

To the Graduate Council:

I am submitting herewith a dissertation written by Robert Graham Reynolds entitled “Islands, metapopulations, and archipelagos: genetic equilibrium and non-equilibrium dynamics of structured populations in the context of conservation.” I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Ecology and Evolutionary Biology.

Dr. Benjamin Fitzpatrick, Major Professor

We have read this dissertation  
and recommend its acceptance:

Dr. James Fordyce

Dr. Graciela Cabana

Dr. Dan Simberloff

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

**Islands, Metapopulations, and Archipelagos**  
**Genetic Equilibrium and Non-equilibrium Dynamics of**  
**Structured Populations in the Context of Conservation**

A Dissertation Presented for  
The Doctor of Philosophy Degree  
The University of Tennessee, Knoxville

Robert Graham Reynolds  
May 2011

Copyright © 2011 by Robert Graham Reynolds  
All rights reserved.

## **ACKNOWLEDGMENTS**

Foremost a debt of gratitude is owed to my wife Jennifer for encouraging me to pursue a lifelong desire to become a scientist. Her support of both her husband and her household was essential to my ability to remain a student until completion of a terminal degree, and I will be forever in her debt. I am also indebted to my advisor, Ben Fitzpatrick, for his mountains of patience and assistance that cannot be efficiently communicated for sake of brevity and lexical forfeit. I am especially grateful to my committee: Jim Fordyce, Dan Simberloff, and Graciela Cabana, for their encouragement and sagacious influence on my development as a scientist. Stan Guffey has been a tremendous pedagogical mentor to me, he is a gifted teacher and I have benefited greatly from his instruction and encouragement. Finally, I am grateful to the many members of the Department of Ecology and Evolutionary Biology who have encouraged or assisted me: especially Chris Boake, Gary McCracken, Darrin Hulsey, Sandy Echternacht, Mark Alston, Brian O'Meara, Phyllis Bice, Cheryl Lynn, Kate Weaver (Core Biology), Pearl Kirkland-Smith, Anne Mintz, and Jaime Call; and to the wonderful custodial and maintenance staff in Hesler Hall who kept my office clean and functional for five years.

I have many individuals to thank for their assistance (including blood and sweat) in the Turks and Caicos Islands: Kelly Bradley, Joe Burgess, Cory Deal, Matt Emery, Glenn Gerber, T.J. Jett, Chuck Knapp, Brian Naqqi Manco, Lindsay Mensen, Naz Missick, Matt Niemiller, Richard Plate, Tarren Wagner, George Waters, and Matt Wilson. The staff, both current and former, of the Department of Environment and Coastal Resources of the Turks and Caicos Islands Government assisted me with mounds

of information and permitting, and I am especially grateful to Brain Naqqi Manco, Brian Riggs, Eric Salamanca, Paul Dickenson, and Marlon Hibbert for their assistance. For logistic support in the Turks and Caicos I am grateful to The Turks and Caicos National Trust, and to Blue Horizon Resort for property access. I am especially grateful to the Turks and Caicos Sporting Club at Ambergris Cay for logistical support, property access, use of the Environmental Center, delicious meals, and air transportation over four seasons of field work; and to Lindsay Mensen and Matt Emery for always making me feel welcome on Ambergris.

Glenn Gerber is owed special thanks for his unofficial advisor status, encouragement, and huge amount of assistance with funding, logistics, and information; and for letting me complete a book chapter even though he was not able to contribute and the chapter suffered the deficit of his knowledge and experience. Glenn also patiently spent many long hours searching for boas at night on Ambergris, ignoring the peril of wobbly boulders surrounded by lacerating cactus catch-nets.

A special thanks is owed to the members of the H.O.F.F. lab group at the University of Tennessee for the years of discussions, manuscript reviews, advice, and encouragement: especially Ben Fitzpatrick, Jim Fordyce, Darrin Hulsey, Brian O'Meara, Matt Niemiller, Stesha Pasachnik, JR Jones, Premal Shah, Philip Hollingsworth, Dylan Dittrich-Reed, Megan Todd-Thompson, Romina Dimarco, Jason Robinson, Amanda Allison, Zach Marion, Barb Banbury, Sharon Clemmensen, and others. Denita Hadžiabdić and the Trigiano Lab assisted greatly in teaching me use of the *eGene* machine for microsatellite analysis and I am grateful for their patience, and I am

appreciative of the assistance and training Joe May provided in the UTK Molecular Biology Resource Facility.

My family and friends are owed special thanks for their support and encouragement and for making my time in graduate school a wonderful experience. To my family: Jennifer Reynolds, Bob and Kim Reynolds, Terry Deal, Blair Reynolds, Taylor Reynolds, and Rick and Margaret Gourdin thank you as always for your love and support: to my grandparents Jack and Betty Reynolds for weekly Sunday home cooked meals and many wonderful afternoons spent together; and to my good friends Matt Niemiller, Premal Shah, Stesha Pasachnik, and Dylan Dittrich-Reed for all the good times, assistance, advice, and field excursions. Thanks as well to Darrin Hulsey for the great trip to Honduras and for the invitation to Malawi I grudgingly had to forfeit.

Funding for my research in graduate school has been provided by the following: The University of Tennessee Department of Ecology and Evolutionary Biology Summer Research Grant Program, The Sigma Xi Grant-in-Aid of Research, The American Philosophical Society Lewis and Clark Fund for Exploration and Research, The American Museum of Natural History Theodore Roosevelt Memorial Fund Grant in Support of Research, The Mohamed bin Zayed Species Conservation Fund, The University of Tennessee W. K. McClure Scholarship for the Study of World Affairs, The San Diego Zoo Institute for Conservation Research and the Offield Family Foundation, and The American Genetics Association.

## Abstract

Understanding complex population dynamics is critical for both basic and applied ecology. Analysis of genetic data has been promoted as a way to reconstruct recent non-equilibrium processes that influence the apportioning of genetic diversity among populations of organisms. In a structured-deme context, where individual populations exist as geographically distinct units, island biogeography theory and metapopulation genetics predict that the demographic processes of extinction, colonization, and migration will affect the magnitude and rate of genetic divergence between demes. New methods have been developed to attempt to detect the influence of non-equilibrium dynamics in structured populations. I challenged two of these methods: decomposed pairwise regression and allele frequency analyses, using simulations of genetic data from structured demes. I found that these methods suffer from a high type II error rate, or failure to reject the null hypothesis of mutation-migration-drift equilibrium for demes experiencing historical demographic events. In addition, island biogeography and metapopulation ecology predict that at equilibrium, some species in a patch will be recent colonists, as equilibrium indicates a balance between colonization of the patch and extinction from the patch. Recent colonists are unlikely to have reached population mutation-migration-drift equilibrium; hence a paradox exists between population and community level equilibrium. I used nuclear and mitochondrial genetic data from populations of two species of reptiles from the Turks and Caicos Islands, British West Indies to test for patterns of equilibrium vs. non-equilibrium. I found unexpected shallow genetic divergence in the Turks Island boa (*Epicrates chrysogaster*), indicating that this

species likely existed as a panmictic population prior to the inundation of the Turks and Caicos Banks during the last glaciation. As the initial methods I tested using simulations proved unreliable, I used methods from phylogeography, landscape genetics, and island biogeography to detect significant non-equilibrium dynamics in the Turks and Caicos curly-tailed lizard (*Leiocephalus psammodromus*), finding evidence for high levels of biased gene flow. I propose that studies of genetic diversity on island archipelagos use tools from all three of these methods to evaluate empirical data in the context of equilibrium and the null hypotheses offered by island biogeography and population genetics theory. I frame the results both in the context of conservation and an understanding of equilibrium and non-equilibrium dynamics.

## TABLE OF CONTENTS

|                                                                                                                                                                                           |            |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>GENERAL INTRODUCTION .....</b>                                                                                                                                                         | <b>1</b>   |
| Literature Cited.....                                                                                                                                                                     | 11         |
| <b>PART I. TESTS OF TWO METHODS FOR INFERENCE OF HISTORICAL<br/>DEMOGRAPHY USING SIMULATED STRUCTURED POPULATIONS<br/>REVEALS SUBSTANTIAL TYPE II ERROR.....</b>                          | <b>15</b>  |
| Abstract .....                                                                                                                                                                            | 16         |
| Introduction .....                                                                                                                                                                        | 17         |
| Methods .....                                                                                                                                                                             | 21         |
| Results .....                                                                                                                                                                             | 28         |
| Discussion .....                                                                                                                                                                          | 31         |
| Literature Cited .....                                                                                                                                                                    | 38         |
| Appendix Ia. Figures and Tables.....                                                                                                                                                      | 41         |
| Appendix Ib. Supplementary Tables and Figures .....                                                                                                                                       | 52         |
| <b>PART II. UNEXPECTED SHALLOW GENETIC DIVERGENCE IN TURKS<br/>ISLAND BOAS (<i>EPICRATES C. CHRYSOGASTER</i>) REVEALS SINGLE<br/>EVOLUTIONARY SIGNIFICANT UNIT FOR CONSERVATION .....</b> | <b>62</b>  |
| Abstract .....                                                                                                                                                                            | 63         |
| Introduction .....                                                                                                                                                                        | 64         |
| Methods .....                                                                                                                                                                             | 67         |
| Results .....                                                                                                                                                                             | 72         |
| Discussion .....                                                                                                                                                                          | 74         |
| Literature Cited .....                                                                                                                                                                    | 84         |
| Appendix IIa. Figures and Tables .....                                                                                                                                                    | 91         |
| Appendix IIb. Primers and Genbank accession numbers.....                                                                                                                                  | 99         |
| <b>PART III. MOLECULAR ISLAND BIOGEOGRAPHY: EQUILIBRIUM VERSUS<br/>NON-EQUILIBRIUM GENETIC STRUCTURE IN AN ARCHIPELAGIC<br/>LIZARD.....</b>                                               | <b>101</b> |
| Abstract .....                                                                                                                                                                            | 102        |
| Introduction .....                                                                                                                                                                        | 103        |
| Methods .....                                                                                                                                                                             | 106        |
| Results .....                                                                                                                                                                             | 117        |
| Discussion .....                                                                                                                                                                          | 120        |
| Literature Cited .....                                                                                                                                                                    | 130        |
| Appendix IIIa. Figures and Tables.....                                                                                                                                                    | 138        |
| Appendix IIIb. Genbank accession numbers.....                                                                                                                                             | 152        |
| <b>GENERAL CONCLUSIONS .....</b>                                                                                                                                                          | <b>154</b> |
| Appendix IV. R Scripts.....                                                                                                                                                               | 158        |
| VITA.....                                                                                                                                                                                 | 184        |

## LIST OF FIGURES

### **PART I. Tests of Two Methods For Inference Of Historical Demography Using Simulated Structured Populations Reveals Substantial Type II Error..... 15**

|                                                                                                                                                                                                                                                      |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure I – 1. Historical demographic event scenarios used in simulations. ....                                                                                                                                                                       | 43 |
| Figure I – 2. Migration scenario used in simulations, where solid arrows indicate exchange of $N_m$ migrant(s)/generation among nearest neighbors and dotted arrows indicate exchange of $N_m/10$ migrants/generation among next-nearest neighbors.. | 44 |
| Figure I – 3. Results of decomposed pairwise regression for bottlenecked focal deme                                                                                                                                                                  | 45 |
| Figure I – 4. Rank of the slope and intercept from the decomposed pairwise regression of the focal deme for 200 simulations of each FDM parameter set.....                                                                                           | 46 |
| Figure I – 5. Decomposed pairwise regression analysis to detect true outliers.....                                                                                                                                                                   | 47 |
| Figure I – 6. Detection of non-equilibrium in the program BOTTLENECK.....                                                                                                                                                                            | 48 |
| Figure I – S1. Average gene diversity per deme for each equilibrium parameter set of the MM.....                                                                                                                                                     | 53 |
| Figure I – S2. Histograms of the combined probability values (SD and Wilcoxon tests) for each deme from the three tests in BOTTLENECK. ....                                                                                                          | 55 |
| Figure I – S3. Histograms of the combined probability values for each deme from the three tests in BOTTLENECK.....                                                                                                                                   | 56 |
| Figure I – S4. $P$ values from the Standardized Differences and Wilcoxon tests plotted by log of the event time in generations prior to the sample.....                                                                                              | 58 |
| Figure I – S5. Plots of $Z_S$ for each simulation from the Standardized Differences and Wilcoxon tests in bottleneck on the y-axis against the mean time since the event in generations. ....                                                        | 60 |

### **Part II. Unexpected Shallow Genetic Divergence in Turks Island Boas (*Epicrates c. chrysogaster*) Reveals Single Evolutionarily Significant Unit For Conservation 62**

|                                                                                                                                                     |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure II – 1. Map of the Turks and Caicos Islands illustrating sampling locations and known distribution of <i>Epicrates c. chrysogaster</i> ..... | 92 |
| Figure II – 2. Phylogeographic genealogical relationships among CytB haplotypes for <i>Epicrates c. chrysogaster</i> .....                          | 93 |
| Figure II – 3. Population-specific haplotypes for <i>Epicrates c. chrysogaster</i> using the nuclear markers NT3 and <i>c-mos</i> ..                | 94 |

### **PART III. Molecular Island Biogeography: Equilibrium versus Non-equilibrium Genetic Structure in an Archipelagic Lizard..... 101**

|                                                                                                                                                       |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure III – 1. Map of the Turks and Caicos Islands, located at the southern terminus of the Bahamian Archipelago roughly 130 km from Hispaniola..... | 139 |
| Figure III – 2. 50% Majority-rule consensus Bayesian gene tree for the mitochondrial ND2 locus for 257 <i>L. psammodromus</i> samples..               | 141 |

|                                                                                                                                                                                            |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure III – 4. Isolation by distance plots of genetic distance ( $\Phi_{ST} / (1 - \Phi_{ST})$ ) versus log-transformed stepping-stone geographic distance for mtDNA and AFLP data.. .... | 143 |
| Figure III – 5. Decomposed pairwise regression of genetic distance and geographic distance for islands on the Caicos Bank.. .....                                                          | 144 |
| Figure III – 6. Distribution of structure groups by sampling location. ....                                                                                                                | 145 |
| Figure III – S1. Simulated distributions of gene diversities for 70 neutral, diallelic markers for three different effective population sizes after 4,000 generations of isolation.....    | 146 |

## LIST OF TABLES

### **Part I. Tests of two methods for inference of historical demography using simulated structured populations reveals substantial type II error**

|                                                                                                                                                                                                                             |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Table I – 1. Simulations of bottlenecked demes used to test parameter sets without migration in the program bottleneck.....                                                                                                 | 50 |
| Table I – 2. Parameters used in FDM and MM simulations..                                                                                                                                                                    | 49 |
| Table I – 3. Results from combined probability analysis using $Z$ transformation..                                                                                                                                          | 51 |
| Table I –S1. Results for <i>ad hoc</i> DPR analysis of 100 simulations of equilibrium metapopulations for each of 8 parameter sets where the focal deme varied in $N_e$ relative to other demes in the metapopulation. .... | 61 |

### **Part II. Unexpected shallow genetic divergence in turks island boas (*epicrates c. chrysogaster*) reveals single evolutionary significant unit for conservation**

|                                                                                                                                                                                                                                          |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Table II – 1. Sampling locations and sampling effort for tissue samples used in this analysis.....                                                                                                                                       | 95  |
| Table II – 2. Haplotypes of <i>Epicrates c. chrysogaster</i> from each population for one mitochondrial gene (CytB) and two nuclear genes (NT3, c-mos). Haplotypes are numbered in bold 1-9 for CytB and 1-2 for each nuclear gene. .... | 96  |
| Table II – 3. Corrected (Tamura-Nei; top diagonal) and uncorrected ( $p$ -distance; bottom diagonal) pairwise genetic divergence between populations of <i>Epicrates c. chrysogaster</i> . Island names abbreviated as per Table 2.....  | 97  |
| Table II – 4. Results of analysis of molecular variance (AMOVA) for various groupings of <i>Epicrates c. chrysogaster</i> . ....                                                                                                         | 98  |
| Appendix IIb. Primers and PCR conditions used in this study. Underlined bases indicate the primer “tail” used to force plus-A addition. ....                                                                                             | 100 |

### **Part III. Molecular island biogeography: equilibrium versus non-equilibrium genetic structure in an archipelagic lizard**

|                                                                                                                                                                   |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Table III – 1. Islands surveyed for <i>L. psammmodromus</i> in the Turks and Caicos and tissue samples acquired from each population.....                         | 147 |
| Table III – 2. Genetic diversity at the mitochondrial ND2 and nuclear AFLP loci... ..                                                                             | 148 |
| Table III – 3. $\Phi_{ST}$ estimates for pairwise comparisons among ten island populations of <i>Leiocephalus psammmodromus</i> based on 10,100 permutations..... | 149 |
| Table III – 4. Results of analysis of molecular variance (AMOVA) for various groupings of <i>Leiocephalus psammmodromus</i> for both AFLP and mtDNA data.....     | 150 |
| Table III – S1. Probable extirpations and islands that lack populations of <i>L. psammmodromus</i> . ....                                                         | 151 |
| Appendix IIIb. Haplotypes of <i>Leiocephalus psammmodromus</i> for ND2 and Genbank accession numbers. ....                                                        | 153 |

## **GENERAL INTRODUCTION**

Islands have played a major role in the investigation of evolutionary processes by serving as natural laboratories to test general evolutionary theories in discrete areas (Whittaker & Fernández-Palacios 2007). Beginning with Darwin (1859) and Wallace (1902), island organisms have factored heavily in the elucidation of mechanisms of evolution, and current genetic techniques have allowed a much deeper investigation of both present and historical evolutionary events. Islands closely resemble models used in population genetics, as they have distinct boundaries and potential for a wide range of migration rates and population sizes. The effects of stochastic evolutionary forces (as opposed to the deterministic process of natural selection) have been shown to influence genetic variation both within populations as well as between populations of organisms (Conner & Hartl 2004). When populations are geographically structured, processes such as mutation, genetic drift, migration or gene flow, local extinction, and recolonization may dramatically influence the apportioning of genetic variation within a structured population (Wright 1977; Slatkin 1977, 1985, 1987; Wade & McCauley 1988; Whitlock & McCauley 1990; Conner & Hartl 2004). Many authors have predicted the effects of these processes on neutral genetic variation in metapopulations (a group of populations (demes) linked by gene flow and undergoing local extinction and recolonization) using theoretical modeling techniques (Wright 1977; Slatkin 1977, 1985, 1987; Maruyama & Kimura 1980; Wade & McCauley 1988; Whitlock & McCauley 1990; Whitlock & Barton 1997; Pannell & Charlesworth 1999; Pannell 2003). In addition, island biogeography theory and metapopulation genetics predict that the processes of extinction,

colonization, and migration will affect the magnitude and rate of genetic divergence between island and mainland species (Johnson *et al.* 2000; Whitlock 2004). Extinction and recolonization rates are assumed to be influenced by axes of island variation such as the size or isolation of the island from other islands (MacArthur & Wilson 1967). In an island archipelago, average genetic diversity is expected to be a function of island area, as larger islands can support larger, less extinction-prone populations and therefore experience a lower rate of genetic drift (Johnson *et al.* 2000). Furthermore, immigration rate is predicted to be greater for islands in close proximity relative to more distant islands, hence reducing divergence and maintaining variation, especially in small populations. Finally, proximity to a “source” that distributes potential immigrants might affect extinction probability (if immigrants effect a rescue) and/or the number (and genetic diversity) of recolonizing individuals after an extinction.

### *Metapopulations*

Metapopulation theory is representative of a specific method of modeling the effects of turnover in spatially divided populations. As such, it involves far more dynamics and evolutionary genetic consequences than a simpler  $2_{(+)}$  population interaction model, where lineages from previously isolated populations anastomose through gene flow. In a most basic sense, metapopulations represent the placement of studies of population biology into a spatial framework (as opposed to the classic Wright-Fisher model), with each local population operating according to a birth/death process analogous to that of individual organisms (Hanski & Gaggiotti 2004). In a classic

deterministic Levins (1969) model, discrete populations exist within an infinite network of habitat patches, each with some probability of going extinct, which are connected by migration of reproductive individuals. This most basic model is assumption-laden, and sets up a necessary migration threshold. Equilibrium within the metapopulation depends on a rate of migration greater than that of extinction; otherwise the metapopulation as a whole will go extinct. This model is directly analogous to the multi-species equilibrium model of MacArthur & Wilson (1967) and the epidemiological models used to predict the spread of pathogens (Grenfell & Harwood 1997). Implicit within these models is the component of local extinction, which leads to a network of occupied and unoccupied habitat patches at any given time.

The origin of the importance of the extinction of local demes can be traced to two original sources. The first is Sewall Wright's (1931, 1940) shifting balance theory, which identified the potential for rapid evolution in spatially separated populations, especially those operating under an extinction/recolonization dynamic. Phases 1 and 2 of shifting balance suggest that subdivided populations are subject to drift, and further that subdivision and hence small population size will allow for favorable mutations to sweep through a local population. Though phase 3, the spread of favorable alleles throughout the metapopulation, has been more contentious, Wright initiated the theoretical framework for understanding the importance of subdivided populations. The second component of early metapopulation theory comes from Andrewartha & Birch (1954), whose influential empirical observations of frequent extinction and recolonization of insect populations led them to conclude that these phenomena may be a common in natural populations.

Subsequent empirical observations (*e.g.*, Sjögren 1994; Hanski *et al.* 1994) have confirmed the predictions of these authors- that populations frequently exist as subdivided groups or demes and that local extinction of demes can be common (Harrison & Taylor 1997; Thrall *et al.* 2000). Subdivision (as opposed to panmixia) may arise through a variety of mechanisms, including the patchiness of suitable habitat in the landscape- a dynamic that is increasingly becoming visible in conservation biology owing to anthropogenic environmental modification; or through a process such as island vicariance due to fluctuating sea levels. Susceptibility to extinction may be either an intrinsic feature of small populations or may be caused by extrinsic stochastic events (Harrison *et al.* 1988; Caughley 1994). Small populations are susceptible to Allee effects (density-dependant birth rates) and demographic stochasticity (fluctuations in population demography, more specifically the number of reproductive females) below a certain size (Allendorf & Luikart 2006), and demes experiencing these effects may either have already declined owing to alternate factors or represent sink populations whose persistence depends on consistent immigration. Extrinsic factors causing local extinction may be environmental stochasticity (random changes in environmental conditions, such as hurricanes) or environmentally deterministic factors (*e.g.*, alteration of habitat by an invasive species or bulldozers).

Importantly, some authors have suggested that equilibrium in a population undergoing turnover is unlikely, or have questioned the general significance of the metapopulation paradigm (Hanski & Gaggiotti 2004). Harrison (1994) states that species “in the balance between the extinction and recolonization of populations is...improbable” This appears to be a reaction to the simplicity and pervasive assumptions of classical

metapopulation theory, which have been largely alleviated by increasingly complex (parameterized) spatially explicit models that may have multiple stable equilibria (Gyllenberg *et al.* 1997) and variation in migration rate and population size.

Subsequent work by Hastings & Harrison (*e.g.*, 1994) has established a more realistic model that amends the classic metapopulation theory, whose “extinction-centered definition is too simplistic.”

Genetic consequences of metapopulation dynamics are influenced by the rates of migration and the nature of the recolonization after local extinction (Bay *et al.* 2008 and references therein). Under a migrant-pool (island) model, where all demes contribute to recolonization, the fraction of global genetic diversity associated with differences among demes ( $F_{ST}$ ) is reduced (assuming colonization is about twice the rate of migration), as extinction removes unique alleles. Conversely, a propagule-pool colonization scenario (all colonists arrive from a single deme) increases interdemic  $F_{ST}$  owing to sampling effect and subsequent drift (again, dependent on the number of colonists and migration rate). Metapopulation models represent a fundamentally different population process from other models of population structure. Island, stepping-stone, and isolation by distance models all include the components of spatial structure and gene flow among demes, and all have significant consequences for partitioning of genetic variation as well as evolutionary change. However, metapopulation models assume an extinction/recolonization dynamic, because empirical observations have supported the existence of this process, and the consequences of this type of dynamic extend beyond what is predicted by less-inclusive models of population structure.

In order to conform to a classic metapopulation model, an empirical system must exhibit extinction/recolonization dynamics, low levels of gene flow (*i.e.*, just enough to permit recolonization but not so much as to swamp local genetic differentiation), and persistence across multiple generations (*i.e.*, be in a stable equilibrium). Few examples exist of populations adhering to this classical model (Harrison & Taylor 1997), but well known classical metapopulations include pool frogs in Sweden (Sjögren 1994), and butterflies in Finland (Hanski *et al.* 1994) and California (Harrison *et al.* 1988). It has been suggested (Harrison & Taylor 1997) that many more species adhere to non-classical metapopulation models, which incorporates a greater range of variable parameters (*i.e.*, variable patch size, explicit spatial relations, variance in migration rate, *etc.*). Recent examples include reef fish in Australia, which exist as discrete populations on coral patches and have been shown to be operating under a metapopulation dynamic (Planes *et al.* 1997; Bay *et al.* 2008).

### *Approach*

**Chapter I.** Prior studies of historical demography have had limited success because the detailed histories of specific populations are varied and subtle. My strategy is different in focusing on general characteristics of a system of populations; the goal is not to diagnose non-equilibrium in specific islands treated independently; rather I seek to evaluate characteristics of an entire archipelago or metapopulation that reflect the importance of non-equilibrium dynamics for patterns of site occupancy, population persistence, and variability. In Chapter I, I evaluate two methods used to infer historical demography using simulations of genetic data from hypothetical structured populations.

These methods have not been previously evaluated for their performance using explicit structured demes in a metapopulation framework.

**Chapters II and III.** Some of the most important insights into the ecological and evolutionary processes of diversification, colonization, and speciation have come from studies of organisms on islands in the Caribbean (Munroe 1948; MacArthur & Wilson 1963, 1967; Schoener *et al.* 2001; Ricklefs & Bermingham 2008). However, the use of multiple genetic markers in the context of evaluating the null hypotheses of island biogeography, metapopulation genetics, and phylogeography is rare in studies of conservation genetics. Hence, the next objective is to evaluate equilibrium and non-equilibrium dynamics in empirical datasets obtained from island populations in the Turks and Caicos Islands, British West Indies. This archipelago was selected because it contains islands of varying size and isolation with which to test methods for inferring non-equilibrium dynamics in the context of metapopulation theory, island biogeography, and conservation genetics.

Lizards have long been model organisms in the Caribbean, as their wide distributions, high diversity, and ease of capture and observation have led to many published studies ranging from behavioral investigations to informing models of speciation and evolution (Roughgarden 1995; Losos 2009; Camargo *et al.* 2010). In contrast, West Indian Boas of the genus *Epicrates* are poorly studied, frequently rare or cryptic, and are increasingly becoming serious conservation concerns (Reynolds 2011).

Our empirical test system consists of populations (herein defined as a group of individuals inhabiting a single island) of two archipelagic squamate species,

*Leiocephalus psammodromus* and *Epicrates chrysogaster*. *Leiocephalus psammodromus*, the Turks and Caicos curly-tailed lizard, is terrestrial, diurnal, and inhabits 18 or more islands within the Turks and Caicos Banks. This species is conspicuous during foraging, easy to capture, and occupies islands of widely varying size (0.6 km<sup>2</sup> to 124 km<sup>2</sup>) and various degrees of isolation (~0.5 km to ~130km) from other islands. *Epicrates c. chrysogaster*, the Turks Island boa, is a medium-sized nocturnal/crepuscular boine and is the largest native terrestrial carnivore in the Turks and Caicos Islands. It is primarily saurophagous and is a major predator of curly-tailed lizards (Reynolds & Niemiller in press). I have located populations of this species on ten islands in the Turks and Caicos Banks, though boas might occur on one additional island (Reynolds 2011). Though *Epicrates* generally are thought to occur in low population densities (Tolson & Henderson 1993), some populations are noted for remarkable densities of 6-12 individuals or more per hectare (Reynolds 2011). Like the curly-tailed lizards, the Turks Island boas occupy a wide range of island sizes, isolation distance, and isolation time. Lizards have long been model organisms in the Caribbean, as their wide distributions, high diversity, and ease of capture and observation have led to many published studies ranging from behavioral investigations to informing models of speciation and evolution (Roughgarden 1995; Losos 2009; Camargo *et al.* 2010 and references therein).

### *Conservation*

Ecology and evolution are dynamic phenomena, and influenced by temporally variable processes. Therefore, although demographic parameters can be estimated at any given time interval, they are subject to a large degree of variation in subsequent or

previous generations (Holsinger 2000). Hence a historical viewpoint is crucial to understanding the mechanisms that contribute to the evolutionary process. Wright (1931) identified small populations as sources of evolutionary potential, and an understanding of how populations persist, change, disappear, and split through time will allow a deeper understanding of natural processes.

Extinction is a demographic process and is influenced by both stochastic and deterministic forces (Holsinger 2000). The study of historical demography allows us to predict the relative impacts of various dynamics on the species, such as whether populations are subject to local extinction. For example, an understanding of historical demography can help to predict a species' time to extinction (Lande 1988) or to parameterize a population viability analysis (Allendorf & Luikart 2006).

Perhaps most relevant to conservationists, management strategies must derive from accurate historical data in order to achieve maximum effectiveness. Simply knowing where the species occurs and its current dynamics represents a snapshot of the populations, and by definition would ignore the *dynamic* component of population biology. In other words, what is occurring in the species now is not necessarily what has occurred in the past and what will occur in the future. Having a broader understanding of historical non-equilibrium dynamics might allow for better prediction of future dynamics, a critical component of any management strategy (Holsinger 2000).

Finally, in order to protect a threatened island-dwelling species, conservation planners must know what types of mechanisms are influencing population dynamics and demography of that species so that appropriate conservation measures may be implemented. If local extinction is common and recolonization is frequent, protecting

only a few islands may not be sufficient for the species to persist. When populations are subdivided and connected by limited gene flow or dispersal, extinction of local demes and subsequent recolonization may greatly influence genetic differentiation within and among populations of organisms. This is of particular importance when one attempts to establish reserves in order to protect sensitive species occupying subdivided populations, as may occur in island archipelagos.

## Literature Cited

Allendorf FW, Luikart G (2007) Conservation and the genetics of populations. Blackwell Publishing, Malden, Massachusetts.

Andrewartha HG, Birch LC (1954) The distribution and abundance of animals. The University of Chicago Press, Chicago, Illinois.

Bay LK, Caley MJ, Crozier RH (2008) Meta-population structure in a coral reef fish demonstrated by genetic data on patterns of migration, extinction and re-colonization. BMC Evol Biol 8:ARTN 248.

Camargo A, Sinervo B, Sites JW Jr (2010) Lizards as model organisms for linking phylogeographic and speciation studies. Mol Ecol 19:3250-3270.

Caughley G (1994) Directions in conservation biology. J Anim Ecol 63:215-244.

Conner JK, Hartl DL (2004) A Primer of ecological genetics. Sinauer Associates, Inc. Massachusetts, MA.

Darwin C (1859) On the origin of species by means of natural selection or the preservation of favored races in the struggle for life. John Murray, London, UK.

Grenfell B, Harwood J (1997) (Meta)population dynamics of infectious diseases. Trends Ecol Evol 12:395-399.

Gyllenberg M, Hanski I, Hastings A (1997) Structured metapopulation models In: Hanski I., Gilpin M.E., (Eds.), Metapopulation Biology. Elsevier Academic Press, Burlington, Massachusetts, pp. 93-122.

Hanski I, Gaggiotti O (2004) Metapopulation biology: past, present, and future. In: Hanski I., Gaggiotti O.E., (Eds.), Ecology, Genetics, and Evolution of Metapopulations. Elsevier Academic Press, Burlington, Massachusetts, pp. 3-22.

Hanski I, Kuussaari M, Nieminen M (1994) Metapopulation structure and migration in the butterfly *Melitaea cinxia*. Ecology 75:747-762.

Harrison S (1994) Metapopulations and conservation. In: Edwards P.J., May R.M., Webb N.R., (Eds.), Large-Scale Ecology and Conservation Biology. Blackwell Publishing, Malden, Massachusetts, pp. 111-128.

Harrison S, Taylor AD (1997) Empirical evidence for metapopulation dynamics. In: Hanski I., Gilpin M.E., (Eds.), Metapopulation Biology. Elsevier Academic Press, Burlington, Massachusetts, pp. 27-42.

Harrison S, Murphy DD, Erlich PR (1988) Distribution of the bay checkerspot butterfly *Euphydryas editha bayensis*: Evidence for a metapopulation model. *Am Nat* 132:360-382.

Hastings A, Harrison S (1994) Metapopulation dynamics and genetics. *Ann Rev Ecol Syst* 25:167-188.

Holsinger KE (2000) Demography and extinction in small populations. In: Young A.C., Clarke G.M., (Eds.), *Genetics, Demography, and Viability of Fragmented Populations*. Cambridge University Press, Cambridge, UK, pp. 55-74.

Johnson KP, Adler FR, Cherry JL (2000) Genetic and phylogenetic consequences of island biogeography. *Evolution* 54:387-396.

Lande R (1988) Genetics and demography in biological conservation. *Science* 241:1455-1460.

Levins R (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull Entomol Soc Am* 15: 237-240.

Losos JB (2009) *Lizard in an evolutionary tree*. University of California Press, Berkeley, California, USA.

MacArthur RH, Wilson EO (1963) An equilibrium theory of insular zoogeography. *Evolution* 17:373-387.

MacArthur, R.H. and E.O. Wilson. 1967. *The theory of island biogeography*. Princeton University Press, Princeton, New Jersey.

Maruyama T, Fuerst PA (1985) Population bottlenecks and non-equilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics* 111:675-689.

Munroe EG (1948) *The geographical distribution of butterflies in the West Indies*. Ph.D. Dissertation, Cornell University, Ithaca, USA.

Pannell JR (2003) Coalescence in a metapopulation with recurrent local extinction and recolonization. *Evolution* 57: 949-961.

Pannell JR Charlesworth B (1999) Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. *Evolution* 53:664-676.

Planes S, Galzin R, Bonhomme F (1996) A genetic metapopulation model for reef fishes in oceanic islands: The case of the surgeonfish *Cantharis triostegus*. *J Evol Bio* 9:103-117.

Reynolds RG (2011) Status, conservation, and introduction of amphibians and reptiles in the Turks and Caicos Islands, British West Indies. In: Hailey, A., Wilson, B.S., Horrocks, J.A. (Eds.), Conservation of Caribbean island herpetofaunas vol. II: Regional Accounts of the West Indies. Brill, Leiden, Netherlands, pp. 377-406.

Reynolds RG, Niemiller ML (In press) *Epicrates chrysogaster chrysogaster*. (Turks Island Boa). Diet. Herpetol Rev.

Ricklefs RE, Bermingham E (2008) The West Indies as a laboratory of biogeography and evolution. Phil Trans Roy Soc B 363:2393-2413.

Roughgarden J (1995) *Anolis* lizards of the Caribbean: ecology, evolution, and plate tectonics. Oxford University Press, New York.

Schoener TW, Spiller DA, Losos JB (2001) Natural restoration of the species-area relation for a lizard after a hurricane. Science 294:1525-1528.

Sjögren Gulve, P (1994) Distribution and extinction patterns within a northern metapopulation case of the pool frog, *Rana lessonae*. Ecology 75:1357-1367.

Slatkin M (1977) Gene flow and genetic drift in a species subject to frequent local extinctions. Theor Pop Biol 12:253-262.

Slatkin M (1985) Gene flow in natural populations. Ann Rev Ecol Syst 16: 393-430.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. Science 236:787-792.

Thrall PH, Burdon JJ, Murray BR (2000) The metapopulation paradigm: a fragmented view of conservation biology. In: Young A.C., Clarke G.M., (Eds.), Genetics, Demography, and Viability of Fragmented Populations. Cambridge University Press, Cambridge, UK, pp. 75-95.

Tolson PJ, Henderson RW (2006) An overview of snake conservation in the West Indies. Appl Herpetol 3:345-356.

Wade MJ, McCauley DE (1988) Extinction and recolonization: Their effects on the genetic differentiation of local populations. Evolution 42: 995-1005.

Wallace AR (1902) Island life. 3<sup>rd</sup> Ed. Macmillan, London.

Whitlock MC (2004) Selection and drift in metapopulations. In: Hanski I., Gaggiotti O.E., (Eds.), Ecology, Genetics, and Evolution of Metapopulations. Elsevier Academic Press, Burlington, Massachusetts, pp. 153-174.

Whitlock MC, Barton NH (1997) The effective size of a subdivided population. *Genetics* 146:427-441.

Whitlock MC, McCauley MJ (1990) Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution* 44:1717-1724.

Whittaker RJ, Fernández-Palacios JM (2007) *Island biogeography*. 2<sup>nd</sup> Ed. Oxford University Press, New York, USA.

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

Wright S (1940) Breeding structure of populations in relation to speciation. *Am Nat* 74:232-248.

Wright S (1977) *Evolution and the genetics of populations*, Vol. 3, experimental results and evolutionary deductions. University of Chicago Press, Chicago, Illinois.

**PART I. TESTS OF TWO METHODS FOR INFERENCE OF  
HISTORICAL DEMOGRAPHY USING SIMULATED  
STRUCTURED POPULATIONS REVEALS SUBSTANTIAL TYPE II  
ERROR**

The following section is a slightly modified version of a paper to be submitted for publication:

Reynolds, R.G., & B.M. Fitzpatrick. Tests of two methods for inference of historical demography using simulated structured populations reveals substantial type II error.

## **Abstract**

Genetic analysis has been promoted as a way to reconstruct recent historical dynamics ("historical demography") by screening for signatures of events that disrupt equilibrium patterns of variation, such as bottlenecks or migration. Such analyses could also identify "metapopulation" processes like extinction and recolonization or source-sink dynamics, but this potential remains largely unrealized. Here I use simulations to test the ability of two currently used strategies to distinguish between a set of interconnected subpopulations (demes) that have undergone bottlenecks or extinction and recolonization events (metapopulation dynamics) from a set of static demes. The first strategy is the recently proposed method of decomposed pairwise regression (DPR), which provides a holistic test for heterogeneity among demes in their patterns of isolation-by-distance. My tests show that this method suffered from a type II error rate of 59-100%, depending on parameter conditions. The second strategy tests for deviations from mutation-drift equilibrium on a deme-by-deme basis to identify sites likely to have experienced recent bottlenecks (or founder effects). Although bottleneck tests have been shown to have good statistical power for single populations with recent population declines, they have not been tested in a metapopulation context with immigration (or colonization) and population recovery. Simulations of hypothetical metapopulations show that population recovery can rapidly eliminate the statistical signature of a bottleneck, and that even

moderate levels of gene flow can lead to rejection of the null hypothesis of mutation-drift equilibrium in the opposite tail even for demes in equilibrium. Although we did cover all possible metapopulation scenarios, the performance of the tests was disappointing. The results indicate that these methods currently available for identifying historical demographic events yield high type II error if population recovery and/or gene flow are influential demographic feature of the deme, and I suggest that inferences of a lack of historical demography using these methods be interpreted cautiously.

## **Introduction**

Many species exist as structured demes (metapopulations *sensu lato*), or a population of subdivided demes interconnected by gene flow and subject to demographic processes, such as extinction and recolonization or fluctuating population sizes (Levins 1969; Hanski & Simberloff 1997; Hanski & Gaggiotti 2004; Wakeley 2004). If we want to infer the history of demographic events such as these, our sole option aside from direct observation is the analysis of genetic variation in the metapopulation. The basic strategy is to assume that variation is not subject to selection so that patterns can be interpreted in terms of mutation and demographic processes, including migration, drift, extinction, and recolonization (Conner & Hartl 2004). Importantly, demographic processes such as bottlenecks and local extinction and recolonization could have profound effects on the apportioning of genetic variation within and among local demes (Wright 1940; Slatkin 1987), and these dynamical processes remain difficult to evaluate given the large time

scales over which they occur. Very few empirical studies have attempted to use molecular data to infer dynamics at the metapopulation level, instead focusing on inferring dynamics in individual populations (but see Manier & Arnold 2005; Koizumi *et al.* 2006), and quantitative methods are not well-established, though a great deal of theoretical work describes metapopulation genealogical history under the coalescent (Hanski & Gaggiotti 2004).

Theoretical models using coalescent simulations predict that the distribution of neutral genetic variation can be used to detect population turnover or other non-equilibrium dynamics, including metapopulation dynamics (Pannell 2003). When organisms are organized into geographically structured demes linked by gene flow, as in a metapopulation context, processes such as mutation, genetic drift, migration or gene flow, local extinction, and recolonization influence the apportioning of genetic variation within the metapopulation and constitute non-equilibrium metapopulation dynamics (Slatkin, 1987; Whitlock & McCauley 1990; Conner & Hartl 2004). In particular, coalescent models have assessed the influence of metapopulation dynamics on measures such as  $F_{ST}$  (Pannell & Charlesworth 1999; Pannell 2003), though these models do not demonstrate diagnostic differences between demes undergoing recurrent extinctions and recolonizations, bottlenecks, or static equilibrium. Application of the rich theoretical background of metapopulation genealogical history to empirical data is rare, but two general strategies can be followed. A "metapopulation strategy" would attempt to infer the existence of demographic processes from a holistic analysis of a metapopulation dataset. A "population-by-population" strategy would test each component deme or sample site individually for signatures of non-equilibrium dynamics. These strategies

could be used to infer the presence of non-equilibrium metapopulation dynamics in structured demes.

I found one example of each kind of strategy in the literature that I wished to challenge with simulations prior to using them for empirical datasets, representing one promising and one popular method. First, Koizumi *et al.* (2006) proposed a new method: decomposed pairwise regression (DPR), based on isolation-by-distance (see: Wright 1940, 1943; Kimura & Weiss 1964). Traditional isolation-by-distance (see: Nei 1972; Slatkin 1993; Rousset 1997; Hutchison & Templeton 1999) is a regression of the response variable genetic distance ( $F_{ST}$  or Nei's  $D$ ) against the independent variable of geographic distance. In a stepping-stone population at equilibrium, there is a nearly linear relationship expected between  $F_{ST}/(1 - F_{ST})$  and geographic distance between pairs of populations (Rousset 1997). However, slope and intercept estimates can be heavily influenced by the inclusion of "outlier" populations that are exceptionally divergent from or similar to other demes in the metapopulation. A recent bottleneck or founder effect might make a population an outlier by disrupting migration-mutation-drift equilibrium. DPR identifies "putative outliers" (Koizumi *et al.* 2006) by comparing the inclusion and exclusion of populations using Akaike's information criterion (AIC) as an *ad hoc* criterion, though Koizumi *et al.* (2006) acknowledged that their use of AIC was not theoretically justified. In fact, because removing or including putative outlier populations changes the data set, use of AIC or any likelihood-based approach is technically invalid. Whether or not DPR effectively identifies outlier populations in practice has not been evaluated.

Second, deme-by-deme strategies include "neutrality tests," such as Tajima's  $D$  and other summary statistics for DNA sequence variation (Tajima 1989; Fu 1997; Ramos-Onsins & Rozas 2002; but see Wares 2009), and similar tests for allele frequency data (Piry *et al.* 1999). These methods test for deviations from patterns expected under mutation-drift equilibrium (MDE). Natural selection can cause such deviations for any specific locus, but a deviation from MDE common to a large sample of loci is better explained by demographic processes, such as bottlenecks or extinction and recolonization, that theoretically should affect all loci equally. Genetic bottlenecks occur when a population of organisms undergoes a moderate to severe reduction in effective population size ( $N_e$ ), and are diagnosed as deviations from population MDE (Wright 1940; Nei *et al.* 1975; Watterson 1984; Cornuet & Luikart 1996). Because bottlenecks cause a loss of alleles at a faster rate than a loss of gene diversity ("heterozygosity excess," *sensu* Nei 1978; Nei & Kumar 2000), recent bottlenecks can be detected by examining allele frequencies at polymorphic loci within an extant population and testing for reduced allele number and excess gene diversity compared to that expected for an equilibrium population (Maruyama & Fuerst 1985; Cornuet & Luikart 1996; Luikart *et al.* 1998; Piry *et al.* 1999; Marth *et al.* 2004; Li *et al.* 2006).

Several methods for detecting bottlenecks are available, including the popular program BOTTLENECK (Piry *et al.* 1999). This program uses allelic data from individuals in a deme to predict whether a recent ( $2 N_e - 4 N_e$  generations previous) bottleneck has occurred by testing for reduced allelic richness relative to the expected heterozygosity (gene diversity) of a deme in MDE (Piry *et al.* 1999). This method has been shown to be effective for single isolated populations under a simple population reduction model (Fig.

I – 1A), but its performance with more complex demographic history (such as that shown in Figures I – 1B and I – 1C) and a structured-deme context, such as in Figure I – 2, has not yet been evaluated. As empirical populations are subject to complex demographic history, it is important to evaluate the performance of these strategies using more realistic data that might violate standing assumptions of these strategies.

To evaluate the use of both DPR and BOTTLENECK analyses in the reconstruction of historical demography in metapopulations, I challenged both approaches using forward-time simulations of varying initial parameter conditions. Given the limitless variety of potential empirical scenarios, my intention is not to provide an exhaustive power analysis for the use of these approaches, but instead to assess their accuracy for a small set of generalized metapopulation scenarios. The approaches and simulations illustrated here could be customized for specific empirical datasets.

## **Methods**

### *Metapopulation Models*

Forward-time simulations were performed for two 12-deme metapopulation scenarios: a Focal Deme Model (FDM) and a Metapopulation Model (MM) (Fig. I – 2). In the FDM, the focal deme is subjected to either one of two historical demographic events- a bottleneck or an extinction followed by recolonization from neighboring demes – or to no demographic event (equilibrium). Other non-focal demes in the model **do not** experience a demographic event. In the MM, all demes undergo one of the three scenarios, as above,

at varying times (# of generations prior to the sample), to model a situation in which each deme is equally likely to experience extinction and recolonization or a bottleneck.

Both the FDM and MM are based on a basic model of population structure (intermediate propagule pool colonization; Whitlock & McCauley 1990) with a migration rate of either one or ten migrants per generation. In this basic model, nearest neighbors exchange one (or ten) migrant(s) per generation on average, and next nearest neighbors exchange an average of one (or ten) migrant(s) per ten generations.

### *Demographic Scenarios*

Three historical demographic scenarios were modeled for both the FDM and the MM: a bottleneck, an extinction and recolonization event, and an equilibrium scenario.

Equilibrium scenarios consisted of demes within a metapopulation exchanging migrants with no historical demographic events, so that the system is expected to be in mutation-migration-drift equilibrium.

A demographic bottleneck consisted of a reduction in effective population size ( $N_e$ ) from the original deme size,  $N_e(o)$ , to  $N_e = 8$  (Luikart *et al.* 1998), and a subsequent recovery following deterministic discrete time logistic growth:

$$N_{t+1} = N_t + r N_t (K - N_t) \quad (1)$$

with  $K = N_e(o)$  (Fig. I – 1B). An extinction and recolonization event consisted of an extinction event in the deme, followed by recolonization of  $N_e = 8$  and subsequent population growth according to the same logistic growth model (Fig. I – 1C). In the

recolonization event, 90% of colonists arrive from each of the two nearest neighbor demes, while the remaining 10% of colonists arrive from the second nearest neighbor demes.

I used equilibrium effective population sizes of either 100 or 1000 individuals per deme. I initially used the program SIMCOAL2 (version 2.1.2; Laval & Excoffier 2004) to generate simulated data; however, at these small effective population sizes the coalescent approximation as implemented in SIMCOAL appears to be inaccurate and led to exceedingly high false positive rates. Therefore I used forward-time simulations with  $10 \times N_e$  (o) generations prior to any bottleneck or extinction events to ensure the system began in mutation-migration-drift equilibrium.

### *Simulations*

All simulations were executed in the program R version 2.11.1 (R Development Core Team 2010) using the R package MetaSim (Fitzpatrick & Reynolds, 2011). Parameters for the simulations were chosen to correspond to high power of detection in an analysis of heterozygosity excess as determined by Cornuet & Luikart (1996) using methods implemented in the program BOTTLENECK. To evaluate whether my chosen parameters yielded similar detection rates to those of Cornuet & Luikart (1996), I initially simulated bottlenecks in 1200 demes experiencing no migration and no recovery of original effective population size  $N_e$  (o). (Parameter sets 1-4; Fig. I – 1A). In addition, I simulated a series of 1200 independent demes having undergone a bottleneck with no migration between them and complete recovery of  $N_e$  (o) (Parameter set 5; Fig. I – 1B). I simulated Simple Sequence Repeat (SSR, or microsatellite) marker types as 20 unlinked loci under

the stepwise mutation model allowing for homoplasy, with a recombination rate of 0.5 and a mutation rate of  $1/N_e$  (resulting in a consistent population mutation parameter  $\theta = 4$ ). Samples from each deme consisted of 20 SSR loci from each of 20 individuals. I varied the parameters  $N_e$  and event time within the range of detection reported in Cornuet & Luikart (1996). Mean event time is the mean time of the event (bottleneck or extinction and recolonization) in generations previous to the sample. Tests for heterozygosity excess included the Sign Test (ST) and the Standardized Differences Test (SD) (Cornuet & Luikart 1996; Piry *et al.* 1999). Detection was calculated as the proportion of demes out of 1200 for which either (or both) of the tests was significant at  $\alpha = 0.05$  and in the direction of heterozygosity excess (Cornuet & Luikart 1996).

Simulations of the metapopulation scenarios under both the FDM and MM were as follows: for non-equilibrium scenarios, twenty samples of diploid individuals were drawn from each of 12 demes for each simulation. Two hundred simulations of each of 12 parameter sets were run, with parameters varying in a hierarchical fashion (Table I – 1), yielding 2400 simulated demes (48,000 individual genotypes) per parameter set. Twenty SSR loci were simulated for each individual, yielding a mean of  $6.9 \pm 1.5$  alleles per locus ( $N_m = 1$ ) and  $9.5 \pm 1.8$  alleles per locus ( $N_m = 10$ ). Original effective population size was simulated as either  $N_e(o) = K = 100$  or 1000. Event times were generated from random draws from an exponential distribution with a mean of  $\frac{1}{2} K$ , corresponding to  $\tau \approx 0.25$ ; where  $\tau$  is the event time in generations divided by  $2N_e$  (Cornuet & Luikart 1996). The demographic model used in Cornuet & Luikart (1996) is slightly different than my deterministic discrete time logistic growth model, hence my estimates of  $\tau$  are approximations to those of these authors. Cornuet & Luikart (1996)

model  $\tau$  as a temporal value relative to the post-bottleneck population size, whereas my model incorporates logistic growth from the bottleneck  $N_e$  back towards  $N_e$  (o).

Post-event population growth rates  $r = 0.001$  and  $0.0002$  (for  $K = 100$  and  $1000$ , respectively) were chosen so that the number of generations spent with  $N < K/10$  was the same (26 generations). Equilibrium simulations consisted of 20 samples drawn from each of 100 focal populations sampled from 100 metapopulations each with 12 demes.

Each simulation yielded a vector of event times for historical events (non-equilibrium simulations only), a pairwise  $F_{ST}$  matrix for all demes in each simulation, and a file containing 240 genotypes per simulation. Plots of the average heterozygosity (gene diversity;  $H_e$ ) per deme for equilibrium parameter sets were examined to ensure that the mutation rates yielded an appropriate amount of heterozygosity (Supporting Information Fig. I – S1).

## **Analyses**

### *Focal Deme Model- Decomposed Pairwise Regression*

Decomposed pairwise regression was implemented in R using the package DPR (Reynolds & Fitzpatrick, 2011). For the FDM, Rousset's (1997) distance measure ( $F_{ST} / (1 - F_{ST})$ ) was computed for each simulation and subjected to isolation by distance regression in R. A geographic distance matrix was constructed in a linear stepping stone model, with an ordinal increase by one with each step away from the focal population. Decomposed pairwise regression (Slatkin 1993; Rousset 1997; Hutchison & Templeton 1999; Koizumi *et al.* 2006) was performed by decomposing the slope and intercept of the IBD analyses for each simulation into the individual regressions of each deme in each

simulation; then recomposing the mean slope and intercept for each simulation. The slope and intercept of the decomposed regression of the focal deme were ranked relative to those of the other demes in each simulation to identify “putative outliers” (Koizumi *et al.* 2006). The frequency of slope and intercept ranks for each of 200 simulations was recorded and graphed in R, where a rank of 1 indicates the highest value and a rank of 12 indicates the lowest value for a focal deme relative to other equilibrium demes in each simulation. According to the arguments of Hutchison & Templeton (1999; see also Slatkin 1993), bottlenecks and founder effects are expected to increase the intercept and decrease the slope of the population-specific regression of genetic vs. geographic distance because genetic drift is expected to dominate the effects of gene flow during such events. Based on their interpretation of Slatkin (1993) and Hutchinson & Templeton (1999), Koizumi *et al.* (2006) predicted that demes with founder events or bottlenecks would have higher intercepts and/or more shallow slopes relative to equilibrium demes.

Following the method of Koizumi *et al.* (2006), outliers were identified in all parameter sets for the FDM using *ad hoc* Akaike’s information criteria ( $AIC_C$ , Burnham & Anderson 2004) implemented in R. Model selection involved determining either the frequency with which the full model received the lowest  $AIC_C$  score under an equilibrium scenario or the frequency with which the model with the focal deme excluded received the lowest  $AIC_C$  score.  $AIC_C$  scores were corrected for the appropriate degrees of freedom following Koizumi *et al.* (2006). Under an equilibrium scenario, a false positive rate was calculated as the frequency with which models other than the full model were selected out of 200 simulations, indicating in those instances a “false signal” for the identification of outliers. For bottleneck and extinction and recolonization scenarios,

detection rate was calculated as the frequency with which the method correctly identified the true outliers by selecting the model with the outlier excluded.

*Metapopulation Model- BOTTLENECK and Combined Probability Analysis*

For MM simulations, files of genotypes for each parameter set were converted to GENEPOP files (file extension: .gen; Raymond & Rousset 1995) using GENALEX 6.1 (Peakall & Smouse 2006) and imported into the program BOTTLENECK 1.2.02 (Piry *et al.* 1999). Tests for heterozygosity excess included the Standardized Differences Test and Wilcoxon Test under the Stepwise Mutation Model. Average gene diversity ( $H_e$ ) per deme was calculated for each equilibrium parameter set, and probability values were obtained for each test in BOTTLENECK. To determine the effect of event time on recovery of a statistical signature of an historical event, times of each event were plotted against the probability values for tests in BOTTLENECK.

I also tested the potential for using a combined probability analysis (CPA) to reject equilibrium in sets of demes using a Z-transform test (Stouffer *et al.* 1949; Whitlock 2005) on the results from BOTTLENECK. This method combines P-values from multiple tests of the same null hypothesis to determine whether their combined weight can suggest the rejection of the null (Whitlock 2005). One-tailed P-values (P) from the Bottleneck tests were transformed to the standard normal deviate ( $Z_i$ ) drawn from a normal distribution with a mean of 0 and a standard deviation of one. The Z-statistic ( $Z_S$ ) is then calculated as:

$$Z_S = \sum_{i=1}^k \frac{Z_i}{\sqrt{k}} \quad (2)$$

where  $k$  is the number of independent tests of the same null hypothesis (Stouffer *et al.* 1949; Whitlock 2005). When the null hypothesis is true,  $Z_S$  is expected to be distributed as a standard normal random variable (mean = 0 and variance = 1). I evaluated detections for simulations of the MM using the combined probability approach. Here, combined probability analysis is a meta-analysis summary statistic consisting of 12 P-values from each of 200 simulations of the parameter set. Detections are rejections of the null hypothesis of equilibrium for simulations of historical bottlenecks or extinction and recolonization events.

## Results

Initial simulations of demes lacking metapopulation structure indicated detection of non-equilibrium, as demonstrated by Cornuet & Luikart (1996). However, the inclusion of recovery to  $N_e$  (o) (parameter set 5) reduced detection by almost 2/3 (Table I – 2).

### *FDM- Decomposed Pairwise Regression*

Examination of decomposed pairwise regression plots of the FDM simulations did not yield obvious “putative outliers” in most cases (Koizumi *et al.* 2006, see Fig. I – 3).

Ranks of the focal deme did not vary considerably for different parameter sets (Fig. I – 4). The expectation for an equilibrium scenario in this analysis is an even distribution of

ranks with a mean of six. Except for 1 out of 200 sets of simulations of the parameter set  $N_e = 1000$ ,  $N_m = 1$ , in none of the other 1,599 sets of simulations do I find the focal deme to have the highest or lowest slope or intercept relative to other demes (those remaining in demographic equilibrium) in that simulation (Fig. I – 4). However, I do see a trend away from an equilibrium expectation, where the slope is generally lower and intercept is generally higher than other equilibrium demes in the simulations. This would correspond to a trend towards a “Pattern 1” scenario in Hutchinson & Templeton (1999) and Koizumi *et al.* (2006), in which it was predicted that the relative influence of drift might be greater than that of gene flow in determining the relationship between genetic and geographic distance. However, it should be noted that for any given dataset, the rank could not be expected to reliably identify the deme as a “putative outlier.”

Analysis using the *ad hoc* AIC method (Koizumi *et al.* 2006) resulted in a low false positive rate (Type 1 error) and generally poor detection rates (Fig. I – 5). False positives indicate that the method selected equilibrium demes as outliers, thus incorrectly rejecting the equilibrium null model. Identification as an outlier indicates that the method correctly identified as the best model the model that excluded only the focal (non-equilibrium) deme. Identification rate for non-equilibrium parameter sets is thus the number of simulations out of 200 where the correct outlier (non-equilibrium) focal deme was identified. The false positive rate was 3% or less for all equilibrium parameter sets, with the parameter set  $N_e = 1000$ ,  $N_m = 10$  achieving a 0% false positive rate (Fig. I – 5C). Bottleneck simulations yielded a maximum identification rate of 10% for the parameter set  $N_e = 1000$ ,  $N_m = 1$ , while the rest of the parameter sets resulted in detection rates less than 3% and as low as 0% (Fig. I – 5A). For simulations of extinction and

recolonization, a maximum identification rate of 41% was found for the parameter set  $N_e = 1000$ ,  $Nm = 1$ , while the rest of the parameter sets resulted in detection rates less than 11% and as low as 0% (Fig. I – 5B). Overall type II error rates ranged from 59—100% using this method.

#### *MM- BOTTLENECK and Combined Probability Analysis*

Bottleneck tests under the stepwise mutation model had difficulty detecting bottlenecks in all parameter sets, and some parameter sets yielded occasional high false positive rates under MDE (Fig. I – 6). Out of 2400 bottlenecked demes for each parameter set, BOTTLENECK correctly detected non-equilibrium in a maximum of only 205 (8.5%) of the demes under the Wilcoxon test for an extinction and recolonization scenario with  $N_e = 100$  and  $Nm = 10$ . High rates of migration tended to increase detection in all parameter sets, though  $N_e$  did not appear to greatly influence detection of historical events (Fig. I – 6). Timing of the event did not appear to have an impact on detection rates, as the few detections occurred across most event times (Supporting information Fig. I – S4).

I calculated  $Z_S$  for each simulation of SSR data from the MM for two tests in BOTTLENECK (Standardized Differences and Wilcoxon). Simulations for which  $Z_S < -1.645$  indicate the signature of demes that are not in MDE (one-tailed test for heterozygosity excess given allele numbers). Combined probability analysis tended to increase rates of detection slightly for parameter sets with high migration, though almost no detection was observed for parameter sets with lower migration (Table I – 3). The average time of the events appeared to have little impact on the ability of combined probability analysis to detect historical events (Supporting information Fig. I – S5).

The tests implemented in BOTTLENECK are based on approximations that are expected to be most problematic with small effective population size (Cornuet & Luikart 1996). Biases in the test statistics might be amplified by the combined probability tests. Therefore, I also examined false positive rates on a deme-by-deme basis, and found rates ranging from 0 to 9% for combined Standardized Differences and Wilcoxon tests (Fig. I – 6C). False positive rates showed a contrasting pattern, such that when  $N_e$  is small, migration decreases false positive rates, while when  $N_e$  is large, dispersal increases false positive rates (Fig. I – 6C).

## Discussion

The use of genetic data in estimating historical demography of metapopulations has a rich theoretical background but few empirical tests. Two recent approaches are the methods of decomposed pairwise regression to detect non-equilibrium demes in a metapopulation context (Koizumi *et al.* 2006) and the analysis of allele frequency spectra to identify whether a specific population has undergone a recent bottleneck (Cornuet & Luikart 1996; Piry *et al.* 1999). A third approach does exist- Manier and Arnold (2005) used the program MIGRATE (Beerli & Felsenstein 1999; 2001) to simultaneously estimate relative population sizes, immigration, and emigration rates in a set of garter snake populations. Based on asymmetrical immigration/emigration and variation in effective population size, they concluded the system was a source-sink metapopulation. Exploration of the program MIGRATE (Beerli & Felsenstein 1999; 2001) revealed that this method requires an extremely large amount of computing power and time, and hence was not used in this

study. However, this may yet represent a fruitful approach to individuals with access to large computing power.

I evaluated the use of decomposed pairwise regression and analysis of allele frequencies for detection of historical demographic events in metapopulations experiencing moderate to high rates of migration. My results are disappointing, with both approaches suffering from high type II error. I conclude that a population genetic test for metapopulation dynamics remains elusive. Moreover, single population studies over the last decade using bottleneck tests and failing to reject the null of equilibrium appear to be unreliable if migration or population expansion post-bottleneck was a demographic component of the deme. Studies using these methods in single-population scenarios should consider that bottleneck tests are sensitive to violation of the assumptions of isolation and no recovery.

### *Decomposed Pairwise Regression*

Decomposed pairwise regression was proposed as a method to detect outlier demes in an isolation by distance regression using model selection (Koizumi *et al.* 2006). To evaluate whether bottlenecks or founder effects reliably caused a deme to be an outlier, I ranked the slope and intercept of the focal deme relative to other demes to determine whether an historical event consistently yielded a deme with a higher or lower slope or intercept, which would correspond to patterns 1, 2, or 4 in an isolation by distance plot (Fig. 1 from Koizumi *et al.* 2006). Koizumi *et al.* (2006) suggested that the pairwise IBD regression of a recently bottlenecked population might have an elevated intercept owing to accentuated differentiation from nearby sites, but a shallow slope owing to a temporarily reduced

importance of gene flow. I found a tendency for non-equilibrium populations to have shallower slopes in addition to higher intercepts (Fig. I – 4). However, in only one out of 200 simulations did I find that the slope and intercept were ranked higher than all other demes in that simulation. This indicates that examination of an IBD plot for any given empirical dataset similar to these parameters would not identify the focal deme as an outlier relative to other demes in the simulation, even for the parameter set with highest detection (*e.g.*, Fig. I – 5B). This is probably because migration from neighboring demes has a homogenizing effect in opposition to the tendency of drift to make demes different.

As a tool for identifying non-equilibrium populations, I found DPR to be unreliable. In the majority of simulations, the rank of the slope and intercept of the focal deme did not yield detection of a putative outlier. In all cases, some signature of both a bottleneck and an extinction and recolonization was detected as a shift in mean ranks away from the equilibrium expectation of six, though this indicates that the signal would likely be unidentifiable in a single dataset. Using the *ad hoc* AIC method, I recorded a maximum detection rate of 41% for an extinction and recolonization scenario. This shows that focal demes were sometimes aberrant in ways not reflected in their fitted IBD slopes and intercepts. Higher detection in extinction and recolonization simulations vs. bottleneck simulations might be due to the severity of the loss of individuals in the deme, as in those cases where  $N_e$  is reduced to zero, though the colonizing propagule size is equivalent to the bottleneck effective population size ( $N_e = 8$ ). Nevertheless, in most cases, detection rates were indistinguishable from false-positive rates ( $P = 0.22$ ; Fig. I – 5). AIC is clearly inappropriate for this approach (which alters the data included for a single model rather than comparing different models applied to a given data set), but my

analysis of the ranks of focal demes suggests that even a formally proper statistical test would rarely identify the true outlier deme as having an exceptional slope or intercept.

Given these results, it is interesting that Koizumi *et al.* (2006) identified 7 of their 17 study populations as outliers using the *ad hoc* AIC method. At face value, we might conclude that the method is rather conservative and therefore Koizumi *et al.*'s results indicate rather strong evidence for metapopulation dynamics in Dolly Varden salmonids. Importantly, several differences between my simulations and real populations are likely to produce stronger detections. First, as suggested by Koizumi *et al.* (2006), barriers to dispersal will distort simple isolation-by-distance regressions. Second, variation in  $N_e$  among demes will increase variance in pairwise  $F_{ST}$  at equilibrium because  $F_{ST}$  depends on both differences between populations and variation within populations (Beaumont 1996). Biologically, this reflects that fact that the effect of gene flow is more a function of numbers of dispersers ( $Nm$ ) than dispersal rate (Wright 1939). To test the second scenario, I simulated equilibrium metapopulations where  $N_e$  of the focal deme varied relative to other demes in the metapopulation. However, DPR analysis of 100 simulations also failed to detect the focal deme as an outlier even if  $N_e$  of the deme was much larger or much smaller than other demes in the metapopulation (Supporting Information Table I – S1). It is unclear why Koizumi *et al.* (2006) have such strong signatures of non-equilibrium in their data from wild Dolly Varden populations, though it is possible that some demes are separated by major barriers to gene flow (not addressed by geographic distance) while others experience some degree of migration each generation. This would indicate that the signatures of non-equilibrium in those analyses could be due to significant variation in gene flow among demes rather than or in addition to non-

equilibrium dynamics. Future work could explicitly test whether these types of demes are consistently recovered as outliers.

#### *BOTTLENECK and Combined Probability Analysis*

Allele frequency analysis has become a widely used method in conservation for detecting bottlenecks in populations of interest, and the original description for the program BOTTLENECK (Piry *et al.* 1999) generated 538 citations on Web of Science as of March 2011. This method tests the null hypothesis of mutation-drift equilibrium for an isolated population, yet the effects of migration on bottleneck detection have not been thoroughly investigated. Cornuet & Luikart (1996) and Piry *et al.* (1999) caution that a large number of variable microsatellite markers should be used, recommending at least 20, though the methods are commonly applied to smaller data sets (10 or fewer microsatellites). Therefore, I performed simulations of 20 microsatellite markers (under a strict stepwise mutation model) to evaluate the performance of the methods for the kind of data I might expect to collect for my own immediate research interests.

Unfortunately, my simulations show that the methods implemented in BOTTLENECK are unreliable at rejecting the null hypothesis of MDE for demes which have experienced some degree of recovery following a reduction in  $N_e$  and experiencing gene flow. Even for very large populations with low amounts of gene flow, the statistical signature of a bottleneck cannot be detected with these methods. This has major implications for the application, both past and future, of this method in empirical populations. Results of the analysis of microsatellite data using the program BOTTLENECK are often used to inform conservation strategies, and if gene flow or recovery have

occurred in the deme, then even major bottlenecks or extinction and recolonization events might go undetected.

It is possible that moderate amounts of gene flow could tend to increase the number of rare alleles within demes relative to the MDE or bottleneck expectation, such that BOTTLENECK tests would tend to reject the null hypothesis of equilibrium in the opposite direction. Histograms of  $P$ -values from the tests in BOTTLENECK showed a trend towards the opposite tail of the distribution ( $P > 0.8$ ) for demes experiencing moderate amounts of gene flow ( $Nm = 1$ ) but not high gene flow ( $Nm = 10$ ) (Fig. I – 6; Supporting Information Fig. I – S2). Equilibrium metapopulations also showed a trend toward the opposite tail, while independent equilibrium demes (no migration) did not (Supporting Information Fig. I – S3). This indicates that gene flow contributes to the observation and might indicate that the Type II error seen in the tests is an artifact of the sensitivity to calculations of heterozygosity excess relative to the number of alleles. However, it is unclear what methods could distinguish between this artifact and actual acceptance of the null hypothesis.

The causes of these problems are not clear, and there is a need for further research and development of consistent methods that are not misleading when applied to typical data sets. Meanwhile, population genetic assessments of historical demography and metapopulation dynamics should be interpreted with caution. There is a possibility, however, that the tests implemented in BOTTLENECK could be modified to account for detection of non-equilibrium in the opposite direction (*i.e.*, the opposite tail of a probability distribution), as moderate gene flow may introduce rare alleles into the deme thus masking the loss of rare alleles caused by a bottleneck or even increasing the number

of rare alleles beyond what is expected at MDE. This possibility demands attention, as it could increase the reliability of the method if it is further understood how gene flow and population recovery influence the detection of non-equilibrium.

## Literature Cited

- Anderson CNK, Ramakrishnan U, Chan YL, Hadly EA (2005) Serial SimCoal: a population genetic model for data from multiple populations and points in time. *Bioinformatics* 21:1733-1734.
- Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152:763-773.
- Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in  $n$  subpopulations by using a coalescent approach. *Proc Nat Acad Sci* 98:4563-4568.
- Burnham KP, Anderson DR (2004) Multimodal inference: understanding AIC and BIC in model selection. *Sociological Methods Res* 33:261-304.
- Conner JK, Hartl DL (2004) *A Primer of ecological genetics*. Sinauer Associates, Inc. Massachusetts.
- Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144:2001-2014.
- Excoffier L, Novembre J, Schneider S (2000) SIMCOAL: a general coalescent program for simulation of molecular data in interconnected populations with arbitrary demography. *J Hered* 91:506-509.
- Excoffier L, Laval G, Schneider S (2005) ARLEQUIN, version 3.0: an integrated software package for population genetics data analysis. *Evol Bioinforma Online* 1:47-50.
- Fitzpatrick BM, Reynolds RG (2011) *MetaSim*- a package to simulate genetic data in structured demes using a forward-time approach. R package version 1.0.
- Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 133:693-709.
- Hanski I, Gaggiotti O, (eds) (2004) *Ecology, genetics, and evolution of metapopulations*. Elsevier Academic Press, London, UK.
- Hanski I, Simberloff D (1997) The metapopulation approach, its history, conceptual domain, and application to conservation. In: Hanski I., Gilpin ME., (Eds.), *Metapopulation Biology*, Academic Press, San Diego, CA, pp. 5-26.

- Hutchinson DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* 53:1898-1914.
- Kimura M, Weiss G (1964) The stepping stone model of population structure and the decrease of genetic correlations with distance. *Genetics* 49:561-576.
- Koizumi I, Yamamoto S, Maekawa K (2006) Decomposed pairwise regression analysis of genetic and geographic distances reveals a metapopulation structure of stream-dwelling Dolly Varden char. *Mol Ecol* 15:3175-3189.
- Laval G, Excoffier L (2004) SIMCOAL version 2.0: A program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics* 20:2485-2487.
- Levins R (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bull Entomol Soc Am* 15:237-240.
- Li H, Meng SJ, Men ZM, Fu YX, Zhang YP (2006) Genetic diversity and population history of golden monkeys (*Rhinopithecus roxellana*). *Genetics* 164:269-275.
- Luikart G, Allendorf FW, Cornuet JM, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent bottlenecks. *J Hered* 89:238-247.
- Marth GT, Czabarka E, Murvai J, Sherry ST (2004) The allele frequency spectrum in genome-wide human variation data reveals signals of different demographic history in three large world populations. *Genetics* 166:351-372.
- Maruyama T, Fuerst PA (1985) Population bottlenecks and non-equilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics* 111:675-689.
- Nei M (1972) Genetic distance between populations. *Am Nat* 106:283-292.
- Nei M (1978) *Molecular evolutionary genetics*. Columbia University Press, New York.
- Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution* 29:1-10.
- Nei M, Kumar S (2000) *Molecular evolution and phylogenetics*. Oxford University Press, New York.
- Pannell JR (2003) Coalescence in a metapopulation with recurrent local extinction and recolonization. *Evolution* 57:949-961.

- Pannell JR, Charlesworth B (1999) Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. *Evolution* 53:664-676.
- Peakall R, Smouse PE (2006) Population genetic software for teaching and research. *Mol Ecol Notes* 6:288-295.
- Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *J Hered* 90:502-503.
- R Development Core Team (2010) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* 19:2092-2100.
- Raymond M, Rousset F (1995) GENEPOP version 1.2: population genetics software for exact tests and ecumenicism. *J Hered* 86:248-249.
- Reynolds RG, Fitzpatrick BM (2011) *DPR*- a package to implement decomposed pairwise regression in an isolation by distance framework. R package version 1.0.
- Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145:1219-1228.
- Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* 236:787-792.
- Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* 47:264-279.
- Stouffer SA, Suchman EA, DeVinney LC, Star SA, Williams RM Jr (1949) *The American soldier, Vol. 1: adjustment during army life*. Princeton University Press, Princeton, NJ.
- Tajima F (1989) Statistical methods for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585-595.
- Wakeley J (2004) Metapopulation models for historical inference. *Mol Ecol* 13:865-875.
- Wares JP (2009) Natural distributions of mitochondrial sequence diversity support new null hypotheses. *Evolution* 64:1136-1142.
- Watterson GA (1984) Allele frequencies after a bottleneck. *Theor Pop Biol* 26:387-407.

Whitlock MC (2005) Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *J Evol Biol* 18:1368-1373.

Whitlock MC, McCauley MJ (1990) Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution* 44:1717-1724.

Wright S (1940) Breeding structure of populations in relation to speciation. *Am Nat* 74:232-248.

Wright S (1943) Isolation by distance. *Genetics* 28:114-138.

## **Appendix Ia. Figures and Tables**



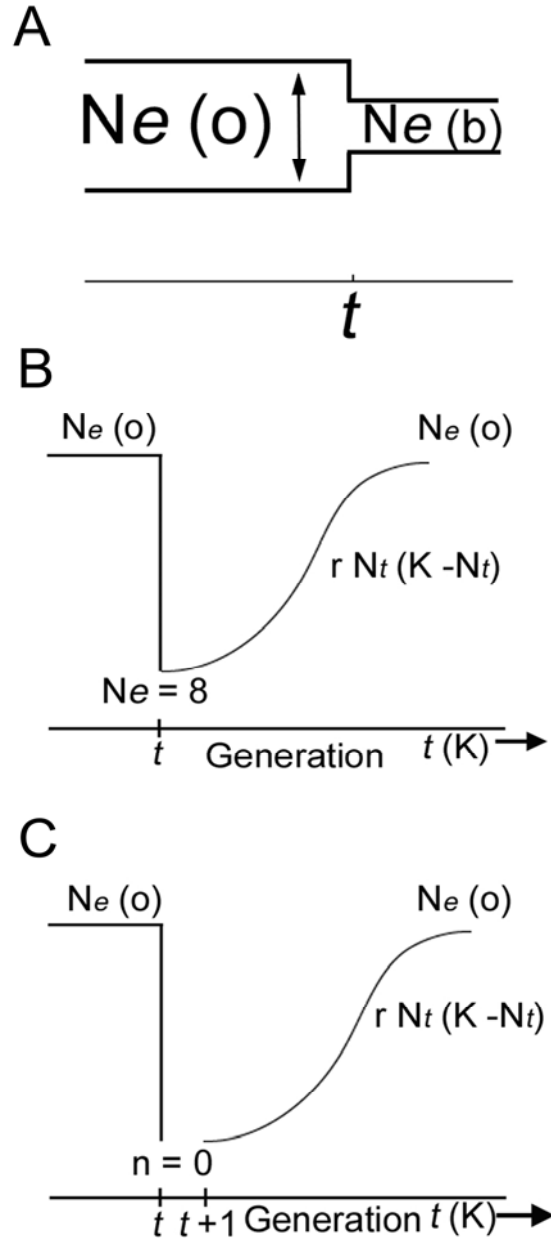


Figure I – 1. Historical demographic event scenarios used in simulations. **A)** Bottleneck scenario used by Cornuet & Luikart (1996). A single population is reduced from  $N_e$  (original) to  $N_e$  (bottleneck) at time  $t$ , with no recovery of  $N_e(o)$ . **B)** Bottleneck scenario used in FDM and MM simulations where population is reduced to  $N_e = 8$  in a single generation at time  $(t)$  and recovers to  $N_e(o)$  at time  $t_K$  following deterministic discrete time logistic growth. **C)** Extinction and recolonization scenario used in FDM and MM simulations where population goes extinct at time  $(t)$ , is recolonized by  $N_e = 8$  individuals from neighboring demes at time  $(t+1)$  and recovers to  $N_e(o)$  at time  $t_K$  following deterministic discrete time logistic growth.

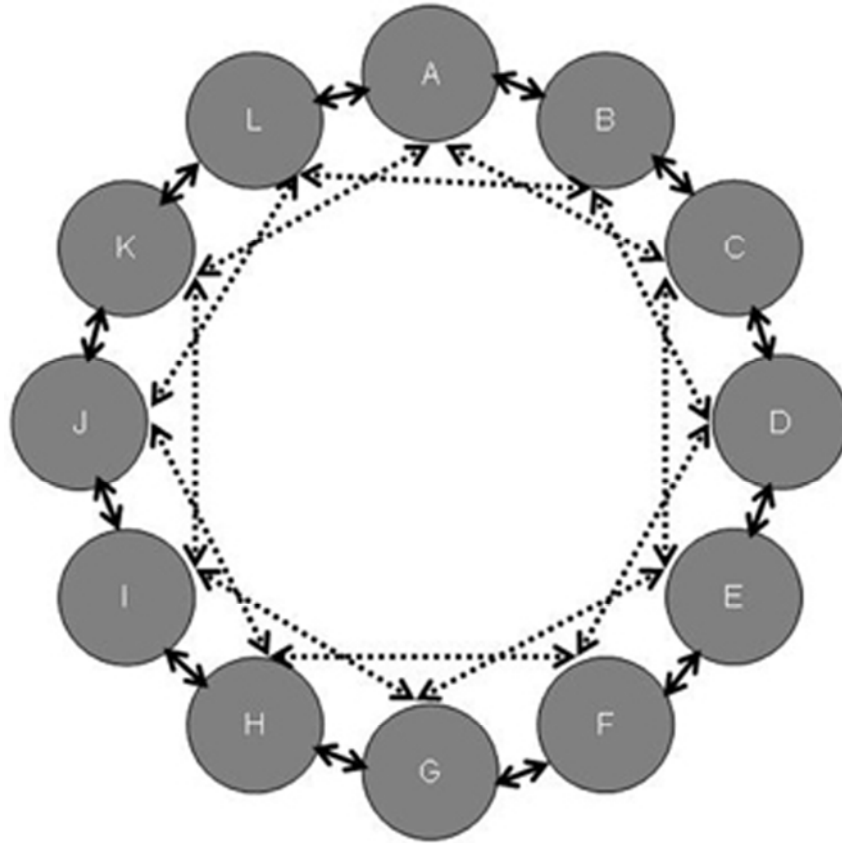


Figure I – 2. Migration scenario used in simulations, where solid arrows indicate exchange of  $Nm$  migrant(s)/generation among nearest neighbors and dotted arrows indicate exchange of  $Nm/10$  migrants/generation among next-nearest neighbors. Focal Deme Model- Intermediate propagule-pool scenario where only deme A (focal deme) undergoes an historical event. Metapopulation Model- Metapopulation scenario where demes A-L are each subject to identical historical events at different times drawn from an exponential distribution.

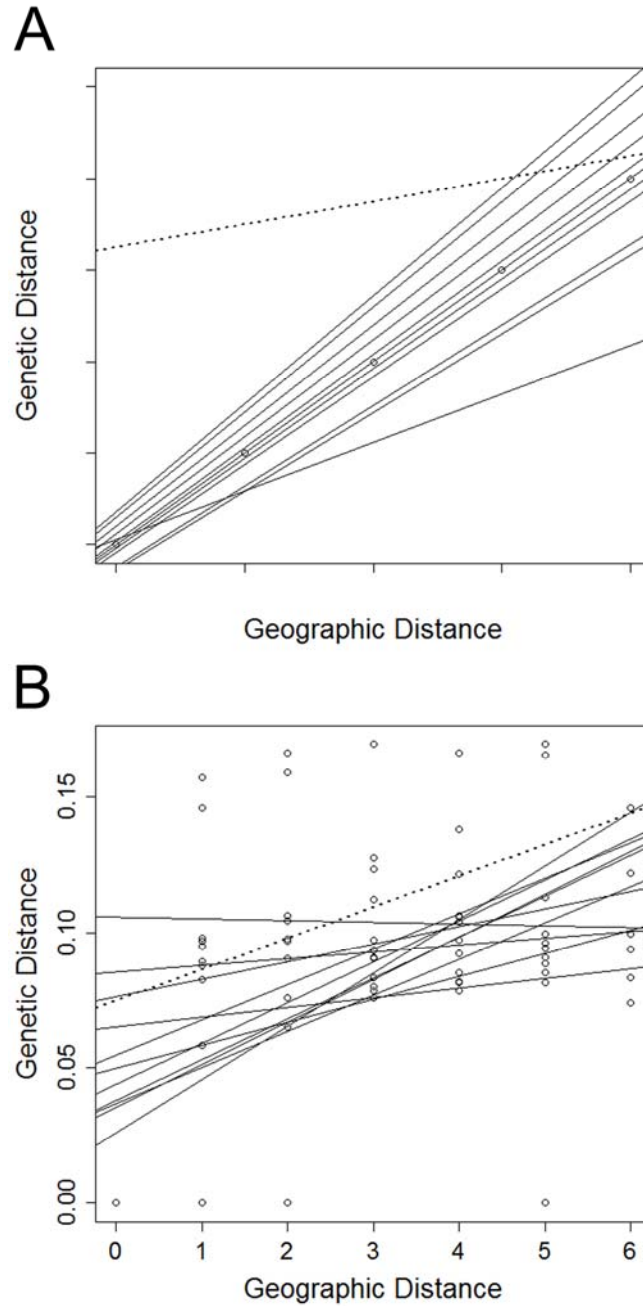


Figure I – 3. Results of decomposed pairwise regression for bottlenecked focal deme. Dotted lines are the individual regressions of the focal demes and solid lines are the remaining 11 non-focal (equilibrium) demes. **A)** Cartoon of the expectation for a non-equilibrium deme as an outlier in the decomposed pairwise regression. **B)** Example output from DPR analysis of the FDM having undergone an extinction-recolonization using parameter set  $N_e(o) = 1000$ ,  $Nm = 1$ . Note that the focal deme is not an outlier, nor does its slope or intercept differ greatly from other demes in the regression.

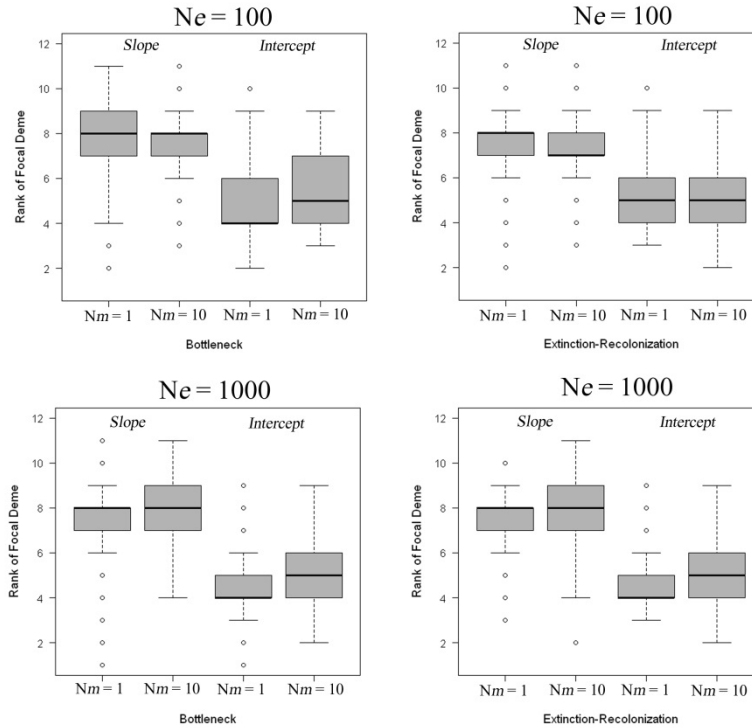


Figure I – 4. Rank of the slope and intercept from the decomposed pairwise regression of the focal deme for 200 simulations of each FDM parameter set. Rank is the numerical rank of the focal deme relative to the other 11 demes in each simulation, where 1 indicates the highest value and 12 indicates the lowest relative to other equilibrium demes. The focal deme has either undergone a bottleneck or extinction and recolonization event. The expectation for an equilibrium scenario is a distribution of ranks with a mean of 6. Note that though they are almost never highest or lowest, the slope is generally lower and intercept is generally higher than other equilibrium demes in the simulations, corresponding to a “Pattern 1” scenario in Hutchinson and Templeton (1999) and Koizumi *et al.* (2006).

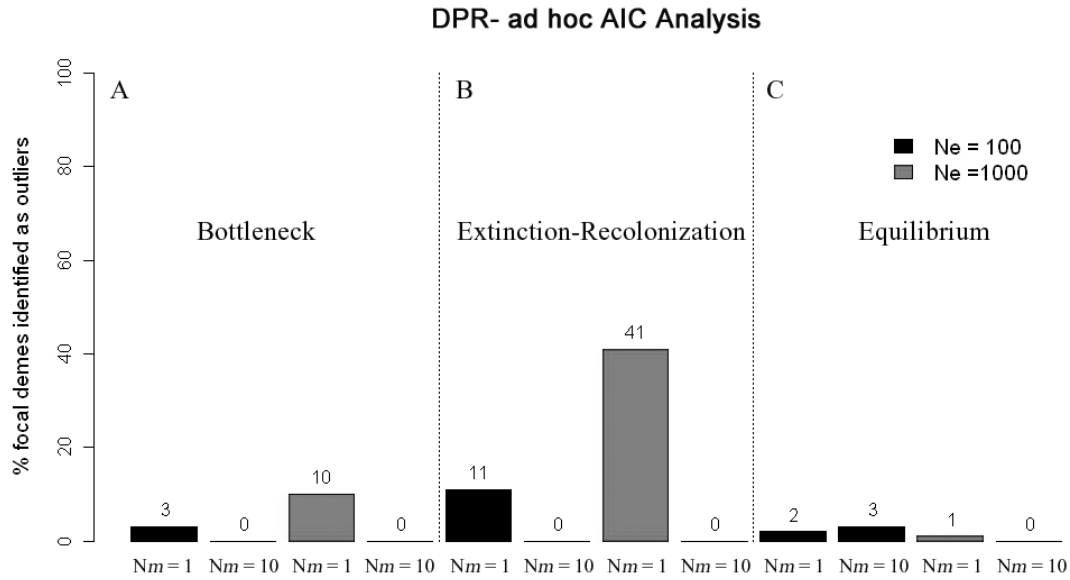


Figure I – 5. Decomposed pairwise regression analysis to detect true outliers. Each bar represents the percentage of focal demes identified as outliers based on 200 simulations (100 simulations for Equilibrium) of a twelve deme focal deme model. Identification as an outlier indicates selection in  $AIC_c$  of the model without the focal deme as a better fit. **A)** Parameter sets where the focal deme underwent a bottleneck result in between 0 and 10% of demes being correctly identified as outliers. **B)** Parameter sets where the focal deme underwent an extinction-recolonization event result in between 0 and 41% of demes being correctly identified as outliers. **C)** False positive rates for equilibrium simulations, where DPR analysis indicated that the removal of one or more populations from the  $AIC_c$  model was a better fit than the complete model.

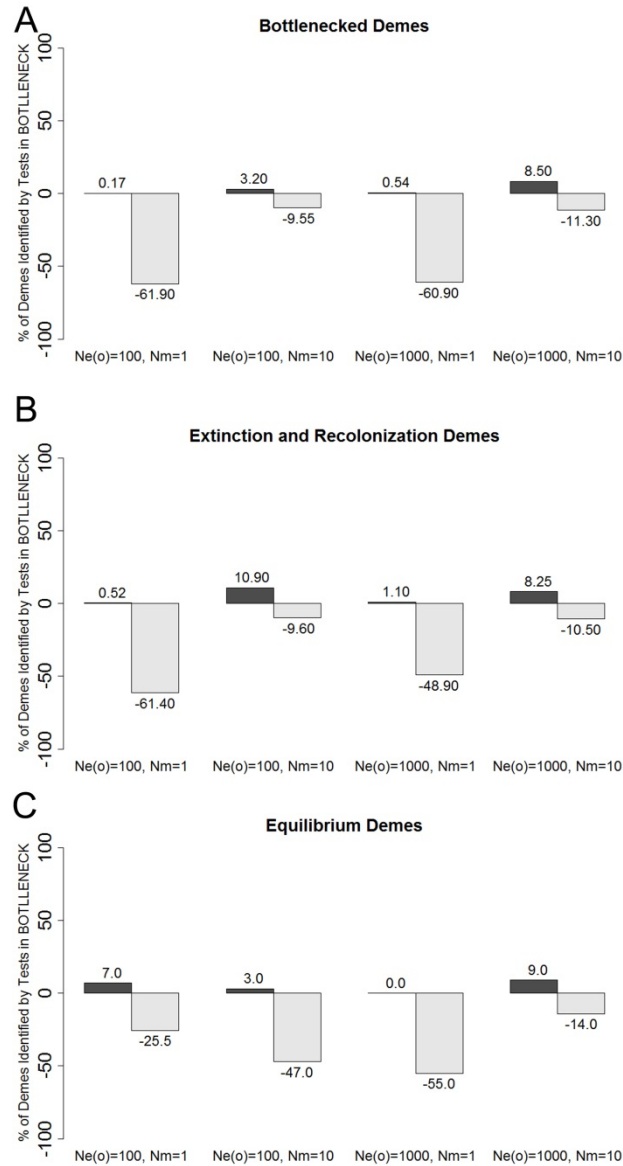


Figure I – 6. Detection of non-equilibrium in the program BOTTLENECK. Bars above the  $x$ -axis indicate detection of heterozygosity excess in one or more of the Standardized Differences and Wilcoxon tests from BOTTLENECK ( $P$ - value of tests  $\leq 0.05$  in the proper direction), while bars below the  $x$ -axis indicate detection of heterozygosity deficiency ( $P \leq 0.05$  in the opposite direction). A) Four parameter sets for demes which have undergone bottlenecks. B) Four parameter sets for demes which have undergone extinction and recolonization events. C) Four parameter sets for demes in mutation-migration-drift equilibrium. Each of the parameter sets for A) and B) consists of 200 simulations of the MM, for a total of 2400 demes per parameter set. Each of the four parameter sets for C) is 200 focal demes drawn from 200 simulations of the MM, for a total of 200 demes per parameter set. Note that detections above the  $x$ -axis in C) represent false positives (Type I error).

Table I – 1. Parameters used in FDM and MM simulations. The parameter space being explored is a 3x2x2 nested set, consisting of event times drawn from an exponential distribution, three historical scenarios, SSR motif markers, and two effective population sizes. Each simulation consisted of a distribution of event times, an event, and 20 SSR loci given a single effective population size. All historical scenarios were simulated for each migration rate and each effective population size for a total of 20 parameter sets. Note that equilibrium scenarios do not require an event time. Equilibrium parameter sets (4) were each simulated 100 times, while non-equilibrium sets (16) were each simulated 200 times, for a total of 3,600 simulations.

| Parameter                   | Value                                                         |
|-----------------------------|---------------------------------------------------------------|
| $K = N_e$ (o)               | 100,1000                                                      |
| $r$                         | 0.001, 0.0002                                                 |
| Migration Rate              | $Nm=1$ , $Nm=10$                                              |
| Historical Scenario         | Equilibrium<br>Bottleneck<br>Extinction and<br>Recolonization |
| Event Time (gens. previous) | $t \sim \exp(\text{rate} = 1 / (0.5K))$                       |
| Marker Type                 | 20 SSR loci                                                   |

Table I – 2. Simulations of bottlenecked demes used to test parameter sets without migration in the program bottleneck. Parameter sets 1-4 experience no recovery of  $N_e$  (o), corresponding to the model of Cornuet & Luikart (1996), while parameter set 5 does experience recovery towards  $N_e$  (o) following deterministic discrete time logistic growth. Detection is reported as the percentage of demes out of 1200 for which bottlenecks were detected in the combined analysis of the Sign Test and Standardized Differences test. Detection of bottlenecks in the absence of migration was similar to that reported for the program by Cornuet & Luikart (1996); however the inclusion of recovery post-bottleneck (parameter set 5) greatly reduced detection relative to parameter set 4.

| Parameter Set   | $N_e$  | Bottleneck Size<br>( $N_e$ ) | Recovery rate       | Mean Event<br>Time<br>(gens. prior) | Detection of bottlenecks | Detection reported by<br>Cornuet & Luikart<br>(1996) |
|-----------------|--------|------------------------------|---------------------|-------------------------------------|--------------------------|------------------------------------------------------|
| 1 (no recovery) | 1000   | 100                          | 0                   | 500                                 | 4.2%                     | 0-20%                                                |
| 2 (no recovery) | 1000   | 100                          | 0                   | 50                                  | 29%                      | 20-40%                                               |
| 3 (no recovery) | 10,000 | 100                          | 0                   | 10,000                              | 71%                      | 40-60%                                               |
| 4 (no recovery) | 100    | 8                            | 0                   | 25                                  | 31%                      | 20-40%                                               |
| 5 (recovery)    | 100    | 8                            | $r^* N_t (K - N_t)$ | 25                                  | 12%                      | —                                                    |

Table I – 3. Results from combined probability analysis using  $Z$  transformation. Each cell contains the percentage of simulations out of 200 for which  $Z_S \leq -1.64$  indicating rejection of the null hypothesis of equilibrium. Two tests from BOTTLENECK under a stepwise mutation model (SMM) were included, with the Wilcoxon test showing higher sensitivity to detection of non-equilibrium, albeit with only a maximum of 21.5% detection for the parameter set  $E/R N_e(o)=100, Nm=10$ .

| Historical Event              | Parameter Set        | Std. Differences | Wilcoxon |
|-------------------------------|----------------------|------------------|----------|
| Bottleneck                    | $N_e(o)=100, Nm=1$   | 0%               | 0%       |
|                               | $N_e(o)=100, Nm=10$  | 3%               | 18%      |
|                               | $N_e(o)=1000, Nm=1$  | 0%               | 0%       |
|                               | $N_e(o)=1000, Nm=10$ | 2%               | 13%      |
| Extinction-<br>Recolonization | $N_e(o)=100, Nm=1$   | 0%               | 0%       |
|                               | $N_e(o)=100, Nm=10$  | 1.5%             | 21.5%    |
|                               | $N_e(o)=1000, Nm=1$  | 0%               | 0.5%     |
|                               | $N_e(o)=1000, Nm=10$ | 0.5%             | 15%      |

## **Appendix Ib. Supplementary Tables and Figures**

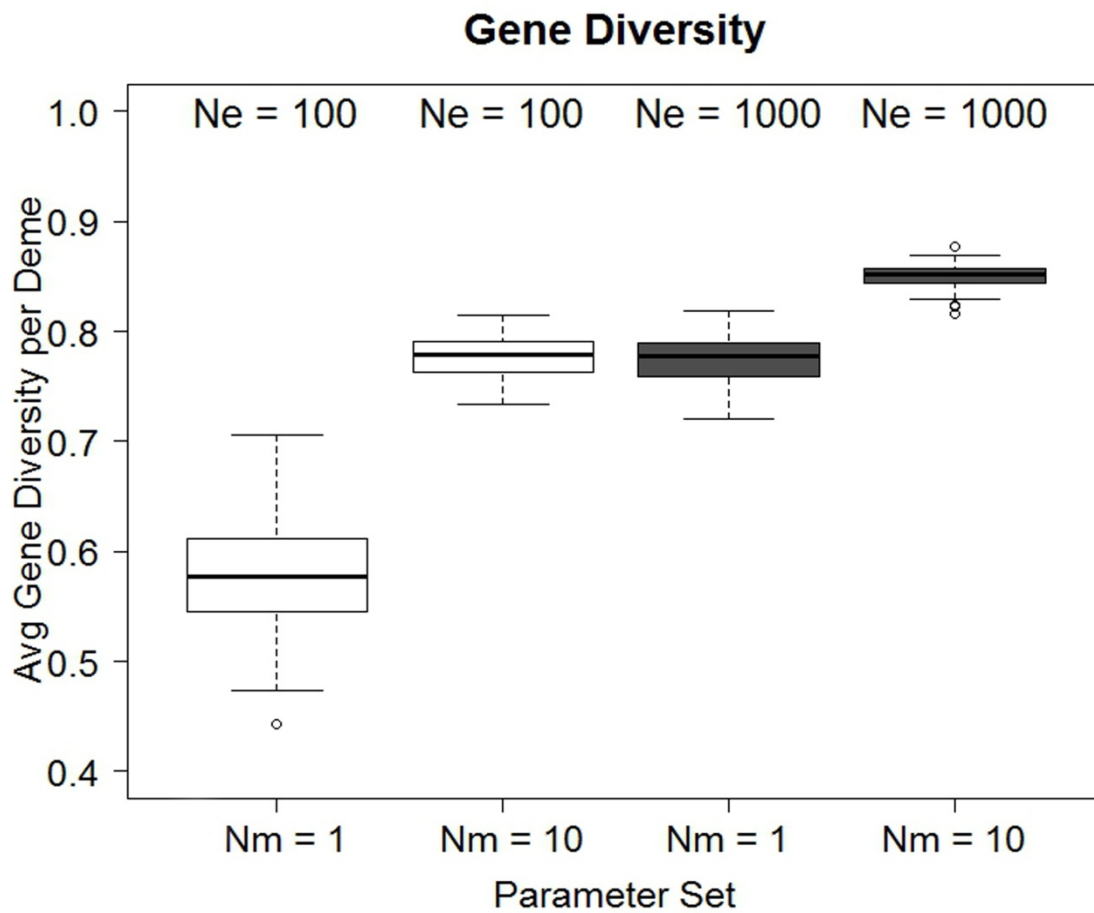
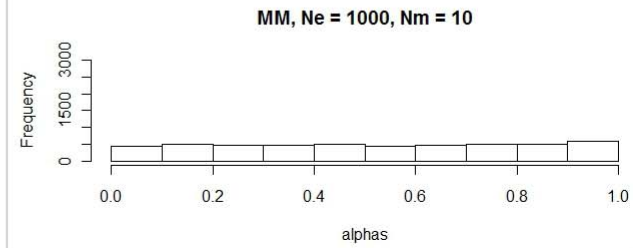
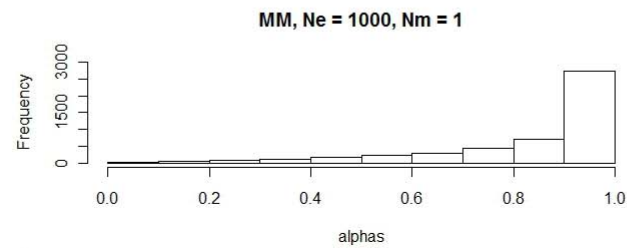
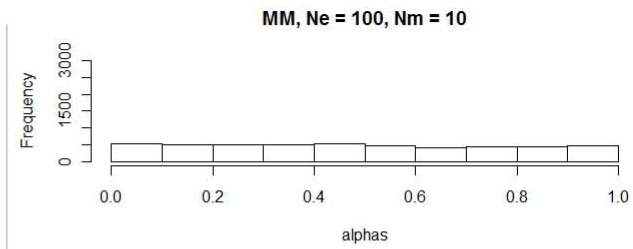
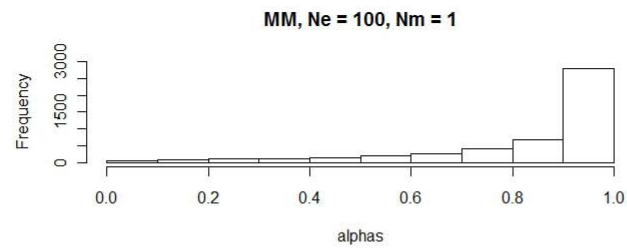


Figure I – S1. Average gene diversity per deme for each equilibrium parameter set of the MM.

# Bottleneck



# Extinction/ Recolonization

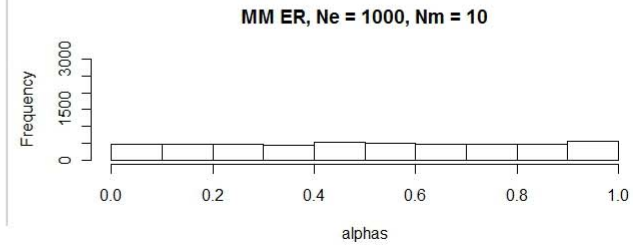
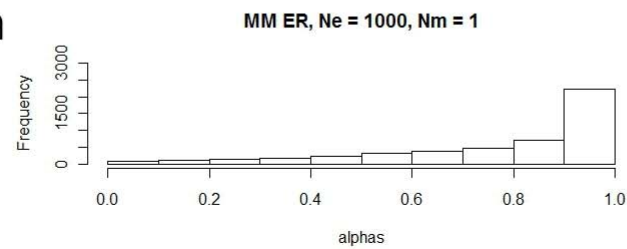
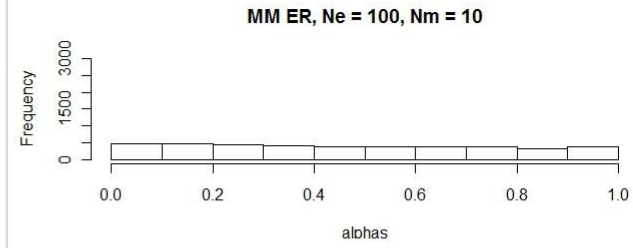
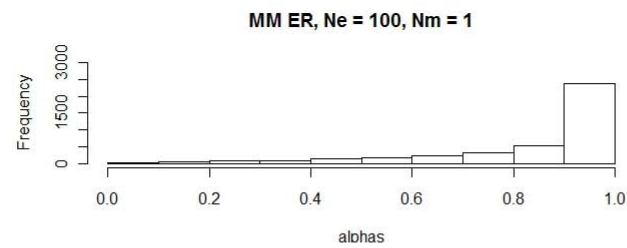


Figure I – S2. Histograms of the combined probability values (SD and Wilcoxon tests) for each deme from the three tests in BOTTLENECK. Each histogram is from 200 simulations of the MM where demes either underwent a bottleneck or an extinction/recolonization event. Note that for moderate levels of gene flow ( $Nm = 1$ ), a significant bias exists towards the opposite end of the distribution of  $P$  values from the tests in bottleneck, while greater levels of gene flow ( $Nm = 10$ ) result in a uniform distribution of  $P$  values.

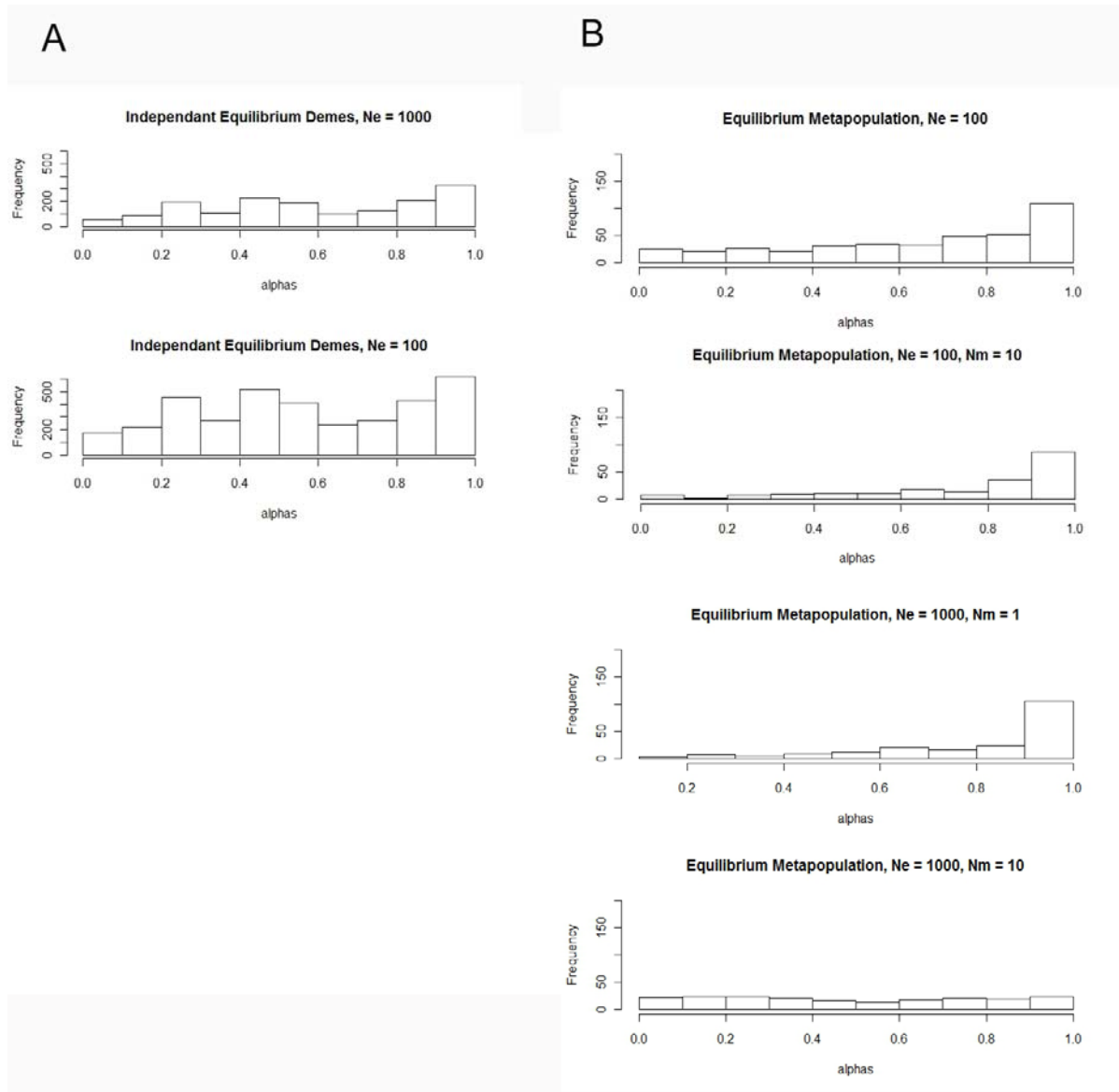
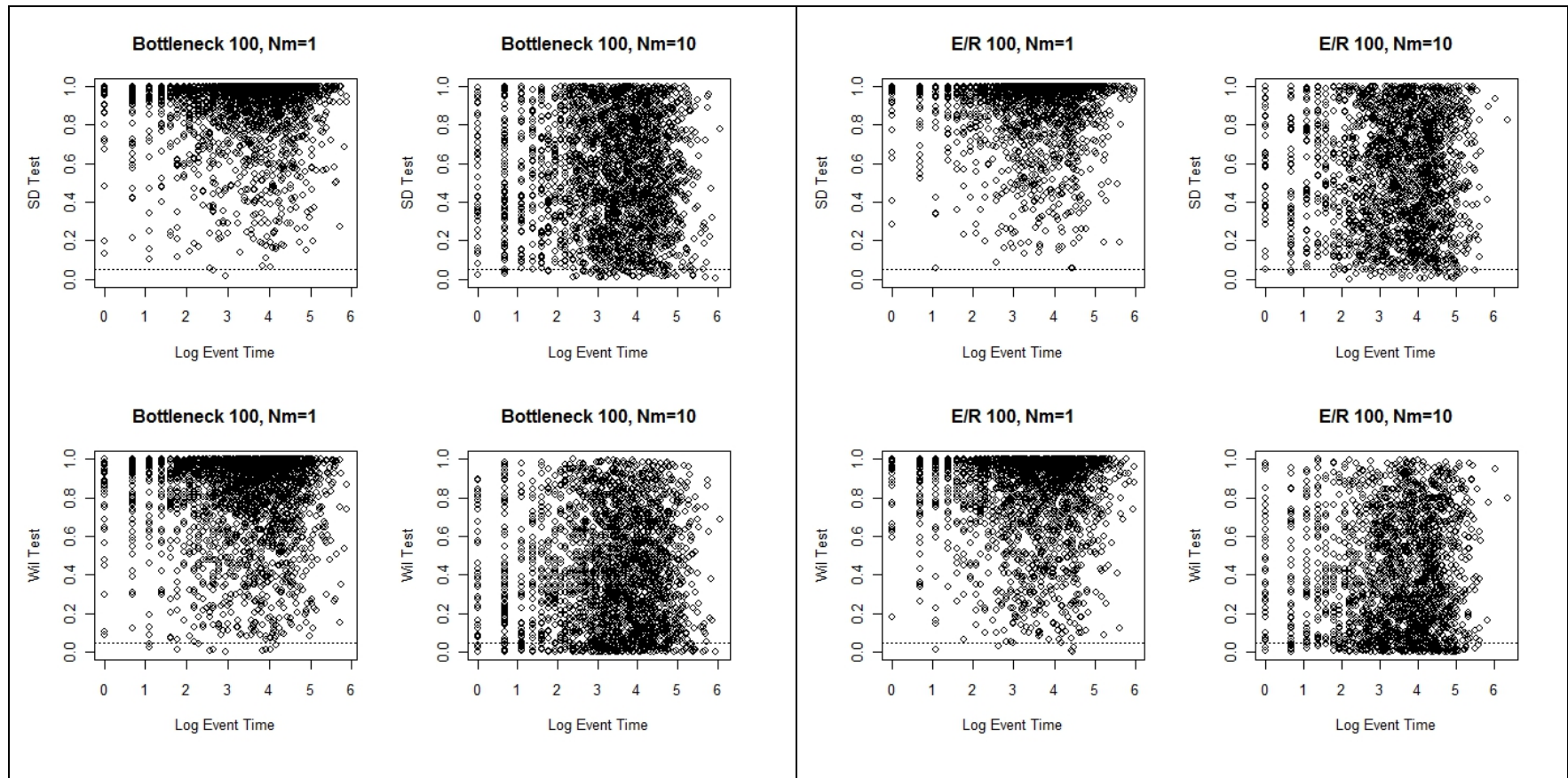


Figure I – S3. Histograms of the combined probability values for each deme from the three tests in BOTTLENECK. A) Independent equilibrium demes with no migration show a more uniform distribution of  $P$  values. B) Equilibrium demes from metapopulations show a trend towards the opposite tail of the  $P$  value distribution ( $< 0.8$ ), except in the case of large population size ( $N_e = 1000$ ) and high migration ( $N_m = 10$ ).



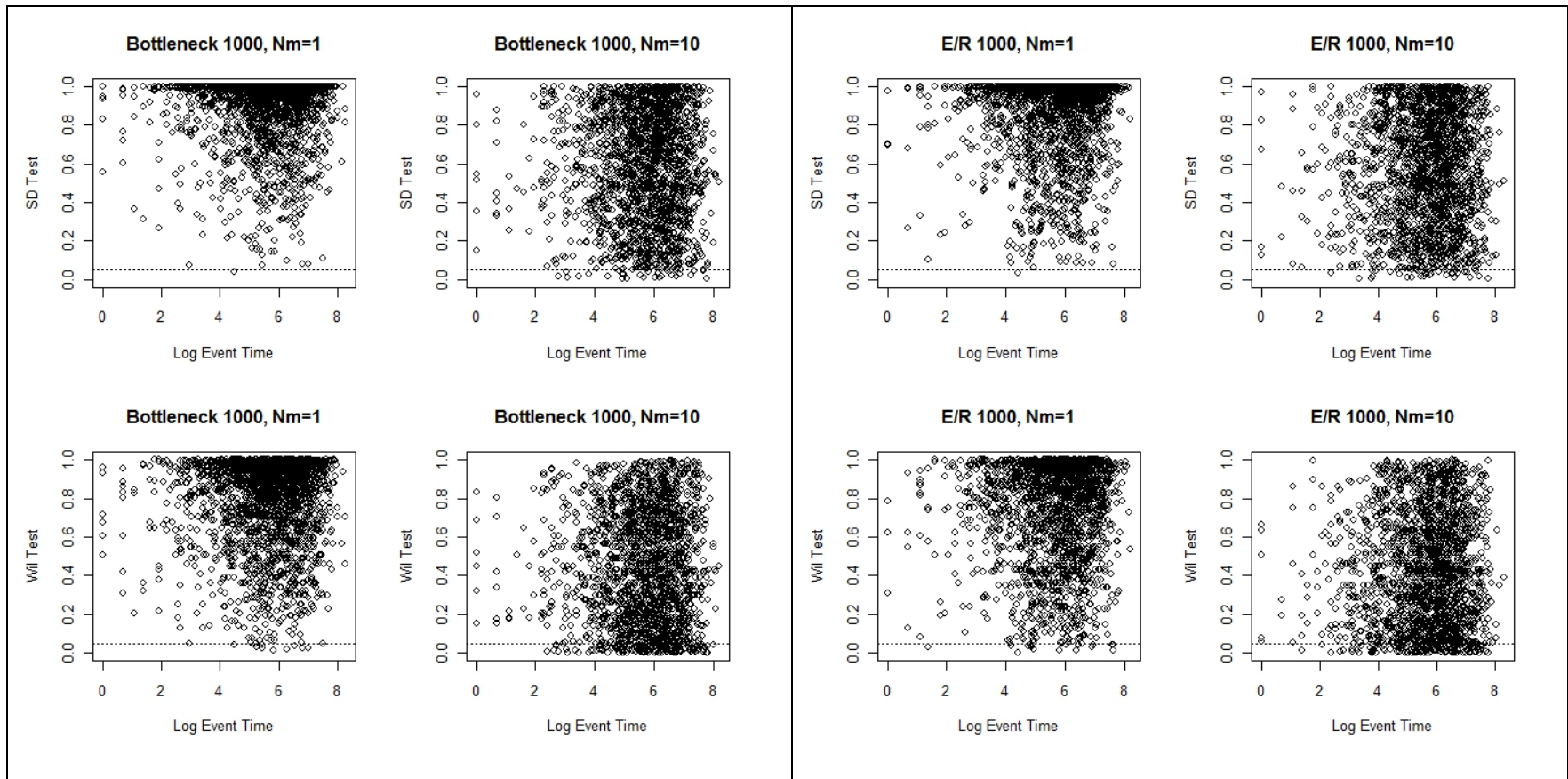
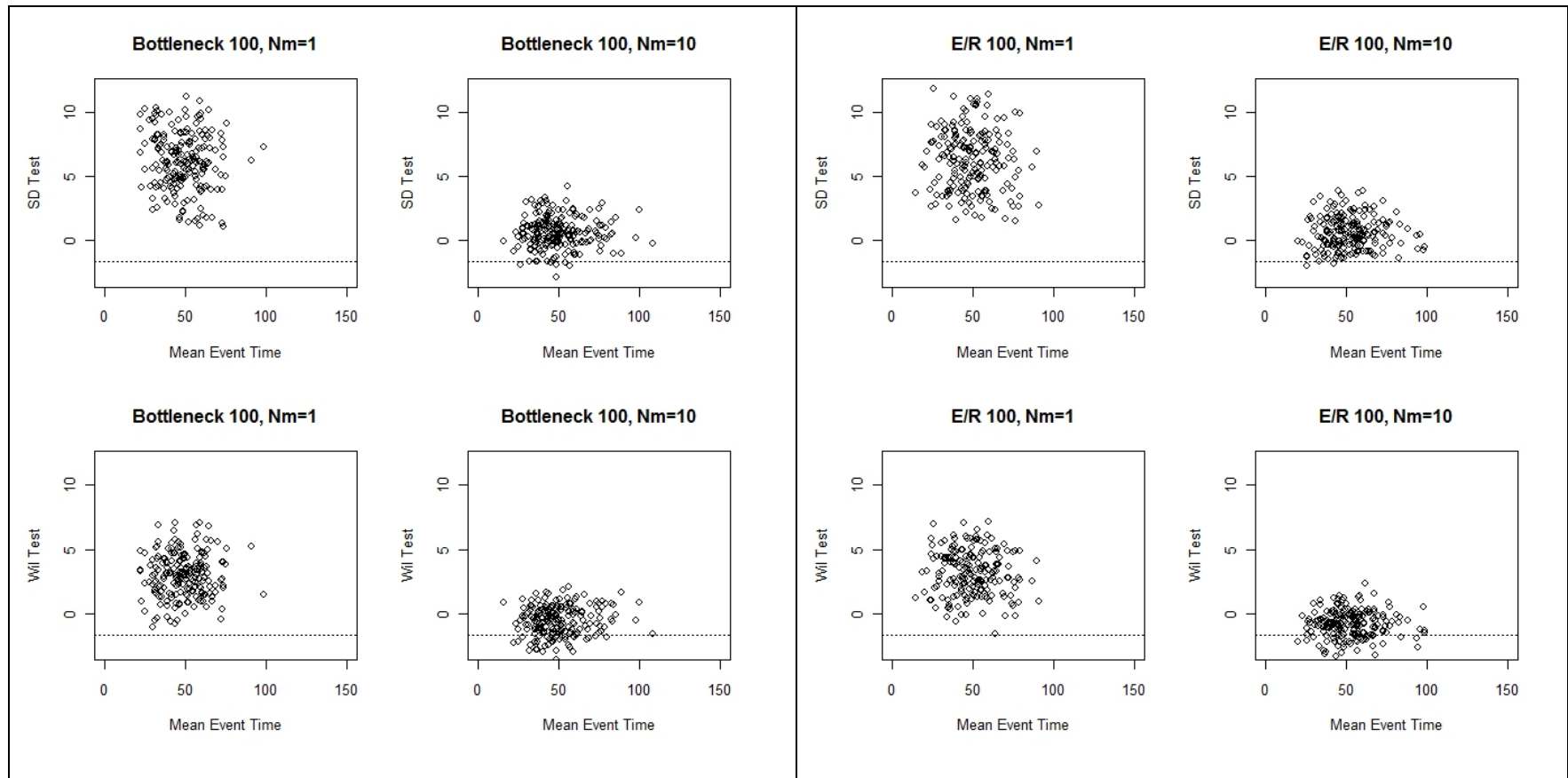


Figure I – S4.  $P$  values from the Standardized Differences and Wilcoxon tests plotted by log of the event time in generations prior to the sample. Note that for moderate migration rates,  $P$  values are distributed towards the opposite tail of the distribution, while high migration rates produce a uniform distribution of  $P$  values. The dotted line represents alpha 0.05.



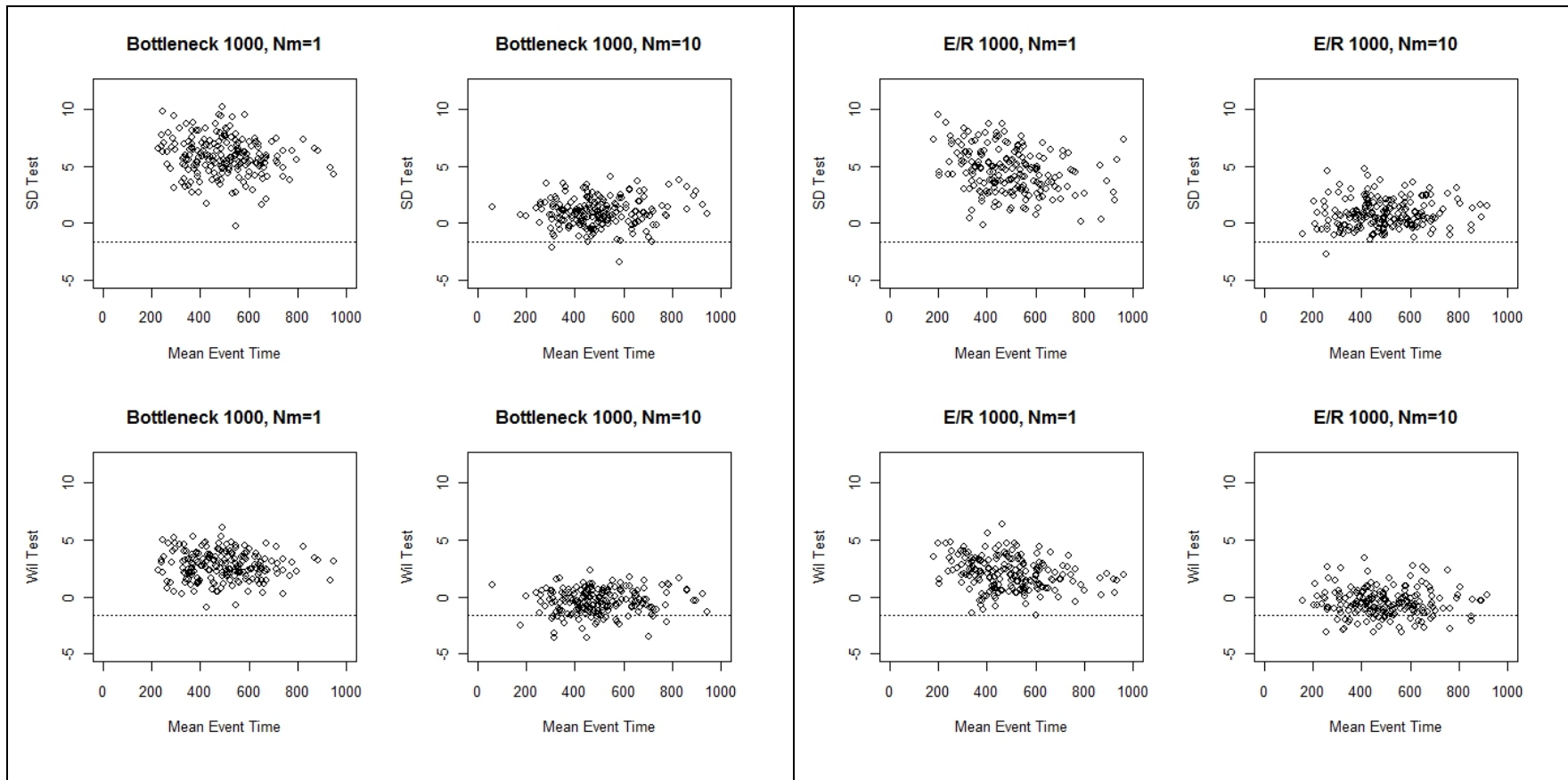


Figure I – S5. Plots of  $Z_S$  for each simulation from the Standardized Differences and Wilcoxon tests in bottleneck on the y-axis against the mean time since the event in generations. A cutoff of  $Z_S = -1.64$  is shown by the dotted line.

Table I –S1. Results for *ad hoc* DPR analysis of 100 simulations of equilibrium metapopulations for each of 8 parameter sets where the focal deme varied in  $N_e$  relative to other demes in the metapopulation.

| Migration Scenario | $N_e$ Focal Deme | $N_e$ Other Demes | % Detection of Focal Deme as an Outlier |
|--------------------|------------------|-------------------|-----------------------------------------|
| No Migration       | 100              | 100               | 11%                                     |
| No Migration       | 1000             | 1000              | 5%                                      |
| $Nm = 1$           | 50               | 100               | 3%                                      |
| $Nm = 10$          | 50               | 100               | 1%                                      |
| $Nm = 1$           | 500              | 100               | 1%                                      |
| $Nm = 10$          | 500              | 100               | 0%                                      |
| $Nm = 1$           | 100              | 1000              | 0%                                      |
| $Nm = 10$          | 100              | 1000              | 4%                                      |

## **PART II. UNEXPECTED SHALLOW GENETIC DIVERGENCE IN TURKS ISLAND BOAS (*EPICRATES C. CHRYSOGASTER*) REVEALS SINGLE EVOLUTIONARILY SIGNIFICANT UNIT FOR CONSERVATION**

### **Preface**

In Chapter I, I explored the use of what we initially identified as promising approaches to identifying historical demographic processes in structured populations. These methods appear to suffer from high type II error rates in demes undergoing expansion or subject to gene flow. My intention was to use these methods to evaluate genetic data in insular reptiles, however population recovery and gene flow could not be ruled out as occurring in these populations. Hence, I attempt to establish a comprehensive approach to using genetic data to estimate population demographics in structured populations using existing methods from phylogeography, population genetics, and island biogeography. Given an ability to collect basic genetic data from a threatened island species, I attempt to extract as much information as possible regarding population dynamics to inform conservation policy for this species.

The following section is a slightly modified version of a paper submitted for publication:

Reynolds, R.G., G.P. Gerber, & B.M. Fitzpatrick. Unexpected shallow genetic divergence in Turks Island Boas (*Epicrates c. chrysogaster*) reveals single evolutionarily significant unit for conservation.

## **Abstract**

The Turks Island boa (*Epicrates c. chrysogaster*) is endemic to the Turks and Caicos Islands and is currently known from only ten islands. The subspecies has likely been extirpated from several islands in its historic range, and all remaining populations are threatened with extirpation owing to habitat loss, introduced feral predators, malicious killing, and vehicle strikes. To assist conservation efforts, I undertook a genetic analysis of 53 individual *E. c. chrysogaster*, representing five island populations, with the goal of identifying existing population structure and genetic diversity. For each snake sampled, I sequenced one mitochondrial and two nuclear genes, resulting in 1,591 bp of sequence, and screened nine microsatellite loci. All individuals were found to be monomorphic at the four microsatellite loci that amplified, and only three individuals were found to vary (by a single nucleotide polymorphism) in either nuclear gene. Nine mitochondrial haplotypes were found, with a maximum sequence divergence of <1%. Taken together, these data indicate shallow genetic divergence in this subspecies, possibly owing to a lack of historical population structure and low population size when the Turks and Caicos Banks were each a single island during the last glacial maximum. *Epicrates c. chrysogaster* appears to represent a single evolutionarily significant unit, a significant finding suggesting that conservation strategies focusing on ecologically intact

populations might be more appropriate than alternate strategies involving near-impossible reversal of declining populations on heavily disturbed islands. In addition, reintroduction programs would likely not disrupt any significant historical population structure.

## **Introduction**

The importance of genetic data in biodiversity conservation is well established, and knowledge of standing genetic variation within and between populations of species is important in designing conservation strategies (Frankham *et al.* 2002; Frankham 2006; Allendorf & Luikart 2007; Caballero *et al.* 2010). The effects of stochastic, demographic, and deterministic (*e.g.*, natural selection) evolutionary forces have been shown to influence genetic variation both within and between populations of organisms (*e.g.*, Conner & Hartl 2004). When populations are separated from each other, as in island archipelagos, processes such as mutation, genetic drift, migration or gene flow, local extinction, and recolonization might dramatically influence the apportioning of genetic variation within geographically structured populations (Wright 1977; Slatkin 1977, 1985, 1987; Wade & McCauley 1988; Whitlock & McCauley 1990; Conner & Hartl 2004). Thus, understanding the distribution of genetic variation in an archipelagic species can shed light on evolutionary processes operating in that system and might be used to develop conservation management strategies that minimize disruption of those processes. In particular, translocation has been an important tool in the recovery of high-profile species such as Peregrine Falcons (Tordoff & Redig 2001; Cade & Burnham 2003), Gray

Wolves (Fritts *et al.* 1997), Bald Eagles (Nye 1988), Seychelles Kestrels (Watson 1989), and Turks and Caicos Rock Iguanas (Gerber 2007). However, many caution against translocations that might disrupt natural genetic structure (Storfer 1999; Avise 2004; Allendorf & Luikart 2007).

The West Indies is one of the world's biodiversity hotspots (Myers *et al.* 2000; Smith *et al.* 2005). The region has a diverse herpetofaunal assemblage and these species are of particular conservation concern owing to their vulnerability to anthropogenic threats, especially habitat destruction and the introduction of invasive mammalian predators such as House Cats (*Felis silvestris catus*), Small Indian Mongoose (*Herpestes auro-punctata*), and Black Rats (*Rattus rattus*), among others (Iverson 1978; Corke 1992; Smith *et al.* 2005; Tolson & Henderson 2006; Hailey *et al.* 2011). The Turks and Caicos Islands (TCI), located at the southern terminus of the Bahamian Archipelago, contain a unique endemic herpetofauna (Reynolds & Niemiller 2010a; Reynolds 2011) that is threatened by increasing development and population growth. The TCI consist of more than 200 islands, ranging in size from < 1 ha to > 12,000 ha, distributed on two shallow banks (Fig. 1). The Turks and Caicos Banks were emergent within the last 8,000-15,000 years, though the two banks are separated by the narrow and deep Turks Island Passage and have never been joined (Lighty *et al.* 1982; Fairbanks 1989; Welch *et al.* 2004). The Turks Island Boa, *Epicrates c. chrysogaster*, is an endemic nocturnal/crepuscular boine and the largest native terrestrial carnivore in the TCI. Boas have been reported from ten islands in the TCI and might occur on at least one additional island (Reynolds 2011). Very little is published about the ecology, distribution, and natural history of this species,

though this information is now becoming available (Reynolds 2011). Importantly, Turks Island Boas face a plenitude of threats, including development and associated habitat loss, introduced predators, direct persecution, and vehicular traffic (Reynolds 2011; Reynolds & Niemiller 2010a). Populations are likely extirpated from several islands, including Grand Turk and South Caicos, while populations on Providenciales and Middle Caicos are likely in a sharp decline (Reynolds 2011). Big Ambergris Cay, a stronghold for the species and home to the highest known density of *Epicrates* in the Bahamian Archipelago (Tolson & Henderson 2006; Reynolds 2011), is currently undergoing extensive development and boas there are being monitored annually.

Other reptiles in the region, such as the Turks and Caicos Rock Iguana (*Cyclura carinata*), are facing similar threats and extirpations (Iverson 1978; Gerber & Iverson 2000; Reynolds 2011). As a result, they are being intensively managed and translocated to some of the outer islands that are free from introduced predators and development (Gerber 2007). To help develop conservation strategies for *Epicrates*, I sampled boas from islands on the Turks Bank and the Caicos Bank to assess standing genetic diversity among populations and to investigate whether separate evolutionarily significant units (ESUs – Ryder 1986, Waples 1991, Dizon *et al.* 1992, Mortiz 1994, Crandall *et al.* 2000), or populations genetically differentiated enough to be managed separately, exist on different islands. In particular, I was interested in the genetic uniqueness of individual populations and whether significant population structure existed on the Caicos Bank or between the Turks and Caicos Banks.

## Methods

*Sample collection and DNA extraction.* I visited seven of the ten islands from which *Epicrates c. chrysogaster* is currently known (Reynolds 2011). Sampling involved diurnal and nocturnal visual encounter surveys conducted between 0700h and 0230h. Diurnal surveys included turning over and replacing cover objects such as rocks, logs, and palm fronds; while nocturnal surveys were conducted by walking in suitable habitat using battery powered headlamps and handheld flashlights. I found boas on five of the seven islands visited, though more focused sampling occurred on some islands owing to logistics (Table 1). Boas proved extremely difficult to find on Providenciales and Middle Caicos, and logistics prevented extensive sampling of East Caicos, Gibbs Cay, and Little Ambergris Cay. Tissue samples were collected from wild *E. chrysogaster* from five different island populations (Table 1, Fig. 1): Big Ambergris Cay (N = 34), Gibbs Cay (N = 3), North Caicos (N = 10), Middle Caicos (N = 1), and Providenciales (N = 3, plus 2 offspring of one sampled adult). Samples were preserved in 95 % ethanol and stored at -20° C. Tissue was obtained by either clipping the distal 3-4 mm of the tail or clipping 3-4 ventral scales. Tails were sanitized before and after clipping and antiseptic dermal adhesive was applied to prevent infection. Tail clips appeared minimally invasive and were found to be preferable to ventral scale clips, as many wild individuals are found with missing tail tips (Reynolds & Gerber submitted), and this method does not have the potential to interfere with locomotion. Whole genomic DNA was extracted using the Wizard SV DNA purification system (Promega, Madison, WI) according to the manufacturer's protocol and stored at -20° C.

*DNA sequencing.* The polymerase chain reaction (PCR) was used to amplify a portion of the mitochondrial genome (617 bp of cytochrome B; CytB) as well as two nuclear intron loci: 518 bp of neutrophin-3 (NT3) and 456 bp of oocyte maturation factor (*c-mos*) (Appendix 1). The CytB fragment was amplified using primers CYTBECF (5'-GCCCCACCAACAAATCCTA-3') and CYTBECR (5'-GGTCTGGTATGGGTGGAATG-3') developed from a congener sequence on Genbank (Campbell 1997; U69803). The NT3 fragment was amplified using primers NT3-F3 (5'-ATATTTCTGGCTTTTCTCTGTGGC-3') and NT3-R4 (5'-GCGTTTCATAAAAATATTGTTTGACCGG-3') from Noonan & Chippindale (2006 *a*, *b*). The *c-mos* fragment was amplified using primers CMOS-Fsnk (5'-GCTGTAAACAGGTGAAGAGATGCAG-3') and CMOS-Rsnk (5'-AGCACGATGGGTGTATGTTCCCCC-3') from Noonan & Chippindale (2006 *a*, *b*). PCR amplifications were conducted in a total volume of 24 µl consisting of: 10.4 µl ddH<sub>2</sub>O, 5.0 µl 10x buffer, 2.5 µl MgCl<sub>2</sub>, 1.5 µl dNTPs, 1.25 µl forward primer (10 µM), 1.25 µl reverse primer (10 µM), 0.125 µl *Taq* polymerase, and 3.0 µl DNA template. PCR reactions were done in an Eppendorf Mastercycler with the following conditions: denaturation at 94°C for 5 min; followed by 35 cycles at 94°C for 30 s, 52°C for 30 s (CytB and NT3) or 61°C for 30 s (*c-mos*), 72°C for 1 min., and final extension at 72°C for 5 min.

PCR products were visualized by gel electrophoresis and purified using exonuclease I/shrimp alkaline phosphatase (ExoSAP). Sequencing reactions were

performed at a volume of 12 µl consisting of: 8 µl ddH<sub>2</sub>O, 2 µl original forward primer, and 2 µl PCR product and resolved on an automated sequencer (Applied Biosystems Inc. ABI 377) at the Molecular Biology Resource Facility at the University of Tennessee, Knoxville (CytB and NT3) or at the High Throughput Genomics Unit at the University of Washington, Seattle (*c-mos*). Sequences were aligned in Sequencher 4.6 (Gene Code Corporation), and ambiguous base calls were manually verified by examining electropherograms. Sequence alignment files were verified and trimmed in MacClade 4.07 (Maddison & Maddison 2005) by comparison to published sequences for *Epicrates*. I will deposit all sequences in GenBank (Appendix 1).

*Microsatellite genotyping.* Polymorphic nuclear markers have not been developed for *E. chrysogaster*, hence PCR was used to screen nine microsatellite markers developed for the congener *E. subflavus* (Tzika *et al.* 2008a) (Appendix 1). Each primer set included a 5' "tail" (GTTTCTT) added to the forward primer to force "plus-A" addition (Brownstein *et al.* 1996). PCR amplifications were conducted in a total volume of 24 µl consisting of: 10.4 µl ddH<sub>2</sub>O, 5.0 µl 10x buffer, 2.5 µl MgCl<sub>2</sub>, 1.5 µl dNTPs, 1.25 µl forward primer (10 µM), 1.25 µl reverse primer (10 µM), 0.125 µl *Taq* polymerase, and 3.0 µl DNA template. PCR reactions were done in an Eppendorf Mastercycler (see Appendix 1 for PCR conditions and primers). PCR products were visualized by gel electrophoresis, and those with distinct bands were genotyped on an EGene<sup>®</sup> multicapillary electrophoresis system at the University of Tennessee, Knoxville. Peaks were cleaned and called using Biocalculator (Qiagen, Crawley, UK). One PCR product for each primer set that was

successfully genotyped was sequenced to ensure that the repeat region was included in the fragment.

*Gene tree analyses.* To eliminate redundant haplotypes I used the online toolbox FaBox (Villesen 2007) to collapse sequences into haplotype groups. A maximum likelihood (ML) tree search was executed using PAUP\* 4.0 (Swofford 2002) and a Bayesian inference was implemented in the program MRBAYES 3.1.2 (Hulsenbeck & Ronquist 2001; Ronquist & Hulsenbeck 2003). The best fit model of sequence evolution was determined with Akaike's information criterion (AIC) implemented in ModelTest Server 3.8 (Posada 2006). The sister taxon *Epicrates exul* from the northern Bahamas was included in the analysis as an outgroup to evaluate genetic divergence among haplotypes of *E. chrysogaster*. ML tree searches were conducted using a heuristic search with 1,000 random-taxon-addition replicates. Confidence at each node was assessed using nonparametric bootstrapping (Felsenstein 1985) with 1000 pseudo-replicates and 10 random-taxon-addition replicates per pseudo-replicate. Bayesian inference was run for  $4 \times 10^6$  generations using the default temperature (0.2) with four Markov chains per generation, sampling trees every 100 generations. Performance of each MCMC run was evaluated by examining convergence in Tracer (Drummond & Rambaut 2007). Nodal support was given as frequency of the recovered clade (posterior probability, Hulsenbeck & Ronquist 2001) after a burn in of  $1.5 \times 10^4$  to reach a stationary distribution of  $-\ln L$  scores. Bayesian analysis was repeated twice to verify that the program recovered

consistent topology and did not get stuck on local optima (Hulsenbeck and Bollback 2001).

A statistical parsimony network was created using default parameters (95% probability criterion) in the program TCS 1.21 (Clement *et al.* 2000). Genetic variation of populations was estimated as nucleotide ( $\pi$ ) and haplotype (h) diversity using ARLEQUIN (version 3.1.1; Schneider *et al.* 2000; Excoffier *et al.* 2005). Tests for neutrality, including Tajima's  $D$  and Fu's  $F_s$ , were conducted for islands with multiple haplotypes using ARLEQUIN, though it is acknowledged that these tests might be biased for some taxa (Wares 2009). Corrected and uncorrected pairwise  $F_{ST}$  were calculated to examine genetic differentiation among populations using MEGA 4 (Tamura *et al.* 2007; Kumar *et al.* 2008).

*Phylogeographic analyses.* To estimate the level of genetic partitioning among islands and between banks, I calculated the traditional  $F$  – statistics using the analysis of molecular variance (AMOVA) framework (Excoffier *et al.* 1992). However, valid statistical hypothesis testing cannot be performed for datasets where the number of populations ( $N_p$ ) is less than seven, as permutation tests will never result in a  $P$ -value of  $< 0.05$  (Fitzpatrick 2009). Though this dataset contains a large number of samples (53),  $N_p = 5$ ; and with one site on the Turks Bank and four on the Caicos Bank there are only five permutations of sites among banks. Therefore I would expect random chance alone to produce the configuration that maximizes the between-bank variance 20% of the time, making it impossible to reject the null hypothesis with  $\alpha < 0.20$  (Fitzpatrick 2009).

However, the kind of statistical inference obtained from permutation tests is not necessarily the most relevant for conservation questions. Given that I have sampled all but two potential *Epicrates* sites on the Caicos bank and the only known site on the Turks bank, the question of whether the Turks and Caicos Banks have genetically distinct hierarchical groups of populations is an unnecessary abstraction. For conservation purposes, the important question is whether the population on Gibbs Cay (or any other population) is genetically distinct enough to warrant independent management as an ESU. Therefore, I discuss my results with respect to proposed ESU criteria. For my data, these include reciprocal monophyly of mtDNA gene trees and significant differentiation of nuclear allele frequencies (Moritz 1994).

## Results

I resolved 617 base pairs (bp) of mitochondrial CytB, 518 bp of nuclear NT3, and 456 bp of nuclear *c-mos* for all 53 individuals. For CytB, there were 13 variable nucleotide positions (2.1% of total nucleotide positions), while for NT3 and *c-mos* there was one variable site per gene (0.19% and 0.21% of total nucleotide positions, respectively). A total of 9 haplotypes were observed for CytB and 2 haplotypes for each nuclear gene (CytB #'s 1-9, nuc genes #'s 1, 2; Table 2). Big Ambergris Cay and North Caicos were the only islands with multiple CytB haplotypes (4 haplotypes represented on each island), and one sequence each from Middle Caicos and North Caicos collapsed into haplotype 1 (Fig. 2). Gibbs Cay and Providenciales each had unique haplotypes for CytB ( #'s 8 and

9, respectively). For the nuclear gene NT3, two haplotypes were observed, one in all populations and the second exclusively on North Caicos (Table 2, Fig. 3). Two haplotypes were observed for *c-mos* as well, with haplotype 2 exclusive to Big Ambergris Cay and haplotype 1 found in all other populations (Table 2, Fig. 3).

Of the nine microsatellite markers screened, only four successfully amplified using PCR. All four markers proved monomorphic for all populations. Sequenced PCR products indicated that the repeat motif was included in each fragment; hence all populations appear monomorphic at these microsatellite loci.

*Gene tree estimation.* Trees were estimated only for the mitochondrial CytB marker, as the nuclear genes yielded extremely low genetic divergence (maximum of two alleles). The ML tree was estimated using the Hasegawa-Kishino-Yano model (proportion of invariant sites = 0; four rate categories) with a log-likelihood of  $-\text{LnL} = 1053.29$ . The ML tree recovered the same topology as the Bayesian analysis with nearly equal nodal support. Both trees and the statistical parsimony network resulted in the same grouping of haplotypes. Statistical parsimony showed a maximum of seven mutational steps separating haplotypes allowing a 95% parsimonious connection (Fig. 2). The sequence divergence among islands ranged from 0.2% to 0.8% (uncorrected *p*-distance and corrected Tamura-Nei) (Table 3). Haplotype diversity was calculated for the only two populations with more than a single haplotype: North Caicos (0.4167) and Big Ambergris Cay (0.6291). Nucleotide diversity was estimated as 0.002 for Big Ambergris Cay and 0.0017 for North Caicos. Tests for deviations from neutrality were non-significant for Big

Ambergris Cay (Tajima's  $D = 1.07$ ; Fu's  $F_s = 1.50$ ) and North Caicos (Tajima's  $D = -1.15$ ; Fu's  $F_s = 0.45$ ).

*AMOVA analysis.* AMOVA (Table 4) revealed an estimated 58% of variation ( $\Phi_{CT} = 0.58$ ) is explained by grouping between the Turks and Caicos Banks, and this configuration maximizes  $\Phi_{CT}$ . Variation among populations within the Caicos Bank accounted for 20.7% of the variation ( $\Phi_{SC} = 0.49$ ). When no partition between banks is made, among population variance (61.1%,  $\Phi_{ST} = 0.61$ ) and within population variance (38.8%) are similar.

## Discussion

The use of population genetics in initial conservation assessment and planning can be a powerful tool for devising conservation strategies that minimize disturbance of evolutionary trajectories for island populations (Frankham 2006). When defining conservation strategies for archipelagic endemics, it is important to recognize that separate ESUs might exist, particularly on islands with deep water isolation and long separation (Frankham *et al.* 2002; Whittaker & Fernández-Palacios 2007). As part of a larger new initiative to promote snake conservation in the TCI, I have started focused studies of the biology, ecology, and distribution of the Turks Island Boa (*E. c.*

*chrysogaster*) in the region (Reynolds 2011). A major component of this initiative was to assess the level of standing genetic variation among islands where this species occurs.

Our surprising result is that Turks Island Boas were found to exhibit less than 1% sequence divergence at the mitochondrial gene CytB, with haplotype sharing between some islands. Private haplotypes were found on Big Ambergris Cay, Providenciales, North Caicos and Gibbs Cay, though these were minimally divergent from other haplotypes (<1%). For nuclear genes, only three individuals were found to be variable, with one unique haplotype on North Caicos for the marker NT3 and one unique haplotype on Big Ambergris Cay for the marker *c-mos*. For both markers, haplotype 1 was shared across all populations, and haplotype 2 differed by a single nucleotide polymorphism. Low polymorphism in nuclear introns is not entirely unexpected in an intraspecific phylogeographic study (Hare 2001; Wang & Shaffer 2008), though my results indicate that what little variation is present is not idiosyncratic of a specific population. Additionally, several populations in the study are represented by only a few individuals and I have limited power to detect shared haplotypes- hence the few private haplotypes I found may well be shared among populations. For instance, Gibbs Cay individuals appear fixed for CytB haplotype 9, though only three unsampled haplotypes separate this population from the rest of the Caicos Bank. Since there are no other known extant populations on the Turks Bank, and since I did not sample boas from the nearest known populations of Long Cay and East Caicos, it is difficult to determine whether the Turks Bank boas harbor a truly unique mtDNA lineage. However, my observation of

very limited divergence from the Caicos haplotypes still stands as likely evidence of recent divergence.

Microsatellite loci developed for the congener *E. subflavus* are subject to ascertainment bias (Ellegren *et al.* 1995, 1997) in *E. chrysogaster*, as only four of nine loci amplified and all four proved monomorphic. Booth *et al.* (2011) screened these same loci in *Boa constrictor* and obtained amplification at only three of the nine loci, while Booth *et al.* (in review) screened these loci in *Epicrates maurus* and found that only one marker amplified and proved to be monomorphic. It is impossible to determine whether the lack of variation in *E. chrysogaster* reflects low effective population size, low mutation rates, or strong purifying selection. Nevertheless, these loci represent four independent nuclear markers, and were found to be monomorphic in *E. chrysogaster* and polymorphic in the related *E. subflavus* and *B. constrictor*.

Taken together, these results indicate extremely shallow genetic divergence between populations of Turks Island Boas, a finding that contrasts greatly with studies of other reptiles in the TCI. Welch *et al.* (2004) found significant divergence (> 5 times estimated divergence) among populations of the endemic *Cyclura carinata*, though divergence was higher within the Caicos Bank than between the Turks and Caicos Banks. This was explained as a relatively recent dispersal from the eastern Caicos Bank to the Turks Bank. Reynolds & Konecny (unpubl.) found high mitochondrial genetic divergence (8.6-10.9 %, ND2) between two species of dwarf geckoes (*Sphaerodactylus caicosensis* and *S. underwoodi*) endemic to the Caicos and Turks banks, respectively. Finally, Reynolds *et al.* (Chapter III, this volume) found partitioning between the Turks

and Caicos Banks of the endemic Turks and Caicos Curly-tailed Lizard (*Leiocephalus psammodromus*), and though some between-bank gene flow likely exists, this stands in stark contrast to the distribution of genetic diversity in *Cyclura*, their larger relatives (Welch *et al.* 2004). Though these species vary considerably in their life histories and dispersal abilities, and hence genetic patterns cannot be equated, they provide a heuristic comparison of the distribution of genetic variation in other squamates in the TCI.

The only other intraspecific phylogeographic study of insular *Epicrates* (Tzika *et al.* 2008b) found significant divergence ( $0.084 < F_{ST} < 0.76$ ) between populations of the Jamaican Boa (*E. subflavus*). Though these populations occur on the same island, uplift of mountains and sea level variation likely contributed to reduced gene flow between populations, and boas have likely occupied Jamaica far longer than the TCI (Tolson 1987; Tziaka *et al.* 2008b).

It is unclear why such low genetic divergence is found in boas, while much higher genetic and morphological divergence occurs in other endemic squamates with similar distributions in the archipelago (Reynolds 2011). I hypothesize that low diversity could be due to the following: 1) low rates of evolution in the markers examined or 2) recent panmixia when banks were emergent. Low rates of molecular evolution are known for some taxa such as turtles (King & Julian 2004); however, recent studies (*e.g.*, Hughes & Mouchiroud 2001) have shown increased rates of evolution at *c-mos* for snakes relative to the orthologous gene in chickens (*Gallus* spp.). The limited availability of markers forced us to select slower-evolving nuclear introns (Hare 2001, Wang & Shaffer 2008); however, CytB has been shown to evolve in snakes at similar rates to ND4, another

commonly used mtDNA marker for intraspecific studies (Kelly *et al.* 2003), and has been used in other intraspecific studies of *Epicrates* (Campbell 1997; Tzika *et al.* 2008b). Furthermore, the proposed negative relationship between body size and rates of molecular evolution (Bromhall 2002) would not explain the difference between iguanas and boas, as the former are much larger by weight as adults. Hence, recent panmixia seems a more plausible explanation for low observed divergence in Turks Island boas. Both the Turks and Caicos Banks were likely emergent as recently as 8,000 years ago (Lighty *et al.* 1982; Fairbanks 1989; Welch *et al.* 2004). These super-islands would have had low topographic relief (Sealey 2006; Reynolds 2011), with the only barriers to gene flow in boas likely being salinas or brackish wetland areas. It is possible that little genetic structure existed on the emergent banks, and hence current populations are samples of the recent panmixia. The low level of differentiation between the Turks and Caicos Banks is slightly more problematic, as frequent dispersal across the Turks Islands Passage seems unlikely even when both banks were fully emergent. Thus a plausible explanation is that one of the banks was recently colonized by individuals from the other. This could either have been accomplished naturally or could have been human-mediated, as is hypothesized to have occurred in rock iguanas (deliberate introduction from the TCI to Booby Cay, Mayaguana; Bryan *et al.* 2007). Humans have occupied the Turks and Caicos Archipelago for at least 1300 years (Keegan 1992; Carlson & Keegan 2006; Marvel 2008) and have used the local reptiles, particularly iguanas (IUCN Iguana Specialist Group 2003) and turtles (Seidel 1996; Lee & Ross 2001), as a food source. Hence, it is possible that people moved boas as well, as remains of at least two individuals have been found in middens of a Native American archeological site on

Grand Turk (Carlson 1999; Newsom & Wing 2004; Reynolds 2011). In addition to the lack of geographic structure, the low level of variation within populations implies that population sizes have been low for many generations. In particular, lack of variation at four microsatellites that were variable in the related Jamaican Boa (Tzika *et al.* 2008a) and distantly related *Boa constrictor* (Booth *et al.* 2011) is exceptional. This result might be a consequence of long-term decline and extirpation, or it might be that these snakes generally occurred in low densities, even under the most natural of circumstances.

Two additional subspecies of Southern Bahamas Boas exist: one in the Inaguas (*E. c. relicquus*) and one in the Crooked-Acklins Islands (*E. c. schwartzi*) of the Bahamas, though almost nothing is known of these populations (Buden 1975; Tolson & Henderson 1993; Henderson & Powell 2009). To my knowledge, no usable tissues exist that would be appropriate for genetic analysis, and the status of these populations, several of which are undoubtedly extirpated or near-extirpation (Barbour 1941), is unclear. The intriguing possibility exists that Turks Island boas are a recent introduction or arrival to the TCI from these Bahamian islands, contrary to previous hypotheses (Buden 1975), and future work should include focused sampling and genetic analysis of these subspecies, a task that will likely be difficult and time-consuming given the apparent rarity of boas on these islands (Buden 1975; Sheplan & Schwartz 1974).

*Turks bank boas.* The recent discovery of Turks Island Boas on the Turks Bank (Reynolds & Niemiller 2010b; Reynolds 2011), where they were thought to have been completely extirpated, raised suspicion that the Turks Bank might harbor a unique

lineage of boas due to the physical separation of deep water and strong currents between the banks. Gibbs Cay is an extremely small ( $\sim 0.07 \text{ km}^2$ ) island less than 1.5 km from Grand Turk and harbors populations of southern Bahamas anoles (*Anolis scriptus*) and Turks and Caicos curly-tailed lizards (*L. psammodromus*) which the adult boas likely feed on almost exclusively (Reynolds & Niemiller in press). Though the Gibbs Cay population of boas is the most divergent and harbors a unique mitochondrial haplotype, separation from the Caicos Bank populations was probably recent. While this population is unique in its habitat and location, management as a separate entity is probably not justified given the shallow divergence of mtDNA and lack of divergence of other markers. That is, I infer that *Epicrates c. chrysogaster* is a single ESU including a small number of insular populations that appear to be genetically exchangeable.

The ESU concept is not perfect (*e.g.*, Waples 2006), however, and I cannot evaluate whether any *Epicrates* population is locally adapted to its island to the point of not being “ecologically exchangeable” with other populations (Crandall *et al.* 2000). Nevertheless, the best available evidence is that all studied populations of the Turks Island boa are genetically similar parts of a single “evolutionary legacy” (Waples 1991).

*Conservation strategies.* My results suggest several conservation strategies for Turks Island boas, including possible translocation and in-situ conservation of ecologically-intact populations. Translocation of reptiles has seen a marked increase in success in the last 20 years, though some controversy still exists regarding appropriateness for many taxa (Germano & Bishop 2008). The federally endangered Virgin Islands boa (*Epicrates*

*monensis granti*) has been successfully reintroduced to at least two islands (Steven Key and Cayo Ratones) on the Puerto Rican bank after a rat and mongoose removal program, and these populations appear to be thriving (USFWS 2009). These results raise the intriguing possibility of relocation of boas to other islands in the TCI, including movement of boas from Caicos Bank populations to islands without boas on the Turks Bank. Rock iguanas (*C. carinata*) have benefited greatly from translocations from the few remaining strong source populations to outlying islands where no introduced predators occur and no human development exists (Gerber 2007). Long Cay, located on the Caicos Bank south of South Caicos, previously had populations of both iguanas and boas, though an introduction of cats caused the extirpation of the iguanas and perhaps boas, the latter of which have not been recorded there in some time (Reynolds 2011). Long Cay was the focus of an intensive cat eradication campaign, which eliminated the invasive predators and allowed re-introduction of iguanas (Mitchell *et al.* 2002). If boas were also extirpated, Long Cay would be an ideal island to relocate boas from other threatened populations.

Big Ambergris Cay harbors the densest population of *Epicrates* in the TCI and the highest known density in the Bahamian Archipelago; with a population size potentially exceeding 2,000 boas and a density as high as five boas/hectare (Reynolds 2011). This estimate represents one of the highest densities of *Epicrates* discovered to date (Tolson & Henderson 2006; Reynolds 2011). As this island currently lacks feral predators, Big Ambergris Cay likely represents one of the few ecologically-intact islands in the TCI. However, this 400 hectare island is currently being heavily developed, including the

recently completed largest private international airport in the Caribbean and a large marina under construction. Development plans include up to 500 homes, numerous recreation facilities, and associated support staff housing and facilities (<http://www.tcsportingclub.com>). This population of boas has undoubtedly already been negatively impacted by development and road construction, and is also subject to stochastic natural disasters such as hurricanes (Reynolds & Gerber submitted). In addition, the island is potentially subject to invasion by predatory mammals via freight ships supplying the island, and recently a feral cat remained loose for several months before it was found and removed (Reynolds 2011). Islands free of development and introduced predators exist on both the Turks and Caicos Banks and, given the low level of genetic diversity across banks, it might be reasonable for future management efforts to include movement of individuals to these islands to reduce the chance of catastrophic loss of remaining populations owing to predator introduction, development, or stochastic natural events. In addition, in situ conservation measures have been established on Big Ambergris Cay to mitigate the loss of habitat to development and prevent the introduction of damaging feral predators. This island therefore remains a focus of applied conservation for both Turks Island boas and Turks and Caicos rock iguanas.

Though additional data might yield further resolution of the minimal genetic structure and divergence in this species, the surprising results provide information valuable to conservation planning efforts for this species. Reasons for low genetic diversity in Turks Island boas are not clear, but the implications for conservation are important. This species should likely be managed as a single unit, including Turks Bank

populations, and the feasibility of future reintroductions should be strongly considered to buffer against the loss of important remaining populations to development and predators.

## Literature Cited

Allendorf FW, Luikart G (2007) Conservation and the genetics of populations. Blackwell Publishing, Oxford, UK.

Awise J (2004) Molecular markers, natural history, and evolution (2nd ed.). Sinauer Associates, Sunderland, Massachusetts, USA.

Barbour T (1941) A new boa from the Bahamas. Proc N Engl Zool Club 18:61-65.

Booth W, Johnson DH, Moore S, Schal C, Vargo EL (2011) Evidence for viable, non-clonal but fatherless Boa constrictors. Biol Lett doi: 10.1098/rsbl.2010.0793.

Bromhall L (2002) Molecular clocks in reptiles: life history influences rate of molecular evolution. Mol Biol Evol 19:302-309.

Brownstein M J, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. Biotechniques 20:1004-1010.

Bryan JJ, Gerber GP, Welch ME, Stephen CL (2007) Re-evaluating the taxonomic status of the Booby Cay Iguana, *Cyclura carinata bartschi*. Copeia 3:734-739.

Buden DW (1975) Notes on *Epicrates chrysogaster* (Serpentes: Boidae) of the southern Bahamas, with description of a new subspecies. Herpetologica 31:166-177.

Caballero A, Rodríguez-Ramilo ST, Ávila V, Fernández J (2010) Management of genetic diversity of subdivided populations in conservation programmes. Cons Genet 11:409-419.

Cade TJ, Burnham W (Eds.) (2003) Return of the peregrine: a North American saga of tenacity and teamwork. Peregrine Fund, Boise, Idaho.

Campbell BN (1997) Hic sunt serpentes- molecular phylogenetics and the Boidae (Serpentes: Booidea). Ph.D. Dissertation, Queen's University, Kingston, Ontario Canada.

Carlson LA (1999) Aftermath of a feast: human colonization of the southern Bahama Archipelago and its effects on the indigenous fauna. Ph.D. Dissertation, University of Florida, Gainesville, Florida.

Carlson LA, Keegan B (2006) Talking taino: boat trips. Times of the Islands Magazine, 76: Fall.

- Clement M, Posada D, Crandall K (2000) TCS: a computer program to estimate gene genealogies. *Mol Ecol* 9:1657-1660.
- Conner JK, Hartl DL (2004) A primer of ecological genetics. Sinauer, Massachusetts.
- Corke D (1992) The status and conservation needs of the terrestrial herpetofauna of the Windward Islands (West Indies). *Biol Cons* 62:47-58.
- Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology: an alternative to 'evolutionarily significant units'. *Trends Ecol Evol* 15:290-295.
- Dizon AE, Lockyer C, Perrin WF, Demaster D, Sisson J (1992) Rethinking the stock concept – a phylogeographic approach. *Cons Biol* 6:24-36.
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.
- Ellegren H, Primmer CR, Sheldon BC (1995) Microsatellite evolution-directionality or bias. *Nat Genet* 11:360-362.
- Ellegren H, Moore S, Robinson N, Byrne K, Ward W, Sheldon BC (1997) Microsatellite evolution- a reciprocal study of repeat lengths at homologous loci in cattle and sheep. *Mol Biol Evol* 14:854-860.
- Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479-491.
- Excoffier L, Laval G, Schneider S (2005) Arlequin ver 3.0: an integrated software package for population genetics data analysis. *Evol Biol Online* 1:47-50.
- Fairbanks RG (1989) A 17,000 year glacio-eustatic sea level record: Influence of glacial melting rates on the Younger Dryas event and deep-ocean circulation. *Nature* 342:637-642.
- Felsenstein J (1985) Confidence-limits on phylogenies- an approach using the bootstrap. *Evolution* 39:783-791.
- Fitzpatrick BM (2009) Power and sample size for nested analysis of molecular variance. *Mol Ecol* 18:3961-3966.
- Frankham R (2006) Evolutionary conservation genetics. In: Fox CW., Wolf JB. (Eds.), *Evolutionary Genetics*. Oxford University Press, New York, pp 502-512.

- Frankham R, Ballou JD, Briscoe DA (2002) Introduction to conservation genetics. Cambridge University Press, Cambridge, UK.
- Fritts SH, Bangs EE, Fontaine JA, Johnson MR, Phillips MK, Koch ED, Gunson JR (1997) Planning and implementing a reintroduction of wolves to Yellowstone National Park and central Idaho. *Rest Ecol* 5:7-27.
- Gerber GP, Iverson JB (2000) Turks and Caicos Iguana, *Cyclura carinata carinata*. In: Alberts AC., (Ed.), West Indian Iguanas: Status Survey and Conservation Action Plan. IUCN the World Conservation Union, Gland, Switzerland.
- Gerber GP (2007) Turks and Caicos iguana translocation program, Bahama Archipelago. *Re-introduction News* 26:53-55.
- Germano JM, Bishop PJ (2008) Suitability of amphibians and reptiles for translocation. *Cons Biol* 23:7-15.
- Hailey A, Wilson BS, Horrocks JA (Eds.). (2011) Conservation of Caribbean island herpetofaunas, vol. I: conservation biology and the wider Caribbean. Brill, Leiden The Netherlands.
- Hare MP (2001) Prospects for nuclear gene phylogeography. *Trends Ecol Evol* 16:700-706.
- Henderson RW, Powell R (2009) Natural history of West Indian reptiles and amphibians. University Press of Florida, Gainesville, Florida.
- Hughes S, Mouchiroud D. (2001) High evolutionary rates in nuclear genes of squamates. *J Mol Evol* 53:70-76.
- Hulsenbeck JP, Bollback JP (2001) Empirical and hierarchical Bayesian estimation of ancestral states. *Syst Biol* 50:351-366.
- Hulsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17:754-755.
- IUCN Iguana Specialist Group (2003) Turks and Caicos Iguana, *Cyclura carinata carinata* conservation and management plan, 2005-2009. Workshop Proceedings from the IUCN/SSC Iguana Specialist Group, Providenciales, Turks and Caicos Islands. Available at: [http://www.iguanafoundation.org/downloads/pdf/TCI-CAMP-4July2007\\_Sml.pdf](http://www.iguanafoundation.org/downloads/pdf/TCI-CAMP-4July2007_Sml.pdf).
- Iverson JB (1978) The impact of feral cats and dogs on populations of the West Indian rock iguana *Cyclura carinata*. *Biol Cons* 14:63-73.

Keegan WF (1992) The people who discovered Columbus. University Press of Florida, Gainesville, Florida, USA.

Kelly CMR, Barker NP, Villet MH (2003) Phylogenetics of advanced snakes (Caenophidia) based on four mitochondrial genes. *Syst Biol* 52:439-459.

King TL, Julian JE (2004) Conservation of microsatellite DNA flanking sequence across 13 Emydid genera assayed with novel bog turtle (*Glyptemys muhlenbergii*) loci. *Cons Genet* 5:719-725.

Kumar S, Dudley J, Nei M, Tamura K (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings Bioinforma* 9:299-306.

Lee DS, Ross JP (2001) The Cat Island turtle, a reptile of problematic origin, including a bibliographic review of the genus *Trachemys* in the West Indian region. In: Clark-Simpson C., Smith GW., (Eds.), *Proceedings of the 8th symposium on the natural history of the Bahamas*. Gerace Research Center, San Salvador, Bahamas.

Lighty RG, Macintyre JG, Stuckenrath R (1982) *Acropora* palmate reef framework: a reliable indicator of sea level in the western Atlantic for the past 10,000 years. *Coral Reefs* 1:25-30.

Maddison DR, Maddison WP (2005) *MacClade 4* release version 4.07 for OSX. Sinauer, Sunderland Massachusetts.

Marvel J (2008) Our first colonists: the pre-columbian people of the Turks and Caicos. In: Mills C., (Ed.), *A History of the Turks and Caicos Islands*. Macmillan, Oxford, UK, pp. 79-86.

Mitchell N, Haeffner R, Veer V, Fulford-Gardner M, Clerveaux W, Veitch CR, Mitchell G (2002) Cat eradication and the restoration of endangered iguanas (*Cyclura carinata*) on Long Cay, Caicos Bank, Turks and Caicos Islands, British West Indies. In: Veitch CR., Clout MN (Eds.), *Turning the Tide: the Eradication of Invasive Species*. IUCN Gland, Switzerland and Cambridge, UK, pp. 206-212.

Moritz C (1994) Defining 'evolutionarily significant units' for conservation. *Trends Ecol Evol* 9: 373-375.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Ke J (2000) Biodiversity hotspots for conservation priorities. *Nature* 403:853-858.

Newsom LA, Wing ES (2004) *On Land and Sea: Native American Uses of Biological Resources in the West Indies*. The University of Alabama Press, Tuscaloosa, Alabama.

Noonan BP, Chippindale PT (2006a) Dispersal and vicariance: the complex evolutionary history of boid snakes. *Mol Phylogenet Evol* 40:347-358.

Noonan BP, Chippindale PT (2006b) Vicariant origin of Malagasy reptiles supports late Cretaceous Antarctic land bridge. *Am Nat* 168:730-741.

Nye P (1988) Bald eagle restoration in northeast United States, 1972-1987. In: Pendelton BG., (Ed.), *Proceedings of the northeast raptor management symposium and workshop*. National Wildlife Federal Scientific and Technical Series No. 13. National Wildlife Federation, Washington, D. C., pp. 264-269.

Posada D (2006) ModelTest Server: a web based tool for the statistical selection of models of nucleotide substitution online. *Nucl Acid Res* 34:W700-W703.

Reynolds RG (2011) Status, conservation, and introduction of amphibians and reptiles in the Turks and Caicos Islands, British West Indies. In: Hailey A., Wilson BS., Horrocks JA., (Eds.), *Conservation of Caribbean Island Herpetofaunas*, vol. II: Regional Accounts of the West Indies. Brill, Leiden, Netherlands, pp. 377-406.

Reynolds RG, Niemiller ML (2010a) Island invaders: introduced amphibians and reptiles in the Turks and Caicos Islands. *Reptil Amphib* 17:117-121.

Reynolds RG, Niemiller ML (2010b) *Epicrates chrysogaster* (Southern Bahamas Boa) distribution. *Caribbean Herpetol* 1:14

Reynolds RG, Niemiller ML (In press) *Epicrates chrysogaster chrysogaster*. (Turks Island Boa). *Diet. Herpetol Rev.*

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

Ryder OA (1986) Species conservation and systematics: the dilemma of subspecies. *Trends Ecol Evol* 1:9-10.

Schneider S, Roessli D, Excoffier L (2000) Arlequin: a software for population genetic data. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Sealey NE (2006) Bahamian landscapes, 3rd edition. Macmillan-Caribbean, Oxford, UK.

Seidel ME (1996) Current status of biogeography of the West Indian turtles in the genus *Trachemys* (Emydidae). In: Powell R., Henderson RW., (Eds.), *Contributions to West Indian Herpetology: a Tribute to Albert Schwartz*. Society for the Study of Amphibians and Reptiles, Ithaca, New York, pp. 169-174.

- Sheplan BR, Schwartz A (1974) Hispaniolan boas of the genus *Epicrates* (Serpentes, Boidae) and their Antillean relationships. *Ann Carnegie Mus Nat Hist* 45:57-143.
- Slatkin M (1977) Gene flow and genetic drift in a species subject to frequent local extinctions. *Theor Popul Biol* 12:253-262.
- Slatkin M (1985) Gene flow in natural populations. *Ann Rev Ecol Syst* 16:393-430.
- Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* 236:787-792.
- Smith ML, Hedges SB, Buck W, Hemphill A, Inchaustegui S, Ivie MA, Martina DA, Maunder M, Ortega JF (2005) Caribbean islands. In: Mittermeier RA., Gill PR., Hoffman M., Pilgrim J., Brooks T., Mittermeier CG., Lamoreaux J., da Fonseca GAB., (Eds.), *Hotspots Revisited: Earth's Biologically Richest and Most Endangered Terrestrial Ecoregions*. CEMEX, S. A. de C. V. Mexico City, Mexico.
- Storfer A (1999) Gene flow and endangered species translocations: a topic revisited. *Biol Cons* 87:173-180.
- Swofford DL (2002) PAUP\*. Phylogenetic analysis using parsimony (\*and other methods). Version 4.0. Sinauer, Sutherland Massachusetts, USA.
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetic analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596-1599.
- Tolson PJ (1987) Phylogenetics of the boid snake genus *Epicrates* and Caribbean vicariance theory. *Occas Papers Mus Zool, University of Michigan* 715:1-68.
- Tolson PJ, Henderson RW (1993) *The natural history of West Indian boas*. R&A Publishing Ltd. Taunton, Somerset, England.
- Tolson PJ, Henderson RW (2006) An overview of snake conservation in the West Indies. *Appl Herpetol* 3:345-356.
- Tordoff HB, Redig PT (2001) Role of genetic background in the success of reintroduced Peregrine Falcons. *Cons Biol* 15:528-532.
- Tzika AC, Remy C, Gibson R, Milinkovitch MC (2008a) Molecular genetic analysis of a captive breeding program: the vulnerable endemic Jamaican yellow boa. *Cons Genet* 10:69-77.
- Tzika AC, Koenig S, Miller R, Garcia G, Remy C, Milinkovitch MC (2008b) Population structure of an endemic vulnerable species, the Jamaican boa (*Epicrates subflavus*). *Mol Ecol* 17:533-544.

US Fish and Wildlife Service (2009) Virgin Islands Tree Boa (*Epicrates monensis granti*) 5 year review: summary and evaluation. US Fish and Wildlife Service Southeast Region Ecological Services, Boquerón, Puerto Rico.

Villesen P (2007) FaBox: an online toolbox for fasta sequences. Mol Ecol Notes 7:965–968.

Wade MJ, McCauley DE (1988) Extinction and recolonization: their effects on the genetic differentiation of local populations. Evolution 42:995-1005.

Wang IJ, Shaffer HB (2008) Rapid color evolution in an aposematic species: a phylogenetic analysis of color variation in the strikingly polymorphic strawberry poison-dart frog. Evolution 62:2742-2759.

Waples RS (1991) Pacific Salmon, *Oncorhynchus* spp., and the definition of “species” under the Endangered Species Act. Mar Fish Rev 53: 11-22.

Waples RS (2006) Distinct population segments. In: Scott JM., Goble DD., Davis FW., (Eds.), The Endangered Species Act at Thirty: Conserving Biodiversity in Human Dominated Landscapes. Island Press, Washington D. C., pp. 127-149.

Wares JP (2009) Natural distributions of mitochondrial sequence diversity support new null hypotheses. Evolution 64:1136-1142.

Watson J (1989) Successful translocation of the endemic Seychelles Kestrel *Falco araea* to Praslin. In: Meyburg B-U., Chancellor RD., (Eds.), Raptors in the Modern World. World Working Group on Birds of Prey, Berlin, Germany, pp. 363-367.

Welch ME, Gerber GP, Davis SK (2004) Genetic structure of the Turks and Caicos rock iguana and its implications for species conservation. In: Alberts AC., Carter RL., Hayes WK., Martins EP., (Eds.), Iguanas, Biology and Conservation. University of California Press, Berkeley, California, pp. 58-70.

Whitlock MC, McCauley DE (1990) Some population genetic consequences of colony formation and extinction: Genetic correlations within founding groups. Evolution 44:1717-1724.

Whittaker RJ, Fernández-Palacios JM (2007) Island biogeography. 2nd ed. Oxford University Press, New York.

Wright S (1977) Evolution and the genetics of populations, Vol. 3, experimental results and evolutionary deductions. University of Chicago Press, Chicago, Illinois.

## **Appendix IIa. Figures and Tables**

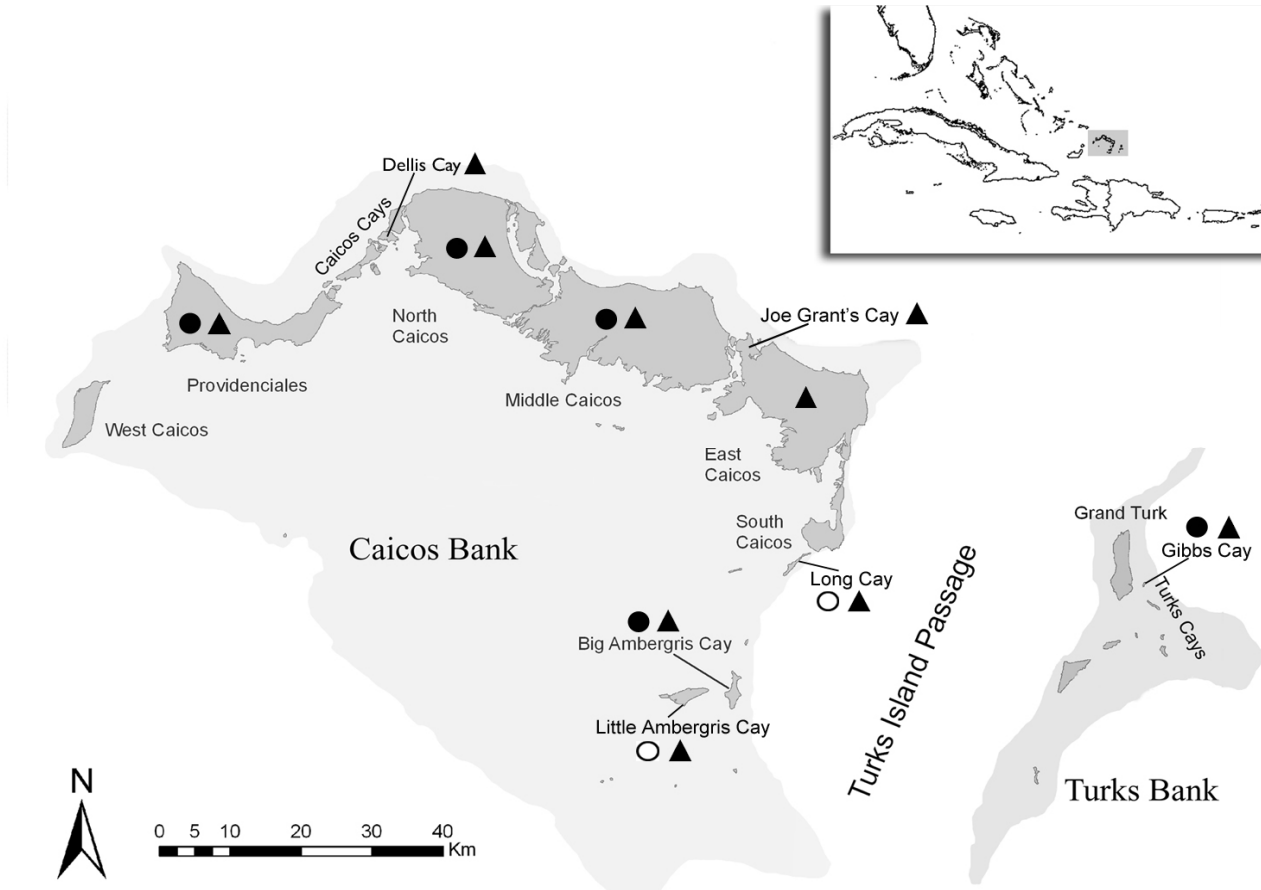


Figure II – 1. Map of the Turks and Caicos Islands illustrating sampling locations and known distribution of *Epicrates c. chrysogaster*. Inset map shows the location of the Turks and Caicos Islands within the greater Caribbean region. Dark outlines indicate the approximate extent of the Caicos and Turks banks. Triangles represent known populations of *Epicrates c. chrysogaster*, whereas circles represent islands sampled during this study. Filled circles are islands from which tissue samples were obtained and open circles are islands that were searched but on which no boas were found. Other islands mentioned in the text are labeled.

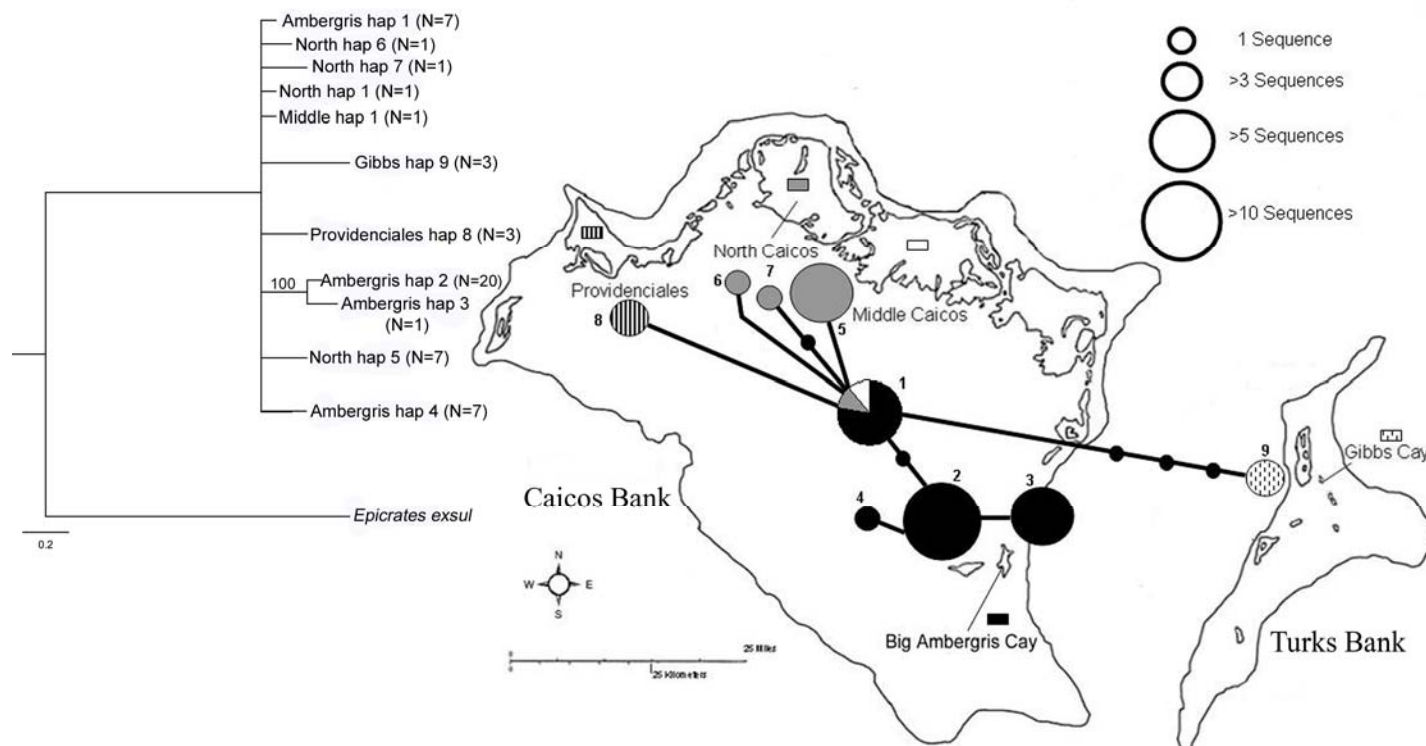


Figure II – 2. Phylogeographic genealogical relationships among CytB haplotypes for *Epicrates c. chrysogaster*. Phylogenetic tree is a consensus from Bayesian inference with posterior probabilities on relevant nodes. Haplotype map is the TCS statistical parsimony network overlaid onto a map of the Turks and Caicos Islands. Circles represent individual haplotypes, the size of which corresponds to the number of sequences each haplotype contains. Circles are divided and shaded to show the proportion of each island population within each haplotype, and each island is assigned a unique shading pattern. Small black circles indicate unobserved haplotypes. Haplotypes are connected with a 95% CL

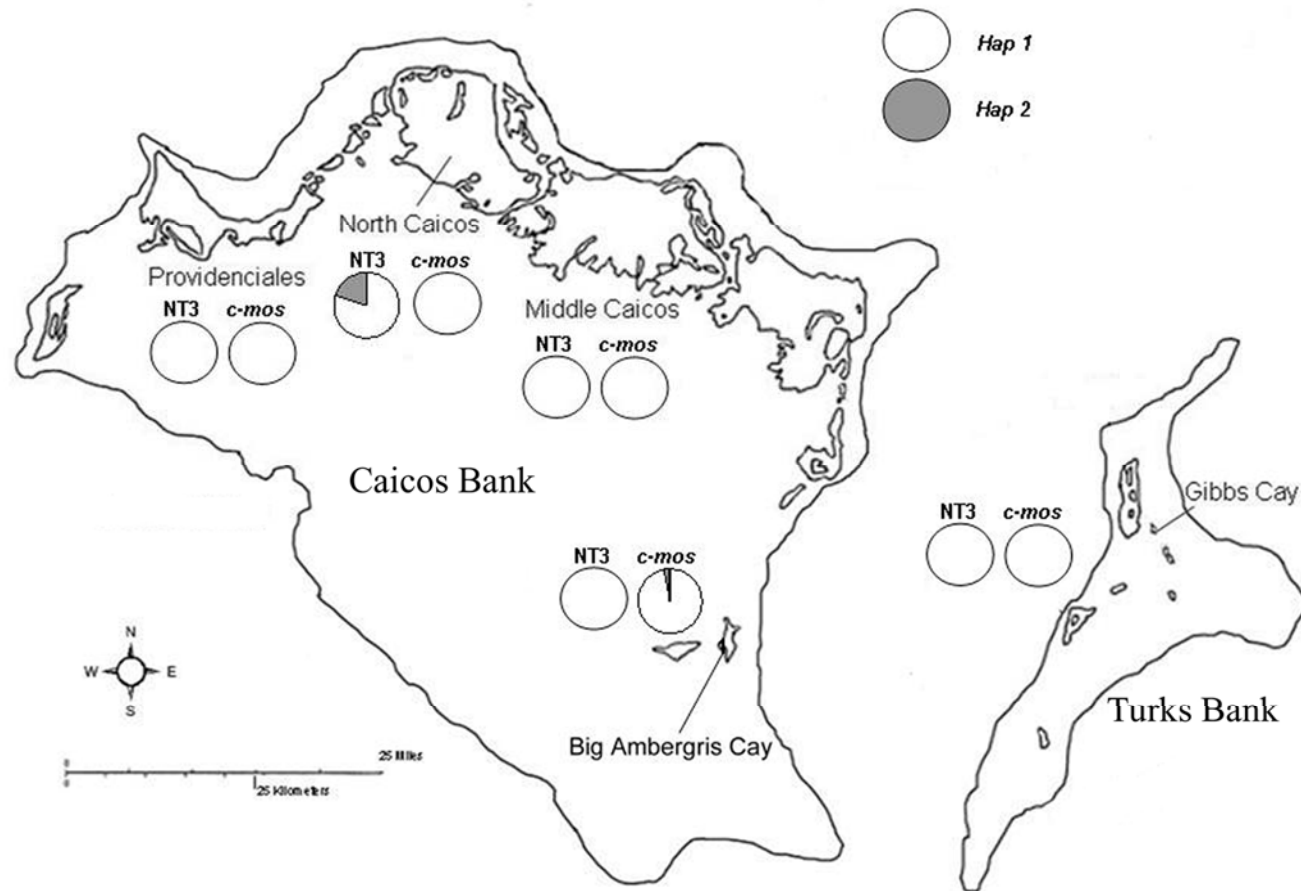


Figure II – 3. Population-specific haplotypes for *Epicrates c. chrysogaster* using the nuclear markers NT3 and *c-mos*. There are two haplotypes for each nuclear gene, designated as Hap 1 (white) and Hap 2 (gray). Circles are divided and shaded to show the proportion of each haplotype for each marker within each island population. Unique haplotypes only occur in North Caicos (NT3, 2 sequences) and Big Ambergris Cay (*c-mos*, 1 sequence).

Table II – 1. Sampling locations and sampling effort for tissue samples used in this analysis. Effort is represented as person hours (p/h), the amount of time spent searching by all individuals that day. Two additional islands (Long Cay and Little Ambergris Cay) where *Epicrates* are thought to occur (Reynolds 2011) were also searched but yielded no specimens. “Captive” indicates that the boas sampled were wild-caught captives being held for educational purposes with known locality information.

| Island               | Date(s)            | Effort    | Captures |
|----------------------|--------------------|-----------|----------|
| Big Ambergris Cay    | 4-10 December 2007 | ~ 54 p/h  | 34       |
| Gibbs Cay            | 6 August 2008      | 12 p/h    | 2        |
|                      | 8 August 2008      | 6 p/h     | 1        |
| North Caicos         | 10 August 2008     | 21 p/h    | 1        |
|                      | 12 August 2008     | 19 p/h    | 1        |
|                      | 12 August 2008     | (captive) | 3        |
|                      | 16 October 2009    | 24 p/h    | 0        |
|                      | 17 October 2009    | 26 p/h    | 1        |
|                      | 18 October 2009    | 23 p/h    | 0        |
|                      | 18 October 2009    | (captive) | 2        |
|                      | 20 October 2009    | 28 p/h    | 0        |
|                      | 21 October 2009    | 27 p/h    | 2        |
|                      | 22 October 2009    | 26 p/h    | 0        |
| Middle Caicos        | 11 August 2009     | 24 p/h    | 0        |
|                      | 13 August 2009     | 12 p/h    | 0        |
|                      | 17 October 2009    | 12 p/h    | 0        |
|                      | 18 October 2009    | 12 p/h    | 0        |
|                      | 19 October 2009    | 18 p/h    | 1        |
| Providenciales       | 16 July 2008       | 13 p/h    | 0        |
|                      | 15 October 2009    | 12 p/h    | 1        |
|                      | 15 October 2009    | (captive) | 4        |
|                      | 13 March 2010      | 3 p/h     | 0        |
| Long Cay (Caicos)    | 19 July 2008       | 5 p/h     | 0        |
| Little Ambergris Cay | 6 December 2007    | 5 p/h     | 0        |
|                      | 19 March 2009      | 4 p/h     | 0        |
| Totals               | 27 days            | 361 p/h   | 53       |

Table II – 2. Haplotypes of *Epicrates c. chrysogaster* from each population for one mitochondrial gene (CytB) and two nuclear genes (NT3, c-mos). Haplotypes are numbered in bold 1-9 for CytB and 1-2 for each nuclear gene. Number of sequences from each island for each haplotype are in parenthesis. Island abbreviations are as follows: BA- Big Ambergris Cay, GC- Gibbs Cay (Turks Bank), MC- Middle Caicos, NC- North Caicos, PR- Providenciales, All- All islands combined. Estimates of haplotype diversity (h), nucleotide diversity ( $\pi$ ), and two neutrality tests were calculated from the mitochondrial CytB marker. Significance is denoted by an asterisk (\*).

| Island     | n  | Haplotypes (n)                                                                                                                                                                                                        | $H$                 | $\pi$               | Tajima's $D$ | Fu's $F_s$ |
|------------|----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|---------------------|--------------|------------|
| <b>BA</b>  | 35 | CytB: <b>1</b> (7); <b>2</b> (20); <b>3</b> (7); <b>4</b> (1)<br>NT3: <b>1</b> (35)<br><i>c-mos</i> : <b>1</b> (34); <b>2</b> (1)                                                                                     | 0.6291 $\pm$ 0.0581 | 0.002 $\pm$ 0.8685  | 1.07         | 1.50       |
| <b>GC</b>  | 3  | CytB: <b>9</b> (3)<br>NT3: <b>1</b> (3)<br><i>c-mos</i> : <b>1</b> (3)                                                                                                                                                | -                   | -                   | -            | -          |
| <b>MC</b>  | 1  | CytB: <b>1</b> (1)<br>NT3: <b>1</b> (1)<br><i>c-mos</i> : <b>1</b> (1)                                                                                                                                                | -                   | -                   | -            | -          |
| <b>NC</b>  | 10 | CytB: <b>1</b> (1); <b>5</b> (7); <b>6</b> (1); <b>7</b> (1)<br>NT3: <b>1</b> (8), <b>2</b> (2)<br><i>c-mos</i> : <b>1</b> (10)                                                                                       | 0.4167 $\pm$ 0.1907 | 0.0017 $\pm$ 0.0014 | -1.15        | 0.45       |
| <b>PR</b>  | 3  | CytB: <b>8</b> (3)<br>NT3: <b>1</b> (3)<br><i>c-mos</i> : <b>1</b> (3)                                                                                                                                                | -                   | -                   | -            | -          |
| <b>All</b> | 53 | CytB: <b>1</b> (9); <b>2</b> (20); <b>3</b> (7); <b>4</b> (1); <b>5</b> (7); <b>6</b> (1); <b>7</b> (1); <b>8</b> (3); <b>9</b> (3)<br>NT3: <b>1</b> (51); <b>2</b> (2)<br><i>c-mos</i> : <b>1</b> (52); <b>2</b> (1) | 0.7934 $\pm$ 0.0387 | 0.0035 $\pm$ 0.0022 | -0.74        | -0.80      |

Table II – 3. Corrected (Tamura-Nei; top diagonal) and uncorrected ( $p$ -distance; bottom diagonal) pairwise genetic divergence between populations of *Epicrates c. chrysogaster*. Island names abbreviated as per Table 2.

|    | BA    | NC    | MC    | PR    | GC    |
|----|-------|-------|-------|-------|-------|
| BA |       | 0.004 | 0.002 | 0.004 | 0.009 |
| NC | 0.004 |       | 0.002 | 0.003 | 0.008 |
| MC | 0.002 | 0.002 |       | 0.002 | 0.007 |
| PR | 0.004 | 0.003 | 0.002 |       | 0.008 |
| GC | 0.009 | 0.008 | 0.006 | 0.008 |       |

Table II – 4. Results of analysis of molecular variance (AMOVA) for various groupings of *Epicrates c. chrysogaster*.

|                                 | Source of variation | df | % variation | $\Phi$ statistic   |
|---------------------------------|---------------------|----|-------------|--------------------|
| Analysis across both banks      | Between banks       | 1  | 58.1        | $\Phi_{CT} = 0.58$ |
|                                 | Within banks        | 3  | 20.7        | $\Phi_{SC} = 0.49$ |
|                                 | Within populations  | 47 | 21.2        |                    |
| Analysis across all populations | Among populations   | 4  | 61.1        | $\Phi_{ST} = 0.61$ |
|                                 | Within populations  | 47 | 38.8        |                    |

## **Appendix IIb. Primers and Genbank accession numbers.**

Appendix IIb. Primers and PCR conditions used in this study. Underlined bases indicate the primer “tail” used to force plus-A addition.

| Primer Names | Primer sequence (5'-3')                    | Annealing Temp. | Source                                  | Result          | Genbank     |
|--------------|--------------------------------------------|-----------------|-----------------------------------------|-----------------|-------------|
| CYTBE CF     | GCCCCACCAACAAATCCTA                        | 52°C            | this study                              | polymorphic     | JF812173-81 |
| CYTBE CR     | GGTCTGGTATGGGTGGAATG                       |                 |                                         |                 |             |
| NT3-F3       | ATATTTCTGGCTTTTCTCTGTGGC                   | 52°C            | Noonan & Chippindale (2006 <i>a,b</i> ) | polymorphic     | JF812171-2  |
| NT3-R4       | GCGTTTCATAAAAATATTGTTTGACCGG               |                 |                                         |                 |             |
| CMOS-Fsnk    | GCTGTAAAACAGGTGAAGAGATGCAG                 | 61°C            | Noonan & Chippindale (2006 <i>a,b</i> ) | polymorphic     |             |
| CMOS-Rsnk    | AGCACGATGGGTGTATGTTCCCCC                   |                 |                                         |                 |             |
| usat1-F      | <u>GTTTCTTT</u> TTTGTTTGTGTTCCCAATGCC      | 60°C            | Tzika <i>et al.</i> 2008                | did not amplify |             |
| usat1-R      | GGTGGAGACTATAAATGCCTTACTT                  |                 |                                         |                 |             |
| usat3-F      | <u>GTTTCTT</u> CCCTCCATACTTCATCTGCTCTT     | 63°C            | Tzika <i>et al.</i> 2008                | monomorphic     |             |
| usat3-R      | CCCGTAGAAAGCCGCCCAA                        |                 |                                         |                 |             |
| usat10-F     | <u>GTTTCTT</u> TTTGGGATGAACATGAGCATGGAAG   | 63°C            | Tzika <i>et al.</i> 2008                | did not amplify |             |
| usat10-R     | ATGGGGTTCACGAAGAGTCGG                      |                 |                                         |                 |             |
| usat11-F     | <u>GTTTCTT</u> GAGACAGTGGAGGACAGGAAGG      | 60°C            | Tzika <i>et al.</i> 2008                | monomorphic     |             |
| usat11-R     | AGCACAAAGAGTTGTAAGAAGTGAGA                 |                 |                                         |                 |             |
| usat13-F     | <u>GTTTCTT</u> CCTTCATTCACATACTTTTCATTC    | 60°C            | Tzika <i>et al.</i> 2008                | did not amplify |             |
| usat13-R     | CGCCCAGAGTCCCTCTTTT                        |                 |                                         |                 |             |
| usat16-F     | <u>GTTTCTT</u> TGTGGGGACTTGTCCA            | 52°C            | Tzika <i>et al.</i> 2008                | did not amplify |             |
| usat16-R     | AAATGTAACTCTAGGCTTT                        |                 |                                         |                 |             |
| usat24-F     | <u>GTTTCTT</u> CCAACCCACCCCTTA ACTCCCAAACC | 65°C            | Tzika <i>et al.</i> 2008                | monomorphic     |             |
| usat24-R     | CATCATCCAGCTTGCAGATTCAGCAGAGA              |                 |                                         |                 |             |
| usat30-F     | <u>GTTTCTT</u> TCTTTTCTTCTGTGCCTTC         | 56°C            | Tzika <i>et al.</i> 2008                | monomorphic     |             |
| usat30-R     | CTCTCATCCATTTCCCCT                         |                 |                                         |                 |             |
| usat36-F     | <u>GTTTCTT</u> TGATACTGGACAGAAAAG          | 51°C            | Tzika <i>et al.</i> 2008                | did not amplify |             |
| usat36-R     | AAATCAAACAGAAGGGAAAG                       |                 |                                         |                 |             |

### **PART III. MOLECULAR ISLAND BIOGEOGRAPHY: EQUILIBRIUM VERSUS NON-EQUILIBRIUM GENETIC STRUCTURE IN AN ARCHIPELAGIC LIZARD**

#### **Preface**

In chapter II, I used genetic data gathered from a threatened island species to evaluate genetic diversity between insular populations to better inform conservation efforts. In that study, I used mtDNA, nuclear introns, and microsatellite loci to reveal a general lack of population genetic structure across the entire range of the species. My intention in chapter III is to expand upon the types of information that can be gathered regarding the origins of genetic diversity in island populations by evaluating genetic data against the null hypotheses of island biogeographic, metapopulation, and population genetic equilibria. As threatened *Epicrates* are very difficult to collect, I choose to collect genetic data from a widely distributed lizard, as lizards have long been model organisms in the Caribbean. However, genetic markers are not well developed for this group, so I chose to use the inexpensive method of AFLPs to generate anonymous nuclear markers.

The following section is a slightly modified version of a paper submitted for publication:

Reynolds, R.G., M.L. Niemiller, & B.M. Fitzpatrick. Molecular island biogeography: equilibrium versus non-equilibrium genetic structure in an archipelagic lizard.

## **Abstract**

Population genetic structure results from a combination of historical and contemporary processes of extinction, colonization, bottleneck and expansion, and gene flow. Studies using both mtDNA and anonymous nuclear markers (*e.g.*, AFLPs) can combine information from historical demography (phylogeography) and landscape (or population) genetics, while the neutral theories of island biogeography and genetic diversity provide the opportunity to test for deviations from null hypotheses using empirical data. I investigated the partitioning of genetic diversity in an endemic lizard, *Leiocephalus psammodromus*, distributed on the Turks and Caicos archipelago using phylogeography, landscape genetics, and island biogeography. Genetic data indicated some divergence between the Turks and Caicos Banks, consistent with the significant marine barrier of the Turks Island Passage, though I also found evidence for biased gene flow from the Turks to the Caicos Bank. Consistent with non-equilibrium dynamics, I found no relationship between genetic diversity and island area, island isolation, or island geographic distance. In addition, I found island-specific deviation from mutation-migration-drift equilibrium. Overall, my results indicate significant evidence of gene flow among islands within banks and across banks, biased migration from the Turks Bank to the Caicos Bank, and non-equilibrium population dynamics.

## Introduction

Islands have been important models in the investigation of evolutionary processes by serving as natural laboratories to test general evolutionary theories in discrete areas (Losos *et al.* 1998; Whittaker & Fernández-Palacios 2007; Losos & Ricklefs 2009). Islands closely resemble models used in population genetics, as they have distinct boundaries and potential for a wide range of migration rates and population sizes (Whittaker & Fernández-Palacios 2007). When populations are separated from each other (a scenario easily envisioned among populations of insular organisms), processes such as mutation, drift, migration or gene flow, local extinction, and recolonization may dramatically influence the apportioning of genetic variation within a geographically structured population (Wright 1977; Slatkin 1987; Wade & McCauley 1988; Whitlock & McCauley 1990). Island biogeography theory and metapopulation genetics predict that the processes of extinction, colonization, and migration will affect the magnitude and rate of genetic divergence between island and mainland species (Johnson *et al.* 2000; Whitlock 2004). Rates of extinction and recolonization are assumed to be influenced by axes of island variation such as the size or isolation of the island from other islands (MacArthur & Wilson 1967). In an island archipelago, average genetic diversity is expected to be a function of island area, as larger islands can support larger, less extinction-prone populations and experience lower rates of drift (Johnson *et al.* 2000). In addition, immigration rate is predicted to be greater for islands in close proximity relative to more distant islands, hence reducing divergence and maintaining variation, especially in small populations. Furthermore, proximity to a “source” that distributes potential

immigrants might affect extinction probability (if immigrants effect a rescue) and/or the number (and genetic diversity) of recolonizing individuals after an extinction.

When analogous assumptions are made, the neutral theories for species diversity and genetic diversity make the same predictions for the equilibrium level of diversity in a given area or population of given size (Ewens 1972; Hubbell 2001). An interesting paradox, however, is that full equilibrium at both population and community levels appears to be mutually exclusive because a dynamic equilibrium for species diversity predicts that some species will often be recent colonists (or newly isolated lineages) and therefore unlikely to have reached population genetic equilibrium. Thus, island biogeography and metapopulation ecology predict deviations from population genetic equilibria, for at least some species on some islands or patches within a system. To fully understand the biological and conceptual relationships between community level and species level island biogeography, theory is required to understand expected patterns and timescales, and empirical studies must be able to test and characterize deviations from equilibrium. Here I use population genetic data to test equilibrium vs. non-equilibrium patterns in a widespread lizard inhabiting the Turks and Caicos, part of an island system (Bahamian Archipelago) that has long been studied in the context of island biogeography.

Some of the most important insights into the ecological and evolutionary processes of diversification, colonization, and speciation have come from studies of organisms on islands in the Caribbean (Munroe 1948; MacArthur & Wilson 1963, 1967; Schoener *et al.* 2001; Ricklefs & Bermingham 2008). Though considerable debate has surrounded the vicariance/dispersal hypotheses of the origin of West Indian biota (Hedges *et al.* 1992; Hedges 1996, 2006; Iturralde-Vinent & MacPhee 1999; Heinicke *et*

*al.* 2007; van Ee *et al.* 2008), important conclusions are being drawn regarding the influence of these processes on more recent shaping of current distributions and island diversity (Ricklefs & Bermingham 2008). In particular, the facility with which modern molecular techniques may be applied is leading to a greater understanding of the partitioning of genetic diversity in the context of both historical demography and landscape genetics.

Lizards have long been model organisms in the Caribbean, as their wide distributions, high diversity, and ease of capture and observation have led to many published studies ranging from behavioral investigations to informing models of speciation and evolution (Roughgarden 1995; Losos 2009; Camargo *et al.* 2010 and references therein). To investigate processes generating genetic structure in an archipelago and to test some predictions of Johnson *et al.* (2000), I sampled curly-tailed lizards (*Leiocephalus psammodromus*; Barbour 1916) from islands of varying sizes and isolation in the Turks and Caicos Islands, located at the south eastern terminus of the Bahamian Archipelago (Fig. 1). *Leiocephalus psammodromus* are conspicuous diurnal omnivores that occupy most habitats without closed canopy and have a catholic diet (Smith 1994, 1995; Iverson & Smith 2006; Reynolds 2009). They occur on most islands larger than 0.42 km<sup>2</sup> (Reynolds 2011); however, some notable extirpations have occurred (Appendix 1), likely aggravated by a short life span (2 years), low clutch sizes, and an extreme susceptibility to predation by feral mammals (Iverson 1978; Smith & Iverson 1993; Reynolds 2011). This species is currently considered to include at least six subspecies distributed on both the Turks and Caicos banks (Fig. 1; Schwartz 1967). The taxonomic distinction of different island populations indicates the possibility of reduced

gene flow (Schwartz 1967) or deeper divergence times similar to other reptile species in the region (Welch *et al.* 2004). Here I examine genetic diversity and allelic partitioning in this lizard in the context of phylogeography, landscape genetics, and island biogeography using mitochondrial DNA (mtDNA) and dominant nuclear marker data. I also discuss the use of genetic data in the context of conservation and evolution in an archipelago. My approach is an attempt to combine population genetics and island biogeography explicitly. I evaluate phylogeographic patterns, hierarchical subdivision, and isolation by distance in addition to testing for associations between genetic diversity, island area, and isolation. I ask whether the system is likely to be in migration-mutation-drift equilibrium and evaluate non-equilibrium alternatives, including relaxation and extinction-recolonization hypotheses.

## **Methods**

### *Study area*

The Turks and Caicos Islands consist of more than 200 islands, ranging in size from < 0.01 km<sup>2</sup> to 124 km<sup>2</sup>, distributed on two shallow banks (Fig. 1). The banks are platforms of marine sediment overlaid with oolitic limestone, and the islands on the banks were formed from oolitic wind-blown deposits that accumulated during peak glaciation events (Keegan 1992; Sealey 2006; Ricklefs & Bermingham 2008; Reynolds 2011). The banks are separated by the narrow (16-40 km) and deep (> 2200 m) Turks Island Passage and have never been joined (Lighty *et al.* 1982; Fairbanks 1989; Welch *et al.* 2004). During the Pleistocene, the banks were periodically inundated and exposed, with the last full

exposure of the Turks and Caicos Banks occurring during the peak of the Wisconsin glaciation (~8,000 to 17,000 years before present). Rising sea levels have since fragmented the banks into the present islands (Morgan 1989; Keegan 1992).

#### *Sample collection and DNA extraction*

I surveyed a total of 19 islands in the Turks and Caicos Islands, 14 on the Caicos Bank and five on the Turks Bank (Figure 1, Table 1). Six of these islands were found not to currently harbor populations of *L. psammodromus* (see Table S1 in Supporting Information). Autotomizable tail tips were collected from 259 wild *L. psammodromus* on 13 different islands in the Turks and Caicos (Table 1), including each of the six populations described as unique subspecies by Schwartz (1967) (Fig. 1). Four of these islands (Little Water Cay, Water Cay, Pine Cay, and Ft. George Cay) are < 160 m apart and are frequently (every few decades to centuries) connected and disconnected by sand spits; hence I consider them here as one population (Caicos Cays). Samples were preserved in 95% ethanol and stored at -20 °C. Whole genomic DNA was extracted using the Wizard SV DNA purification system (Promega, Madison, WI) according to the manufacturer's protocol and stored at -20° C.

#### *Mitochondrial DNA sequencing*

The polymerase chain reaction (PCR) was used to amplify an 800 bp portion of the mitochondrial genome (NADH dehydrogenase subunit 2; ND2), using primers LND2F (5'-AAGCTATCGGGCCCATAACC-3') and LND2R (5'-AAAGTGTTTGAGTTGCATTCAG-3') developed from Macey *et al.* (1997). PCR

amplifications were conducted in a total volume of 24 µl consisting of: 10.4 µl ddH<sub>2</sub>O, 5.0 µl 10X *Taq* buffer, 2.5 µl MgCl<sub>2</sub>, 1.5 µl dNTPs (10mM), 1.25 µl forward primer (10 µM), 1.25 µl reverse primer (10 µM), 0.125 µl *Taq* polymerase, and 3.0 µl DNA template. PCR reactions were done in an Eppendorf Mastercycler with the following conditions: denaturation at 94°C for 5 min; followed by 35 cycles at 94°C for 30 s, 52°C for 30 s, 72°C for 1 min., and final extension at 72°C for 5 min. PCR products were visualized by gel electrophoresis and purified using exonuclease I/shrimp alkaline phosphatase (ExoSAP). Sequencing reactions were performed at a volume of 12 µl consisting of 8 µl ddH<sub>2</sub>O, 2 µl original forward primer, and 2 µl PCR product and resolved on an automated sequencer (Applied Biosystems Inc. ABI 377) at the Molecular Biology Resource Facility at the University of Tennessee, Knoxville or at the High Throughput Genomics Unit at the University of Washington, Seattle. Sequences were aligned in SEQUENCHER 4.6 (Gene Code Corporation), and ambiguous base calls were manually verified by examining electropherograms. Sequence alignment files were verified and trimmed in MACCLADE 4.07 (Maddison & Maddison, 2005) by comparison to *Leiocephalus* sequences on GenBank. All mtDNA haplotypes generated were deposited in GenBank (GenBank accession nos. JF812182- JF81254).

#### *Mitochondrial DNA analyses*

Redundant haplotypes were collapsed into haplotype groups using the online toolbox FABOX (Villesen 2007). I conducted a Bayesian analysis to infer the mtDNA gene tree in the program MRBAYES 3.1.2 (Hulsenbeck & Ronquist 2001; Ronquist & Hulsenbeck 2003). The best fit model of sequence evolution was determined with Akaike's

information criterion (AIC) implemented in ModelTest Server 3.8 (Posada 2006).

*Leiocephalus carinatus* collected from Boyton Beach, Florida (introduced; native to the Little Bahama Bank) were included in the analysis as an outgroup to evaluate genetic divergence among haplotypes of *L. psammodromus*. Bayesian inference was run for  $1 \times 10^7$  generations using the default temperature (0.2) with four Markov chains per generation, sampling trees every 100 generations. The analysis was repeated twice to verify that the program recovered consistent topology and did not get stuck on local optima (Hulsenbeck & Bollback 2001). Performance of each MCMC run was evaluated by examining convergence in Tracer (Drummond & Rambaut 2007). Nodal support was given as frequency of the recovered clade (posterior probability; Hulsenbeck & Ronquist 2001) after a burn-in of  $2 \times 10^4$  to reach a stable equilibrium of  $-\ln L$  scores.

I used ARLEQUIN (Excoffier *et al.* 2005) to estimate the number of haplotypes per island ( $n$ ), haplotype diversity ( $h$ ), and nucleotide diversity ( $\pi$ ) of island populations. To test for historical demographic events, I used Tajima's (1989)  $D$  and Fu's (1997)  $F_S$  tests to detect departures from neutrality, though it is acknowledged that these tests might be biased in some taxa (Wares 2009). To estimate the level of genetic partitioning among islands and between banks, I calculated  $\Phi$ -statistics in an analysis of molecular variance (AMOVA) framework (Excoffier *et al.* 1992; Excoffier 2003) and population pairwise  $\Phi_{ST}$  estimates implemented in ARLEQUIN. Significance of  $\Phi_{ST}$  values was determined via 50,175 data permutations.

#### *AFLP genotyping*

Amplified fragment length polymorphism (AFLP) markers were obtained following the general procedure described in Vos *et al.* (1995) with modifications from Fitzpatrick *et al.* (2008), Pasachnik (2010), and Pasachnik *et al.* (in press). Whole genomic DNA was digested with the restriction endonucleases *EcoRI* and *MseI* and ligated to *EcoRI* and *MseI* adaptors in an 11 µl reaction at 37°C for 2 hr. The restriction-ligation master mix contained 3.9 µl ddH<sub>2</sub>O, 1.1 µl T4 DNA ligase buffer (New England Biolabs, Beverly, Massachusetts), 0.55 µl BSA (1mg/mL; NEB), 1.1 µl NaCl (0.5M), 0.1 µl *MseI* (NEB), 0.42 µl *EcoRI* (NEB), 0.33 µl T4 DNA ligase (NEB), 1 µl *MseI* adaptor (100 µM), 1 µl *EcoRI* adaptor (10 µM) and 1.5 µl template DNA. The product was then diluted with 60 µl of 1X TE Buffer.

One pre-selective amplification was performed with a 20 µl reaction mix containing 8.1 µl ddH<sub>2</sub>O, 5.0 µl 5X *Taq* buffer, 1.6 µl MgCl<sub>2</sub>, 0.3 µl BSA (1 mg/ml), 0.5 µl dNTPs, (10mM) 0.5 µl *MseI* pre-selective primer (20 µM), 0.5 µl *EcoRI* pre-selective primer (20 µM), 0.5 µl *Taq* polymerase, and 4.0 µl of the diluted restriction/ligation product. Pre-selective product was diluted with 40 µl of 1X TE Buffer and 200 µl ddH<sub>2</sub>O. Two primer combinations (Applied Biosystems) were used for the selective PCR stage (fluorescent dye in parenthesis): *EcoRI* (6-FAM)-ACA / *MseI*-CTAG-GATGAGTCCTGAGTAACTAG and *EcoRI* (NED)-AGC / *MseI*-CTAC-GATGAGTCCTGAGTAACTAC (Fitzpatrick *et al.* 2008; Pasachnik 2010). Both selective amplifications were performed at a reaction volume of 20 µl containing 6.1 µl ddH<sub>2</sub>O, 4.0 µl 5X *Taq* buffer, 1.6 µl MgCl<sub>2</sub>, 0.3 µl BSA (1 mg/ml), 0.5 µl dNTPs (10mM), 2.0 µl *MseI* selective primer (5 µM), 2.0 µl *EcoRI* selective primer (1 µM), 0.5 µl *Taq* polymerase, and 3.0 µl of the diluted pre-selective product. Both pre-selective and

selective reactions were done in an Eppendorf Mastercycler with conditions from Vos *et al.* (1995).

Fragment analysis was performed at a volume of 15  $\mu$ l. One  $\mu$ l of each fluorescently labeled PCR product was multiplexed by individual, along with 0.5  $\mu$ l GeneScan ROX 350 internal size standard (Applied Biosystems) and 12.5  $\mu$ l HiDi formamide (Applied Biosystems), and run on an ABI 3100 automated capillary sequencer at the University of Tennessee. A complete technical replicate was run for a subset of 40 randomly selected individuals (15% of the total) from the restriction-ligation stage. Four negative controls were run at four different times to test for systematic contamination and dye blobs. Loci at which two or more negative controls showed a peak were removed from downstream analyses.

AFLP profiles were aligned and called using GeneMarker 1.91 software (SoftGenetics LLC). Bin sizes were manually adjusted where needed, similarly to the method described in Whitlock *et al.* (2008), to center bins over peaks and to remove overlapping bins (potential homoplasy). Default parameter settings in GeneMarker have been shown to introduce more error in automated scoring (Holland *et al.* 2008), so I used the following settings: peak detection threshold (min 50, max 30000); minimum fragment length (50); local and global detection percentages (0%); and stutter peak filter (off). A matrix of peak heights was then exported to an Excel file.

Raw AFLP data were scored using AFLPSCORE 1.4b (Whitlock *et al.* 2008), implemented in the statistical environment R *x64* 2.11.1 (R Development Core Team 2010). Data were normalized to the median peak height and filtered with a 200 relative florescence unit (rfu) locus selection threshold and a relative genotype calling threshold

of 1%. A range of locus (50 to 900) and genotype (0 to 100%) thresholds were tested to find values that minimized the mismatch error rate and the  $\varepsilon_{1,0}$  error rate while still retaining as many loci as possible. The mismatch error rate for the optimized thresholds was 5.71% and  $\varepsilon_{1,0}$  error rate was 0.31%, and the filtered dataset consisted of a matrix of binary AFLP profiles.

### *AFLP Analyses*

The MCMC clustering algorithm STRUCTURE 2.3.3 (Pritchard *et al.* 2000) was used to investigate population structure in the AFLP dataset, which consisted of 19 sample locations from ten islands (Fig. 1). Twenty independent runs were conducted for each value of  $K = 1$  to  $K = 10$ , with 100,000 generations of burn-in and 1,000,000 post-burn-in replicates using the admixture model. Values of  $K$  were estimated by the  $\Delta K$  method, an *ad hoc* statistic that estimates the optimal value of  $K$  across multiple iterations of the STRUCTURE analysis (Evanno *et al.* 2005). STRUCTURE results were then visualized using DISTRUCT (Rosenberg 2004). Because clustering algorithms use stochastic simulation algorithms, different analyses might result in different clustering outcomes; hence I used the program BAPS 5 (Corander *et al.* 2004) to check for similar clustering in my dataset. BAPS uses a Bayesian mixture model and separate joint probability distributions across loci to characterize variation in subgroups and to assign individuals to optimal clusters of genetic similarity (Corander *et al.* 2004). I performed the “clustering of individuals” analysis with 1-10 maximum clusters and 1000 iterations to estimate the admixture coefficient for each individual.

Jaccard distances were calculated from binary AFLP data using the *vegdist* function implemented in the R 2.11.1 package *vegan* 1.17 (Oksanen *et al.* 2010). Jaccard distances are considered more appropriate for dominant data because they do not attribute meaning to the coincidence of band absence and instead rely only on the shared presence of bands (Mattioni *et al.* 2002; Lowe *et al.* 2004). In this case, shared absence is not considered in estimates of similarity from dominant data (Rieseberg 1996), and these distances do not assume that band absence indicates homology. To estimate the level of genetic partitioning among islands and between banks, I calculated  $\Phi$ -statistics in the analysis of molecular variance (AMOVA) framework (Excoffier *et al.* 1992) using Jaccard distance data implemented in ARLEQUIN. I also calculated pairwise values of  $\Phi_{ST}$  between all populations and estimated the number of haplotypes ( $n$ ) of island populations. It should be noted that this method of calculating  $\Phi_{ST}$  makes no assumption of Hardy-Weinberg equilibrium; we treat the AMOVA as a general method of partitioning a pairwise distance matrix into hierarchical components (Excoffier *et al.* 1992).

Gene diversity ( $H_e$ ) was calculated using Zhivotovsky's (1999) Bayesian estimator implemented in R (Appendix S2), which is superior to the method of Lynch & Milligan (1994) for dominant data with low variation (Zhivotovsky 1999). The number of segregating sites and fragment frequencies were calculated using AFLP-SURV 1.0 (Vekemans *et al.* 2002). Permutation tests were implemented in R (Appendix S2) to test for a difference between the distribution of observed genotypes in a sample and that expected based on allele frequencies and the number of individuals in the sample. The null hypothesis is that each lizard genotype is a random sample of alleles, which is then

evaluated across 10,000 permutations of random draws from the population given the fragment frequencies of each locus in the population.

### *Island Biogeography*

I evaluated four predictions about genetic variation within and between islands derived from migration-mutation-drift equilibrium: 1) I expect a positive relationship between genetic diversity and island area, assuming population size is associated with island area; 2) I expect a negative relationship between genetic diversity and isolation; 3) I expect a consistent positive relationship between genetic divergence and geographic distance between islands – the pattern of “isolation by distance” (Wright 1943; Rousset 1997; Koizumi *et al.* 2006; and 4) I expect no island-specific deviations from standard “neutrality” tests, which also indicate demographic stability (Finn *et al.* 2009).

Alternatively, in a “relaxation” archipelago characterized by zero dispersal since island formation (Brown 1971), I expect a strong diversity-area relationship but no effect of geographic distances on diversity or divergence, and no evidence of recent colonization (*e.g.*, signatures of population expansion). In a metapopulation characterized by extinction and recolonization, I expect the pattern of isolation by distance to be disrupted (Koizumi *et al.* 2006) and frequent violations of expected neutral distributions of allele frequencies.

To test for patterns of isolation by distance (IBD) I calculated genetic distances as Rousset’s (1997) distance measure ( $\Phi_{ST} / (1 - \Phi_{ST})$ ) for mtDNA and AFLP (Jaccard distance) data. These were then regressed against the log-transformed geographic distance, measured using ARCGIS v9 (ESRI Inc., Redland CA), in the online program IBD

v3.16 (Jensen *et al.*, 2005) using 30,000 randomizations. To control for separation between banks, I established an indicator variable matrix to distinguish between the Turks and Caicos Banks and plotted the reduced major axis regression. Correlations were then evaluated with partial Mantel tests (Wright 1943; Mantel 1967; Legendre & Legendre 1998). I partitioned the data in two ways: First, I included all samples from an island together as a single population, and second I separated samples from distinct sampling sites (Fig. 1) within islands into separate “populations.” For the first partition, two types of genetic distance matrices were constructed for both mtDNA and AFLP data: 1) Geographic distance, which is the Euclidean distance between island centroids; and 2) Stepping-stone distance, which is the sum of island centroid distances for extant islands separating samples. For the separated sampling localities partition, I used geographic distance matrices 1 and 2; as well as a third geographic distance matrix: island distance, which is an ordinal increase by 1 with each step between islands, such that populations on the same island are separated by “0” while populations on nearest-neighbor islands are separated by “1,” next-nearest neighbors by “2,” and so forth.

It is possible that one or more islands with reduced gene flow or non-equilibrium dynamics could drive a signal of IBD within the Caicos Bank (Slatkin 1993; Hutchison & Templeton 1999). To test for outlier islands, I performed a decomposed pairwise regression (DPR; Koizumi *et al.* 2006) of genetic distance ( $\Phi_{ST} / (1 - \Phi_{ST})$ ) versus stepping-stone geographic distance using the R package *DPR* (Reynolds & Fitzpatrick 2011) for AFLP data. This method identifies outliers by comparing the inclusion and exclusion of populations using Akaike’s information criterion ( $AIC_C$ ; Burnham & Anderson 2004) as an *ad hoc* criterion. Although DPR appears to have rather low

statistical power in simple metapopulation models (Chpt. I), Koizumi *et al.* (2006) were able to identify several outlier populations in salmonids whose deviance from the modal pattern could be explained by known recent events such as river dam construction.

To investigate predictions of Johnson *et al.* (2000), I tested for correlations between gene diversity ( $H_e$ ), island area ( $\text{km}^2$ ), and island isolation (km) in a multiple regression framework implemented in R. Separate analyses were done for mtDNA and AFLP data. Island area was estimated from data available from the Turks and Caicos Department of Economic Planning and Statistics (2011) that include the total land area of each island, including tidal marshes. Regrettably, more detailed spatial vegetation maps do not exist for the Turks and Caicos Islands; hence I am unable to precisely calculate the potential area where populations of *L. psammodromus* are most likely to occur, especially for islands with heterogeneous habitat such as Providenciales, North Caicos, and Middle Caicos. However, diurnal visual encounter transects (Reynolds unpubl.) indicate that this species is widespread across these larger islands and occupies most habitats, indicating that total island area is a reasonable estimate of the realized range of *L. psammodromus* on each island. Island isolation was calculated as the proximity index patch metric (Gustafson and Parker 1992):

$$\sum_{s=1}^n \frac{a_{tjs}}{h_{tjs}}$$

where  $a_{ijs}$  is the area ( $\text{km}^2$ ) of patch  $ijs$  in the neighborhood of patch  $ij$ , and  $h_{ijs}$  is the distance (km) between patch  $ijs$  and patch  $ij$ . The proximity index was calculated from the Euclidean geographic distance between island centroids. Here, a high index value indicates closer proximity to neighboring patches, as weighed by patch area.

Finally, although there are currently no models for testing deviations from mutation-drift equilibrium in AFLP data, the standard coalescent-based tests for mtDNA (Tajima's  $D$ , Fu's  $F_S$ ) can be used as indicators of recent demographic instability (Finn *et al.* 2009).

## Results

### *Mitochondrial DNA analyses*

I resolved 795 bp of mitochondrial ND2 for 257 individuals, for which there were 87 variable nucleotide positions (11% of total nucleotide positions). A total of 80 haplotypes were observed, 67 from the Caicos Bank and 13 from the Turks Bank (Table 2). The Bayesian tree was estimated using the Hasegawa-Kishino-Yano (HKY+I+G) model (proportion of invariant sites = 0.7) with a log-likelihood of  $-\text{LnL} = 1984$  (Fig. 2). Pairwise  $\Phi_{ST}$  among islands ranged from 0.11 (North Caicos-Middle Caicos) to 0.98 (Long Cay Turks Bank-Little Ambergris Cay) (Table 3). Haplotype diversity was lowest for Little Ambergris Cay ( $h = 0.25 \pm 0.13$ ) and highest for North Caicos ( $h = 0.92 \pm 0.03$ ), while nucleotide diversity was lowest for Gibbs Cay ( $\pi = 0.0007 \pm 0.18$ ) and highest for Big Ambergris Cay ( $\pi = 0.25 \pm 0.00$ ) (Table 2).

AMOVA (Table 4) revealed that 72.5% of variation ( $\Phi_{CT} = 0.72$ ) is explained by grouping between the Turks and Caicos Banks. Variation among populations within the Caicos Bank accounted for 14.8% of the variation ( $\Phi_{SC} = 0.53$ ), while variation within populations accounted for 12.7% of the variation ( $\Phi_{ST} = 0.87$ ).

Tests for neutrality using Tajima's  $D$  were mostly non-significant, though I found evidence of a significant departure from neutrality on Middle Caicos ( $D = -1.89$ ,  $P = 0.01$ ) and North Caicos ( $D = 1.46$ ,  $P = 0.049$ ). Fu's  $F_S$  statistic, which is considered a more powerful test to detect population expansion (Ramos-Onsins & Rozas 2002), was significantly negative for three islands: North Caicos ( $F_S = -8.6$ ,  $P = 0.002$ ), Long Cay Caicos Bank ( $F_S = -3.41$ ,  $P = 0.023$ ), and Big Ambergris Cay ( $F_S = -4.1$ ,  $P = 0.018$ ). These results suggest a trend of demographic expansion on some islands on the Caicos Bank.

#### *AFLP Analysis*

I scored 70 AFLP fragments in 255 individuals ranging from 50 to 352 base pairs, 33 (47.1%) of which were segregating. I found between two (Big Sand Cay) and 15 (North Caicos and Long Cay Caicos) genotypes for each island (Table 2). Permutation tests revealed that for four populations (North, Big Ambergris, Little Ambergris, and Gibbs Cay) the number of unique genotypes in each population was significantly less than the expectation given the number of individuals sampled (Table 1). Pairwise  $\Phi_{ST}$  among islands ranged from 0.01 to 0.69 (Table 3). Gene diversity was highest for Big Ambergris Cay ( $H_e = 0.135$ ) and lowest for Long Cay on the Turks Bank ( $H_e = 0.058$ ) (Table 2).

AMOVA (Table 4) revealed an estimated 46.3% of variation ( $\Phi_{CT} = 0.46$ ) is explained by grouping between the Turks and Caicos Banks (Table 4). Variation among populations within the Caicos Bank accounted for 6.32% of the variation ( $\Phi_{SC} = 0.12$ ), while variation within populations accounted for 47.3% of the variation ( $\Phi_{ST} = 0.53$ ).

Structure analysis returned  $K = 3$  as the optimal value for  $K$  for the combined dataset, with two groups partitioned by bank (Caicos = white, Turks = grey; Fig. 3) and one group (Shared = black) shared between the Turks and Caicos Banks (Fig. 3). However, one individual on the Caicos Bank shares more than 50% of its genome with the Turks Bank group. BAPS returned a nearly identical clustering of individuals to that given by the STRUCTURE analysis, identifying three clusters with high posterior probability ( $\text{Pr}(K = 3) = 1.00$ ). The only difference between STRUCTURE and BAPS results was the inclusion of two Gibbs Cay individuals in the Shared group in STRUCTURE (black, Fig. 2), while BAPS found no individuals from the northern Turks Bank clustering with the Shared group.

### *Island Biogeography*

Regression of Rousset's (1997) genetic distances on log-transformed geographic distances revealed little signal of isolation by distance on either the Turks or Caicos Banks. For samples grouped by island, Mantel tests indicated a lack of correlation between genetic and geographic distance for mtDNA ( $r = -0.09$ ,  $P = 0.75$ ) and AFLP ( $r = -0.06$ ,  $P = 0.63$ ) data, as well as for stepping stone distance for mtDNA ( $r = -0.13$ ,  $P = 0.82$ ) and AFLP ( $r = -0.02$ ,  $P = 0.52$ ) data.

For mtDNA samples grouped by sampling location, partial Mantel tests (one-tailed  $P$  values) indicated no significant correlation between genetic and geographic distance ( $r = 0.12$ ,  $P = 0.12$ ), or stepping stone distance ( $r = 0.11$ ,  $P = 0.11$ ) and island distance ( $r = 0.14$ ,  $P = 0.05$ ). For AFLP samples grouped by sampling location, geographic ( $r = 0.25$ ,  $P = 0.03$ ) and stepping-stone distances ( $r = 0.22$ ,  $P = 0.03$ ) showed

correlations with genetic distance, while island distance did not show significant correlation ( $r = 0.12$ ,  $P = 0.09$ ). Overall, grouping samples by specific locality rather than lumping samples within islands resulted in a modest pattern of positive correlation.

Decomposed pairwise regression analysis identified Long Cay (Caicos Bank) as an outlier from the rest of the islands on the Caicos Bank with a generally elevated level of differentiation from other Caicos islands (Fig. 5), indicating possible reduced gene flow or non-equilibrium dynamics in this population.

Multiple regression of gene diversity on island area and island isolation (proximity index) (Table 1) indicated no significant association between diversity and either island area (Slope = 0.0003,  $T = 0.2$ ,  $P = 0.8$ ) or island isolation (Slope = -0.007,  $T = -0.5$ ,  $P = 0.6$ ). The same result was found for haplotype diversity in the mtDNA dataset, with both island area (Slope = 0.009,  $T = 0.34$ ,  $P = 0.76$ ) and island proximity (Slope = 0.131,  $T = 0.47$ ,  $P = 0.65$ ) showing no relationship to genetic diversity.

Multicollinearity between variables did not appear to influence multiple regression, as simple linear regressions of gene diversity on island area indicated no correlation for mtDNA (Slope = 0.014,  $T = 0.56$ ,  $P = 0.6$ ) or AFLP (Slope = 0.0,  $T = 0.05$ ,  $P = 0.9$ ) data. Proximity was also not correlated with gene diversity for mtDNA (Slope = 0.16,  $T = 0.68$ ,  $P = 0.52$ ) or AFLP (Slope = -0.006,  $T = -0.5$ ,  $P = 0.6$ ) data.

## **Discussion**

Genetic variation within and among populations of organisms is the foundation for evolutionary processes, such as speciation and adaptation (Conner & Hartl 2004), and the

study of population genetics in discrete spatially-distributed organisms might lend understanding to such important mechanisms of evolution (Thorpe 2005). Current genetic structure results from a combination of historical and contemporary population processes of extinction, colonization, bottleneck and expansion, and gene flow. Studies using both mtDNA and anonymous nuclear markers (AFLPs) can combine information of historical demography (phylogeography) and landscape (or population) genetics (Thomson *et al.* 2010; Wang 2010). In addition, the neutral theories of island biogeography and genetic diversity provide the opportunity to test for deviations from null hypotheses using empirical data. I investigated the partitioning of genetic diversity in an endemic lizard distributed on an island archipelago using the methods of phylogeography, landscape genetics, and island biogeography. For islands in mutation-migration-drift equilibrium, I expected to find a positive relationship between genetic diversity and island area as well as geographic distance between islands, a negative relationship between genetic diversity and island isolation, and demographic stability of island populations. Contrary to these expectations, I found little or no relationship between genetic diversity and island area, island isolation, or island geographic distance (IBD); and I found evidence for island-specific deviation from neutrality, indicating the presence of non-equilibrium dynamics. Overall, my results indicate significant gene flow both among islands on the same banks as well as across banks, biased gene flow from the Turks Bank to the Caicos Bank, and a departure from mutation-migration-drift equilibrium.

## **Phylogeographic Analysis**

### *Mitochondrial DNA analyses*

Phylogeographic analyses of a portion of mitochondrial ND2 revealed a distinct haplogroup on the Turks Bank (Fig. 2; Table 4). Some genetic structure was observed within banks, but many haplotypes are minimally divergent from each other (Fig. 2). Three haplotypes from Big Sand Cay grouped with Caicos Bank haplotypes in the Bayesian analysis, indicating incomplete lineage sorting or ongoing gene flow. Neutrality tests revealed some signature of population expansion on Middle and North Caicos, Big Ambergris Cay, and Long Cay (Caicos Bank). Middle and North Caicos have been recently deforested (100-250 years before present; Reynolds 2011), increasing the available habitat for *L. psammodromus* across the interiors of these islands. Indeed, *L. psammodromus* is absent only from areas of dense coppice or where little open ground exposed to direct sunlight occurs. It is likely that this species was largely confined to the beach and scrub areas of these islands until fairly recently, and they are now found throughout each of these large islands. Long Cay was recently plagued by invasive cats (*Felis silvestris catus*), which led to the extirpation of rock iguanas (Mitchell *et al.* 2002) and potentially boas (Reynolds 2011). Cats will readily consume curly-tailed lizards (Iverson 1978) and could have severely reduced this population. Cats were removed from the island in 1999 and iguanas were subsequently reintroduced and have successfully established (Mitchell *et al.* 2002); and Reynolds *et al.* (in review) have suggested that boas be reintroduced as well. The signature of expansion could be a result of the recovery of this population since 1999. Big Ambergris Cay has remained more or less stable as a privately protected island since the early 20<sup>th</sup> century, and it was occupied in the 19<sup>th</sup> century as a whaling outpost and small plantation; however, it is unclear whether this

would have impacted *L. psammodromus*, and the signature of expansion could be due to naturally driven non-equilibrium dynamics.

## **Population Genetic Analysis**

### *AFLP analysis*

Clustering analysis of anonymous nuclear markers (AFLPs) in both STRUCTURE and BAPS indicated the presence of three *L. psammodromus* clusters or putative ancestral lineages in the Turks and Caicos. Group 1 (Fig. 3; white) and Group 2 (gray) are idiosyncratic of the Caicos and Turks Banks, respectively, except for a single individual of the Turks group that was sampled from the Caicos Cays on the Caicos Bank (Fig. 3). This individual shares the majority of its alleles with the Turks Bank group and with Group 3 (black), which is found on both the Turks and Caicos Banks. Superficially, lineage separation between the Turks and Caicos banks seems reasonable, as dispersal across a water gap (minimum > ~16 km) is the only means of accomplishing gene flow between these groups. However, I find that two groups occur on the Turks Bank, a more northern group (group 2; Fig. 3) exclusive to the Turks Bank and a group at higher frequency on the southern island of Big Sand Cay, which is shared with the Caicos Bank. Though this shared group is not idiosyncratic of a particular island or population, it does appear to occur at higher frequencies in some populations as opposed to others (Fig. 6). These clusters of shared group genotypes appear to be non-randomly distributed within islands, which might be a consequence of episodic dispersal of larger groups of individuals instead of diffuse individual dispersal.

Hurricane tracks, wind, and water currents in the region generally move from

southeast to northwest (Keegan 1992; Sealey 2006), with an occasional current set to the southwest (National Geospatial Intelligence Agency 2004); hence dispersal (which is passive) would be expected to move from the Turks Bank to the Caicos Bank. However, for individuals leaving from Big Sand Cay, the Caicos Bank is a much bigger target heading northwest and they have a more reasonable chance of making landfall prior to exiting the Turks Island Passage to the north, a situation Keegan (1992) refers to as the “target effect.” If the horizon relative to islands on the Turks Bank is divided into arc segments that define the probabilities of landfall on the Caicos Bank spanning 90° west to north, then individuals leaving from Big Sand Cay (or the southern Turks Bank) have an 86% probability of making landfall on the Caicos Bank, while individuals leaving from Gibbs Cay (or the northern Turks Bank) have a 57% probability, assuming current and wind trajectories between due west and due north.

Given my data, it appears that biased gene flow exists from the Turks Bank to the Caicos Bank. Group 3 occurs in a higher frequency on the southern Turks Bank, hence a likely explanation is that most propagules to the Caicos Bank arrive from the southern Turks Bank. I detected one individual on the Caicos Bank (Caicos Cays) that clustered with the Turks Bank, consistent with some gene flow from the Turks Bank group.

Permutation tests for the expected number of genotypes in a population revealed that four populations had fewer unique genotypes than expected given the number of individuals sampled (Table 1). This result would be expected when significant linkage disequilibrium caused association of alleles at various loci and could be due to small population size (drift), selective sweeps, or gene flow (Allendorf & Luikart 2007).

## Island Biogeography Analyses

I found no significant signal of isolation by distance between islands on either the Turks or Caicos Bank. I also found that between-island isolation by distance is no greater than within-island isolation by distance; indeed the only signature of IBD was for AFLP data on Caicos Bank islands driven largely by the relatively large divergence between Long Cay and other Caicos islands (Tables 1, 3). In a simple island-mainland or stepping-stone scenario, isolation is a sufficient description of the distance between populations.

However, an archipelago is a more complex arrangement of isolation distances between populations, hence a proximity index better characterizes the actual isolation in an archipelago. Interestingly, and contrary to the suggestions of Johnson *et al.* (2000), I found no correlation between island area or isolation and genetic diversity. This observation could be due to the relatively recent inundation of the Turks and Caicos Banks and a slow attrition of alleles from small populations. Alternatively, this may be an effect of ongoing gene flow among islands, which would serve to homogenize populations and maintain diversity on smaller islands (Pannell 2003; Bay *et al.* 2008).

The expectation for mutation-drift in isolated populations after vicariance is a gradual loss of alleles and approach to a new mutation-drift equilibrium negatively related to island area. To verify that this relaxation model predicts a diversity-area relationship, I simulated 70 neutral, diallelic markers in isolated island populations of three effective population sizes (assumed to be a function of island area) and varying initial gene diversity and mutation rate. This provides a model of attrition of alleles post-vicariance on the Turks and Caicos banks, which occurred as recently as 8,000 years ago (Morgan 1989; Keegan 1992). Given a generation time in *L. psammodromus* of two years

(Smith & Iverson 1993), I simulated 4,000 generations of drift in hypothetical island populations and compared mean expected gene diversity in isolated populations to the observed gene diversity calculated from AFLP profiles. These simulations (Fig. S1) indicated that gene diversity declines substantially after 4,000 generations of isolation in small populations ( $N_e = 100$ ; 1000) but not in larger populations ( $N_e = 10,000$ ), as expected for an island-area effect. My empirical results, however, are not consistent with this prediction- I found no relationship between island area and gene diversity, despite a range of areas spanning five orders of magnitude. Therefore we reject the relaxation hypothesis and favor the idea that gene flow between islands is driving gene diversity in each population, and not attrition of alleles in isolation.

In a stepping-stone population at equilibrium, there is a nearly linear relationship expected between  $F_{ST}/(1 - F_{ST})$  and geographic distance between pairs of populations (Rousset 1997). However, slope and intercept estimates can be heavily influenced by the inclusion of "outlier" populations that are exceptionally divergent from or similar to other demes in the archipelago. I found that Long Cay (Caicos Bank) was identified as an outlier in decomposed pairwise regression analysis, indicating non-equilibrium dynamics in this population. This observation is supported by the results from Fu's  $F_S$ , which suggests that the population has recently expanded. Examination of the decomposed regression (Fig. 5) indicates that this island has a higher intercept than other populations, a situation expected for populations where the relative strength of drift is much greater than that of gene flow (Hutchinson & Templeton 1999; Koizumi *et al.* 2006). Drift would be expected to be far stronger in populations with a small effective population size. Since *L. psammodromus* has been extirpated from the nearby (0.75 km) island of South Caicos,

Long Cay has the lowest proximity to any other island in the Caicos Bank (Table 1).

Taken together, these results suggest that this population has experienced little gene flow owing to the extirpation of *L. psammodromus* from South Caicos and has increased in size since the eradication of cats in 1999.

### **Systematics and Conservation**

Information on taxonomy and distribution is important in the context of understanding evolutionary relationships and identifying conservation priorities (Frankham *et al.* 2002; Frankham 2006). In the Turks and Caicos Islands, little is known about the ten native reptile species except for the rock iguana (*Cyclura carinata*) and Turks Island boa (*Epicrates chrysogaster*) (Reynolds 2011). Of particular concern are the ongoing threats to the dominant terrestrial herpetofauna, including loss of habitat and the introduction of feral mammalian predators (Iverson 1978; Corke 1992; Smith *et al.* 2005; Tolson & Henderson 2006; Reynolds 2011), as well as non-native species (Reynolds & Niemiller 2010; Reynolds 2011). *Leiocephalus psammodromus* is not immune to anthropogenic activities, and several notable extirpations appear to have occurred on South Caicos, Grand Turk, Cotton Cay, and Salt Cay (Table S1; Reynolds 2011). The current accepted taxonomy of this species includes six subspecies, two on the Turks Bank and four on the Caicos Bank (Fig. 1; Schwartz 1967). Though the utility of subspecific appellations is controversial in taxonomy and systematics (Manier 2004), it might prove more useful as a heuristic tool (Mayr 1982), especially if it corresponds to reciprocally monophyletic groups (Zink 2004). In particular, differences in gross morphology could indicate meaningful population differentiation in an archipelago and provide context for the

justification of independent conservation initiatives (Frankham *et al.* 2002). If the current classification for this species is relevant, then at least two subspecies, *mounax* and *aphretor*, have been extirpated from the majority of their range (Fig. 1). However, my analysis that includes individuals from all six purported subspecies does not support current taxonomy. Mitochondrial DNA suggests that the Turks Bank harbors a unique haplogroup, while AFLP data suggest that the species is composed of three groups, two of which are idiosyncratic to their respective bank and another that occurs across the Turks and Caicos Islands. It is not my intention here to reorganize the taxonomy of this species; however it is clear that the current taxonomy does not correspond to unique or independent genetic lineages and hence should likely not be recognized.

### **A More Comprehensive Approach for Island Archipelagos**

Although progress has been made toward unification of metapopulation ecology and genetics (*e.g.*, Hanski & Gaggiotti 2004), cross-fertilization has only begun and the potential for insights and approaches from one to enrich the others is not yet fully realized. Studies of genetic distribution in archipelagos frequently use only the methods of phylogeography (*e.g.*, gene trees and haplotype networks) and landscape genetics (*e.g.*, estimates of gene flow and recent historical demography), sometimes confusing the two approaches (Wang 2010). However, Johnson *et al.* (2000) outlined an approach for predicting the distribution of genetic diversity on islands using methods from island biogeography, in particular island-mainland scenarios. My analysis of genetic diversity across and island archipelago indicates that a more complete understanding of processes influencing genetic partitioning is derived from multiple methods of inference. In the

study system, vicariance phylogeography caused by sea level fluctuation influences the partitioning of mitochondrial haplotypes, while the population genetic process of gene flow owing to dispersal across aquatic barriers leads to unidirectional gene flow in anonymous nuclear markers. Finally, island axes of variation (area and isolation) were found to have no relationship to genetic diversity within banks, a departure from neutral equilibrium expectations (Johnson *et al.* 2000) and evidence of non-equilibrium processes such as extinction-recolonization or frequent bottlenecks. Unless all three approaches are investigated in an archipelago-structured population, processes influencing the genetic structuring in the archipelago might be missed, a problem of even more significance if the species is of major conservation concern. I propose that studies of genetic diversity on island archipelagos use methods from all three schools: phylogeography, landscape genetics, and island biogeography, to evaluate empirical data in the context of equilibrium and the null hypotheses of island biogeography and population genetics.

## Literature Cited

Allendorf FW, Luikart G (2007) Conservation and the genetics of populations. Blackwell Publishing, Oxford, UK.

Barbour T (1916) Additional notes on West Indian reptiles and amphibians. *Proc Biol Soc Washington* 29:215-220.

Bay LK, Caley MJ, Crozier RH (2008) Meta-population structure in a coral reef fish demonstrated by genetic data on patterns of migration, extinction and re-colonization. *BMC Evol Biol* 8:248.

Brown JH (1971) Mammals on mountaintops: non-equilibrium insular biogeography. *Am Nat* 105:467-478.

Burnham KP, Anderson DR (2004) Multimodal inference: understanding AIC and BIC in model selection. *Sociological Methods Res* 33:261-304.

Camargo A, Sinervo B, Sites JW Jr (2010) Lizards as model organisms for linking phylogeographic and speciation studies. *Mol Ecol* 19:3250-3270.

Conner JK, Hartl DL (2004) A primer of ecological genetics. Sinauer, Massachusetts, USA.

Corander J, Marttinen P (2006) Bayesian identification of admixture events using multi-locus molecular markers. *Mol Ecol* 15:2833-2843.

Corke D (1992) The status and conservation needs of the terrestrial herpetofauna of the Windward Islands (West Indies). *Biol Cons* 62:47-58.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.

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

Ewens W (1972) The sampling theory of selectively neutral alleles. *Theor Pop Biol* 3:87-112.

Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479-491

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

Fairbanks RG (1989) A 17,000 year glacio-eustatic sea level record: Influence of glacial melting rates on the Younger Dryas event and deep-ocean circulation. *Nature* 342:637-642.

Finn DS, Bogan MT, Lytle DA (2009) Demographic stability metrics for conservation prioritization of isolated populations. *Cons Biol* 23:1185-1194.

Fitzpatrick BM, Placyk JS, Niemiller ML, Casper GS, Burghardt GM (2008) Distinctiveness in the face of gene flow: hybridization between specialist and generalist gartersnakes. *Mol Ecol* 17:4107-4117.

Frankham R (2006) Evolutionary conservation genetics. In: Fox, C.W., Wolf, J.B. (Eds.), *Evolutionary Genetics*. Oxford University Press, New York, pp 502-512.

Frankham R, Ballou JD, Briscoe DA (2002) *Introduction to conservation genetics*. Cambridge University Press, Cambridge.

Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking, and background selection. *Genetics* 147:915-925.

Gustafson EJ, Parker GR (1992) Relationship between landcover proportion and indices of landscape spatial pattern. *Landsc Ecol* 7:101-110.

Hanski I, Gaggiotti OE (2004) *Ecology, genetics, and evolution of metapopulations*. Elsevier Academic Press, Burlington, Massachusetts.

Hedges SB (1996) The origin of West Indian amphibians and reptiles. In: Powell, R., Henderson, R.W. (Eds.), *Contributions to West Indian herpetology: a Tribute to Albert Schwartz*. Society for the Study of Amphibians and Reptiles, Ithaca, New York, pp 95-128.

Hedges SB (2006) Paleogeography of the Antilles and the origin of West Indian terrestrial vertebrates. *Ann Missouri Bot Gard* 93:231-244.

Hedges SB, Hass CA, Maxon LR (1992) Caribbean biogeography: molecular evidence for dispersal in West Indian terrestrial vertebrates. *Proc Nat Acad Sci USA* 89:1909-1913.

Heinicke MP, Duellman WE, Hedges SB (2007) Major Caribbean and Central American frog faunas originated by ancient oceanic dispersal. *Proc Nat Acad Sci USA* 104:10092-10097.

Holland BR, Clarke AC, Meudt HM (2008) Optimizing automated AFLP scoring parameters to improve phylogenetic resolution. *Syst Biol* 57:347-366.

- Hubbell SP (2001) The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, USA.
- Hulsenbeck JP, Bollback JP (2001) Empirical and hierarchical Bayesian estimation of ancestral states. *Syst Biol* 50:351-366.
- Hulsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17:754-755.
- Hutchinson DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* 53:1898- 1914.
- Iturralde-Vinent MA, MacPhee RDA (1999) Paleogeography of the Caribbean region: implications for Cenozoic biogeography. *Bull Am Mus Nat Hist* 238:1-95.
- Iverson JB (1978) The impact of feral cats and dogs on populations of the West Indian rock iguana *Cyclura carinata*. *Biol Cons* 14:63-73.
- Iverson JB, Smith GR (2006) *Leiocephalus psammmodromus*. (Turks and Caicos curly-tailed Lizard). Necrophagy. *Herpetol Rev* 37:345-346.
- Jensen JL, Bohonak AJ, Kelly ST (2005) Isolation by distance, web service, v.3.16. *BMC Genet* 6:13.
- Johnson KP, Adler FR, Cherry JL (2000) Genetic and phylogenetic consequences of island biogeography. *Evolution* 54:387-396.
- Keegan WF (1992) The people who discovered Columbus. University Press of Florida, Gainesville, Florida, USA.
- Koizumi I, Yamamoto S, Maekawa K (2006) Decomposed pairwise regression analysis of genetic and geographic distances reveals a metapopulation structure of stream-dwelling Dolly Varden char. *Mol Ecol* 15:3175-3189.
- Legendre P, Legendre L (1998) Numerical ecology. 2<sup>nd</sup> edition. Elsevier Academic Press, New York, USA.
- Lighty RG, Macintyre JG, Stuckenrath R (1982) Acropora palmate reef framework: a reliable indicator of sea level in the western Atlantic for the past 10,000 years. *Coral Reefs* 1:25-30.
- Losos JB (2009) Lizard in an evolutionary tree. University of California Press, Berkeley, California, USA.

Losos JB, Jackman TR, Larson A, de Queiroz K, Rodríguez-Schettino L (1998) Historical contingency and determinism in replicated adaptive radiations of island lizards. *Science* 279:2115-2118.

Losos JB, Ricklefs RE (2009) Adaptation and diversification on islands. *Nature* 457:830-836.

Lowe A, Harris S, Ashton P (2004) *Ecological genetics*. Blackwell Science Ltd, Oxford, UK.

Lynch M, Milligan BG (1994) Analysis of population genetic structure with RAPD markers. *Mol Ecol* 3:91-99.

MacArthur RH, Wilson EO (1963) An equilibrium theory of insular zoogeography. *Evolution* 17:373-387.

MacArthur RH, Wilson EO (1967) *The theory of island biogeography*. Princeton University Press, New Jersey, USA.

Macey JR, Larson A, Ananjeva NB, Fang ZL, Papenfuss TJ (1997) Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. *Mol Biol Evol* 14:91-104.

Maclean WP, Kellner R, Dennis H (1977) Island lists of West Indian amphibians and reptiles. *Smithsonian Herpetol Inf Serv No.* 40, Washington, DC.

Maddison DR, Maddison WP (2005) *MacClade 4* release version 4.07 for OSX. Sinauer, Sunderland, Massachusetts, USA.

Manier MK (2004) Geographic variation in the long-nosed snake, *Rhinocheilus lecontei* (Colubridae): beyond the subspecies debate. *Biol J Linn Soc* 83:65-85.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* 27:209-220.

Mattioni C, Casasoli M, Gonzalez M, Ipinza R (2002) Comparison of ISSR and RAPD markers to characterize three Chilean *Nothofagus* species. *Theor Appl Genet* 104:1064-1070.

Mayr E (1982) Of what use are subspecies? *The Auk* 99:593-595.

Mitchell N, Haeffner R, Veer V, Fulford-Gardner M, Clerveaux W, Veitch CR, Mitchell G (2002) Cat eradication and the restoration of endangered iguanas (*Cyclura carinata*) on Long Cay, Caicos Bank, Turks and Caicos Islands, British West Indies. In: Veitch CR.,

Clout MN (Eds.), Turning the Tide: the Eradication of Invasive Species. IUCN Gland, Switzerland and Cambridge, UK, pp. 206-212.

Morgan GS (1989) Fossil Chiroptera and Rodentia from the Bahamas, and the historical biogeography of the Bahamian mammal fauna. In: Woods, CA (Ed.), Biogeography of the West Indies: Past, Present, and Future. CRC Press, Boca Raton, Florida, pp 685-740.

Munroe EG (1948) The geographical distribution of butterflies in the West Indies. Ph.D. Dissertation, Cornell University, Ithaca, USA.

National Geospatial Intelligence Agency (2004) Sailing directions (Enroute), vol. 1. Caribbean Sea 8<sup>th</sup> ed. ProStar Publications, Inc. Annapolis, Maryland, USA.

Newsom LA, Wing ES (2004) On land and sea: Native American uses of biological resources in the West Indies. The University of Alabama Press, Tuscaloosa, USA.

Oksanen J, Blanchet FG, Kindt R, Legendre P, O'Hara RB, Simpson GL, Solymos P, Henry M, Stevens H, Wagner H (2010) vegan: community ecology package. R package version 1.17.4.

Pannell JR (2003) Coalescence in a metapopulation with recurrent local extinction and recolonization. *Evolution* 57:949-961.

Pasachnik SA (2010) Conservation genetics and the *Ctenosaura palearis* clade. Ph.D. Dissertation, University of Tennessee, Knoxville, USA.

Pasachnik SA, Echternacht AC, Pineda EA, Corniel J, Montgomery C, Fitzpatrick BM (In Press) Population genetics of *Ctenosaura melanosterna*: implications for conservation and management. *Divers Distrib*.

Posada D (2006) ModelTest Server: a web based tool for the statistical selection of models of nucleotide substitution online. *Nuc Acid Res* 34:W700-W703.

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

R Development Core Team (2010) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* 19:2092-2100.

Reynolds RG (2009) *Leiocephalus psammodromus*. (Turks and Caicos curly-tailed Lizard). Conspecific necrophagy. *Herpetol Rev* 40:88-89.

Reynolds RG (2011) Status, conservation, and introduction of amphibians and reptiles in the Turks and Caicos Islands, British West Indies. In: Hailey, A., Wilson, B.S., Horrocks, J.A. (Eds.), Conservation of Caribbean island herpetofaunas vol. II: Regional Accounts of the West Indies. Brill, Leiden, Netherlands, pp. 377-406.

Reynolds RG, Fitzpatrick BM (2011) *DPR*- a package to implement decomposed pairwise regression in an isolation by distance framework. R package version 1.0.

Reynolds RG, Fitzpatrick BM (In Prep.) Tests of two methods for inference of historical demography using simulated structured populations reveals substantial type II error. Mol Ecol.

Reynolds RG, Gerber GP, Fitzpatrick BM (In Review) Unexpected shallow genetic divergence in Turks Island Boas (*Epicrates c. chrysogaster*) reveals single evolutionary significant unit for conservation. Herpetologica.

Reynolds RG, Niemiller ML (2010) Island invaders: introduced reptiles and amphibians of the Turks and Caicos Islands. Reptil Amphib 17:117-121.

Ricklefs RE, Bermingham E (2008) The West Indies as a laboratory of biogeography and evolution. Phil Trans Roy Soc B 363:2393-2413.

Rieseberg LH (1996) Homology among RAPD fragments in interspecific comparisons. Mol Ecol 5:99-105.

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

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes 4:137-138.

Roughgarden J (1995) *Anolis* lizards of the Caribbean: ecology, evolution, and plate tectonics. Oxford University Press, New York.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145:1219-1228.

Schoener TW, Spiller DA, Losos JB (2001) Natural restoration of the species-area relation for a lizard after a hurricane. Science 294:1525-1528.

Schwartz A (1967) The *Leiocephalus* (Lacertilia, Iguanidae) of the southern Bahama islands. Ann Carnegie Mus 39:153-185.

Sealey NE (2006) Bahamian landscapes, 3<sup>rd</sup> edition. Macmillan-Caribbean, Oxford, UK.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* 236:787-792.

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* 47:264-279.

Smith GR (1994) Observations on activity in two lizards (*Anolis scriptus* and *Leiocephalus psammodromus*). *Bull Maryland Herpetol Soc* 30:120-125.

Smith GR (1995) Observations on perch use in two lizards (*Anolis scriptus* and *Leiocephalus psammodromus*). *Herpetol J* 5:285-286.

Smith GR, Iverson JB (1993) Reproduction in the curly-tailed lizard *Leiocephalus psammodromus* from the Caicos Islands. *Canadian J Zool* 71:2147-2151.

Smith ML, Hedges SB, Buck W, Hemphill A, Inchaustegui S, Ivie MA, Martina DA, Maunder M, Ortega JF (2005) Caribbean islands. In: Mittermeier RA., Gill PR., Hoffman M., Pilgrim J., Brooks T., Mittermeier CG., Lamoreaux J., da Fonseca GAB., (Eds.), *Hotspots Revisited: Earth's Biologically Richest and Most Endangered Terrestrial Ecoregions*. CEMEX, S. A. de C. V. Mexico City, Mexico.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis of DNA polymorphism. *Genetics* 123:585-595.

Thomson RC, Wang IJ, Johnson JR (2010) Genome-enabled development of DNA markers for ecology, evolution, and conservation. *Mol Ecol* 19:2184-2195.

Thorpe RS (2005) Population evolution and island biogeography. *Science* 310:1778-1779.

Tolson PJ, Henderson RW (2006) An overview of snake conservation in the West Indies. *Appl Herpetol* 3:345-356.

Van Ee B, Berry PE, Riina R, Gutierrez-Amaro JE (2008) Molecular phylogenetics and biogeography of the Caribbean-centered *Croton* subgenus *Moacroton* (Euphorbiaceae s.s.). *Bot Rev* 74:132-165.

Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of relationship between degree of homoplasy and fragment size. *Mol Ecol* 11:139-151.

Villesen P (2007) FaBox: an online toolbox for fasta sequences. *Mol Ecol Notes* 7:965-968.

Vos P, Hogers M, Bleeker M, Reijans M, Van De Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA finger printing. *Nuc Acid Res* 23:4407-4414.

Wade MJ, McCauley DE (1988) Extinction and recolonization: their effects on the genetic differentiation of local populations. *Evolution* 42:995-1005.

Wang IJ (2010) Recognizing the temporal distinctions between landscape genetics and phylogeography. *Mol Ecol* 19:2605-2608.

Wares JP (2009) Natural distributions of mitochondrial sequence diversity support new null hypotheses. *Evolution* 64:1136-1142.

Welch ME, Gerber GP, Davis SK (2004) Genetic structure of the Turks and Caicos rock iguana and its implications for species conservation. In: Alberts AC., Carter RL., Hayes WK., Martins EP., (Eds.), *Iguanas, Biology and Conservation*. University of California Press, Berkeley, California, pp. 58-70.

Whitlock MC (2004) Selection and drift in metapopulations. In: Hanski I., Gaggiotti O.E., (Eds.), *Ecology, Genetics, and Evolution of Metapopulations*. Elsevier Academic Press, Burlington, Massachusetts, pp. 153-174.

Whitlock MC, McCauley MJ (1990) Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. *Evolution* 44:1717-1724.

Whitlock R, Hipperson H, Mannarelli M, Butlin RK, Burke T (2008) An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Mol Ecol Res* 8:725-735.

Whittaker RJ, Fernández-Palacios JM (2007) *Island biogeography*. 2<sup>nd</sup> Ed. Oxford University Press, New York, USA.

Wright S (1943) Isolation by distance. *Genetics* 28:114-138.

Wright S (1977) *Evolution and the genetics of populations*, Vol. 3, experimental results and evolutionary deductions. University of Chicago Press, Chicago, Illinois.

Zhivotovsky L (1999) Estimating population structure in diploids with multilocus dominant DNA markers. *Mol Ecol* 8:907-913.

Zink RM (2004) The role of subspecies in obscuring avian biological diversity and misleading conservation policy. *Proc Roy Soc B* 271:561-564.

## **Appendix IIIa. Figures and Tables**

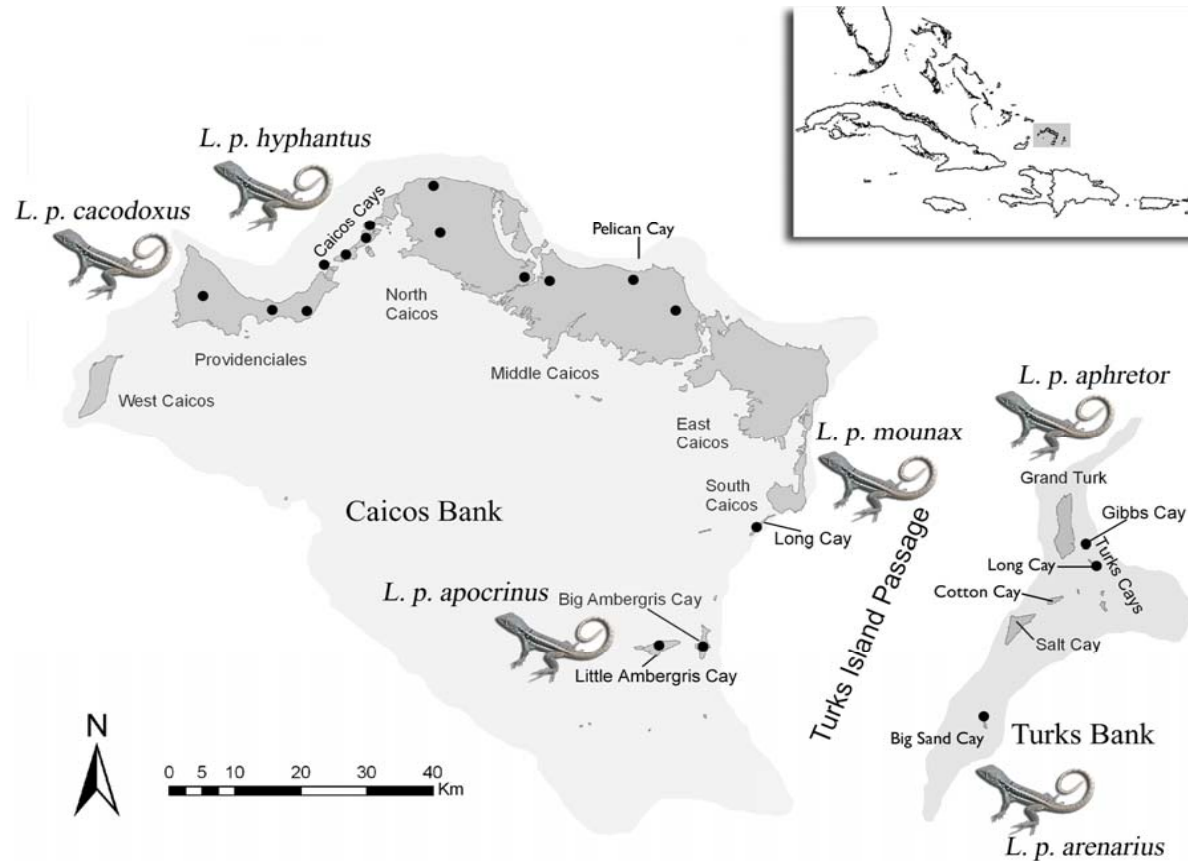


Figure III – 1. Map of the Turks and Caicos Islands, located at the southern terminus of the Bahamian Archipelago roughly 130 km from Hispaniola. Light gray outlines indicate the approximate extent of the Caicos and Turks banks. Islands surveyed for *Leiocephalus psammodromus* are labelled, and dark circles indicate sampling locations. Six previously described subspecies (Schwartz 1967) are shown: *L. p. aphretor* (Grand Turk (extirpated) and Turks Cays), *L. p. apocrinus* (Big and Little Ambergris Cays), *L. p. arenarius* (Big Sand Cay), *L. p. cacodoxus* (Providenciales), *L. p. hyphantus* (Caicos Cays), and *L. p. mounax* (South Caicos (extirpated) and Long Cay).

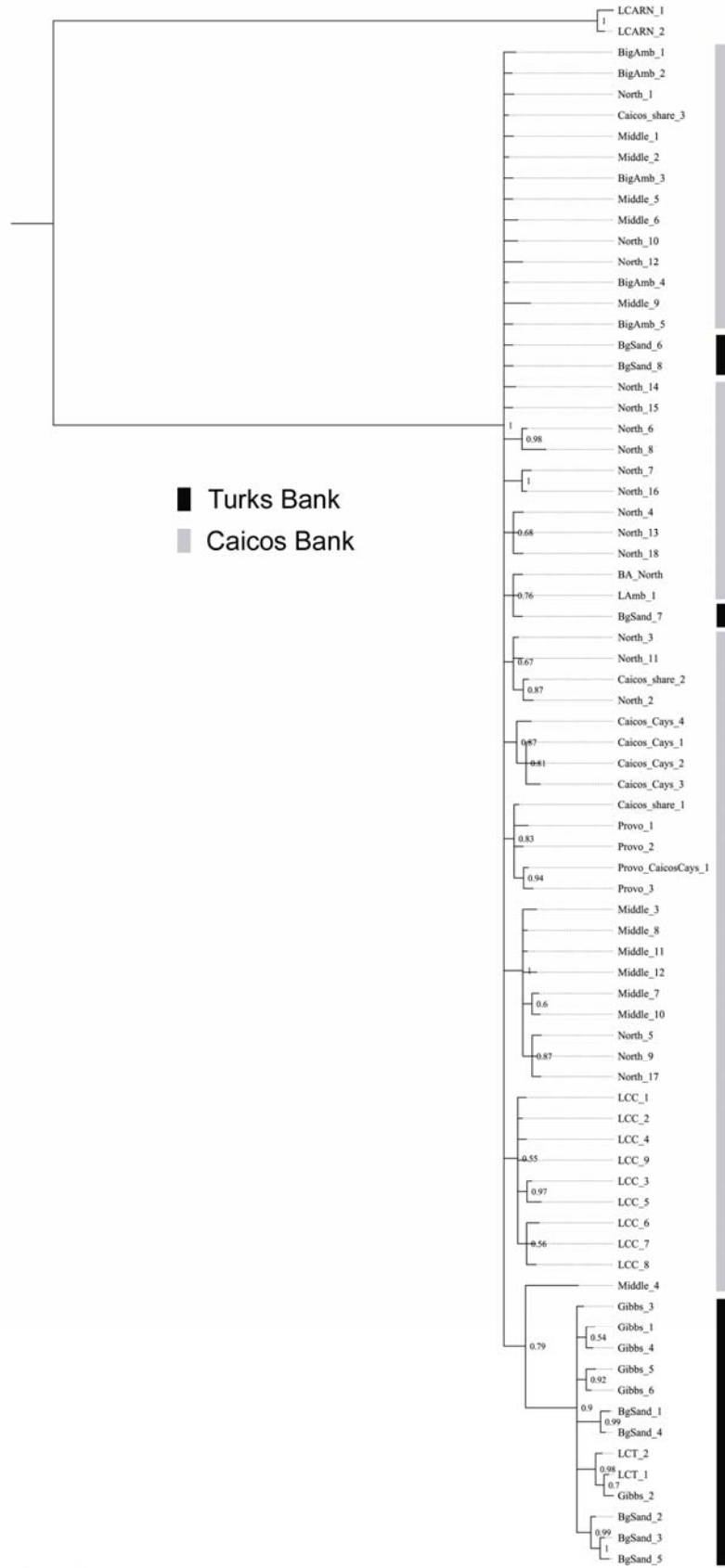


Figure III – 2. 50% Majority-rule consensus Bayesian gene tree for the mitochondrial ND2 locus for 257 *L. psammodromus* samples. Nodal support is given as the posterior probability at each node. *Leiocephalus carinatus* from Boyton Beach, Florida is included as an outgroup. Bank identity is indicated by the gray bars.

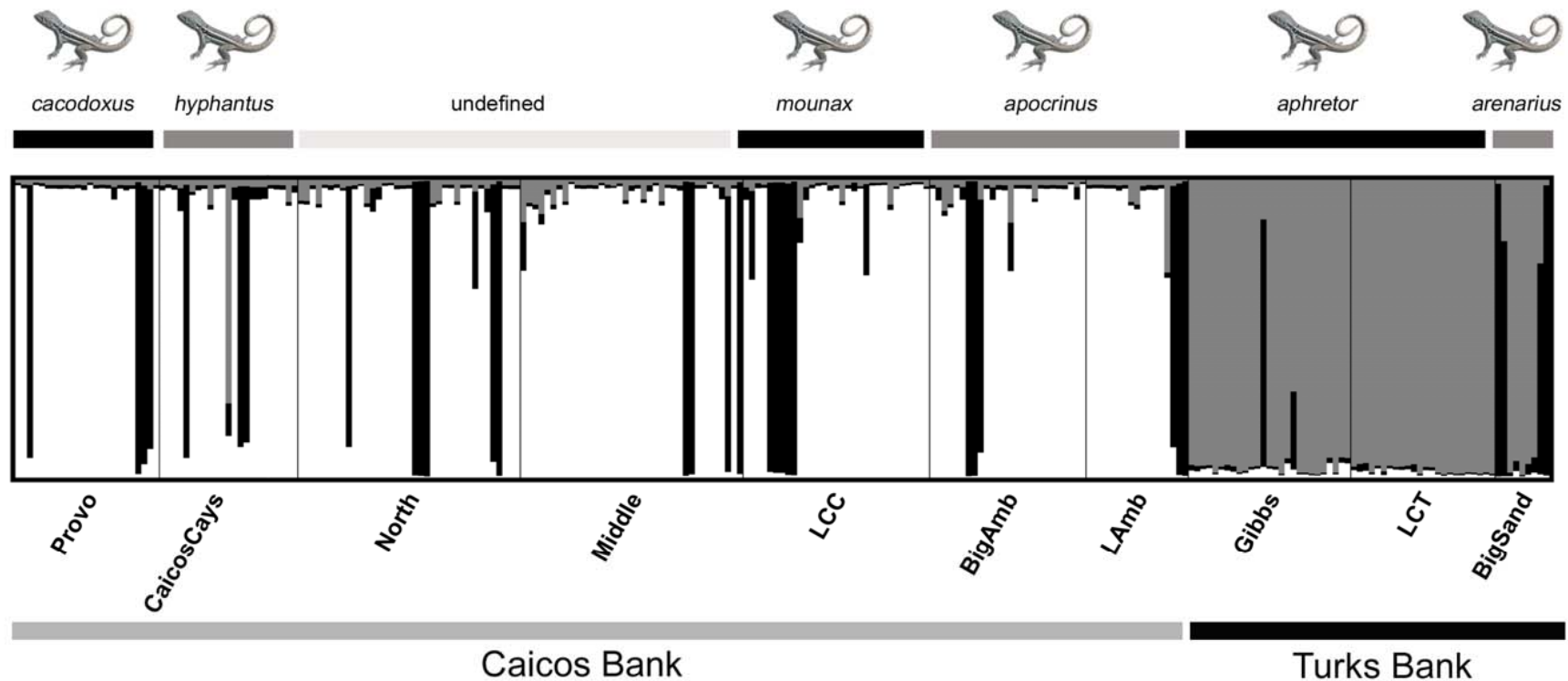


Figure III – 3. Results of STRUCTURE analysis visualized in DISTRUCT. Colours are arbitrarily assigned and represent putative populations. Columns represent individuals grouped by the island from which the sample originated. Scale on the y-axis represents the proportion of an individual's genetic variation assigned to each cluster. Individuals mostly cluster with either the Turks or Caicos Bank, though considerable random admixture between the banks exists. Note that the previously designated subspecies do not correspond to any detected genetic structuring, and at least one individual on the Caicos Bank (Caicos Cays) shares the majority of its genetic variation with the Turks Bank cluster.

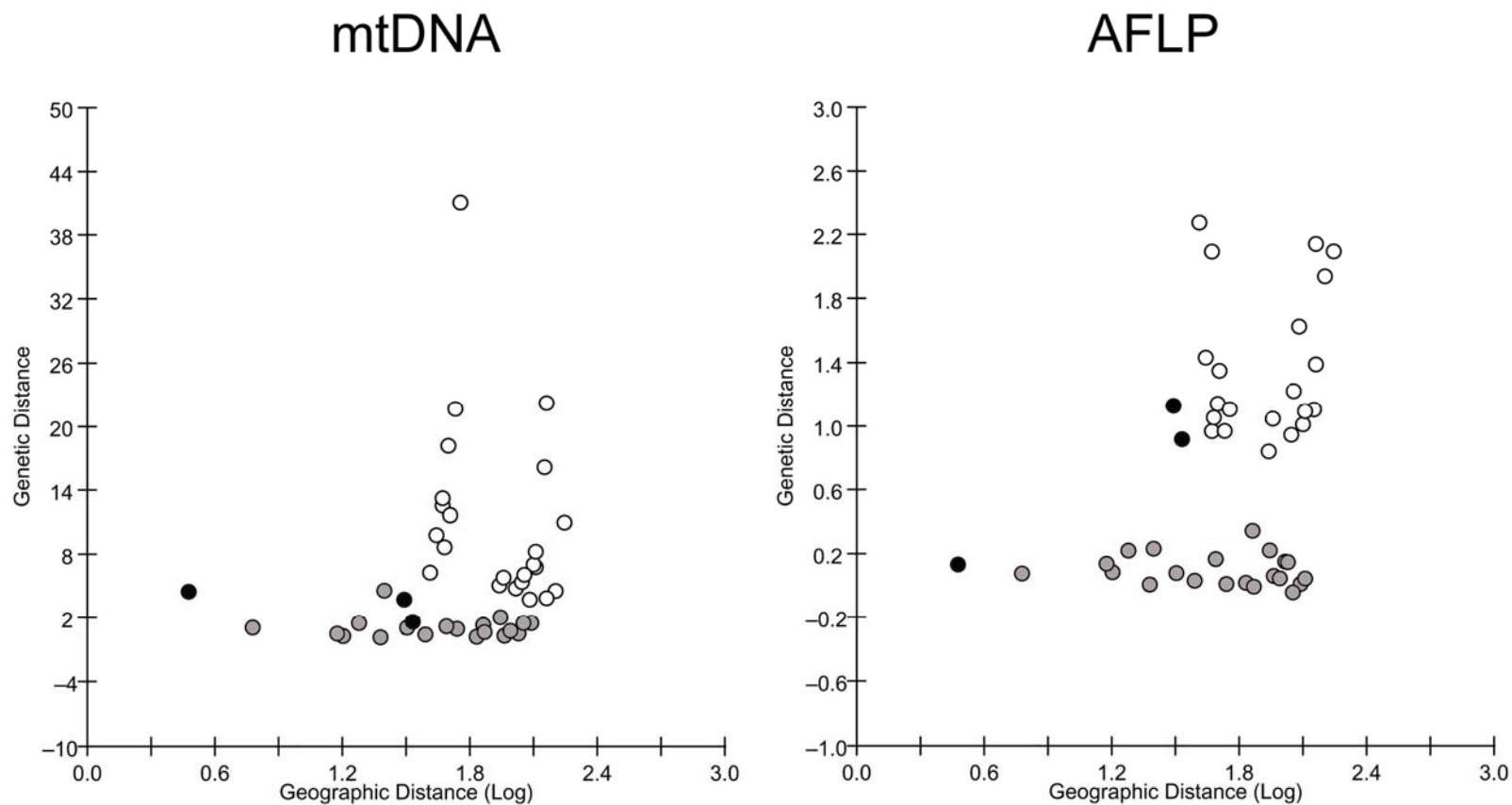


Figure III – 4. Isolation by distance plots of genetic distance ( $\Phi_{ST} / (1 - \Phi_{ST})$ ) versus log-transformed stepping-stone geographic distance for mtDNA and AFLP data. Regressions were constructed by controlling for Bank identity using an indicator matrix. Partial Mantel tests indicated no correlation between genetic and geographic distance for mtDNA ( $r = -0.13$ ,  $P = 0.82$ ) or AFLP ( $r = -0.02$ ,  $P = 0.5$ ) data. Gray circles are pairwise comparisons within the Caicos Bank, black circles are within the Turks Bank, and white circles are between banks.

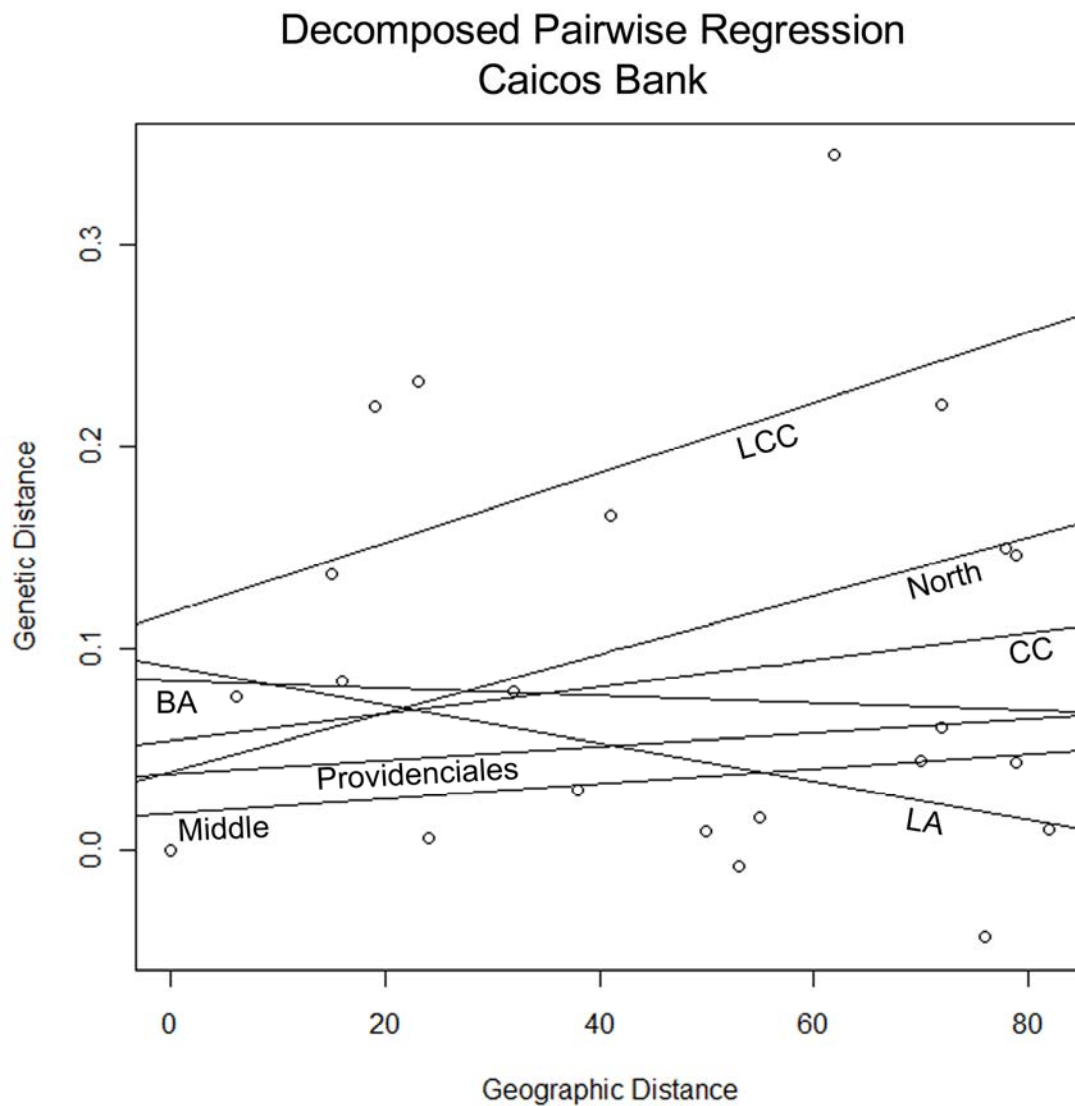


Figure III – 5. Decomposed pairwise regression of genetic distance and geographic distance for islands on the Caicos Bank. Regressions consist of pairwise comparisons for each island (seven data points per regression) with the name of the island under the respective regression line.

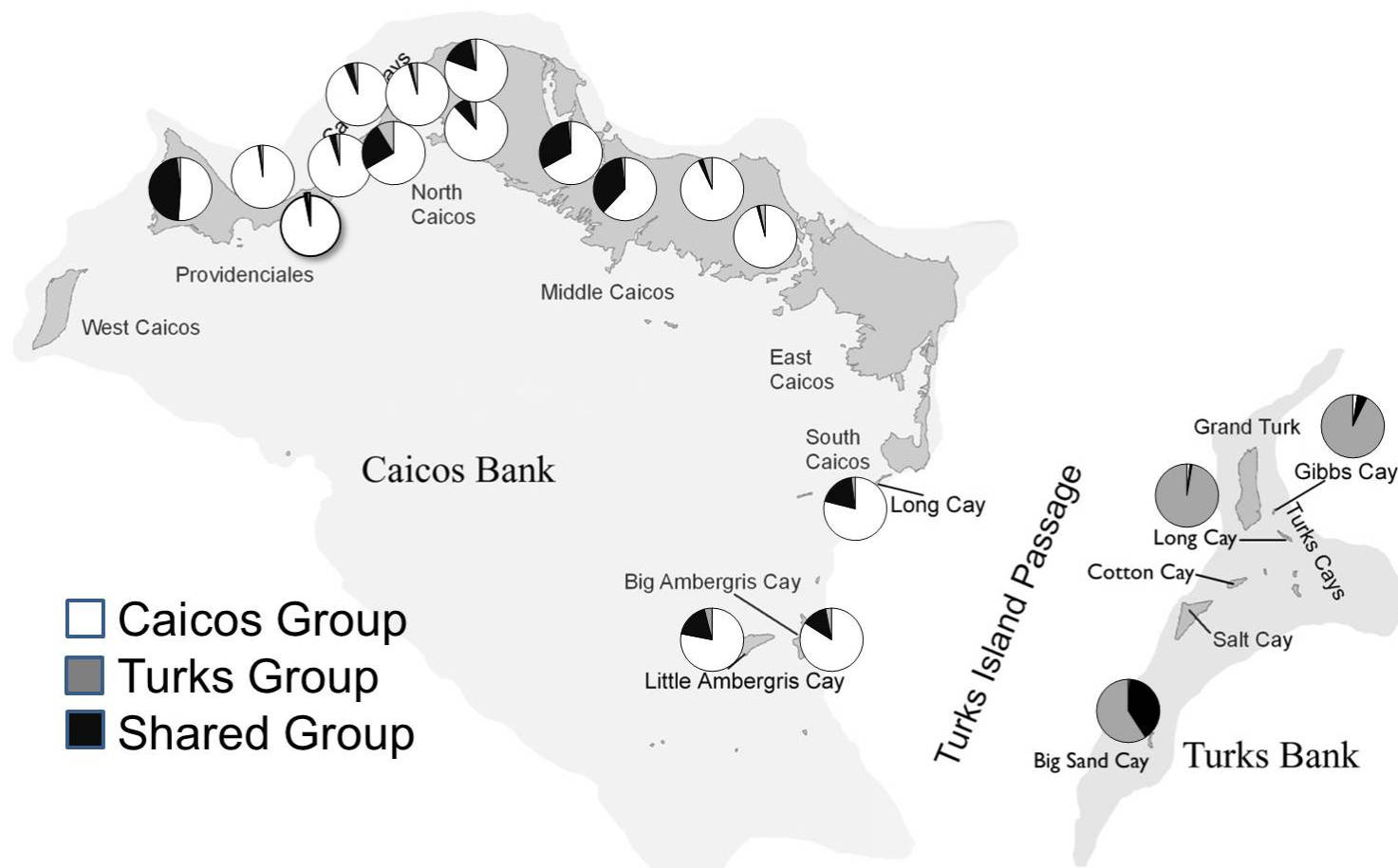


Figure III – 6. Distribution of structure groups by sampling location. Each pie represents the average proportion of the genome reopresented by each of the three groups for individuals from that sampling location.

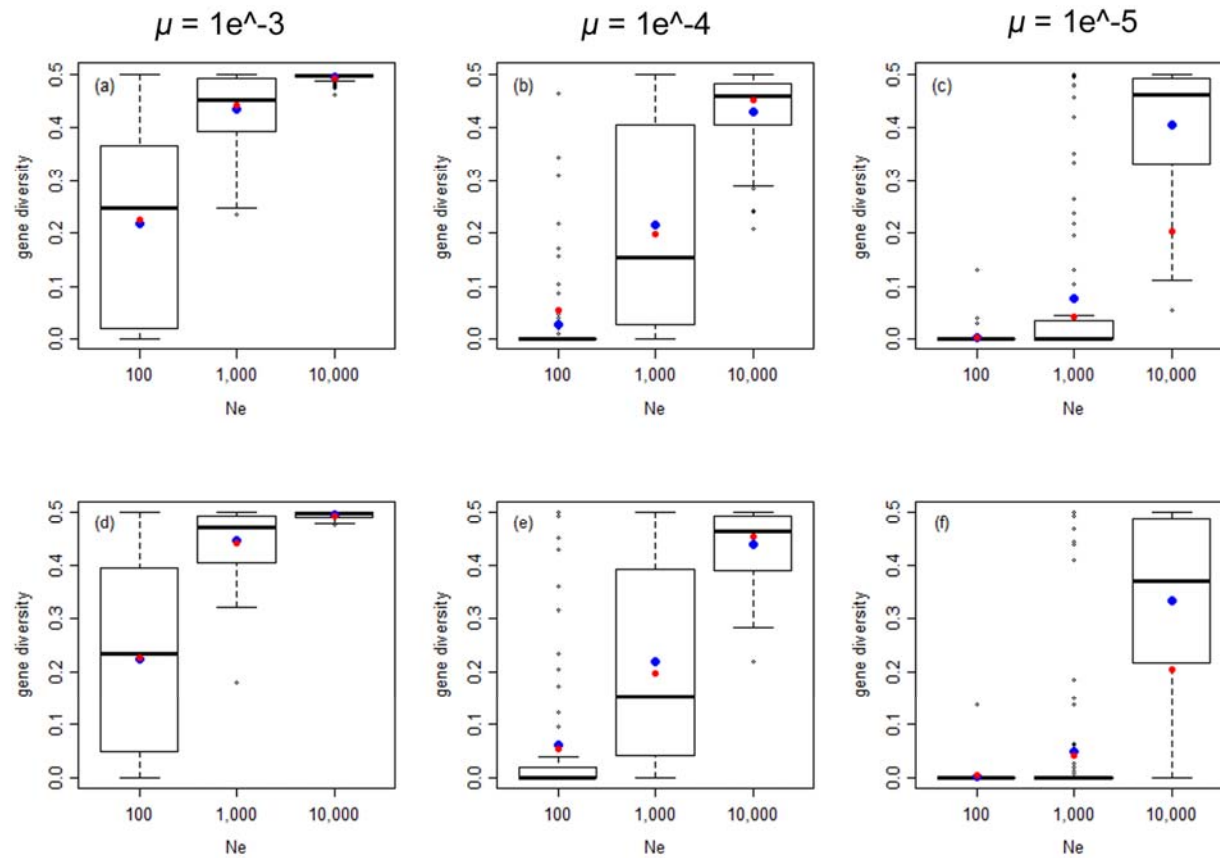


Figure III – S1. Simulated distributions of gene diversities for 70 neutral, diallelic markers for three different effective population sizes after 4,000 generations of isolation. Panels show different parameter combinations: (a) initial gene diversity ( $H_o$ ) = 0.5, mutation rate ( $\mu$ ) =  $1e^{-3}$ , (b)  $H_o$  = 0.25,  $\mu$  =  $1e^{-3}$ , (c)  $H_o$  = 0.5,  $\mu$  =  $1e^{-4}$ , (d)  $H_o$  = 0.25,  $\mu$  =  $1e^{-4}$ . Boxplots show medians (bold horizontal lines), interquartile ranges (the "box"), "whiskers" out to 1.5x the interquartile range, and individual markers more extreme than the whiskers. The large blue circle illustrates the mean of the 70 randomly simulated markers and the red circle illustrates the expected mean at equilibrium assuming a beta distribution of allele frequencies with shape parameters equal to  $4N_e\mu$ .

Table III – 1. Islands surveyed for *L. psammodromus* in the Turks and Caicos and tissue samples (*n*) acquired from each population. See Appendix 1 for additional islands surveyed and found to have no populations of *L. psammodromus*.

| Bank   | Island                         | Abbr. | Date(s) Surveyed               | Approximate area (km <sup>2</sup> ) | Proximity Index | <i>n</i> |
|--------|--------------------------------|-------|--------------------------------|-------------------------------------|-----------------|----------|
| Caicos | Big Ambergris Cay              | BA    | 4-10 December 2007             | 4.45                                | 0.266           | 27       |
|        | Caicos Cays (Ft. George Cay)   | CC    | 16 March 2007                  | 0.28                                | 1.004           | 5        |
|        | Caicos Cays (Pine Cay)         | CC    | 15 March 2007                  | 3.23                                | 1.004           | 2        |
|        | Caicos Cays (Water Cay)        | CC    | 15 March 2007                  | 3.7                                 | 1.004           | 12       |
|        | Caicos Cays (Little Water Cay) | CC    | 11 March 2007                  | 0.6                                 | 1.004           | 4        |
|        | Long Cay                       | LCC   | 17-20 July 2008                | 1.0                                 | 0.147           | 31       |
|        | North Caicos                   | N     | 10-13 August 2008              | 106                                 | 0.365           | 38       |
|        | Middle Caicos                  | M     | 8-11 August 2008               | 124                                 | 0.240           | 37       |
|        | Little Ambergris Cay           | LA    | 6 December 2007; 22 March 2009 | 6.6                                 | 0.211           | 17       |
|        | Providenciales                 | PR    | 11-17 March 2007; 16 July 2008 | 114                                 | 0.186           | 24       |
| Turks  | Big Sand Cay                   | BS    | 26 February 2007               | 0.45                                | 0.042           | 10       |
|        | Gibbs Cay                      | G     | 6, 8 August 2008               | 0.06                                | 0.070           | 28       |
|        | Long Cay (Turks Bank)          | LCT   | 7 August 2008                  | 0.23                                | 0.048           | 24       |

Table III – 2. Genetic diversity at the mitochondrial ND2 and nuclear AFLP loci. Abbreviations are: Number of individuals ( $N$ ), number of haplotypes ( $n$ ), haplotype ( $h$ ) and nucleotide ( $\pi$ ) diversity mean ( $\pm$  SD), number of genotypes ( $g$ ), and gene diversity ( $H_e$ ) test estimates from ten populations of curly-tailed lizards in the Turks and Caicos.  $P$  represents 1-tailed  $P$ -values for the unique genotypes test, representing the fraction of replications where random draws of genotypes from a population were less than the observed number of genotypes ( $g$ ).

| Island | ND2 |     |                 |                   | AFLP |     |       |        |
|--------|-----|-----|-----------------|-------------------|------|-----|-------|--------|
|        | $N$ | $n$ | $h$             | $\pi$             | $N$  | $g$ | $H_e$ | $P$    |
| PR     | 24  | 5   | $0.49 \pm 0.12$ | $0.001 \pm 0.02$  | 24   | 12  | 0.080 | 0.23   |
| CC     | 22  | 6   | $0.74 \pm 0.07$ | $0.004 \pm 0.04$  | 23   | 14  | 0.085 | 0.46   |
| M      | 34  | 14  | $0.84 \pm 0.04$ | $0.005 \pm 0.04$  | 37   | 23  | 0.092 | 0.19   |
| N      | 38  | 20  | $0.92 \pm 0.03$ | $0.005 \pm 0.03$  | 37   | 17  | 0.098 | 0.008* |
| LCC    | 32  | 9   | $0.79 \pm 0.04$ | $0.002 \pm 0.02$  | 31   | 22  | 0.093 | 0.14   |
| BA     | 27  | 11  | $0.81 \pm 0.05$ | $0.25 \pm 0.00$   | 26   | 13  | 0.096 | 0.03*  |
| LA     | 15  | 2   | $0.25 \pm 0.13$ | $0.01 \pm 0.22$   | 17   | 10  | 0.085 | 0.04*  |
| G      | 28  | 5   | $0.65 \pm 0.06$ | $0.0007 \pm 0.18$ | 27   | 14  | 0.083 | 0.03*  |
| LCT    | 24  | 3   | $0.54 \pm 0.06$ | $0.004 \pm 0.35$  | 24   | 14  | 0.079 | 0.46   |
| BS     | 10  | 5   | $0.67 \pm 0.16$ | $0.03 \pm 0.03$   | 9    | 9   | 0.113 | 0.32   |

\*significant at  $P < 0.05$

Table III – 3.  $\Phi_{ST}$  estimates for pairwise comparisons among ten island populations of *Leiocephalus psammodromus* based on 10,100 permutations.  $\Phi_{ST}$  values for mtDNA are given below the diagonal, while  $\Phi_{ST}$  values for AFLP loci, calculated using Jaccard distances, are given above the diagonal.

| Island | PR    | CC    | M     | N     | LCC   | BA    | LA    | G     | LCT   | BS    |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| PR     | —     | 0.08* | 0.07* | 0.01  | 0.13* | 0.01  | 0.04  | 0.52* | 0.58* | 0.68* |
| CC     | 0.20* | —     | 0.12* | 0.03  | 0.18* | 0.13* | -0.05 | 0.50* | 0.52* | 0.66* |
| M      | 0.51* | 0.32* | —     | 0.01  | 0.26* | 0.06* | 0.04  | 0.49* | 0.55* | 0.68* |
| N      | 0.48* | 0.29* | 0.11* | —     | 0.14* | 0.02  | -0.01 | 0.46* | 0.51* | 0.62* |
| LCC    | 0.82* | 0.67* | 0.56* | 0.54* | —     | 0.18* | 0.19* | 0.49* | 0.53* | 0.59* |
| BA     | 0.60* | 0.33* | 0.22* | 0.16* | 0.60* | —     | 0.07  | 0.51* | 0.57* | 0.69* |
| LA     | 0.87* | 0.60* | 0.42* | 0.39* | 0.82* | 0.51* | —     | 0.49* | 0.53* | 0.68* |
| G      | 0.92* | 0.88* | 0.84* | 0.83* | 0.93* | 0.90* | 0.96* | —     | 0.12* | 0.48* |
| LCT    | 0.96* | 0.89* | 0.86* | 0.85* | 0.95* | 0.92* | 0.98* | 0.82* | —     | 0.53* |
| BS     | 0.94* | 0.82* | 0.79* | 0.79* | 0.91* | 0.86* | 0.93* | 0.61* | 0.79* | —     |

\*significant at  $P < 0.05$

Table III – 4. Results of analysis of molecular variance (AMOVA) for various groupings of *Leiocephalus psammodromus* for both AFLP and mtDNA data.

| Marker       | Source of Variation            | d.f. | Variance component | % total variance | $\Phi$ statistics  | $P$       |
|--------------|--------------------------------|------|--------------------|------------------|--------------------|-----------|
| <i>AFLPs</i> | Between Banks                  | 1    | 0.89               | 46.34            | $\Phi_{CT} = 0.46$ | $< 0.001$ |
|              | Among populations within banks | 8    | 0.12               | 6.32             | $\Phi_{SC} = 0.12$ | $< 0.001$ |
|              | Within populations             | 179  | 0.90               | 47.33            | $\Phi_{ST} = 0.52$ | $< 0.001$ |
| <i>mtDNA</i> | Between Banks                  | 1    | 6.55               | 72.5             | $\Phi_{CT} = 0.72$ | $< 0.001$ |
|              | Among populations within banks | 8    | 1.33               | 14.8             | $\Phi_{SC} = 0.53$ | $< 0.001$ |
|              | Within populations             | 244  | 1.15               | 12.7             | $\Phi_{ST} = 0.87$ | 0.009     |

Table III – S1. Probable extirpations and islands that lack populations of *L. psammodromus*.

| Bank   | Island                     | Dates surveyed  | Island Area (km <sup>2</sup> ) | Status      | Reference                           |
|--------|----------------------------|-----------------|--------------------------------|-------------|-------------------------------------|
| Caicos | Mangrove Cay (Caicos Cays) | 16 March 2007   | 0.2                            | Not present | Maclean et al. 1977; This study     |
|        | South Caicos               | 17-18 July 2008 | 21.2                           | Extirpated  | Maclean et al. 1977; Reynolds 2011  |
|        | Middleton Cay              | —               | 0.05                           | Not present | Reynolds 2011                       |
|        | Six Hills Cay              | —               | 0.09                           | Not present | Maclean et al. 1977                 |
| Turks  | Grand Turk                 | 7-9 August 2008 | 18                             | Extirpated  | Newsom and Wing 2004; Reynolds 2011 |
|        | Cotton Cay                 | 8 August 2008   | 1.12                           | Extirpated  | Maclean et al. 1977; This study     |
|        | Salt Cay                   | 4-5 August 2008 | 6.7                            | Extirpated  | Maclean et al. 1977; This study     |

### **Appendix IIIb. Genbank accession numbers.**

Appendix IIIb. Haplotypes of *Leiocephalus psammodromus* for ND2. Genbank accession numbers are JF812182 through 254.

| Haplotype(s)      | Island(s)                                      |
|-------------------|------------------------------------------------|
| LCARN1            | Boyton Beach, FL                               |
| LCARN 2           | Boyton Beach, FL                               |
| Provo_CaicosCays1 | Providenciales, Water Cay, Little Water Cay    |
| Provo 1-3         | Providenciales                                 |
| Caicos_Share 1-2  | Providenciales, Water Cay, North Caicos        |
| Caicos_Share 3    | North Caicos, Middle Caicos, Big Ambergris Cay |
| Caicos_Cays 1     | Ft. George Cay, Pine Cay                       |
| Caicos_Cays 2-3   | Ft. George Cay                                 |
| Caicos_Cays 4     | Pine Cay                                       |
| BA_North          | Big Ambergris Cay, North Caicos                |
| BigAmb 1-5        | Big Ambergris Cay                              |
| Lamb 1            | Little Ambergris Cay                           |
| North 1-18        | North Caicos                                   |
| Middle 1-12       | Middle Caicos                                  |
| LCC 1-9           | Long Cay (Caicos Bank)                         |
| Gibbs 1-6         | Gibbs Cay                                      |
| LCT 1-2           | Long Cay (Turks Bank)                          |
| BgSand 1-8        | Big Sand Cay                                   |

## GENERAL CONCLUSIONS

### Preface

This dissertation explores the use of various methods for evaluating population dynamics in structured populations using genetic data that can be readily gathered at a relatively low cost. Chapter I is a test of what we identified to be promising methods for detecting population dynamics in metapopulations, however these methods suffered from high type II error. Given this important conclusion, I then attempted to use genetic data gathered from structured island populations (metapopulations *sensu lato*) to determine the extent of population dynamics that can be recovered using standard molecular markers. Chapter II is explicitly intended to diagnose genetic variation and historical demographic processes in a threatened boa, and the results are being applied directly in conservation planning for this species. In chapter III, I evaluate a comprehensive approach for estimating the contributions of stochastic demographic forces to genetic diversity in structured populations. I term this approach “archipelagic genetics,” and suggest that though limited, this approach of testing for non-equilibrium dynamics is preferable to using only one or two approaches in island archipelagos. Overall I show how genetic data can be used to understand population processes influencing genetic variation on structured island populations in the context of conservation.

## **PART I. TESTS OF TWO METHODS FOR INFERENCE OF HISTORICAL DEMOGRAPHY USING SIMULATED STRUCTURED POPULATIONS REVEALS SUBSTANTIAL TYPE II ERROR**

- The objective of this research was to test the performance of two strategies for distinguishing between structured demes that are in mutation-migration-drift equilibrium from those that are not.
- I used simulations of two metapopulation models- a focal deme model (FDM) and a metapopulation model (MM) to generate microsatellite data that was used in the methods of allele frequency analysis and decomposed pairwise regression to identify non-equilibrium demes in a metapopulation.
- Decomposed pairwise regression consisted of analysis of focal demes in an isolation by distance framework, where the slope and intercept of each focal deme were ranked relative to other demes in the simulation. Outlier demes are expected to have higher or lower slopes and intercepts relative to equilibrium demes. This

is followed by detection of outliers using *ad hoc* AIC model selection to identify demes that are not in MDE.

- Allele frequency analysis consisted of tests for heterozygosity excess relative to allele frequencies implemented in the program BOTTLENECK. P-values from these tests were obtained for 2400 simulations of the MM for each of 12 different parameter sets. Combined probability analysis (CPA) was also used to test for increase power of detection of non-equilibrium in simulated demes.
- Decomposed pairwise regression failed to identify outliers consistently in an isolation-by-distance framework.
- *Ad hoc* AIC analysis to detect outliers resulted in type II error rates of 59 – 100% across parameter sets.
- The deme-by-deme strategy of testing for bottlenecks using the program bottleneck demonstrated that population recovery can rapidly eliminate the statistical signature of a bottleneck and that moderate levels of gene flow can lead to the rejection of MDE in the opposite direction.
- When population recovery or gene flow are occurring in the deme of interest, the sensitivity of these methods may lead to failure to reject the null hypothesis of MDE in non-equilibrium demes.
- These results call into question potentially hundreds of unpublished studies that have failed to detect non-equilibrium, a finding with significant implications for conservation genetics, as the results of these tests are often used to guide conservation policy.

## **PART II. UNEXPECTED SHALLOW GENETIC DIVERGENCE IN TURKS ISLAND BOAS (*EPICRATES C. CHRYSOGASTER*) REVEALS SINGLE EVOLUTIONARILY SIGNIFICANT UNIT FOR CONSERVATION**

- The objective of this research was to evaluate genetic diversity in the boid *Epicrates c. chrysogaster*, an endemic subspecies, across the Turks and Caicos Islands.
- This species is a serious conservation concern, and the relationships of various island populations, including the newly discovered Turks Bank population, were unknown.

- I collected samples from 53 individuals from five island populations during 361 hours of field work
- I sequenced a region of the mitochondrial gene cytochrome B, and well as a region of the nuclear genes NT3 and c-mos, resulting in over 1,500 bp of sequence data per individual.
- I screened individuals at nine newly-isolated microsatellite markers from the species *Epicrates subflavus*.
- Overall sequence divergence was  $< 1\%$ , and only nine haplotypes were found.
- Significant ascertainment bias existed with the microsatellite markers, as only four amplified PCR product and all four were found to be monomorphic for all individuals.
- These data suggest shallow genetic divergence in this species, suggesting little genetic structure prior to vicariance post-glaciation and possible non-equilibrium dynamics.
- I suggest that *Epicrates chrysogaster* represents a single evolutionarily significant unit for conservation, and that programs such as reintroduction or relocation to protected islands might assist in the conservation of this species as it faces heavy development of and the introduction of feral predators to many of the remaining island populations.

### **PART III. MOLECULAR ISLAND BIOGEOGRAPHY: EQUILIBRIUM VERSUS NON-EQUILIBRIUM GENETIC STRUCTURE IN AN ARCHIPELAGIC LIZARD**

- The objective of this research was to evaluate genetic diversity in an archipelagic endemic lizard in order to combine population genetics and island biogeography explicitly as a more inclusive approach to investigating genetic diversity in island archipelagos.
- I collected samples from 259 *Leiocephalus psammodromus* from 13 islands in the Turks and Caicos archipelago, sequenced 796 bp of the mitochondrial gene ND2,

and obtained anonymous nuclear markers (AFLPs) using two primer combinations.

- I used a combination of phylogeographic, landscape genetic, and island biogeography analyses to evaluate equilibrium versus non-equilibrium dynamics in this species.
- Phylogenetic analyses, including haplotype diversity, neutrality tests, and AMOVA analyses indicated that both the Turks Bank and the Caicos Bank contained unique mitochondrial haplogroups and that some islands on the Caicos Bank exhibited signatures of recent demographic expansion, implying the presence of non-equilibrium dynamics.
- Landscape genetic analyses, including clustering algorithms and permutation tests indicated that three groups existed: a Caicos group, a northern Turks Bank group, and a southern Turks Bank group that was shared across the Caicos Bank. I infer that intermittent gene flow is occurring across the Turks Island passage, a permanent marine barrier.
- Island biogeographic analyses, including decomposed pairwise regression, isolation by distance, and island gene diversity relative to island area and isolation revealed substantial evidence for ongoing gene flow across each bank.
- I found no evidence for isolation by distance or relationships between genetic diversity and axes of island variation, indicating departure from mutation-migration-drift equilibrium in the archipelago.
- This species was thought to comprise at least six subspecies distributed across the Turks and Caicos, an hypothesis not supported by genetic data. I suggest that this species consists of a single taxonomic unit.
- I propose that studies of island archipelagos incorporate the methods of phylogeography, landscape genetics, and island biogeography to inclusively evaluate empirical genetic data in the context of equilibrium and non-equilibrium dynamics.

## APPENDIX IV. R SCRIPTS

### Index:

#### Simulations

|                                                                             |            |
|-----------------------------------------------------------------------------|------------|
| <b>Step 1) Function File .....</b>                                          | <b>159</b> |
| <b>Step 2) Population Model Files .....</b>                                 | <b>162</b> |
| I. Focal Deme Model- Bottleneck .....                                       | 162        |
| II. Focal Deme Model- Extinction - Recolonization .....                     | 162        |
| III. Focal Deme Model with Varying $N_e$ .....                              | 163        |
| IV. Metapopulation Model- Equilibrium .....                                 | 163        |
| V. Metapopulation Model- Bottleneck .....                                   | 164        |
| VI. Metapopulation Model- Extinction - Recolonization .....                 | 165        |
| VII. Cornuet and Luikart Model .....                                        | 166        |
| <b>Step 3) Migration Matrices .....</b>                                     | <b>166</b> |
| <b>Step 4) Simulation File- input parameters and # of simulations .....</b> | <b>167</b> |
| I. Focal Deme and Metapopulation Models .....                               | 167        |
| II. Focal Deme Model with varying $N_e$ .....                               | 167        |
| III. Cornuet and Luikart Model .....                                        | 168        |

#### Analyses

|                                                                                                                |            |
|----------------------------------------------------------------------------------------------------------------|------------|
| <b>Decomposed Pairwise Regression Putative Outlier using the method of Koiuzumi <i>et al.</i> (2006).....</b>  | <b>169</b> |
| <b>Decomposed Pairwise Regression using the <i>ad hoc</i> AIC method of Koiuzumi <i>et al.</i> (2006).....</b> | <b>173</b> |
| <b>Unique Genotypes Test.....</b>                                                                              | <b>175</b> |
| <b>Simulation of Expected Gene Diversity at Equilibrium.....</b>                                               | <b>176</b> |
| <b>Gene Diversity for AFLP Data .....</b>                                                                      | <b>181</b> |
| <b>References.....</b>                                                                                         | <b>183</b> |

## METAPOPULATION SIMULATIONS

### Notes:

Notations are in [UTK Orange](#)

R code is in [black](#)

Functions are in [light blue](#)

### Step 1) Function File

#### R file for simulation of population data

This file sets up the functions to simulate population models from Chapter I

Functions: [LG](#); [SMM](#); [pop](#); [smd](#); [migrate](#); [Fst](#), [MMD](#)

Step 1) Execute this file

Step 2) Add a Population Model File: either MM (Bottleneck or Extinction and Recolonization) or FDM (Bottleneck or Extinction/Recolonization)

Step 3) Add and adjust the migration matrix

Step 4) Add Simulation File- input parameters and # of simulations

Execute all of the above

Notations are in [UTK Orange](#)

R code is in [black](#)

Functions are in [light blue](#)

[# LG is the logistic growth function](#)

```
LG=function(N,r,K,T){  
  plot(0,0,type="n",xlim=c(0,T),ylim=c(0,2*K),xlab="generation",ylab="N")  
  for(i in 1:T){  
    N=N+r*N*(K-N)  
    points(i,N)  
    print(N)  
  }  
}
```

[#Function SMM simulates the stepwise mutation model for SSR data](#)

```
SMM=function(n,N,theta,T,P){  
  mu=theta/(4*N)  
  for(i in 1:T){  
    mutations=sample(c(-  
1,1),size=2*N,replace=TRUE)*sample(c(0,1),size=2*N,replace=TRUE,prob=c(1-mu,mu))  
    P=sample(P,2*N,TRUE)  
    P=P+mutations  
  }  
  cbind(sample(P,n),sample(P,n))  
}
```

[# make genepop file](#)

```
pop=function(L=20,n=20,N=1000,theta=4,T=10000,P=rep(50,2000)){  
  X=SMM(n,N,theta,T,P)  
  for(i in 2:L){  
    X=cbind(X,SMM(n,N,theta,T,P))  
  }  
}
```

```

X
}

#Function smd simulates a stepping stone metapopulation
smd=function(N,mu,P){
  #one generation of stepwise mutation (no drift)
  mutations=sample(c(-
1,1),size=2*N,replace=TRUE)*sample(c(0,1),size=2*N,replace=TRUE,prob=c(1-mu,mu))
  P+mutations
}

#Function migrate is for one generation of migration among pops according to migration matrix m
migrate=function(pops,N,m){
  #pops is a matrix of k populations and L allele frequencies
  #N causes each population's contribution of migrants to be proportional to its size
  #npops is then the expected number of gametes with each allele in each population
  k=nrow(pops)
  npops=pops
  for(i in 1:k){
    npops[i,]=t(m[i,]*N)%*%pops
  }
  npops
}

# Function Fst calculates pairwise Fst with no bias correction
Fst=function(x){
  # assume pops is a matrix with population membership in the first column and two columns of
  alleles
  tab=table(pop=rep(x[,1],2),allele=c(x[,2],x[,3]))
  k=nrow(tab)
  n=rowSums(tab)
  tab=tab/n
  f=matrix(NA,k,k)
  for(i in 1:k){
    for(j in 1:k){
      ptab=tab[c(i,j),]
      Ht=1-sum(colMeans(ptab)^2)
      Hs=1-sum(colMeans(ptab^2))
      f[i,j]=(Ht-Hs)/Ht
    }
  }
  f
}

#Function MMD simulates T generations of mutation, migration, and drift in a set of populations
#this version works for exactly 12 populations
#pops is a table of allele frequencies
#N is a vector of initial population sizes
#r is the intrinsic growth rate
#K is a vector of carrying capacities
MMD=function(pops,N,m,mu,T,r,K,n){
  k=nrow(pops)
  alleles=colnames(pops)
  #plot(rep(0,k),N,xlim=c(0,T),ylim=c(0,2*max(K)),type="n",xlab="generation",ylab="N")

```

```

for(i in 1:T){
  #migration
  popsm=migrate(pops,N,m)
  N=rowSums(popsm)
  #drift+mutation
  N=N+r*N*(K-N)
  N=round(N)
  N=replace(N,N<1,1)
  #points(rep(i,k),N)

  P=list(rep(NA,2*N[1]),rep(NA,2*N[2]),rep(NA,2*N[3]),rep(NA,2*N[4]),rep(NA,2*N[5]),rep(NA,2*N[6]),rep(NA,2*N[7]),rep(NA,2*N[8]),rep(NA,2*N[9]),rep(NA,2*N[10]),rep(NA,2*N[11]),rep(NA,2*N[12]))

  pnames=list(rep(1,2*N[1]),rep(2,2*N[2]),rep(3,2*N[3]),rep(4,2*N[4]),rep(5,2*N[5]),rep(6,2*N[6]),rep(7,2*N[7]),rep(8,2*N[8]),rep(9,2*N[9]),rep(10,2*N[10]),rep(11,2*N[11]),rep(12,2*N[12]))
  for(j in 1:k){

    P[[j]]=smd(N[j],mu,as.numeric(sample(alleles,size=2*N[j],replace=TRUE,prob=popsm[j,])))
    }
    alleles=names(table(unlist(P)))
    pops=table(unlist(pnames),unlist(P))/(2*N)
    }
  n=rep(n,k)
  for(g in 1:12){if(n[g]>N[g]) n[g]=N[g]}
  p=list(rep(NA,2*n[1]),rep(NA,2*n[2]),rep(NA,2*n[3]),rep(NA,2*n[4]),rep(NA,2*n[5]),rep(NA,2*n[6]),rep(NA,2*n[7]),rep(NA,2*n[8]),rep(NA,2*n[9]),rep(NA,2*n[10]),rep(NA,2*n[11]),rep(NA,2*n[12]))
  pnames=list(rep(1,n[1]),rep(2,n[2]),rep(3,n[3]),rep(4,n[4]),rep(5,n[5]),rep(6,n[6]),rep(7,n[7]),rep(8,n[8]),rep(9,n[9]),rep(10,n[10]),rep(11,n[11]),rep(12,n[12]))
  for(j in 1:k){
    p[[j]]=sample(alleles,size=2*n[j],replace=TRUE,prob=pops[j,])
    }
  sam=matrix(as.numeric(unlist(p)),ncol=2,byrow=T)
  f=Fst(cbind(unlist(pnames),sam))
  list(pops=pops, sam=sam, Fst=f, N=N,pnames=unlist(pnames))
}

```

#Step 2) Add a Population Model File: either MM (Bottleneck, Extinction/Recolonization, Equilibrium), or FDM (Bottleneck, Extinction/Recolonization, Varying Ne).

#Step 3) Add and adjust the Migration Matrix File or use the no migration matrix

#Step 4) Add Simulation File- input parameters and # of simulations

#Execute all of the above

## Step 2) Population Model Files

**Focal Deme Model- Bottleneck**

**Focal Deme Model- Extinction - Recolonization**

**Focal Deme Model- Varying  $N_e$**

**Metapopulation Model (MM)- Equilibrium**

**Metapopulation Model (MM)- Bottleneck**

**Metapopulation Model (MM)- Extinction/Recolonization**

### **I. Focal Deme Model- Bottleneck**

*# function for focal deme model with bottleneck*

```
FDMbot=function(r,K,L,m,mu,time,n,b){
  pops=table(1:12,rep(50,12))
  N1=rep(K,12)
  Exdat=as.vector(matrix(rep(c("A","B","C","D","E","F","G","H","I","J","K","L"),n),ncol=12,byrow=TRUE))
  pnames=Exdat
  f=matrix(0,ncol=12,nrow=12)
  for(i in 1:L){
    # step 1: simulate for 10N generations to achieve equilibrium
    bpops=MMD(pops=pops,N=N1,m=m,mu=mu,T=10*K,r=r,K=K,n=n)$pops
    # step 2: pop1 drops to b individuals
    N2=c(b,rep(K,11))
    #step 3: simulate more generations
    npops=MMD(pops=bpops,N=N2,m=m,mu=mu,T=time,r=r,K=K,n=n)$sam
    Exdat=cbind(Exdat,npops)
    f=f+Fst(cbind(pnames,npops))
  }
  list(data=Exdat,Fst=f/L)
}
```

### **II. Focal Deme Model- Extinction - Recolonization**

*# function for focal deme model with extinction - recolonization*

```
FDMer=function(r,K,L,m,mu,time,n){
  pops=table(1:12,rep(50,12))
  N1=rep(K,12)
  Exdat=as.vector(matrix(rep(c("A","B","C","D","E","F","G","H","I","J","K","L"),n),ncol=12,byrow=TRUE))
  pnames=Exdat
  f=matrix(0,ncol=12,nrow=12)
  for(i in 1:L){
    # step 1: simulate for 10N generations to achieve equilibrium
    eqpops=MMD(pops=pops,N=N1,m=m,mu=mu,T=10*K,r=r,K=K,n=n)$pops
    # step 2: pop1 goes extinct
    bpops=eqpops
    bpops[1,]=0
    N2=c(1,rep(K,11))
    #step 3: simulate more generations
    npops=MMD(pops=bpops,N=N2,m=m,mu=mu,T=time,r=r,K=K,n=n)$sam
    Exdat=cbind(Exdat,npops)
  }
}
```

```

        f=f+Fst(cbind(pnames,npop))
    }
    list(data=Exdat,Fst=f/L)
}

```

### III. Focal Deme Model with Varying $N_e$

# function for focal deme model with different  $N_e$  for focal deme

```

FPMN=function(r,K=Ne,N=Nb,L,m,mu,time,n,b=Nb){
  pops=table(1:12,rep(50,12))
  N1=c(N,rep(K,11))
  Exdat=as.vector(matrix(rep(c("A","B","C","D","E","F","G","H","I","J","K","L"),n),ncol=12,byrow=TRUE))
  pnames=Exdat
  f=matrix(0,ncol=12,nrow=12)
  for(i in 1:L){
    # step 1: simulate for 10N generations to achieve equilibrium
    bpop=MMDc(pops=pops,N=N1,m=m,mu=mu,T=10*K,n=n)$pops
    # step 2: pop1 (Focal Deme) drops to b individuals
    N2=c(b,rep(K,11))
    #step 3: simulate more generations
    npop=MMDc(pops=bpop,N=N2,m=m,mu=mu,T=time,n=n)$sam
    Exdat=cbind(Exdat,npop)
    f=f+Fst(cbind(pnames,npop))
  }
  list(data=Exdat,Fst=f/L)
}

```

### IV. Metapopulation Model- Equilibrium

#function for Metapopulation Model with constant population size (no demographic event)

```

MMeq=function(pops,N,m,mu,T,n){
  #simulate T generations of mutation, migration, and drift in a set of populations
  #this version works for exactly 12 populations
  #pops is a table of allele frequencies
  #N is a vector of initial population sizes
  k=nrow(pops)
  alleles=colnames(pops)
  for(i in 1:T){
    #migration
    popsm=migrate(pops,N,m)
    #drift+mutation

    P=list(rep(NA,2*N[1]),rep(NA,2*N[2]),rep(NA,2*N[3]),rep(NA,2*N[4]),rep(NA,2*N[5]),rep(NA,2*N[6]),rep(NA,2*N[7]),rep(NA,2*N[8]),rep(NA,2*N[9]),rep(NA,2*N[10]),rep(NA,2*N[11]),rep(NA,2*N[12]))

    pnames=list(rep(1,2*N[1]),rep(2,2*N[2]),rep(3,2*N[3]),rep(4,2*N[4]),rep(5,2*N[5]),rep(6,2*N[6]),rep(7,2*N[7]),rep(8,2*N[8]),rep(9,2*N[9]),rep(10,2*N[10]),rep(11,2*N[11]),rep(12,2*N[12]))
    for(j in 1:k){
      P[[j]]=smd(N[j],mu,as.numeric(sample(alleles,size=2*N[j],replace=TRUE,prob=popsm[j,])))
    }
  }
}

```

```

alleles=names(table(unlist(P)))
pops=table(unlist(pnames),unlist(P))/(2*N)
}

p=list(rep(NA,2*n),rep(NA,2*n),rep(NA,2*n),rep(NA,2*n),rep(NA,2*n),rep(NA,2*n),rep(NA,2*
n),rep(NA,2*n),rep(NA,2*n),rep(NA,2*n),rep(NA,2*n))
pnames=list(rep(1,n),rep(2,n),rep(3,n),rep(4,n),rep(5,n),rep(6,n),rep(7,n),rep(8,n),rep(9,n),rep(10,n
),rep(11,n),rep(12,n))
for(j in 1:k){
  p[[j]]=sample(alleles,size=2*n,replace=T,prob=pops[j,])
}
sam=matrix(as.numeric(unlist(p)),ncol=2,byrow=T)
f=Fst(cbind(unlist(pnames),sam))
list(pops=pops, sam=sam, Fst=f, N=N, pnames=unlist(pnames))
}

```

## V. Metapopulation Model- Bottleneck

# Function MPMb is for a Metapopulation Model (MM) with bottlenecks

```

MPMbot=function(r,K,L,m,mu,Time,n,b){
  pops=table(1:12,rep(50,12))
  N1=rep(K,12)
  f=matrix(0,ncol=12,nrow=12)
  events=length(Time)
  island=NA
  # first locus
  # step 1: simulate for 10N generations to achieve equilibrium
  npop=MMD(pops=pops,N=N1,m=m,mu=mu,T=10*K,r=r,K=K,n=n)
  # step 2: set population to bottleneck size
  for(j in 1:events){
    EX=j
    island[j]=EX
    bpop=npop$pops
    N2=npop$N
    N2[EX]=b
    #step 3: simulate more generations
    npop=MMD(pops=bpop,N=N2,m=m,mu=mu,T=Time[j],r=r,K=K,n=n)
  }
  Exdat=cbind(npop$pnames,npop$sam)
  f=f+Fst(cbind(npop$pnames,npop$sam))
  for(i in 2:L){
    # step 1: simulate for 10N generations to achieve equilibrium
    npop=MMD(pops=pops,N=N1,m=m,mu=mu,T=10*K,r=r,K=K,n=n)
    # step 2: a random population goes extinct at each Time
    for(j in 1:events){
      EX=j
      island[j]=EX
      bpop=npop$pops
      N2=npop$N
      N2[EX]=b
      #step 3: simulate more generations
      npop=MMD(pops=bpop,N=N2,m=m,mu=mu,T=Time[j],r=r,K=K,n=n)
    }
  }
}

```

```

        Exdat=cbind(Exdat,npop$sam)
        f=f+Fst(cbind(npop$pnames,npop$sam))
    }
    list(data=Exdat,Fst=f/L,extinctions=data.frame(island,Time))
}

```

## VI. Metapopulation Model- Extinction - Recolonization

# Function MPM is for a Metapopulation Model (MM) with extinction-recolonization

```

MPMer=function(r,K,L,m,mu,Time,n){
  pops=table(1:12,rep(50,12))
  N1=rep(K,12)
  Exdat=as.vector(matrix(rep(c("A","B","C","D","E","F","G","H","I","J","K","L"),n),ncol=12,byrow=TRUE))
  f=matrix(0,ncol=12,nrow=12)
  events=length(Time)
  island=NA
  # first locus
  # Step 1: simulate for 10N generations to achieve equilibrium
  npop=MMD(pops=pops,N=N1,m=m,mu=mu,T=10*K,r=r,K=K,n=n)
  # Step 2: a random population goes extinct at each Time
  for(j in 1:events){
    EX=j
    island[j]=EX
    bpop=npop$pops
    bpop[EX,]=0
    N2=npop$N
    N2[EX]=1
    #Step 3: simulate more generations
    npop=MMD(pops=bpop,N=N2,m=m,mu=mu,T=Time[j],r=r,K=K,n=n)
  }
  Exdat=cbind(npop$pnames,npop$sam)
  f=f+Fst(cbind(npop$pnames,npop$sam))
  for(i in 2:L){
    # Step 1: simulate for 10N generations to achieve equilibrium
    npop=MMD(pops=pops,N=N1,m=m,mu=mu,T=10*K,r=r,K=K,n=n)
    # Step 2: a random population goes extinct at each Time
    for(j in 1:events){
      EX=j
      island[j]=EX
      bpop=npop$pops
      bpop[EX,]=0
      N2=npop$N
      N2[EX]=1
      #Step 3: simulate more generations
      npop=MMD(pops=bpop,N=N2,m=m,mu=mu,T=Time[j],r=r,K=K,n=n)
    }
    Exdat=cbind(Exdat,npop$sam)
    f=f+Fst(cbind(npop$pnames,npop$sam))
  }
  list(data=Exdat,Fst=f/L,extinctions=data.frame(island,Time))
}

```

## VII. Cornuet and Luikart Model

# function for 12 simultaneous bottlenecks and recovery to arbitrary size

```
Bottle12=function(r,K,L,m,mu,time,n,b,r2,K2){
  pops=table(1:12,rep(50,12))
  N1=rep(K,12)
  Exdat=as.vector(matrix(rep(c("A","B","C","D","E","F","G","H","I","J","K","L"),n),ncol=12,byrow=TRUE))
  pnames=Exdat
  f=matrix(0,ncol=12,nrow=12)
  for(i in 1:L){
    # step 1: simulate for 10N generations to achieve equilibrium
    bpops=MMD(pops=pops,N=N1,m=m,mu=mu,T=10*K,r=r,K=K,n=n)$pops
    # step 2: pop1 drops to b individuals
    N2=rep(b,12)
    #step 3: simulate more generations
    npops=MMD(pops=bpops,N=N2,m=m,mu=mu,T=time,r=r2,K=K2,n=n)$sam
    Exdat=cbind(Exdat,npops)
    f=f+Fst(cbind(pnames,npops))
  }
  list(data=Exdat,Fst=f/L)
}
```

## Step 3) Migration Matrices

# No Migration Matrix

# simulate isolated populations

```
m=0
x=c(
  1-m,.9*m/2,.1*m/2,0,0,0,0,0,0,.1*m/2,.9*m/2,
  .9*m/2,1-m,.9*m/2,.1*m/2,0,0,0,0,0,.1*m/2,
  .1*m/2,.9*m/2,1-m,.9*m/2,.1*m/2,0,0,0,0,0,0,
  0,.1*m/2,.9*m/2,1-m,.9*m/2,.1*m/2,0,0,0,0,0,
  0,0,.1*m/2,.9*m/2,1-m,.9*m/2,.1*m/2,0,0,0,0,
  0,0,0,.1*m/2,.9*m/2,1-m,.9*m/2,.1*m/2,0,0,0,
  0,0,0,0,.1*m/2,.9*m/2,1-m,.9*m/2,.1*m/2,0,0,
  0,0,0,0,0,.1*m/2,.9*m/2,1-m,.9*m/2,.1*m/2,0,
  0,0,0,0,0,0,.1*m/2,.9*m/2,1-m,.9*m/2,.1*m/2,
  .1*m/2,0,0,0,0,0,0,.1*m/2,.9*m/2,1-m,.9*m/2,
  .9*m/2,.1*m/2,0,0,0,0,0,0,.1*m/2,.9*m/2,1-m)
m=matrix(x,nrow=12,ncol=12)
```

#Migration Matrix

#Adjust m for migration amount relative to original effective population size

#Example is for  $N_m = 1$ ,  $N_e(o) = 1000$

```
m=1/1000
x=c(
  1-m,.9*m/2,.1*m/2,0,0,0,0,0,0,.1*m/2,.9*m/2,
  .9*m/2,1-m,.9*m/2,.1*m/2,0,0,0,0,0,0,.1*m/2,
  .1*m/2,.9*m/2,1-m,.9*m/2,.1*m/2,0,0,0,0,0,0,
  0,.1*m/2,.9*m/2,1-m,.9*m/2,.1*m/2,0,0,0,0,0,
  0,0,.1*m/2,.9*m/2,1-m,.9*m/2,.1*m/2,0,0,0,0,
  0,0,0,.1*m/2,.9*m/2,1-m,.9*m/2,.1*m/2,0,0,0,
  0,0,0,0,.1*m/2,.9*m/2,1-m,.9*m/2,.1*m/2,0,0,
  0,0,0,0,0,.1*m/2,.9*m/2,1-m,.9*m/2,.1*m/2,0,
  0,0,0,0,0,0,.1*m/2,.9*m/2,1-m,.9*m/2,.1*m/2,
  .1*m/2,0,0,0,0,0,0,.1*m/2,.9*m/2,1-m,.9*m/2,
  .9*m/2,.1*m/2,0,0,0,0,0,0,.1*m/2,.9*m/2,1-m)
```

```

0,0,0,.1*m/2,.9*m/2,1-m,.9*m/2,.1*m/2,0,0,0,0,
0,0,0,0,.1*m/2,.9*m/2,1-m,.9*m/2,.1*m/2,0,0,0,
0,0,0,0,0,.1*m/2,.9*m/2,1-m,.9*m/2,.1*m/2,0,0,
0,0,0,0,0,0,.1*m/2,.9*m/2,1-m,.9*m/2,.1*m/2,0,
0,0,0,0,0,0,0,.1*m/2,.9*m/2,1-m,.9*m/2,.1*m/2,
.1*m/2,0,0,0,0,0,0,.1*m/2,.9*m/2,1-m,.9*m/2,
.9*m/2,.1*m/2,0,0,0,0,0,0,.1*m/2,.9*m/2,1-m)

```

```
m=matrix(x,nrow=12,ncol=12)
```

## Step 4) Simulation File- input parameters and # of simulations

### *I. Focal Deme and Metapopulation Models*

#Metasim runs the simulations for the following Population Models: FDMer, FDMbot, MPMer, or MPMbot

#Input is as follows:

#s=number of sims

#r=intrinsic rate of growth

#K=Effective population size (carrying capacity) for each deme

#L=number of loci drawn from each deme

#m=user defined migration matrix

#mu=mutation rate

#n=number of samples from each deme

#MM=the metapopulation model (either FDMer, FDMbot, MPMer, or MPMbot)

```

Metasim=function(s,r,K,L,m,mu,n,MM){
#Open blank files: for genotype file, Fst file, and event times file
write.table(NA,file="Genotypes.txt")
write.table(NA,file="Fst.txt")
write.table(NA,file="Times.txt")
for(i in 1:s){
  Times=rexp(10,1/500)
  Times=round(Times)
  Times=Times[order(Times)]
  Times=c(Times[1],diff(Times))
  Times=replace(Times,Times<1,2) #Time=0 gives 2 generations
  Sims=MM(r,K,L,m=m,mu,Time=Times,n)
  #Fill Data Tables
  write.table(Sims$data,file="Genotypes.txt",append=TRUE)
  write.table(Sims$Fst,file="Fst.txt",append=TRUE)
  write.table(Sims$extinctions,file="Times.txt",append=TRUE)
}
}
#Example:
#Metasim(200,0.0002,1000,20,m,1/1000,20,MPMb)

```

### *II. Focal Deme Model with varying $N_e$*

#unequalNe runs the simulations for the following Population Model: FPMN

```

#Input is as follows:
#s=number of sims
#r=intrinsic rate of growth
#Ne=effective populaton size of all demes
#Nb=effective population size for Deme 1
#K=Effective population size (carrying capacity) for each deme
#L=number of loci drawn from each deme
#m=user defined migration matrix
#mu=mutation rate
#n=number of samples from each deme
unequalNe=function(s,r,Ne,Nb,L,m,mu,time,n){
  for(i in 1:s){
    B1000=FPMN(r,K=Ne,N=Nb,L,m,mu,time,n,b=Nb)
    write.table(B1000$data,file="Geotypes.txt",append=TRUE)
    write.table(B1000$Fst,file="Fst.txt",append=TRUE)
  }
}

```

```

#Example:
#unequalNe(1,.001,10,1,20,m,1/100,10,20)

```

### ***III. Cornuet and Luikart Model***

```

#For Cornuet and Luikart Model use the simulation file below
#Input is as follows:
#r=intrinsic rate of growth
#Ne=effective populaton size of all demes
#Nb=effective population size for Deme 1
#L=number of loci drawn from each deme
#m=user defined migration matrix
#mu=mutation rate
#n=number of samples from each deme
#MM=the metapopulation model (either FDMer, FDMbot, MPMer, or MPMbot)
CL=function(r,Ne,Nb,L,m,mu,time,n,alpha=Ne/Nb){
  #input tau
  tau=0.25
  times <- rep(tau*Nb,8) # not clear for Cornuet and Luikart whether this should be Ne or Nb
  write.table(data.frame(alpha,tau,times),file="CL_Genotypes.txt")
  write.table(data.frame(alpha,tau,times),file="CL_Fst.txt")

  for(i in 1:length(times)){
    B1000=Bottle12(r,K=Ne,L,m,mu,time=times[i],n,b=Nb,r2=r,K2=Ne)
    write.table(B1000$data,file="CLR100.8.txt",append=TRUE)
    write.table(B1000$Fst,file="CLR100.8.Fst.txt",append=TRUE)
  }
}

#For population recovery post-bottleneck, add a value >0 for r
#Example: CL(.001,100,8,20,m,1/100,times,20,alpha)

```

## Decomposed Pairwise Regression Putative Outlier using the method of Koiuzumi *et al.* (2006)

This method is available in the package DPR (Reynolds & Fitzpatrick, 2011)

Notations are in [UTK Orange](#)

R code is in [black](#)

Functions are in [light blue](#)

n=number of populations in your dataset

x=number of sims

y=total number of populations. This version works for 12 unless you modify the function make.dm, see documentation

z = the deme to be ranked relative to other demes (putative outlier)

Input GM saved as GM.txt

Input  $F_{ST}$  matrix, saved as fdist.txt

```
DPR1=function(n,x,y,z){
#Begin by inputting the geographic distance matrix
GM<-read.table("GM.txt")
GM=as.matrix(as.dist(GM))

#View Geographic Distance Matrix
cat("View Geographic Distance Matrix","\n")
print(GM)
cat("","\n")
#Input is a symmetrical  $F_{ST}$  Distance Matrix between demes across all simulations
Fdist<-read.table("fdist.txt")
#Create a data object as a list of  $F_{ST}$  values for each simulation
fst.mat=as.matrix(Fdist)
#Input the number of rows to correspond to the number of simulations
fst=matrix(t(fst.mat),nrow=x,byrow=TRUE)
#Check that matrix is properly formatted
write.table(fst, "fst.mat.txt")

#Input  $F_{ST}$  Matrix and convert to a full square matrix
cat("This works for exactly twelve populations","\n")
cat("If your dataset contains fewer, fill in the extra rows and columns of both the Fst and the Geographic distance matrices with NA to create a full twelve by twelve square matrix","\n")
cat("Alternatively, or if your dataset contains more than twelve populations, check the documentation to refit the function make.dm","\n")
Data<-read.table("fst.mat.txt")

make.dm=function(x,ncol=y){
dm=rbind(
c(x[1],rep(NA,11)),
c(x[2:3],rep(NA,10)),
c(x[4:6],rep(NA,9)),
c(x[7:10],rep(NA,8)),
```

```

c(x[11:15],rep(NA,7)),
c(x[16:21],rep(NA,6)),
c(x[22:28],rep(NA,5)),
c(x[29:36],rep(NA,4)),
c(x[37:45],rep(NA,3)),
c(x[46:55],rep(NA,2)),
c(x[56:66],rep(NA,1)),
c(x[67:78],rep(NA,0)))
}
make.dm(Data[1,])

#Extract Simulation #1 for the decomposed IBD Regression
cat("Extract Simulation #1 for the decomposed IBD Regression","\n")
cat("","\n")
sim1=as.matrix(as.dist(make.dm(Data[1,])))

#Transform  $F_{ST}$  Matrix to Rousset's (1997) Distance Measure  $F_{ST} / (1 - F_{ST})$ 
sim1.fst<-sim1/(1-sim1)

#Isolation by Distance Regressions
#as.dist takes matrix and turns it into a distance object
#res1 is the linear model (lm) for the regression
res1=lm(as.dist(sim1.fst) ~ as.dist(GM))
cat("","\n")
cat("res1 is the linear model (lm) for the regression","\n")
print(summary(res1))

# Plot the IBD Regression of simulation #1
#Red regression line is the mean IBD regression for all 12 demes in the simulation
plot(GM,sim1.fst, ylab="Genetic Distance",xlab="Geographic Distance", main="Decomposed
Pairwise Regression for Simulation 1")

#Decomposed IBD regressions for each deme in simulation #1
# Dotted line is the regression of the focal population (Deme 0)
abline(lm(sim1.fst[,1]~GM[,1]), lwd=2, lty=3)
for(i in 2:n){
  abline(lm(sim1.fst[,i]~GM[,i]))
}
title(sub="Red regression line is the mean IBD regression for all 12 demes in the simulation
Dotted line is the regression of the focal population (Deme 0)",cex.sub=0.5)

# Now compute the decomposed IBD slopes and intercepts
#Recompose into means
#Draw a table of the following for each simulation: 1)mean intercept 2)mean slope 3)variance of the
#intercept 4)variance of the slope.
#Input # of simulations and the # of populations in each simulation
sims=x
pops=y

#Create an empty matrix with nrow = the number of demes and ncol = the number of simulations
Slope=Intercept=matrix(NA,nrow=y,ncol=x)
# Draw an empty table of the following for each simulation: 1)mean intercept 2)mean slope 3)variance of
#the #intercept 4)variance of the slope.
Means=cbind(mean.I=rep(NA,sims),var.I=rep(NA,sims),mean.S=rep(NA,sims),var.S=rep(NA,sims))

```

#Loop to fill Slope and Intercept matrices as well as Means table

```
for(i in 1:sims){
  sim=as.matrix(as.dist(make.dm(Data[i,])))
  sim=sim/(1-sim)
  coefs=lm(as.dist(sim)~as.dist(GM))$coef
  for(j in 1:pops){
    coefs=rbind(coefs,lm(sim[j,]~GM[j,])$coef)
  }
  Means[i,1]=mean(coefs[2:y+1,1])
  Means[i,2]=var(coefs[2:y+1,1])
  Means[i,3]=mean(coefs[2:y+1,2])
  Means[i,4]=var(coefs[2:y+1,2])
  Slope[,i]=coefs[2:(y+1),2]
  Intercept[,i]=coefs[2:(y+1),1]
}
```

```
cat("Barplots are only useful for multiple datasets (simulations)","\\n")
```

#Rank the Intercept for each simulation and place in a matrix with nrow = number of demes per simulation  
#and ncol = number of simulations

```
rank.I=matrix(NA,nrow=y,ncol=x)
vectorranks.I = c()
for (i in 1:sims){
  rank.I[,i]=order(Intercept[,i])
}
```

#Rank the Intercept of the focal deme (deme 1)

```
vectorranks.I.2=rank.I[z,]
```

#View graph of Intercept Ranks

```
windows()
barplot(table(vectorranks.I.2), ylim=c(0,x), col="white", xlab="Rank of Focal Deme Relative to Other
Demes in Each Simulation", ylab="Frequency",
  main="Intercept Rank of Focal Deme", sub="1 is highest, 12 is lowest",axis.lty=1)
```

#Rank the Slope for each simulation and place in a matrix with nrow = number of demes per simulation  
#and ncol = number of simulations

```
rank.S=matrix(NA,nrow=y,ncol=x)

vectorranks.S = c()
for (i in 1:sims){
  rank.S[,i]=order(Slope[,i])
}
```

#Rank the Slope of the focal deme (deme 1)

```
vectorranks.S.2=rank.S[z,]
```

#View graph of Slope Ranks

```
windows()
barplot(table(vectorranks.S.2), ylim=c(0,x), col="white", xlab="Rank of Focal Deme Relative to Other
Demes in Each Simulation", ylab="Frequency",
  main="Slope Rank of Focal Deme", sub="1 is highest, 12 is lowest", axis.lty=1)
```

#View tables of Slope and Intercept Ranks

#Top row is the Rank (from 1 to the maximum # of demes in simulations; example is 1-12) where 1 is the highest and 12 is the lowest

#Bottom row is the number of focal demes that fall under each rank class across all simulations

#Recall that each focal deme is ranked relative to others in its simulation (metapopulation)

```
cat("Table of Interept Ranks","\n")
```

```
cat("Top row is the Rank (from 1 to the maximum # of demes in simulations; example is 1-12) where 1 is the highest and 12 is the lowest","\n")
```

```
cat("Bottom row is the number of focal demesthat fall under each rank class across all simulations","\n")
```

```
cat("Recall that each focal deme is ranked relative to others in its simulation (metapopulation)","\\n")
```

```
print(table(vectorranks.I.2))
```

```
cat("", "\\n")
```

```
cat("Table of Slope Ranks")
```

```
cat("Top row is the Rank (from 1 to the maximum # of demes in simulations; example is 1-12) where 1 is the highest and 12 is the lowest","\n")
```

```
cat("Bottom row is the number of focal demes that fall under each rank class across all simulations","\n")
```

```
cat("Recall that each focal deme is ranked relative to others in its simulation (metapopulation)","\\n")
```

```
print(table(vectorranks.S.2))
```

```
cat("", "\\n")
```

#Write tables to text files

```
cat("", "\\n")
```

```
cat("Two text files of each table are written to the directory","\n")
```

```
write.table(vectorranks.S.2,sep="\\t","Slope.txt",col.names=NA)
```

```
write.table(vectorranks.I.2,sep="\\t","Intercept.txt",col.names=NA)
```

```
cat("", "\\n")
```

```
cat("Three graphs should be generated, one regression plot and two barplots","\n")
```

```
cat("", "\\n")
```

```
}
```

#Call function as: DPR1(n,x,y,z)

Example data: Generate geographic distance matrix and  $F_{ST}$  matrix

```
GM=matrix(c(0:6,5:1,1),ncol=12,nrow=12)
```

```
write.table(GM,sep="\\t","GM")
```

```
fdist=matrix(rbeta(144,.2,.5),ncol=12,nrow=12)
```

```
write.table(fdist,sep="\\t","fdist")
```

```
DPR1(12,1,12,1)
```

## Decomposed Pairwise Regression using the *ad hoc* AIC method of Koizumi *et al.* (2006)

This method is available in the package DPR (Reynolds & Fitzpatrick, 2011)

Notations are in [UTK Orange](#)

R code is in [black](#)

Functions are in [light blue](#)

x=number of sims

y=total number of populations. This version works for 12 unless you modify the function make.dm, see documentation

```
DPR2=function(x,y){  
  #Begin by inputting the geographic distance matrix  
  GM<-read.table("GM.txt")  
  GM=as.matrix(as.dist(GM))  
  
  #View Geographic Distance Matrix  
  cat("View Geographic Distance Matrix","\n")  
  print(GM)  
  cat("\n","\n")  
  
  #Input is a symmetrical FST Distance Matrix between demes across all simulations  
  Data<-read.table("fdist.txt")  
  #Create a data object as a list of FST values for each simulation  
  fst.mat=as.matrix(Data)  
  #Input the number of rows to correspond to the number of simulations  
  fst=matrix(t(fst.mat),nrow=x,byrow=TRUE)  
  #Check that matrix is properly formatted  
  write.table(fst, "fst.mat.txt")  
  
  #Input the number of simulations in the  $F_{ST}$  matrix and the number of demes in each simulation  
  sims=x  
  pops=y  
  
  #Input the number of rows and columns in the  $F_{ST}$  matrix file ("fst.mat")  
  #A file of 12 demes will have nrow=13  
  sim.k=sim.l=matrix(NA,nrow=y+1,ncol=x)  
  
  #best.k is to identify the demes from each simulation that are outliers in the Koizumi et al. (2006) analysis  
  #best.c is to identify the demes from each simulation that are outliers in the AICC model selection from  
  #Burnham and Anderson (2004)  
  best.k=best.c=NA  
  
  #Label demes  
  deme=0:y  
  
  #Vr is an empty matrix with ncol= # of demes plus 1, and nrow= # of simulations  
  #Vr will be filled by the following loop  
  #LL.K is the Log Likelihood from the linear models  
  #AIC.K is the AIC for the Koizumi analysis, defined as  $AIC = 2K + n \ln(RSS/n)$   
  #where K is the # of parameters, n is the # of populations, and RSS is the residual sum of squares
```

```

#AIC.C is the AIC analysis with a correction for small sample size from Burnham & Anderson (2004).
Vr=LL.K=AIC.K=AIC.C=matrix(NA,ncol=pops+1,nrow=sims)

#Loop analysis for all simulations
for (i in 1:sims){
  sim=matrix(fst[i,],ncol=y)
  #Transform FST Matrix to Rousset's (1997) Distance Measure FST/(1-FST)
  sim.fst<-sim/(1-sim)
  Fit.0=lm(as.dist(sim.fst)~as.dist(GM))
  vr=var(Fit.0$res)
  ll.koiz=logLik(Fit.0)
  aic.koiz=2*2+pops*log(sum(Fit.0$res^2)/pops)+2*2*(2+1)/(pops-2-1)
  aic.c=-2*logLik(Fit.0)*2/(pops-1)+4+12/(pops-3)
  for (j in 1:pops){
    fit.d=lm(as.dist(sim.fst[-j,-j])~as.dist(GM[-j,-j]))
    vr=c(vr,var(fit.d$res))
    ll.koiz=c(ll.koiz,logLik(fit.d))
    aic.koiz=c(aic.koiz,4+(pops-1)*log(sum(fit.d$res^2)/(pops-1))+12/((pops-1)-3))
    aic.c=c(aic.c,-2*logLik(fit.d)*2/(pops-2)+12/(pops-4))
  }
  sim.k[,i]=order(aic.koiz)
  best.k[i]=deme[order(aic.koiz)][1]
  best.c[i]=deme[order(aic.c)][1]
  LL.K[i,]=ll.koiz
  AIC.K[i,]=aic.koiz
  AIC.C[i,]=aic.c
  Vr[i,]=vr
}

cat("Print AIC.K scores","\n")
print(AIC.K)
cat("", "\n")
#Best.k is a table showing the deme number on the top row and the number of simulations where that deme
#was identified as an outlier in the bottom row
#In the example, there are 12 demes, identified as 1-12, and 200 simulations
#Demes which have undergone an historical event should be identified as outliers
#For example, if deme #1 underwent a dramatic bottleneck in 200 independent simulations, this table will
#show how many times deme 1 was identified as an outlier relative to other demes in the simulation
cat("Best.k is a table showing the deme number on the top row and the number of simulations where that
deme was identified as an outlier in the bottom row","\n")
cat("Demes which have undergone an historical event should be identified as outliers","\n")
cat("For example, if deme #1 underwent a dramatic bottleneck in 200 independent simulations, this table
will show how many times deme 1 was identified as an outlier relative to other demes in the
simulation","\n")
print(table(best.k))
cat("", "\n")

#Write the table of outliers as a text file
cat("Table of outliers written as a text file (best.k)", "\n")
write.table (best.k, "best.k.txt")
cat("", "\n")
}
#Call function as: DPR2(x,y)
#Use example data from DPR1

```

## Unique Genotypes Test

This test gives the 1 tailed  $P$ -value to test the fraction of replications where random draws of genotypes from a population are less than the observed number of genotypes ( $g$ ). Use a minimum of 10,000 replications, and repeat to ensure repeatability.

Notations are in [UTK Orange](#)

R code is in [black](#)

Functions are in [light blue](#)

```
NG.test <- function(X,N,n, reps){
  # X is a vector of band frequencies
  # N is the diploid sample size
  # n is the observed number of unique genotypes
  L <- length(X) # determine number of markers
  G <- numeric()
  for(i in 1:reps){
    genos <- matrix(NA,N,L)
    for(j in 1:L){
      genos[,j] <- sample(c(0,1),size=N,replace=TRUE,prob=c(1-X[j],X[j]))
    }
    geno.c <- numeric()
    for(j in 1:N){
      geno.c[j] <- paste(genos[,j],sep="",collapse="")
    }
    G[i] <- length(unique(geno.c))
  }
  G
}
```

### Example

#### **Sample Data:**

Simulate band frequencies from a beta distribution, with 29 samples and shape parameters of 0.2

```
X <- rbeta(29,.2,.2)
hist(X)
sum(X*(1-X)>1/29)
N <- 29
n <- 15
reps <- 10000
xx <- NG.test(X=X,N=N,n=n, reps=reps)
hist(xx)
sum(xx<n)/reps
```

## Simulation of Expected Gene Diversity at Equilibrium

This script estimates population gene diversity ( $H_e$ ) across simulations of 70 neutral, diallelic markers in isolated island populations of three effective population sizes (proxy for island area) and varying initial gene diversity. This provides a null model of attrition of alleles from a population post-vicariance.

Notations are in [UTK Orange](#)

R code is in [black](#)

Functions are in [light blue](#)

[# equilibrium heterozygosity](#)

```
eq.H <- function(N,mu){
  4*N*mu/(1+4*N*mu)
}
N <- 10^(1:6)
plot(N,eq.H(N,1/1000),log="xy")
```

[# expected number of alleles in sample of size n \(Ewens 1972 in Cornuet and Luikart\)](#)

```
E.K <- function(n,N,mu){
  M <- N*mu
  xx <- 1:n
  yy <- M/(M+xx-1)
  sum(yy)
}
K.N <- numeric()
N <- 10:50000
for(i in 1:length(N)){
  K.N[i] <- E.K(N[i],N[i],1/1000)
}
plot(N,K.N)
```

[# heterozygosity in for a diallelic locus](#)

```
p.dist <- function(N.e, mu, nu, lty) {
  x <- seq(0.001, 0.999, by=0.001)
  shape1 <- 4*N.e*nu
  shape2 <- 4*N.e*mu
  y <- dbeta(x, shape1, shape2)
  lines(x, y, lty=lty)
}
H.2 <- function(N.e,mu,nu,L){
  shape1 <- 4*N.e*nu
  shape2 <- 4*N.e*mu
  y <- rbeta(L,shape1,shape2)
  H <- 2*y*(1-y)
  mean(H)
}
```

[#expectation](#)

```
H.2.e <- function(N.e,mu,nu){
  shape1 <- 4*N.e*nu
  shape2 <- 4*N.e*mu
  y <- shape1/(shape1+shape2)
  2*y*(1-y)
```

```

    }

H2 <- numeric()
for(i in 1:length(N)){
  H2[i] <- H.2.e(N[i],1/1000,1/1000)
}
plot(N,H2)

for(i in 1:length(N)){
  H2[i] <- H.2(N[i],1/1000,1/1000,20)
}
plot(N,H2,log="xy")

mu <- c(1e-3,1e-4,1e-5,1e-6)
H2 <- numeric()
for(i in 1:length(mu)){
  H2[i] <- H.2(1000,mu[i],mu[i],70)
}
plot(mu,H2)

```

# how many generations since isolation? 8K? If this is a "relaxation" metapopulation, do we expect a #Heterozygosity-Area relationship?

```

EH.r <- function(Ne,time=4000,p0=0.25,mu=1e-4){
  p.t=p0
  for(i in 1:time){
    ptt <- rbinom(1,as.integer(2*Ne),p.t)/(2*Ne)
    mut1 <- rbinom(1,as.integer(ptt*2*Ne),mu)
    mut2 <- rbinom(1,as.integer((1-ptt)*2*Ne),mu)
    p.t <- (ptt*2*Ne-mut1+mut2)/(2*Ne)
  }
  2*p.t*(1-p.t)
}

```

#The following creates four boxplot charts of He estimates for varying effective population size, time, original #gene diversity, and mutation rate  
 #Each boxplot script gives three boxplots for three effective population sizes (Ne=100,1000,10,000)  
 #Boxplots show medians (bold horizontal lines), interquartile ranges (the "box"), "whiskers" out to 1.5x the #interquartile range, and individual markers more extreme than the whiskers. The large blue circle #illustrates the mean of the 70 randomly simulated markers and the red circle illustrates the expected mean #at equilibrium assuming a beta distribution of allele frequencies with shape parameters equal to  $4N_e\mu$ .

#### #Boxplot Group 1

```

H.r100 <- numeric()
for(i in 1:70){
  H.r100[i] <- EH.r(Ne=100,time=4000,p0=0.5,mu=1e-3) #Input parameters
}
H.r1000 <- numeric()
for(i in 1:70){
  H.r1000[i] <- EH.r(Ne=1000,time=4000,p0=0.5,mu=1e-3) #Input parameters
}
H.r10000 <- numeric()
for(i in 1:70){
  H.r10000[i] <- EH.r(Ne=10000,time=4000,p0=0.5,mu=1e-3) #Input parameters
}

```

#Create a 6 panel display

```
windows(width=10,height=7)
par(mfrow=c(2,3))
```

```
boxplot(list(H.r100,H.r1000,H.r10000),names=c("100","1,000","10,000"),xlab="Ne",ylab="gene
diversity",cex.lab=1.2,cex.axis=1.2)
text(.65,.48,"(a)",cex=1.2)
```

#Label the mean gene diversity for the simulated markers in blue

```
points(1:3,lapply(list(H.r100,H.r1000,H.r10000),mean),pch=16,col="blue",cex=1.5)
```

#Expected mean gene diversity for each value of mu

```
t.mean3 <- c(H.2(100,1e-3,1e-3,70),H.2(1000,1e-3,1e-3,70),H.2(10000,1e-3,1e-3,70))
t.mean4 <- c(H.2(100,1e-4,1e-4,70),H.2(1000,1e-4,1e-4,70),H.2(10000,1e-4,1e-4,70))
t.mean5 <- c(H.2(100,1e-5,1e-5,70),H.2(1000,1e-5,1e-5,70),H.2(10000,1e-5,1e-5,70))
```

#Label the expected mean gene diversity at equilibrium in red

```
points(1:3,t.mean3,pch=16,col="red")
```

#Boxplot Group 2

```
H.r100 <- numeric()
for(i in 1:70){
  H.r100[i] <- EH.r(Ne=100,time=4000,p0=0.5,mu=1e-4)
}
H.r1000 <- numeric()
for(i in 1:70){
  H.r1000[i] <- EH.r(Ne=1000,time=4000,p0=0.5,mu=1e-4)
}
H.r10000 <- numeric()
for(i in 1:70){
  H.r10000[i] <- EH.r(Ne=10000,time=4000,p0=0.5,mu=1e-4)
}
```

```
boxplot(list(H.r100,H.r1000,H.r10000),names=c("100","1,000","10,000"),xlab="Ne",ylab="gene
diversity",cex.lab=1.2,cex.axis=1.2)
text(.65,.48,"(b)",cex=1.2)
points(1:3,lapply(list(H.r100,H.r1000,H.r10000),mean),pch=16,col="blue",cex=1.5)
points(1:3,t.mean4,pch=16,col="red")
```

#Boxplot Group 3

```
H.r100 <- numeric()
for(i in 1:70){
  H.r100[i] <- EH.r(Ne=100,time=4000,p0=0.5,mu=1e-5)
}
H.r1000 <- numeric()
for(i in 1:70){
  H.r1000[i] <- EH.r(Ne=1000,time=4000,p0=0.5,mu=1e-5)
}
H.r10000 <- numeric()
for(i in 1:70){
  H.r10000[i] <- EH.r(Ne=10000,time=4000,p0=0.5,mu=1e-5)
}
```

```
boxplot(list(H.r100,H.r1000,H.r10000),names=c("100","1,000","10,000"),xlab="Ne",ylab="gene
diversity",cex.lab=1.2,cex.axis=1.2)
```

```
text(.65,.48,"(c)",cex=1.2)
points(1:3,lapply(list(H.r100,H.r1000,H.r10000),mean),pch=16,col="blue",cex=1.5)
points(1:3,t.mean5,pch=16,col="red")
```

#### #Boxplot Group 4

```
H.r100 <- numeric()
for(i in 1:70){
  H.r100[i] <- EH.r(Ne=100,time=4000,p0=0.25,mu=1e-3)
}
H.r1000 <- numeric()
for(i in 1:70){
  H.r1000[i] <- EH.r(Ne=1000,time=4000,p0=0.25,mu=1e-3)
}
H.r10000 <- numeric()
for(i in 1:70){
  H.r10000[i] <- EH.r(Ne=10000,time=4000,p0=0.25,mu=1e-3)
}
```

```
boxplot(list(H.r100,H.r1000,H.r10000),names=c("100","1,000","10,000"),xlab="Ne",ylab="gene
diversity",cex.lab=1.2,cex.axis=1.2)
text(.65,.48,"(d)",cex=1.2)
points(1:3,lapply(list(H.r100,H.r1000,H.r10000),mean),pch=16,col="blue",cex=1.5)
points(1:3,t.mean3,pch=16,col="red")
```

#### #Boxplot Group 5

```
H.r100 <- numeric()
for(i in 1:70){
  H.r100[i] <- EH.r(Ne=100,time=4000,p0=0.25,mu=1e-4)
}
H.r1000 <- numeric()
for(i in 1:70){
  H.r1000[i] <- EH.r(Ne=1000,time=4000,p0=0.25,mu=1e-4)
}
H.r10000 <- numeric()
for(i in 1:70){
  H.r10000[i] <- EH.r(Ne=10000,time=4000,p0=0.25,mu=1e-4)
}
```

```
boxplot(list(H.r100,H.r1000,H.r10000),names=c("100","1,000","10,000"),xlab="Ne",ylab="gene
diversity",cex.lab=1.2,cex.axis=1.2)
text(.65,.48,"(e)",cex=1.2)
points(1:3,lapply(list(H.r100,H.r1000,H.r10000),mean),pch=16,col="blue",cex=1.5)
points(1:3,t.mean4,pch=16,col="red")
```

#### #Boxplot Group 6

```
H.r100 <- numeric()
for(i in 1:70){
  H.r100[i] <- EH.r(Ne=100,time=4000,p0=0.25,mu=1e-5)
}
H.r1000 <- numeric()
for(i in 1:70){
  H.r1000[i] <- EH.r(Ne=1000,time=4000,p0=0.25,mu=1e-5)
}
H.r10000 <- numeric()
```

```

for(i in 1:70){
  H.r10000[i] <- EH.r(Ne=10000,time=4000,p0=0.25,mu=1e-5)
}

boxplot(list(H.r100,H.r1000,H.r10000),names=c("100","1,000","10,000"),xlab="Ne",ylab="gene
diversity",cex.lab=1.2,cex.axis=1.2)
text(.65,.48,"(f)",cex=1.2)
points(1:3,lapply(list(H.r100,H.r1000,H.r10000),mean),pch=16,col="blue",cex=1.5)
points(1:3,t.mean5,pch=16,col="red")

# after 4000 generations, we expect isolated populations to be very close to mutation drift equilibrium and
#therefore to see a rather strong relationship between gene diversity and island area

```

## Gene Diversity for AFLP Data

This method uses Zhivotovsky (1999) Bayesian estimator, which is superior to Lynch and Milligan's (1994) method as implemented in AFLP-SURV 1.0 (Vekemans *et al.*, 2002).

Notations are in [UTK Orange](#)

R code is in [black](#)

Functions are in [light blue](#)

```
Zhit.fn <- function(dat){
  # estimate gene diversity, etc with Zhivotovsky's (1999) Bayesian estimator
  # this is superior to Lynch and Milligan because they simply toss markers with fewer than 3 0's,
  # leading to weird biases in data sets with little variation
  # dat is a matrix or data frame containing an indicator for population in column 1 and fragment
  # presence/absence (1's and 0's) in subsequent columns
  pop <- dat[,1]
  npops <- length(table(pop))
  X <- dat[,-1]
  L <- ncol(X)
  Q <- VQ <- H <- matrix(NA,nrow=npops,ncol=L)
  q.fn <- function(n,m,a,b){
    beta(m+a+0.5,n-m+b)/beta(m+a,n-m+b)
  }
  Vq.fn <- function(n,m,a,b){
    beta(m+a+1,n-m+b)/beta(m+a,n-m+b)-(beta(m+a+0.5,n-m+b)/beta(m+a,n-m+b))^2
  }
  for(i in 1:L){
    # estimate non-uniform priors
    xx <- na.omit(data.frame(pop=pop,x=X[,i]))
    xx0 <- xx[xx$x==0,]
    ni <- table(xx$pop)
    mi <- ni*tapply(xx$x,xx$pop,sum)
    n <- sum(ni)
    m <- sum(mi)
    Ri <- mi/ni
    Rbar <- m/n
    VR <- sum((ni/n)*(Ri-Rbar)^2)
    a <- Rbar*((Rbar*(1-Rbar))/VR-1)
    b <- (1-Rbar)*((Rbar*(1-Rbar))/VR-1)
    for(j in 1:npops){
      # estimate allele frequencies
      Q[j,i] <- q.fn(ni[j],mi[j],a,b)
      # variances of estimates
      VQ[j,i] <- Vq.fn(ni[j],mi[j],a,b)
    }
    # gene diversities
    H[,i] <- (ni/(ni-1))*2*Q[,i]*(1-Q[,i])
    if(m==0){H[,i] <- Q[,i] <- rep(0,npops)}
    if(m==n){H[,i] <- rep(0,npops); Q[,i] <- rep(1,npops)}
  }
  list(freq=Q,Var.f=VQ,H=H)
}
```

### Sample Data

| Island | order | Locus 1 | Locus 2 | Locus 3 | Locus 4 |
|--------|-------|---------|---------|---------|---------|
| Provo  | 1     | 1       | 1       | 1       | 1       |
| North  | 2     | 1       | 0       | 1       | 1       |

Example using sample data

```
aflp <- read.table("sampledata.txt",header=TRUE)
```

```
GS.Z <- Zhit.fn(aflp[,2:6])
```

```
rowMeans(GS.Z$H)
```

## References

Fitzpatrick BM, Reynolds RG (2011) *MetaSim*- a package to simulate genetic data in structured demes using a forward-time approach. R package version 1.0.

Lynch M, Milligan BG (1994) Analysis of population genetic structure with RAPD markers. *Mol Ecol* 3:91-99.

Reynolds RG, Fitzpatrick BM (2011) *DPR*- a package to implement decomposed pairwise regression in an isolation by distance framework. R package version 1.0.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145:1219-1228.

Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of relationship between degree of homoplasy and fragment size. *Mol Ecol* 11:139-151.

Zhivotovsky L (1999) Estimating population structure in diploids with multilocus dominant DNA markers. *Mol Ecol* 8:907-913.

## **VITA**

Robert Graham Reynolds was born in Salt Lake City, Utah in 1981 and raised in Asheville, North Carolina. He attended Carolina Day School in Asheville and graduated in 2000 having received the school's Science Award and the Faculty Prize. He completed a senior honors thesis studying the mammalian community of the Pisgah National Forest in Madison County, North Carolina. While in high school Graham participated in the program QUEST, which allowed high school juniors to finish coursework one month early in order to pursue an internship or apprenticeship. Graham enrolled in an Earthwatch Institute program working with Drs. Carlos López Gonzáles, Alberto Gonzáles Romero, and Mircea Hildalgo-Mihart of Arizona State University and the Instituto de Biología, Universidad Nacional Autónoma de México on carnivore community ecology in the tropical dry forest of the Chamela-Cuixmala Biosphere Reserve. Graham credits this experience with solidifying his interest in science and the tropics, and returned to the project as an intern in the subsequent two summers.

Graham enrolled in Duke University in Durham, North Carolina in 2000, majoring in Biology and volunteering in the lab of Dr. Susan Alberts. While there he participated in an Organization for Tropical Studies semester abroad in Costa Rica, where he studied tropical biology, conservation, and Latin American culture. Upon his return, he wrote an undergraduate research proposal to study mate choice behavior in poison dart frogs in Panama, receiving funding from the Duke University Howard Hughes Fellowship in Biology, the Deans' Summer Fellowship, and the Mellon Foundation Latin American Travel Grant. Graham graduated from Duke in 2004 and moved to San Diego, California for two years while his future wife Jennifer attended graduate school, a period

of time spent traveling and exploring, especially in California, Baja California, and the Sea of Cortés. During this time he spent five months backpacking in Central America, South America, and Africa, where he attempted to see as much wildlife as possible and fulfill lifelong dreams of diving in the Indian Ocean, exploring the Amazon by open canoe, and climbing Mt. Kilimanjaro.

Graham enrolled in the doctoral program at the University of Tennessee in 2006 to work with Dr. Ben Fitzpatrick. During this time he received enough grant money to travel at least annually to the Turks and Caicos for field work and became active in promoting conservation in that region. In 2008, he received the W.K. McClure Scholarship for the Study of World Affairs from the University of Tennessee to examine the interface of conservation, politics, and the local community in the Turks and Caicos, a pursuit which, combined with his field work, recently led to a book chapter in *Conservation of Caribbean Island Herpetofaunas* (Brill, Netherlands). In addition to his dissertation work and numerous side projects, Graham co-wrote and co-edited the book *The Amphibians of Tennessee* (University of Tennessee Press) with Matthew Niemiller. Graham's favorite activity is spending time with his wife Jennifer (MPH, CHES) and their dog Cheyenne.