendly graphical user interface. *Bioinformatics* **24**, 1406-1407.

Hardy OJ, Charbonnel N, Freville H, Heuertz M (2003) Microsatellite allele sizes: A simple test to assess their significance on genetic differentiation. *Genetics* **163**, 1467-1482.

Hardy OJ, Vekemans X (2002) SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**, 618-620.

Hastie T, Tibshirani R, Friedman J (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.*, Second edn., New York, NY.

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

Hokit DG, Ascunce M, Ernst J, Branch LC, Clark AM (2010) Ecological metrics predict connectivity better than geographic distance. *Conservation Genetics* **11**, 149-159.

Holderegger R, Kamm U, Gugerli F (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology* **21**, 797-807.

Holderegger R, Wagner HH (2008) Landscape genetics. *Bioscience* **58**, 199-207.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* **9**, 1322-1332.

Jenkins D, Carey M, Czerniewska J, *et al.* (2010) A meta-analysis of isolation by distance: relic or reference standard for landscape genetics? *Ecography* **33**, 315-320.

Jensen JB, Camp C.D., Gibbons W., Elliott M. (2008) Amphibians and Reptiles of Georgia, p. 561. The University of Georgia Press, Athens, GA.

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, Web Service. *BMC Genetics* **6 (13)**, 1-6.

Kozak KH, Graham CH, Wiens JJ (2008) Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology & Evolution* **23**, 141-148.

Krauss J, Steffan-Dewenter I, Tscharntke T (2003) Local species immigration, extinction, and turnover of butterflies in relation to habitat area and habitat isolation. *Oecologia* **137**, 591-602.

Lannoo M (2005) Amphibian Declines: The conservation status of United States Species, p. 1094. University of California Press, Berkeley and Los Angeles, California.

Lee-Yaw JA, Davidson A, McRae BH, Green DM (2009) Do landscape processes predict phylogeographic patterns in the wood frog? *Molecular Ecology* **18**, 1863-1874.

Liaw A, Wiener M (2002) Classification and Regreesion by randomForest. *R News* **2 (3)**, 18-22.

Lowe WH, Likens GE, McPeek MA, Buso DC (2006) Linking direct and indirect data on dispersal: Isolation by slope in a headwater stream salamander. *Ecology* **87**, 334-339.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**, 189-197.

Manier MK, Arnold SJ (2006) Ecological correlates of population genetic structure: a comparative approach using a vertebrate metacommunity. *Proceedings of the Royal Society B-Biological Sciences* **273**, 3001-3009.

Mantel N (1967) Detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, 209-&.

McRae BH (2006) Isolation by resistance. *Evolution* **60**, 1551-1561.

Minch E, Ruiz-Linares A, Goldstein D, Feldman M, Cavalli-Sforza LL (1996) MICROSAT (version 1.5b): A Computer Program for Calculating Various Statistics on Microsatellite Allele Data, p. program available from: http://hpgl.stanford.edu/projects/microsat/.

Murphy MA, Evans JS, Cushman SA, Storfer A (2008) Representing genetic variation as continuous surfaces: an approach for identifying spatial dependency in landscape genetic studies. *Ecography* **31**, 685-697.

Murphy MA, Evans JS, Storfer A (2010) Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics. *Ecology* **91**, 252-261.

NatureServe (2006) International classification of ecological communities: terrestrial vegetation of the United States. Arlington, VA: NatureServe.

Pavlacky DC, Goldizen AW, Prentis PJ, Nicholls JA, Lowe AJ (2009) A landscape genetics approach for quantifying the relative influence of historic and contemporary habitat heterogeneity on the genetic connectivity of a rainforest bird. *Molecular Ecology* **18**, 2945-2960.

Pierson JC, Allendorf FW, Saab V, Drapeau P, Schwartz MK (2010) Do male and female black-backed woodpeckers respond differently to gaps in habitat? *Evolutionary Applications* **3**, 263-278.

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

Raufaste N, Rousset F (2001) Are partial mantel tests adequate? *Evolution* **55**, 1703-1705.

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

Rice WR (1989) Analyzing tables of statistical tests. *Evolution* **43**, 223-225.

Richards-Zawacki CL (2009) Effects of slope and riparian habitat connectivity on gene flow in an endangered Panamanian frog, *Atelopus varius*. *Diversity and Distributions* **15**, 796-806.

Rousset F (2002) Partial Mantel tests: reply to Castelloan and Balletto. *Evolution* **56**, 1874-1875.

Rousset F (2008) GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.

Sacks BN, Bannasch DL, Chomel BB, Ernest HB (2008) Coyotes demonstrate how habitat specialization by individuals of a generalist species can diversify populations in a heterogeneous ecoregion. *Molecular Biology and Evolution* **25**, 1384-1394.

Sambrook J, Russel D (2001) *Molecular Cloning: a Laboratory Manual*, 3 edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Semlitsch RD (1998) Biological delineation of terrestrial buffer zones for pond-breeding salamanders. *Conservation Biology* **12**, 1113-1119.

Semlitsch RD, Bodie JR (2003) Biological criteria for buffer zones around wetlands and riparian habitats for amphibians and reptiles. *Conservation Biology* **17**, 1219-1228.

Siroky DS (2009) Navigating Random Forest and related advances in algorithmic modeling. *Statistics Surveys* **3**, 147-163.

Smouse PE, Long JC, Sokal RR (1986) Multiple-regression and correlation extensions of the mantel test of matrix correspondence. *Systematic Zoology* **35**, 627-632.

Soares TN, Chaves LJ, Telles MPD, Diniz-Filho JAF, Resende LV (2008) Landscape conservation genetics of *Dipteryx alata* ("baru" tree : Fabaceae) from Cerrado region of central Brazil. *Genetica* **132**, 9-19.

Sokal RR, Oden NL (1978a) Spatial Autocorelation in Biology 2. Some Biological Implication and 4 Applications of Evolutionary and Ecological interest. *Biological Journal of the Linnean Society* **10**, 229-249.

Sokal RR, Oden NL (1978b) Spatial autocorrelation in biology 1. methodology. *Biological Journal of the Linnean Society* **10**, 199-228.

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology* **15**, 4261-4293.

Spear SF, Peterson CR, Matocq MD, Storfer A (2005) Landscape genetics of the blotched tiger salamander (*Ambystoma tigrinum melanostictum*). *Molecular Ecology* **14**, 2553-2564.

Spear SF, Storfer A (2008) Landscape genetic structure of coastal tailed frogs (*Ascaphus truei*) in protected vs. managed forests. *Molecular Ecology* **17**, 4642-4656.

Spear SF, Storfer A (2010) Anthropogenic and natural disturbance lead to differing patterns of gene flow in the Rocky Mountain tailed frog, *Ascaphus montanus*. *Biological Conservation* **143**, 778-786.

Steele CA, Baumsteiger J, Storfer A (2009) Influence of life-history variation on the genetic structure of two sympatric salamander taxa. *Molecular Ecology* **18**, 1629-1639.

Steele CA, Storfer A (2007) Phylogeographic incongruence of codistributed amphibian species based on small differences in geographic distribution. *Molecular Phylogenetics and Evolution* **43**, 468-479.

Stevens VM, Verkenne C, Vandewoestijne S, Wesselingh RA, Baguette M (2006) Gene flow and functional connectivity in the natterjack toad. *Molecular Ecology* **15**, 2333-2344.

Stewart A, Komers PE, Bender DJ (2010) Assessing landscape relationships for habitat generalists. *Ecoscience* **17**, 28-36.

Storfer A, Eastman JM, Spear SF (2009) Modern molecular methods for amphibian conservation. *Bioscience* **59**, 559-571.

Storfer A, Murphy MA, Evans JS, *et al.* (2007) Putting the 'landscape' in landscape genetics. *Heredity* **98**, 128-142.

Svetnik V, Liaw A, Tong C, *et al.* (2003) Random forest: A classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Sciences* **43**, 1947-1958.

Svetnik V, Wang T, Tong C, *et al.* (2005) Boosting: An ensemble learning tool for compound classification and QSAR modeling. *Journal of Chemical Information and Modeling* **45**, 786-799.

Trenham PC, Shaffer HB (2005) Amphibian upland habitat use and its consequences for population viability. *Ecological Applications* **15**, 1158-1168.

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

Van Zant JL, Wooten MC (2007) Old mice, young islands and competing biogeographical hypotheses. *Molecular Ecology* **16**, 5070-5083.

Vandergast AG, Gillespie RG, Roderick GK (2004) Influence of volcanic activity on the population genetic structure of Hawaiian Tetragnatha spiders: fragmentation, rapid population growth and the potential for accelerated evolution. *Molecular Ecology* **13**, 1729-1743.

Wang IJ (2009) Fine-scale population structure in a desert amphibian: landscape genetics of the black toad (*Bufo exsul*). *Molecular Ecology* **18**, 3847-3856.

Windes K (2010) Treefrog (*Hyla squirella*) responses to rangeland management in semi-tropical Florida, USA, MS thesis. University of Central Florida.

With KA, Crist TO (1995) Critical thresholds in species responses to landscape structure. *Ecology* **76**, 2446-2459.

With KA, Gardner RH, Turner MG (1997) Landscape connectivity and population distributions in heterogeneous environments. *Oikos* **78**, 151-169.

Wright AH (2002) *Life-Histories of the Frogs of Okefinokee Swamp, Georgia.* Cornell University Press, Ithaca, NY.

Wright AH, Wright AA (1995) *Handbook of Frogs and Toads of the United States and Canada* Comstock Publishing Company, Inc. Ithaca, NY.

Wright S (1943) Isolation by distance. *Genetics* **28**, 139-156.

Zhao S, Dai Q, Fu J (2009) Do rivers function as genetic barriers for the plateau wood frog at high elevations? *Journal of Zoology* **279**, 270-276.