2017

USING GENOMICS TO UNDERSTAND POPULATION DEMOGRAPHICS IN THE CONTEXT OF AMPHIBIAN CONSERVATION

Schyler O. Nunziata
University of Kentucky, schyler.nunziata@uky.edu
Digital Object Identifier: https://doi.org/10.13023/ETD.2017.390

USING GENOMICS TO UNDERSTAND POPULATION DEMOGRAPHICS IN THE
CONTEXT OF AMPHIBIAN CONSERVATION

_____

DISSERTATION

_____

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

By

Schyler Olivia Nunziata

Lexington, Kentucky

Director: Dr. David Weisrock, Professor of Biology

Lexington, Kentucky

2017

ABSTRACT OF DISSERTATION

USING GENOMICS TO UNDERSTAND POPULATION DEMOGRAPHICS IN THE
CONTEXT OF AMPHIBIAN CONSERVATION

Understanding the demography of species over recent history (e.g., < 100 years) is critical in studies of ecology and evolution, but records of population history are rarely available. Large single nucleotide polymorphism datasets generated with restriction-site associated DNA sequencing (RADseq), in combination with demographic inference methods, are improving our ability to gain insights into the population history of both model and non-model species. However, to assess the performance of genetic methods it is important to compare their estimates of population history to known demography, in both simulation and empirical settings. Here, I used a simulation approach to examine the potential for RADseq datasets to accurately estimate effective population size (N_e) in Wright-Fisher populations over the course of stable and declining population trends, and distinguish stable from steadily declining populations over a contemporary time scale (20 generations). Overall, my results reveal that demographic inference using genome-wide data can be successfully applied to estimate N_e, and the detection of population-size declines. Next, I assess these methods in an empirical study from a wetland with 37 years of amphibian mark-recapture data to study the utility of genetically-based demographic inference on salamander species with documented population declines (*Ambystoma talpoideum*) and expansions (*A. opacum)*. For both species, demographic model inference supported population size changes that corroborated mark-recapture data. To further validate these findings, I used individual-based population models of the pond-breeding salamander, *Ambystoma opacum,* with life-history parameters estimated from a long-term dataset, over a 50 year projection. My results demonstrate that genetically estimated N_e is positively correlated with census size in isolated and subdivided *A. opacum* populations. Finally, I investigated metapopulation patterns of genomic diversity in *A. opacum* and *A. talpoideum* and how migration may impact N_e estimation. I found strong patterns of subpopulation structuring, signatures of migration between subpopulations, and differences in N_e at the subpopulation level in both species. Overall, my findings suggest the ability of genomic data to reconstruct recent demographic changes, which can have important applications to conservation biology, and

ultimately can help us elucidate the effects of environmental disturbances in the demography of endangered or declining species.

Schyler Olivia Nunziata

June 2, 2017

USING GENOMICS TO UNDERSTAND POPULATION DEMOGRAPHICS IN THE
CONTEXT OF AMPHIBIAN CONSERVATION

By

Schyler Olivia Nunziata

David W. Weisrock, Ph.D.

Director of Dissertation

David F. Westneat, Ph.D.

Director of Graduate Studies

June 2, 2017

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

Chapter Six: Conclusions and future directions

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

GENERAL INTRODUCTION: GENOMICS AND N_e ESTIMATION FOR ASSESSING
RECENT POPULATION SIZE TRENDS

Introduction

Estimating the impacts of environmental change on natural populations is critical in

ecology, evolution, and conservation biology. Genetic approaches of estimating the effective

population size (N_e) can provide information on population abundance and demographic history,

and also insight into the evolutionary history and genetic viability of populations (Frankham *et

al.* 2014). However, before this information can be applied to assess the current and projected

response of populations to environmental stressors, it is necessary to understand the evolutionary

processes and mechanisms regulating genetic diversity of populations. Especially important is

the understanding of how recent demographic changes influence estimates of N_e at an ecological

timescale (i.e., tens of generations), because this is the scale at which conservation and

management is focused.

A range of evolutionary processes occur at an ecological time scale (Allendorf & Hard

2009; Schoener 2011), including short-term changes in N_e (Allendorf *et al.* 2010; McCoy *et al.*

2014). However previous attempts to estimate contemporary changes in N_e using microsatellite

markers have frequently been associated with high uncertainty, as the signal of drift is often

overwhelmed by noise in the dataset (Nunziata *et al.* 2015; Therkildsen *et al.* 2013). Genome-

scale datasets have the potential to overcome these limitations by reducing sampling error and

yielding insight into eco-evolutionary processes, revealing not only short-term population size

trends but also the evolutionary potential of populations. Both the response of species to

anthropogenic influences and their potential for future genetic adaptation is critical information

in the conservation of populations (Franklin 1980; Mace *et al.* 2008; Soule & Wilcox 1980).

Genome-scale SNP datasets are currently generated through a range of reduced genome

representation sequencing methods ideal for non-model species of conservation concern, as

many methods don't require any prior genomic information (Andrews *et al.* 2016; Davey *et al.*

2011). Some of the most popular methods of marker discovery and sequencing include:

restriction-site associated DNA sequencing (RADseq) (Baird *et al.* 2008; Peterson *et al.* 2012),

sequencing of RNA (RNAseq) (Wang *et al.* 2009), and targeting known regions of the genome

for sequencing through sequence capture (Grover *et al.* 2012). The development of these

methods and advances in speed and cost of next-generation sequencing are shifting eco-

evolutionary studies of non-model species to a genomic scale.

Genomic tools have the potential to revolutionize the conservation and management of

species of concern, and offer the ability to identify traits under selection and their associated

genes, population responses to anthropogenic stressors, and assessments of species status and

gene flow among populations (Allendorf *et al.* 2010; Stapley *et al.* 2010). Here, I focus on the

potential of genomics to provide robust estimates of N_e and demographic inferences at a

contemporary time scale. Although the concept of N_e is well accepted, its estimation becomes

complicated at short time scales with fluctuations in demography. The potential of genomic

datasets for accurately estimating contemporary N_e has been outlined in a number of reviews

(Allendorf *et al.* 2010; Pujolar *et al.* 2013), but only now are we seeing an accumulation of these

datasets in published studies (see Table 1.1 for a summary of recent studies).

Estimation of effective size

Effective population size is defined as the size of an ideal population (discrete generations, random mating, equal sex ratio, and random variation in reproductive success) that experiences the same rate of drift as the observed population. I review two methods of estimating N_e using genome-scale data from single-season samples: inbreeding N_e using a linkage-disequilibrium (LD) approach, and coalescent N_e. While variance N_e is a popular estimator of contemporary N_e, it requires samples from at least two time points (Wang 2005), and here I focus on single-sample estimators. Many studies of contemporary N_e focus solely on inbreeding and/or variance N_e and ignore coalescent N_e, as it is often thought to be reflective of N_e strictly at a historical time scale (i.e., hundreds to thousands of generations). However, in small populations where coalescent events are clustered in the recent past, coalescent N_e may provide insight to demographic processes at a contemporary timescale. In the following sections I discuss the history of these methods in population genetics, review studies that have applied N_e estimation to characterize recent demographic history, evaluate the potential of these methods with genomic datasets, and identify questions regarding the utility of these methods at a contemporary timescale.

LD-based estimations of recent demographic history

With finite population size and a limited number of parents, nonrandom associations of alleles at different genetic markers occur (i.e. linkage disequilibrium), even without any physical linkage on a chromosome (Hill 1981; Waples 1991). Because LD is a function of drift, it can be used to estimate N_e. Estimates of LD-based N_e are reflective of the past 1-2 generations, as linkage requires several generations to decay even as N_e fluctuates (Waples 2005). Estimates of

LD-based N_e have been found to be robust when applied to species with a complex life-history, including species with population structure and migration up to 5-10% of the total population (Waples & England 2011), and with overlapping generations (Waples & Do 2010).

Estimation of LD-based N_e in population and conservation genetics has become an important and widely used tool, with multiple well-established and easy to use software packages available for estimation (Do *et al.* 2014; Waples & Do 2010). While widely used, the method is best applied to small populations, as the signal of drift that generates LD in small populations will be weak in large populations. Several empirical studies employing microsatellite loci have found positive correlations with LD-based N_e and census size, demonstrating the potential benefits of genetic monitoring (Charlier *et al.* 2012; Osborne *et al.* 2010). While several other studies have failed to detect any correlation with N_e and census size, even with observed population declines (Duong *et al.* 2013; Whiteley *et al.* 2015). One major limitation of these studies employing microsatellites is that sampling is often inadequate for robust estimation of N_e, especially in large populations where the signal of drift is small.

Although large genomic datasets are predicted to increase the resolution of LD-based N_e estimates, few studies have used the method with these large datasets, and those that have, revealed varied results. In a recent study on a population of Iberian pigs Saura *et al.* (2015) evaluated the use of genomic datasets and LD-based N_e estimation to document observed recent population size trends. The intense 26 generation study of a closed population of Iberian pigs allowed for precise estimates of pedigree-based N_e for a comparison to LD-based estimates using genome-scale data (Saura *et al.* 2015). Estimates of LD-based N_e were correlated with pedigree estimates, accurately documenting the founding event and subsequent expansion, and validating the utility of this method for characterizing recent population size trends with genomic datasets.

The house finch is another species with well documented changes in demographic history, providing an opportunity to assess genomic estimates of demography(Shultz *et al.* 2016). House finch populations were introduced to the Eastern US and Hawaii from their native range in the Western US in the 1940's and 1870's, respectively. After which, introduced populations rapidly expanded until being introduced to a novel pathogen that caused severe declines across the Eastern US. Temporal genetic sampling prior to pathogen outbreak and during the subsequent decline provided the opportunity to evaluate the impacts of recent population decline on genetic diversity. LD-based N_e estimates reflected signatures of the population introductions, but did not show declines with recent population decline. This is likely the result of the still large size of contemporary populations, allowing for maintenance of large N_e despite population size declines. These results show that LD-based estimates of N_e will not always track contemporary census-size trajectories, even with the application of tens of thousands of SNP markers.

Coalescent-based estimations of recent demographic history

While LD-based methods offer the appeal of quick and straight-forward calculations, they provide little other insight into demographic history, such as level of migration and timing and duration of demographic events. Demographic inference in a model testing framework has been a popular tool in population genetics for a number of years, with many of these approaches built on the backbone of coalescent theory (Kingman 1982). The principle of the coalescent is that when tracing the ancestry of a sample of *n* individuals, each pair of individuals that have the same parent are said to coalesce. As lineages progress back in time, the number of distinct lineages (or potential parents) decreases until reaching one, the most recent common ancestor of

the sample. Because the rate of coalescence depends on the size of the population, this model can

be used approximate N_e and model demographic events backward in time.

Up until the past several years, the use of coalescent based demographic inference has

been limited by computationally demanding frameworks such as approximate Bayesian

computation (ABC) (Beaumont *et al.* 2002). This had limited the size of the dataset and the

complexity of the models tested to balance the computational demand. New, computationally

efficient approaches of demographic inference are able to take advantage of large genomic

datasets, and test increasingly complex models of demographic history (Excoffier *et al.* 2013;

Gutenkunst *et al.* 2009). However, most studies incorporating these methods have estimated

events at historical time scales (i.e., thousands of generations; see (Lanier *et al.* 2015; Moura *et*

al. 2014; Papadopoulou & Knowles 2015), and not a contemporary time scale. Coalescence

theory states that the probability of coalescence t generations ago is $(1-(1/2N))^{t-1}(1/2N)$, with the

coalescent N_e estimated as the expected time of coalescence in generations, T, or $T = 2N_e$

(Nordborg & Krone 2002; Wakeley & Sargsyan 2009). Given these equations, when N_e is small

enough, as is often the case in species of conservation concern, large sample sizes (individuals

and/or loci) may be effective in estimating coalescent N_e at a contemporary scale as coalescent

events will be clustered in the recent past (Robinson *et al.* 2014).

Two recent studies demonstrate the use of coalescent-based demographic inference

methods to estimate timing of known population introductions in the past several decades. In a

study of the checkerspot butterfly (*Euphydryas gilletti*), McCoy et al. (2014) accurately

characterized a recent (~ 30 generations ago) introduction event and associated bottleneck with

the coalescent modeling program δaδi (Gutenkunst et al. 2009). Similarly, the timing of

introduction of a series of guppy populations introduced in the past ~20-40 years was accurately estimated in the program δαδι (Fraser *et al.* 2015).

Although the above studies highlight the potential of these methods for small populations, no studies have successfully applied them to characterize a population decline. A comparative study of two species of bumblebees, one with temporal stability in census size and the other which is in decline, used the program δαδι in an attempt to infer these recent population histories (Lozier 2014). There was no detected difference in genetic diversity between the species, and both species displayed a historic signature of population expansion. Although the study highlights the importance of historical processes in shaping current genomic diversity, demographic inference results provided little insight into contemporary demographic processes.

Major differences between LD and coalescent Ne

Coalescent-based and LD-based estimates of N_e differ in important ways, especially in the time scale they are reflective of. Both methods are based on the principle of inbreeding N_e, the probability of two gene copies within a population being identical by descent. However, the coalescent N_e is a long-term estimate of N_e integrated backward in time to the MRCA, while LD-based N_e is based on associations of genetic markers only 1-2 generations in the past, providing an estimate of contemporary N_e. Given the principles behind these two approaches, if the population remains stable over time both methods should give congruent estimates of N_e. With fluctuations in N_e, the methods may give different estimates as they are reflecting drift at different time scales (Charlesworth 2009). Studies rarely take a comparative approach of these two methods, and it is unclear under what circumstances they will be comparable or which may give the better insights into populations experiencing recent changes in census size.

Demographic inference in complex and dynamic systems

While the above review highlights the potential resolution offered by a genomic perspective, a number of questions still exist regarding the impacts of life-history and population dynamics on estimates of recent demography. The dynamic properties of real populations create complex challenges in the use of demographic inference methods to track contemporary processes and in the interpretation of N_e in empirical systems. In the review above I point out that some studies have found correlations between genomic estimates of population history and recent census size, while others find no correlation. These disparate results lead to a number of unanswered questions such as, how many generations of decline will need to occur before a genetic signature of the event becomes apparent with either LD-based or coalescent-based N_e estimation. As well as, under fluctuating population size what time scale is reflected in N_e estimates. Another long-standing question is the impact of population structure and migration on N_e estimates (Neel *et al.* 2013). Migration can contribute to changes in allele frequencies over time, which if left unaccounted for in models, will be attributed to drift and bias N_e estimates (Gilbert & Whitlock 2015). High levels of migration between subpopulations will also lead to N_e estimates reflective of the global or metapopulation N_e instead of the local population (Neel *et al.* 2013). Therefore, accurate genetically based estimates of N_e within local populations should account for population structure.

In other words, even with the increased precision of genomic estimates of demography and N_e, important questions still exist regarding the geographic and temporal scale reflected in these estimates. These are critical questions if these methods are going to be of any use in a

conservation and management, where the spatial and temporal scales of demographic changes are essential information.

Case Studies with a focus on amphibians--Chapters 2 through 5

Further empirical and simulation based studies are important for understanding the circumstances under which genomic data in combination with N_e estimation and demographic inference can be applied to understand recent population history. In this dissertation I attempt to fill in these gaps in knowledge using simulations of ideal populations, as well as assessing current and projected responses of pond-breeding amphibian populations to climate change in a series of case studies. I discuss the application of genome-wide SNP markers in various N_e estimation methods and evaluate sampling requirements in a simulation study, and then assess the application of N_e estimates to infer recent demography in two complex and dynamic salamander populations at a subpopulation and metapopulation level. Estimation of demographic history in amphibians is complicated not only by metapopulation structuring across landscapes, but also stochastic reproductive success that is tightly linked to environmental condition. Before genetic estimates of demography can be applied to amphibian conservation it is necessary to evaluate if detection of overall population growth trends is achievable with this complex demography.

Rainbow Bay, an ephemeral wetland in eastern South Carolina, provides a unique opportunity for this work because amphibian species have been continuously sampled over the last 39 years, providing a valuable time series across which climatological and ecological conditions have changed significantly (Todd *et al.* 2010). The Rainbow Bay wetland was completely encircled by a drift fence in 1978, with amphibians and reptiles entering and leaving

the wetland censused daily since initiation. All captured animals have been cohort marked, sexed, and released on the opposite side of the drift fence. Importantly, genetic samples have been collected across years. Findings from monitoring the RB amphibian community found that increased droughts from 1978-2004 were negatively correlated with census size in amphibian species adapted to long hydroperiods, while other amphibian species exhibited significant population expansions with decreased hydroperiod (Daszak *et al.* 2005). This comprehensive dataset allows for direct calculations of census-size, and accurate estimates of life-history parameters for modeling N_e with stochastic reproduction. Accurate and validated estimates of N_e are vital in ecology, evolution, and conservation, as they measure the evolutionary potential and viability of populations.

In this dissertation, I assess the application of genomic datasets in coalescent and LD-based estimation of contemporary N_e. Next, I use the Rainbow Bay study system in a number of simulation and empirical studies to investigate the ability of genome-scale data to estimate N_e and track population size changes. Overall, I aimed to address the following questions in this dissertation:

1.) Can genome-scale datasets and N_e estimators be applied to infer population abundance and population size trends over a contemporary time scale? What are the impacts of the various aspects of population demography (initial population size and the number of generations since λ began) on estimation, as well as the impacts of alternative N_e estimators, sampling conditions, and data filtering?

2.) Can recent population growth or decline of empirical salamander populations be detected using genome-wide SNP data and coalescent-based demographic inference? Do non-

temporally sampled data have similar power as temporal samples to detect recent

population trends?

3.) How are N_e and genetic diversity impacted by complex life-history characteristics and

stochastic reproduction common of amphibian populations? How does metapopulation

structure mediate population viability in amphibian populations and impact N_e estimates

at the subpopulation level?

4.) What spatial scale is reflected in N_e estimates of amphibians breeding populations, the

subpopulation (wetland) or metapopulation? Do subpopulation census size changes

impact the connectivity of the metapopulation as a whole?

Table 1.1 Summary of recent studies using genomic datasets to estimate contemporary N_e and recent demographic history.

Species	Data Collection	Approach	Major Findings	Comparison with Empirical Data	Citation
Bumble bees	RADseq	δaδi	No difference in genetic diversity between stable and declining species. Both show signal of historic population growth, no signature of recent decline in declining species.	Yes	Lozier 2014
Checkerspot Butterfly	RNAseq	δaδi	Accurate timing of population introduction	Yes	McCoy et al. 2014
Guppy	ddRADseq	δaδi	Accurate timing of population introduction	Yes	Fraser et al. 2015
Iberian pigs	SNPchip	LD using custom scripts	Correlation between pedigree and LD Ne over 26 generations	Yes	Saura et al. 2015
Waterflea	SNPchip	Temporal genetic diversity	Stability of genetic diversity despite cyclical population size	Limited	Orsini et al. 2016
House finch	ddRADseq	LD using custom scripts	Ne remained stable over ~8 generations with decline in Nc, although overall population size remained large. Human-induced founder events detected.	Yes	Shultz et al. 2016
Brook Charr	RADseq	LD in NeEstimator	No difference in N_b between founder and source populations	Limited	Narum et al. 2017
Chinook Salmon	RADseq	LD in NeEstimator	No difference in N_b between founder and source populations	Limited	Narum et al. 2017

CHAPTER TWO

ESTIMATION OF CONTEMPORARY EFFECTIVE POPULATION SIZE AND
POPULATION DECLINES USING RAD SEQUENCE DATA

Abstract

Large single nucleotide polymorphism (SNP) datasets generated with restriction-site

associated DNA sequencing (RADseq), in combination with demographic inference methods, are

improving our ability to gain insights into the population history of both model and non-model

species. I used a simulation approach to examine the potential for RADseq datasets to accurately

estimate effective population size (N_e) over the course of stable and declining population trends,

and accurately distinguish stable from steadily declining populations over a contemporary time

scale (20 generations). Using a linkage disequilibrium-based analysis, individual sampling had

the greatest effect on N_e estimation and the detection of population-size declines (i.e., $n \geq 30$),

with declines reliably detected across scenarios approximately 10 generations after they began.

Coalescent-based inference required fewer sampled individuals (i.e., $n = 15$), and instead was

most influenced by the size of the SNP dataset, with tens of thousands of SNPs required for

accurate detection of population trends and at least 20 generations after decline began. The

number of samples available and targeted number of RADseq loci are important criteria when

choosing between these methods. With an understanding of the limitations and biases of these

approaches, researchers can make more informed decisions when designing their sampling and

analyses protocol. Overall, my results reveal that demographic inference using RADseq data can

be successfully applied to population parameters that are changing on a very recent time scale

and may be important tools for population monitoring and conservation biology.

Introduction

One of the most important parameters in wildlife management and conservation biology is effective population size (N_e), with estimates providing insight into the demographic history and extinction risk of populations. Although N_e is informative about population viability and broadly applicable in ecology, conservation, and evolution, it is notoriously difficult to estimate (Luikart *et al.* 2010). Rarely is enough demographic information available from natural populations to directly estimate N_e, making indirect genetic estimates of considerable use, especially given their ease of generation relative to direct demographic methods (Luikart *et al.* 2010; Schwartz *et al.* 2007). It is now possible to generate population genomic data for almost any species for the investigation of population and evolutionary history (Andrews *et al.* 2016; Narum *et al.* 2013). The increase in power and precision offered by a genomic approach is poised to greatly improve estimates of demographic history, including N_e, the timing of demographic events, and migration estimates. However, as emphasized in a recent review, the application of genomic techniques in conservation studies has been rare (Shafer *et al.* 2015). One obvious, but unanswered, question is whether genomic data coupled with demographic inference methods have the ability to accurately characterize population history over a contemporary time scale (e.g., tens of generations).

Genomic data and demographic inference methods have become increasingly popular for gaining insight into population history (Excoffier *et al.* 2013; Gutenkunst *et al.* 2009). However, it is unclear how well these inference methods perform when applied to demographic events occurring at a contemporary time scale. Simulation studies have suggested that microsatellite markers have the ability to detect bottlenecks and population size trends at a contemporary scale, but require sample sizes of 60 or more individuals and are not accurate with large (≥ 1000)

population sizes (Antao *et al.* 2011; Tallmon *et al.* 2010). These studies either did not use single nucleotide polymorphism (SNP) data that would be common in contemporary population genomic studies, or they simulated a small number (100 - 1000) of SNP markers (Antao *et al.* 2011; Hollenbeck *et al.* 2016). It is possible that the increased power offered by large genomic datasets can result in accurate estimates of demography over short time scales while using smaller numbers of sampled individuals.

Recent empirical studies have shown that coalescent based demographic inference can accurately date documented introductions of populations occurring in the past few decades (Fraser *et al.* 2015; McCoy *et al.* 2014). Coalescence theory states that the probability of coalescence t generations ago is $(1-(1/2N))^{t-1}(1/2N)$, with the coalescent N_e estimated as the expected time of coalescence in generations, T, or $T = 2N_e$ (Nordborg & Krone 2002; Wakeley & Sargsyan 2009). Given these equations, when N_e is small enough, as is often the case in species of conservation concern, large sample sizes (individuals and/or loci) may be effective in estimating coalescent N_e at a contemporary scale as coalescent events will be clustered in the recent past (Robinson *et al.* 2014). Before these methods can be applied to real-world conservation biology, vigorous exploration is needed to estimate their accuracy with realistic sampling conditions to gain an understanding of implicit limitations and biases (Shafer *et al.* 2015).

Restriction site-associated DNA sequencing (RADseq) is arguably the most popular method for generating genome-wide population genetic data from a reduced subset of the genome (Andrews *et al.* 2016; Davey *et al.* 2011). While RADseq can yield many thousands or tens of thousands of shared orthologous loci across individuals and populations, it also has inherent properties that lead to allele dropout, and consequently, missing data that may create

biases in population genetic results. Allele dropout via mutations in restriction cut sites and the

shotgun nature of Illumina sequencing, which randomly under-sequences loci or alleles can lead

to either missing genotypes for loci, or the misinterpretation of null alleles as homozygous at

heterozygous loci. Both of these scenarios can result in skewed estimation of allele frequencies

(Arnold *et al.* 2013). Simulation studies have highlighted the downstream effects of these biases

in commonly estimated population genetic summary statistics (Arnold *et al.* 2013; Gautier *et al.*

2013) and in phylogenetic inferences (Huang & Knowles 2014). However, the effect of allele

dropout in RADseq-based studies of N_e and contemporary population size trends has not been

investigated.

Here, I use an approach similar to (Tallmon *et al.* 2010) and assess the ability of

RADseq-generated SNP data and different N_e estimators to infer population abundance and

population size trends (λ) over a contemporary time scale. I simulated ideal Wright-Fisher (W-F)

populations over a range of known census sizes (N_C) and with either stable population size, or a

steadily declining population. In ideal W-F populations $N_C = N_e$, so that estimates of N_e can be

directly compared to the simulated N_C. Using both a linkage disequilibrium-based analysis, and a

coalescent-based analysis, I assess the estimation of N_e and population size trends. In doing so, I

also evaluate the impacts of the various aspects of the population model (initial population size

and the number of generations since λ began) on estimation, as well as the impacts of sampling

and data filtering.

Methods

Data simulation

I conducted simulations of RADseq data for populations with both stable and declining

population sizes using the Python program simuPOP v1.1.4 (Peng & Kimmel 2005), a forward-

time and individual-based population genetic modeling program. Initial haploid allele

frequencies were generated with the coalescent simulator fastsimcoal2 v2.5.2.21 (fsc2)

(Excoffier *et al.* 2013) for 20,000 150 base pair (bp) loci using a diploid N_e of 1000. A mutation

rate (μ) was randomly assigned to each locus from a log-normal distribution with a mean μ of

2.5E-8 and a log standard deviation of 1.3. This mutation rate has been robustly estimated in

humans (Nachman & Crowell 2000) and similarly used in other RADseq simulation studies

(Huang & Knowles 2014). Loci were generated as Arlequin-formatted files and were

subsequently converted to Phylip format using the program PGDSpider v2.0.5.1 (Lischer &

Excoffier 2012). Initial diploid genotypes for individuals in the simuPOP population were

generated by pairing the fsc2-simulated alleles for each locus using random sampling with

replacement, which approximated W-F populations. Diploid populations were constructed with

initial population sizes of $N = 250$, 500, and 1000, with 100 replicates constructed for each initial

population size. Throughout the subsequent simulations, populations maintained an average sex

ratio of 1 with random mating. Under these conditions N_C should be approximately equal to N_e.

A $\mu = 2.5E-8$ was fixed across all loci. All simulated populations went through an equilibrium

phase of 10 generations to reach Hardy-Weinberg equilibrium (Antao *et al.* 2011; Tallmon *et al.*

2010; Waples 2006), after which each replicate diploid population evolved for one generation (t_{-1}) according to two separate deterministic growth rates that approximated a stable population ($\lambda = 1.0$) and a declining population ($\lambda = 0.9$). Data collection began at generation t_0 as the

population evolved at the same λ for 20 generations as in (Tallmon *et al.* 2010). In each simulation, genotypes from all loci were recorded after 0, 5, 10, 15, and 20 generations. Sample collection began one generation after the initiation of the deterministic growth rate because inbreeding N_e estimates are reflective of the number of parents in the parental generation (Waples 2005). To assess the effect of the sample size of individuals, we sampled 15, 30 and 60 individuals from each of the specified generations.

In silico RADseq mutations and Data Filtering

Using custom Python scripts, I filtered RADseq loci from sampled individuals to mimic empirical RADseq data recovery and filtering conditions typically used in population genomic studies. To simulate allelic dropout as a result of a mutation in the restriction enzyme cutting site, all individual sequences were deleted containing a mutation in the first 8 bp. To simulate missing data as a result of variation in sequencing coverage, I simulated the number of reads for each individual allele by drawing randomly from a Poisson distribution with a mean of 10 (Huang & Knowles 2014). I imposed a sequencing coverage cutoff of 10, which is considered an efficient sequencing coverage cutoff for diploids. Individuals must have a coverage ≥ 5 reads per allele for a given locus to be included in the data matrix. If one allele had a coverage ≥ 10 reads and the other had < 5, the locus was recorded as homozygous for the recovered allele due to allele dropout. Loci below these coverage cutoffs were recorded as missing data. All other sources of missing data and biases from sequencing errors, coverage cut-offs, and alignment errors were ignored here as they are not the focus of our study. These have been thoroughly reviewed in other studies, and are expected to cause general biases in all sequencing projects (Huang & Knowles 2014; Pool *et al.* 2010; Rokas & Abbot 2009).

I next filtered the simulated RADseq data using criteria specific to the two analytical programs used in demographic estimation.

LD-based estimation – Linkage disequilibrium methods for N_e estimation assume unlinked loci. To remove the inclusion of linked sites within a RADseq locus, I used only the first SNP in a locus in all LD-based data sets. To examine if the LD-based method produced unbiased N_e estimates with perfect detection of allele dropout, I analyzed datasets that removed all loci with missing data exclusively due to RADseq cut site mutations, hereafter referred to as the *LD RAD mutation* dataset. I further examined how LD-based N_e estimation would be affected by the combined impacts of missing data from allele dropout due to RADseq cut site mutation and low sequencing coverage. For these analyses, I generated two filtered datasets that removed loci with ≥ 10% and ≥ 50% missing data; hereafter referred to as the *10%* missing and *50%* missing datasets, respectively.

fsc2 – In fastsimcoal2, the use of linked SNPs should not bias parameter estimation, so all datasets analyzed in this study used all SNPs in a locus. However, the inclusion of loci with missing data is expected to lead to a biased site-frequency spectrum (SFS) and result in inaccurate parameter estimates (Excoffier *et al.* 2013). Therefore I included only loci with no missing data across all sampled individuals. To examine the potential effects of allele dropout on N_e estimation in the program fsc2, I analyzed the simulated RADseq data under a range of filtering strategies that accounted for allele dropout due to mutations in restriction cut sites and insufficient sequencing coverage. First, I analyzed an unfiltered data matrix with no allele dropout. Here, the SFS was constructed using the complete 20,000 locus (3,000,000 bp) simulated dataset, and is hereafter referred to as the *fsc2 complete* dataset. Next, I examined the performance of N_e estimation in fsc2 when accounting for the perfect detection of allele dropout

due to restriction cut site mutations. Here, the SFS was constructed after removal of all loci with a restriction cut site mutation, hereafter referred to as the *fsc2 RAD mutation* dataset. Finally, I examined the performance of N_e estimation in fsc2 when allowing for allele dropout due to both cut-site mutation and low sequencing coverage, hereafter referred to as the *fsc2 RAD mutation and coverage* dataset.

N_e estimation and demographic inference

I used the program NeEstimator v2.01 (Do *et al.* 2014) to estimate N_e using the linkage disequilibrium method (Hill 1981). This method measures the deviation from the expected genotype frequency based on allele frequencies in the population (Waples & Do 2010). I estimated N_e from all sampled generations of our temporally simulated populations, employing all three LD-based data filtering scenarios described above. In addition, I assessed the effect of filtering data using minor allele frequency cutoffs (MAF_{cut}). For all data sets, I separately applied a MAF_{cut} of 0.01, 0.02, and 0.05. An allele frequency cutoff of 0.02 has been recommended to balance precision and bias (Waples & Do 2010), although 0.05 is a common value used in SNP-based studies.

I used fsc2 to perform demographic inference using the joint SFS generated from serial samples taken at generations 0 (t_0) and 20 (t_{20}) in the temporally simulated populations. For all fsc2 analyses, I used a simple model of a single population with N_e at t_0 fixed at the known starting value and N_e in subsequent generations allowed to vary according to the model. Fixing N_e at t_0 allowed me to reduce the number of parameters estimated from the model, scale N_e estimation without a mutation rate, and ignore invariant sites in the SFS. Defined parameter ranges were uniformly distributed with N_e ranging from 1 to 10,000. A total of 100,000

simulations were performed to estimate the SFS, with a minimum and maximum of 10 and 100

loops (ECM cycles). The stopping criterion was defined as the minimum relative difference in

parameters between two iterations, and was set to 0.001. A total of 50 replicate fsc2 runs were

performed for each replicate simulation of a demographic scenario, and for each of the three fsc2

filtering options described above. The overall maximum likelihood run across all 50 fsc2

replicates was retained as a point estimate for N_e^{t20}. Due to computational limitations, for each

combination of initial population size and population growth rate, only the first 40 temporally

simulated replicates (out of 100) were analyzed with fsc2.

Accuracy assessments

The performance of each N_e estimation method was evaluated for the overall accuracy of

N_e estimates. To characterize the accuracy of N_e estimates across simulation replicates, I

measured the root mean squared error (RMSE) calculated after removing infinitely large

estimates by

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\frac{1}{\widehat{N}_{ei}} - \frac{1}{N_e})^2} \, ,$$

where $\widehat{N}_{ei}$ is the estimated N_e in the ith (i = 1-100) replicate, and N_e is the simulated N_e. The

RMSE was not calculated if over 50% of the estimates of $\widehat{N}_{ei}$ reached infinity.

Detection of population size change

To estimate population size trends, I calculated $\hat{\lambda}$ as the slope of a linear regression of the

log transformation of N_e estimates from current and historical samples within a simulated

replicate and I compared these to known λ. I performed these calculations for results generated

from both NeEstimator and fsc2 using all simulated demographic scenarios, data filtering

scenarios, and MAF_{cut} levels. Following (Tallmon *et al.* 2010), I recorded the proportion of times

$\hat{\lambda} < 0.95$ when true $\lambda = 0.9$. This is a practical conservation scenario to identify populations that

are declining by at least 5% per generation. I also assessed how often a stable population was

incorrectly identified as declining as the proportion of times $\hat{\lambda} < 0.95$ when true $\lambda = 1.0$ (false

positive rate).

Results

The number of SNPs generated in the simulation depended on the initial population size,

imposed lambda, and the post-simulation filtering scenario used (LD-based data: Table 2.1; fsc2

data: Table 2.2). Consistent with theoretical expectations, in the LD-based SNP datasets larger

populations generally had more SNPs and lost genetic diversity less rapidly due to drift, and

declining populations lost genetic diversity more rapidly than stable populations. The mean

number of SNPs in the joint-SFS was highly dependent on data-filtering method, with the

number of shared SNPs between t_0 and t_{20} declining with allele dropout from both RADseq

mutation and insufficient sequencing coverage. Although the number of SNPs will vary with

study design, such as the number of individuals multiplexed in an Illumina sequencing lane, and

coverage cutoffs, the number of SNPs we recovered in our simulations is comparable to

empirical RADseq studies.

Stable population size estimation

LD-based estimation – Here, I focus on results from estimation of $\hat{N}_e$ at t_{20} under a $\lambda =$

1.0, where the accuracy of $\hat{N}_e$ estimation was most influenced by the number of individuals

sampled and the MAF_{cut} employed (Fig. 2.1; Fig. A.1). Estimates of $\widehat{N}_e$ at time points t_0 through t_{15} were nearly identical to $\widehat{N}_e$ at t_{20}, and are not presented here. RMSE calculations yielding the lowest measures of error for all simulated demographic and filtering scenarios are presented in Table 2.3. The lowest individual sample size (n = 15) only produced meaningful results at a simulated population size of N = 250 and a $MAF_{cut} = 0.05$, with the majority of replicates at higher simulated population sizes and/or different filtering methods yielding either infinite $\widehat{N}_e$ or very wide ranges of parameter estimates. A full summary of the proportion of replicate estimates that reached infinity can be found in Tables A.1 to A.3. In contrast, increased individual sampling (n = 30 and n = 60) produced more accurate and certain estimates of $\widehat{N}_e$ over most demographic and data-filtering scenarios. Analyses of the *LD RAD mutation* dataset generated $\widehat{N}_e$ estimates with the greatest accuracy and least variance; however, datasets with *10%* and *50%* missing data due to both cut-site mutations and insufficient read coverage also generated similarly accurate $\widehat{N}_e$ estimates under many simulated population sizes and MAF_{cut} levels. The MAF_{cut} level yielding the most accurate results varied with the number of individuals sampled and simulated population size. Generally, including low frequency alleles with an $MAF_{cut} = 0.01$ appeared to have the largest effect by upwardly biasing $\widehat{N}_e$ and yielding the greatest variance (Figure A.1).

  fsc2 – Estimation of $\widehat{N}_e$ at t_{20} under a $\lambda = 1.0$ population model was most influenced by the allele dropout filtering scenario used (Figure 2.2A-C). The *fsc2 complete* dataset, which also has the greatest number of SNPs (Table 2.2), generally yielded the most accurate and precise estimates of $\widehat{N}_e$, particularly in the N = 250 and N = 500 models. This data set yielded the most precise estimates in the N = 1000 model, but also had a negative bias in its estimates. Increased individual sampling had a slight improvement on accuracy and/or precision under all three

population size models when using the *fsc2 complete* dataset, and generally had a negligible effect under the other allele dropout filters. However, analysis of 60-individual data sets with the *fsc2 RAD mutation and coverage* dataset yielded a very wide range of estimates under all three population size models, with highly inaccurate and negatively biased estimates under a $N = 250$ model. In general, accuracy and precision in all scenarios proportionally decreased with the number of SNPs in the data set (Table 2.2).

The *fsc2 complete* dataset yielded the lowest RMSE values for the $N = 250$ and $N = 500$ population models, and when sampling 60 individuals in the $N = 1000$ model (Table 2.4). Overall, *fsc2 RAD mutation* and *fsc2 RAD mutation and coverage* datasets had similar RMSE values.

Declining population size estimation

LD-based estimation – The number of generations since the beginning of a population decline was the biggest factor affecting the accuracy and precision of $\widehat{N}_e$ estimation (Figure 2.3, Figures A.1 to A.3), with the variance in estimates decreasing over time as population size declined. Individual sampling also affected results, with an $n = 15$ yielding a greater estimation variance, particularly in earlier generations of the decline. Estimation using an $n = 30$ or 60 produced highly certain estimates of $\widehat{N}_e$ in t_{10} through t_{20}. In general, $\widehat{N}_e$ estimation over time was only minimally affected by the initial population size, the missing data filter used, or the MAF_{cut} used. However, with individual samples size of $n = 15$ a MAF_{cut} of 0.05 lead to a greater proportion of finite $\widehat{N}_e$ (Table A.6 to A.8).

Similarly, estimation of $\hat{\lambda}$ over different time intervals was most influenced by the number of generations passing between sampling events. The data filter used had minimal

33

impact on the accuracy of $\hat{\lambda}$ estimation and we present results from analyses of the *10%* missing

data here (Table 2.5) with results from analysis of additional allele dropout datasets presented in

Tables A.4 to A.5. When sampling 30 to 60 individuals, the MAF_{cut} did not have a large impact

on population trend detection, but with an individual samples size of 15, a MAF_{cut} of 0.05

improved population trend detection. For example, when sampling 15 individuals, population

declines with an initial $N \leq 500$ were detected 67% of the time when at least ten generations

passed, and increased to 85% of the time when 20 generations passed. With n = 15 and an initial

N = 1000, at least 20 generations must pass for population declines to be detected 64% of the

time. However, with n = 15 using a MAF_{cut} of 0.05 also increased the false positive rate, where

stable populations were incorrectly identified as declining with $\hat{\lambda}$ estimates of < 0.95 across

many replicates (Table 2.6, Tables A.9 to A.10). Increased individual sampling greatly improved

the correct identification of a declining population. For example, under an N = 1000 model,

sampling 60 individuals resulted in the correct identification of a population decline $> 95\%$ of

the time when 10 generations passed and correct identification $> 71\%$ of the time after just five

generations.

      fsc2 –The accuracy of $\hat{N}_e$ at t_{20} was most influenced by the allele dropout filter used

(Figures 2.2 D-F), which itself influenced the number of SNPs included in the joint SFS.

Estimates of $\hat{N}_e$ at t_{20} were positively biased across all datasets, with greater bias in datasets with

fewer numbers of SNPs. Similarly, estimation of $\hat{\lambda}$ was most influenced by the allele dropout

filter used. Population declines of $\hat{\lambda} < 0.95$ were detected across all 40 analyzed replicates using

the *fsc2 complete* dataset (Fig. 2.2 D-F). In contrast, none of the replicates for either *fsc2 RAD*

mutation, or *fsc2 RAD mutation and coverage* datasets meet our criteria of $\hat{\lambda} < 0.95$, although

most qualitatively indicated decline relative to N_e at t_0.

Discussion

My results demonstrate that RADseq data have the potential to improve the inference of population demography and the detection of population declines on a very recent time scale. The linkage disequilibrium and coalescent methods I applied to estimate N_e use largely different sources of information from genomic datasets. The relative performance of these methods was influenced by different factors related to the study design, such as the number of individuals sampled (very important for LD-based estimation) and the amount of variable data generated (very important for coalescent estimation). Given that the accuracy and precision of N_e estimators hinge on aspects of the study design and the underlying population history, I further discuss these influences and provide guidelines for inferring N_e and population size trends. While I compare and contrast the performance of both estimators, combining results from both methods in empirical studies may be the best approach to develop an encompassing view of overall population demographic history, as suggested by (Waples 2016).

Performance of estimators

In my analysis of RADseq data, LD-based demographic inference generally outperformed coalescent-based inference for N_e estimation and the detection of population declines. However, there were limitations with LD-based inference, most notably with the number of sampled individuals required to provide both accurate and precise results. Sampling of 15 individuals led to large variance in estimates. This was most evident under a stable population size and in the early generations of a population decline, particularly when population size was large (e.g., N = 1000). In contrast, increasing sampling to 30 individuals greatly increased the

35

accuracy and precision of N_e estimates. This may be discouraging from the perspective of sampling, as many population genetic studies sample far fewer than 30 individuals per population. However, in light of microsatellite-based simulations showing that 30 individuals resulted in largely biased N_e estimation (Tallmon *et al.* 2010), LD-based analysis of RADseq appears to provide new opportunities for accurate demographic inference. This method also has potential for quick and accurate detection of population declines, generally after only 10 generations from initiation of a decline.

In contrast, the coalescent-based N_e estimation (using fsc2) was not greatly affected by the number of individuals sampled, with highly precise N_e estimates produced using as few as 15 individuals. The most significant limitation for the coalescent approach was the data filter used, and therefore the number of SNPs in the joint-SFS. Use of the *fsc2 complete* data set consistently produced more accurate N_e estimates compared to data sets in which loci were removed due to cut-site mutations and insufficient sequencing coverage. All data sets yielded a consistent upward bias in N_e estimation in the declining populations (Figures 3.2D-F), and I am not sure what drives this estimation bias, but it was most pronounced in the data sets affected by allele dropout (and fewer SNPs). Despite this positive bias population declines were obvious using the complete data set at 20 generations from initiating declines. Due to the intense computational needs inherent to fsc2, N_e was not estimated at other time points. While complete data sets like the ones used here are not attainable in empirical research, the positive correlation between numbers of SNPs and accurate coalescent-based N_e estimation is encouraging. Technological improvements and sequencing costs continue to increase our ability to generate more complete genome-wide SNP data, even when factoring in allele dropout. In contrast, increasing sample size, especially temporally, will remain difficult for many species. The use of true N_e as a prior

for one of the sampled years is also unlikely to be available in most study systems, which would

further model complexity and add analytical time to an already computationally challenging set

of analyses. Ultimately, coalescent-based demographic inference using a joint SFS-based method

may be a great option for a more limited set of studies with access to large SNP data sets, and

prior population information, as has been illustrated in a number of empirical studies (Fraser *et

al.* 2015; McCoy *et al.* 2014; Nunziata *et al.* 2015).

Allele dropout and data filtering

Missing data via allele dropout in RADseq studies has been shown to affect a number of

population genetic summary statistics, including measures of genetic diversity and population

structure (Arnold *et al.* 2013; Gautier *et al.* 2013). My results from parameter estimators for N_e

are therefore encouraging, as increasing levels of missing data via allele dropout had little impact

on LD-based N_e estimation and were generally comparable to the dataset with no null alleles.

Interestingly, while LD-based estimation was robust to the effects of allele dropout and missing

data, the MAF_{cut} influenced $\widehat{N}_e$ accuracy and precision, particularly under a model of stable

population size. These results are consistent with other studies (Waples & Do 2010), where the

inclusion of low frequency alleles created a positive bias, while the exclusion of these alleles

created a slightly negative bias, particularly at the lowest sample size (Waples & Do 2010). Also

consistent with the guidelines outlined in Waples and Do (2010), when low individual samples

sizes were used (n = 15) a MAF_{cut} of 0.05 yielded the most finite and accurate estimates, as it is

the only MAF_{cut} that screened out singletons, which can bias $\widehat{N}_e$.

In contrast to the LD-based analyses, the allele dropout filter used in the fsc2 analyses did

affect the results. Because these analyses preclude the use of loci with missing data, the direct

impact of filtering loci by allele dropout was a major reduction of the number of SNPs included in the joint-SFS. Contemporary population declines purge rare alleles, creating a predictable signature in the SFS (Gattepaille *et al.* 2013; Nei *et al.* 1975), with the likelihood of detecting this signature increasing with the number of SNPs included in the dataset. I found that N_e estimation was most accurate, and declines were only detected, using my most inclusive dataset containing >150,000 SNPS. However, this was an empirically unrealistic dataset generated with perfect recovery of loci without any allele dropout. The generation of empirical datasets robust enough to detect population declines may, therefore, require increased sequencing efforts to offset the effects of allele dropout by increasing the number of loci sampled and their coverage. Maybe counter intuitively, increased individual sampling does not solve this problem as adding individuals increases the probability of allele dropout through a cut site mutation or insufficient sequencing coverage, creating a smaller SNP matrix and decreasing precision in $\widehat{N}_e$ (Fig. 2.2). Potentially, this result can be overcome by subsampling individuals for non-missing data (Papadopoulou & Knowles 2015).

Allele dropout often goes undetected in many studies, and my preliminary exploration suggests that although biases were observed in SFS-based demographic inference, the underlying population history of either stable or declining populations were recovered and point estimates were almost always within an order of magnitude of real N_e. Previous simulation work has revealed that non-equilibrium demography, such as a population decline, can cause low N_e and result in fewer loci with missing data and more accurate allele frequency estimation (Arnold *et al.* 2013). Therefore, my findings should not be interpreted as applicable across systems, since I may have modeled scenarios (i.e., low N_e, steadily declining) that limit allele dropout and create evident signatures in the SFS at a contemporary time scale.

Practical considerations

Many additional factors influence N_e that I have not modeled here, including selection, migration, and overlapping generations (Slatkin 2008). In real populations N_e rarely equals N_C, and changes in N_e could track of any number of demographic changes, not exclusively N_C (Palstra & Ruzzante 2008). Further simulations are needed under more realistic scenarios to determine the application of evaluated methods across systems. One factor that must be considered with RADseq datasets and the LD-based approach is that although pairwise r^2 values (correlation of genes within individuals) increase with number of loci, SNPs on the same chromosome are not independent and will reduce precision of $\hat{N}_e$ because LD will be the result of physical linkage and not drift (Waples *et al.* 2016). The use of linked SNPs could be corrected for by using known genomic architecture (Waples *et al.* 2016); however further research is needed to assess the importance of this issue in the application of LD-based N_e estimation to RADseq data.

When inferring $\hat{\lambda}$ from $\hat{N}_e$ for conservation purposes, false positives can lead to a waste of management resources when stable populations are misidentified as declining (Schwartz *et al.* 2007). The absence of any false positives in the fsc2-based λ estimation, and the lower number of individuals required, is promising for its application in conservation studies. However, the failure to detect a decline in all fsc2 analyses while accounting for allele dropout highlights the need for very large SNP datasets, as well as temporal sampling, especially if quick detection of population declines is a goal. False positives for LD-based λ estimates were also low, although this typically required larger sample sizes of at least 30 individuals. With large resources available to researchers, the application of both methods for demographic inference will be the

ideal approach to take, but given constraints on sampling or sequencing, the results here can be useful for guiding decisions about how to design a conservation genetic study aimed at detecting recent population declines. Finally, even when temporal sampling is unavailable, N_e is itself an important indicator of population viability and evolutionary potential and RADseq data can serve as a valuable source of information for this parameter.

Table 2.2. Number of SNPs used for LD-based analysis resulting from simulations in simuPop. SNPs were calculated as the average across the 100 replicates for each simulated scenario and are presented for the initial generation (t_0) after the 10-generation equilibrium phase and from the final generation (t_{20}). SNP levels are broken down across the different initial population sizes (N), population growth rates (λ) and data-filtering methods (LD RAD mutation, 10% missing, and 50% missing data sets).

N	λ	Initial variation at t_0			Final variation at t_{20}		
		RAD mutation	10%	50%	RAD mutation	10%	50%
		# SNPs	# SNPs	# SNPs	# SNPs	# SNPs	# SNPs
250	1.0	3660	3938	4527	3521	3695	4223
500	1.0	3718	4054	4657	3629	3876	4451
1000	1.0	3756	4130	4742	3703	4013	4611
250	0.9	3915	3941	4676	3099	2965	3457
500	0.9	3719	4053	4661	3392	3493	3975
1000	0.9	3756	4128	4739	3570	3773	4327

Table 2.2. Mean number of SNPs in the joint-SFS for generations t0 and t20 resulting from simulations in simuPop. SNPs were calculated as the average across 40 replicate simulations for each simulated scenario and are broken down across the different initial population sizes (N), population growth rates ($\lambda = 0.9$, 1.0), and individual sample sizes (n). Results are further parsed by data-filtering method, including the fsc2 complete dataset, (2) fsc2 RAD mutation (RAD mut.), and (3) fsc2 RAD mutation and coverage (mut. & cov.).

N	n	Stable population $\lambda = 1.0$			Declining population $\lambda = 0.9$		
		complete	RAD mut.	mut. & cov.	complete	RAD mut.	mut. & cov.
250	15	158,824	14,797	7,494	155,197	14,649	7,550
250	30	165,545	15,192	3,121	-	-	-
250	60	170,077	15,386	624	-	-	-
500	15	165,308	15,197	7,110	161,361	15,171	8,193
500	30	173,444	15,377	3,565	169,989	15,578	3,570
500	60	179,068	15,765	548	-	-	-
1000	15	168,248	15,353	7,832	165,730	14,817	7,985
1000	30	177,853	15,812	3,600	175,501	15,170	3,305
1000	60	185,150	16,226	772	182,895	15,667	592

For some parameter combinations, there were insufficient numbers of individuals for target n (-).

Table 2.3. RMSE values for all filtering scenarios for LD-based analysis in NeEstimator under a stable population ($\lambda = 1.0$). Bold values identify the lowest RMSE for a particular combination of population size (N), individual sampling level (n), and minor allele frequency cutoff (MAF_{cut}). Bold and italicized values identify the lowest RMSE for a particular population size. RMSE was not estimated if over 50% of the estimates of $\widehat{N}_{ei}$ reached infinity for a particular parameter combination (Inf.).

		LD RAD mutation (& MAF_{cut} level)			10% missing data (& MAF_{cut} level)			50% missing data (& MAF_{cut} level)		
		0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05
N = 250	n = 15	2.73E-03	2.73E-03	2.95E-03	3.41E-03	3.41E-03	2.73E-03	3.68E-03	3.68E-03	**2.70E-03**
	n = 30	1.41E-03	**1.10E-03**	1.25E-03	1.55E-03	1.11E-03	1.18E-03	1.532E-03	1.11E-03	1.23E-03
	n = 60	6.88E-04	6.17E-04	*6.10E-04*	1.55E-03	1.11E-03	1.18E-03	1.53E-03	1.11E-03	1.23E-03
N = 500	n = 15	Inf.	Inf.	1.60E-03	Inf.	Inf.	**1.54E-03**	Inf.	Inf.	1.62E-03
	n = 30	9.65E-04	7.98E-04	8.41E-04	1.16E-03	**7.79E-04**	8.06E-04	1.16E-03	7.79E-04	8.15E-04
	n = 60	4.59E-04	4.63E-04	*4.40E-04*	4.99E-04	4.56E-04	4.54E-04	5.02E-04	4.47E-04	4.43E-04
N = 1000	n = 15	Inf.	Inf.	1.99E-03	Inf.	Inf.	**1.84E-03**	Inf.	Inf.	1.87E-03
	n = 30	**6.65E-04**	7.76E-04	8.90E-04	8.03E-04	7.44E-04	7.87E-04	8.23E-04	7.32E-04	8.14E-04
	n = 60	*3.01E-04*	3.14E-04	3.04E-04	8.03E-04	7.44E-04	7.87E-04	8.23E-04	7.32E-04	8.14E-04

Table 2.4. RMSE values for all allele dropout filtering scenarios using fastsimcoal2 under a stable population ($\lambda = 1.0$). Bold values identify the lowest RMSE for a particular combination of population size (N) and individual sampling level (n). Bold and italicized values identify the lowest RMSE for a particular population size.

		fsc2 complete	fsc2 RAD mutation	fsc2 RAD mutation & coverage
N = 250	n = 15	**2.54E-04**	8.74E-04	8.69E-04
	n = 30	**1.82E-04**	9.96E-04	1.04E-03
	n = 60	***1.40E-04***	7.80E-04	2.90E-03
N = 500	n = 15	**2.68E-04**	3.53E-04	3.58E-04
	n = 30	**1.59E-04**	4.90E-04	4.54E-04
	n = 60	***1.06E-04***	4.87E-04	1.17E-03
N = 1000	n = 15	4.42E-04	1.76E-04	**1.21E-04**
	n = 30	2.11E-04	2.56E-04	**1.94E-04**
	n = 60	***1.01E-04***	2.61E-04	3.16E-04

Table 2.5. Number of times that a declining population trend was correctly identified out of 100 replicate runs for LD-based analysis in NeEstimator under a declining population model ($\lambda = 0.9$). Results are presented for increasing intervals of time and by the combination of population size (N), individual sampling level (n), and minor allele frequency cutoff (MAF_{cut}). Results are based on the analysis of datasets with 10% missing data.

		$t_0 - t_5$ (& MAF_{cut} level)			$t_0 - t_{10}$ (& MAF_{cut} level)			$t_0 - t_{15}$ (& MAF_{cut} level)			$t_0 - t_{20}$ (& MAF_{cut} level)		
		0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05
N = 250	n = 15	49	49	68	63	63	83	71	71	95	72	72	99
	n = 30	86	79	79	99	94	93	100	100	100	-	-	-
	n = 60	94	94	94	100	100	100	-	-	-	-	-	-
N = 500	n = 15	8	8	48	18	18	67	25	25	79	29	29	85
	n = 30	79	73	72	94	84	86	100	99	98	100	99	99
	n = 60	80	81	84	97	98	98	100	100	100	-	-	-
N = 1000	n = 15	0	0	35	0	0	40	2	2	53	2	2	64
	n = 30	57	65	66	70	80	79	76	91	91	78	97	97
	n = 60	71	73	74	96	95	96	100	100	100	100	100	100

For some parameter combinations, there were insufficient numbers of individuals for target n (-).

Table 2.6. Number of times that a population trend was incorrectly identified as declining out of 100 replicate runs for LD-based analysis in NeEstimator under a stable population model ($\lambda = 1.0$). Results are presented for increasing intervals of time and by combination of population size (N), individual sampling level (n), and minor allele frequency cutoff (MAF_{cut}). Results are based on the analysis of datasets with 10% missing data.

		$t_0 - t_5$ (& MAF_{cut} level)			$t_0 - t_{10}$ (& MAF_{cut} level)			$t_0 - t_{15}$ (& MAF_{cut} level)			$t_0 - t_{20}$ (& MAF_{cut} level)		
		0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05
N = 250	n = 15	26	26	38	24	24	29	24	24	24	20	20	10
	n = 30	29	27	25	20	13	12	15	9	8	3	1	1
	n = 60	8	9	8	1	1	1	1	1	1	0	0	0
N = 500	n = 15	1	1	38	0	0	32	2	2	29	3	3	23
	n = 30	47	41	43	36	21	22	29	16	15	20	8	8
	n = 60	26	24	27	6	6	6	1	1	1	0	0	0
N = 1000	n = 15	1	1	20	0	0	12	0	0	10	0	0	9
	n = 30	16	36	37	11	30	28	16	27	27	7	18	18
	n = 60	21	18	25	14	13	12	4	4	4	1	1	0
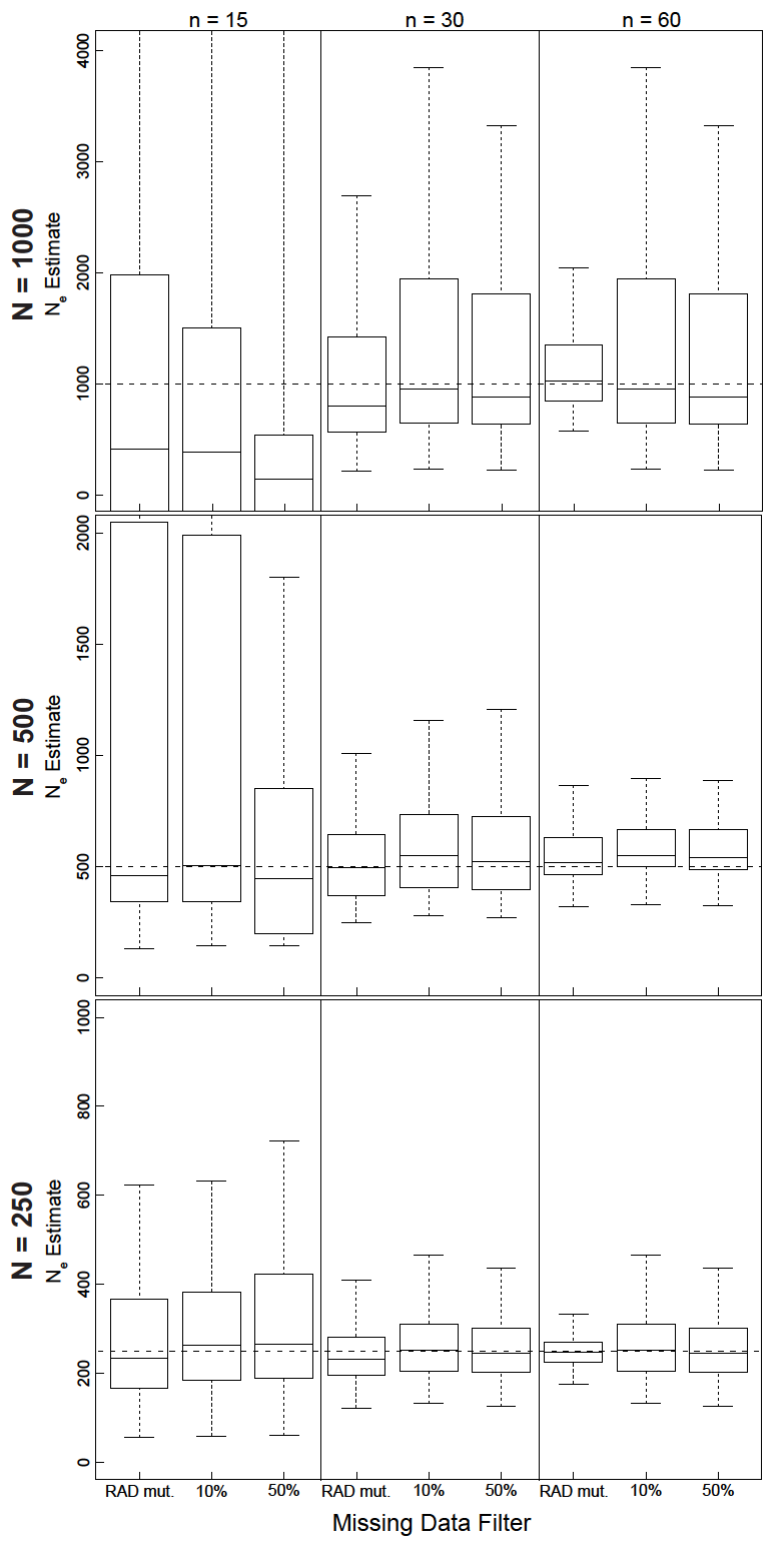
Figure 2.1. Boxplots of the distribution of $\widehat{N}_e$ estimates from 100 replicate simulations for LD-based estimation at generation 20 from temporal simulations under stable population sizes ($\lambda = 1.0$) with a MAFcut = 0.05. Dashed lines represent true N_e for the three population size models (1000, 500, and 250). Different missing data filtering strategies are shown at the bottom of the figure. The number of individuals sampled is shown at the top.
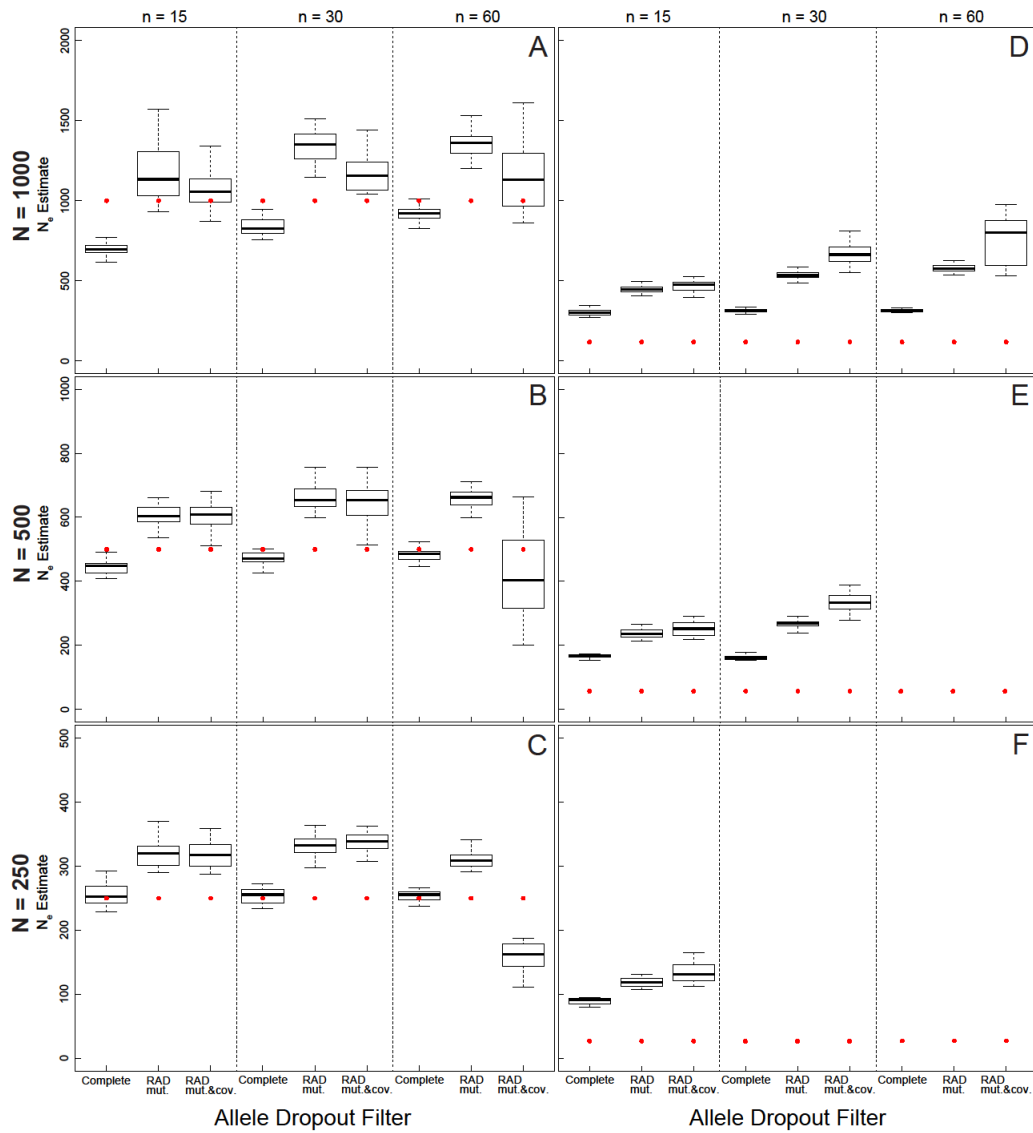
Figure 2.2. Boxplots of the distribution of fastsimcoal2 estimates of $\widehat{N}_e$ at t20 from 40 replicate temporal simulations. (A-C) $\widehat{N}_e$ estimates under a stable population size ($\lambda = 1.0$). (D-F) $\widehat{N}_e$ estimates under declining population size ($\lambda = 0.9$). Red dots represent the true N_e at t_{20}. Results are broken down across the number of individuals sampled (identified at the top of each panel) and the different data filtering strategies (identified at the bottom of each panel). Data filtering abbreviations are as follows: fsc2 complete dataset (Complete), fsc2 RAD mutation (RAD mut.),

and fsc2 RAD mutation and coverage (RAD mut. & cov.) For some parameter combinations,

there were insufficient numbers of individuals for target n.
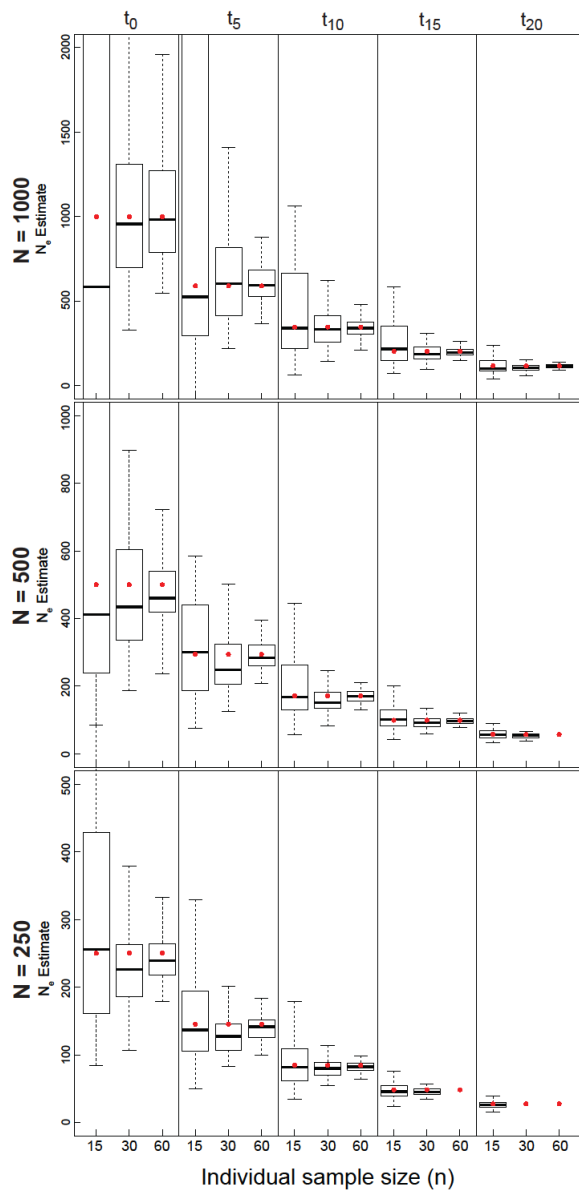
Figure 2.3. Boxplots of the distribution of point estimates from 100 replicate simulations for LD-based Ne estimation from five temporal sampling points (t0 - t20) under declining population growth model ($\lambda = 0.9$) using the 10% missing dataset and a MAFcut = 0.05. Red dots represent true N_e over time, starting from an initial N of 1000 (top), 500 (middle) or 250 (bottom). Results are also broken down across different levels of individual sample size (n = 15, 30, or 60). For some parameter combinations, there were insufficient numbers of individuals for target n.

CHAPTER THREE

GENOMIC DATA DETECT CORRESPONDING SIGNATURES OF POPULATION SIZE
CHANGE ON AN ECOLOGICAL SCALE IN TWO SALAMANDER SPECIES

Abstract

Understanding the demography of species over recent history (e.g., < 100 years) is

critical in studies of ecology and evolution, but records of population history are rarely available.

Surveying genetic variation is a potential alternative to census-based estimates of population

size, and can yield insight into the demography of a population. However, to assess the

performance of genetic methods it is important to compare their estimates of population history

to known demography. Here, I leveraged the exceptional resources from a wetland with 37 years

of amphibian mark-recapture data to study the utility of genetically-based demographic inference

on salamander species with documented population declines (*Ambystoma talpoideum*) and

expansions (*A. opacum*); patterns that have been shown to be correlated with changes in wetland

hydroperiod. I generated ddRAD data from two temporally sampled populations of *A. opacum*

(1993, 2013) and *A. talpoideum* (1984, 2011) and used coalescent-based demographic inference

to compare alternate evolutionary models. For both species, demographic model inference

supported population size changes that corroborated mark-recapture data. Parameter estimation

in *A. talpoideum* was robust to our variations in analytical approach, while estimates for *A.

opacum* were highly inconsistent, tempering my confidence in detecting a demographic trend in

this species. Overall, my robust results in *A. talpoideum* suggest that genome-based demographic

inference has utility on an ecological scale, but researchers should also be cognizant that these

methods may not work in all systems and evolutionary scenarios. Demographic inference may be

an important tool for population monitoring and conservation management planning.

Introduction

Detecting shifts in demography within populations over recent time scales (e.g., < 100 years) is important for understanding evolutionary responses to intrinsic and extrinsic factors. This knowledge is essential to identify the biological, ecological, or environmental drivers of population-size change, and to devise informed conservation management plans when needed. Surveying genetic variation can potentially provide an effective alternative to field-intensive census-based estimates of population size, yielding insight into the demographic history of a population, including migration events, population structure, expansions, and bottlenecks (Gutenkunst *et al.* 2009; Schwartz *et al.* 2007; Steiner *et al.* 2013). However, most studies incorporating genetically based demographic inference have estimated events at historical time scales (i.e., thousands of generations) (Lanier *et al.* 2015; Moura *et al.* 2014; Papadopoulou & Knowles 2015), and not at an ecological time scale (tens of generations or fewer) (Lozier 2014; McCoy *et al.* 2014).

Understanding demographic changes at an ecological time scale is particularly important in the conservation of species impacted by recent climate and environmental change. Climate change is predicted to have devastating impacts on ecosystems and communities in the upcoming century, affecting population demography and community dynamics (Palut & Canziani 2007). An important step towards understanding future impacts of climate change on species is assessing current population responses to environmental variables (Blois *et al.* 2013; Lanier *et al.* 2015). Demography and genetic diversity are essential parameters in population health and viability, and the surveying of genetic variation combined with new demographic inference techniques can potentially yield important insights in rapidly changing ecosystems (Shafer *et al.* 2015).

However, there is a lack of research validating genetically-based demographic parameter estimates with detailed census-based demographic information at an ecological time scale (but see (McCoy *et al.* 2014). Previous studies attempting to estimate recent demographic history with genetic data have been limited by the lack of temporally sampled individuals or small sample sizes, the sole use of mitochondrial DNA data or small numbers of microsatellite loci, or the indirect validation of demography using fossil evidence or simulations (Glenn *et al.* 1999; Weber *et al.* 2000). The combination of advances in sequencing technology producing large genomic datasets and the development of parameterized demographic inference models may now provide the opportunity to accurately estimate demographic parameters with smaller sample sizes and degraded museum samples (Bi *et al.* 2013; McCoy *et al.* 2014; Robinson *et al.* 2014). At the leading edge of these new opportunities is the need to directly validate the use of genomically-based demographic parameter estimation with known demography. However, comprehensive historic population census data are rarely available, especially in combination with documented ecological change.

Rainbow Bay (RB), an ephemeral wetland in south central South Carolina (Figure 3.1), provides a unique opportunity to assess the accuracy of genetically-based demographic inferences in comparison to results based on long-term capture-mark-recapture data (CMR). Rainbow Bay was completely encircled by a drift fence in 1978, with amphibians and reptiles entering and leaving the wetland censused daily since initiation. This continuous sampling has provided a valuable time series over 37 years that has been used to document recent population expansion and decline events for multiple amphibian species. Furthermore, climatological and ecological conditions have changed significantly across this time period and have been shown to be correlated with the changing amphibian community dynamics at RB (Todd *et al.* 2010). This

extensive knowledge base and the resources available at RB provide an ideal opportunity to assess the ability of genetically-based demographic inference to provide informative estimates of population demographic shifts over a very recent time scale.

Here, I focus on two salamander species exhibiting contrasting population growth trends based on CMR census data: the marbled salamander (*Ambystoma opacum*), which colonized the RB wetland in 1980 and expanded steadily since then, and the mole salamander (*A. talpoideum*), which had an established large population (>1000 females) at RB when monitoring began, and has steadily declined towards local extinction over the 37-year period (Daszak *et al.* 2005). These population trajectories at RB have been correlated with increased drought and shortened hydroperiod — the length of time a wetland holds water — and they illustrate the rapid response of populations to environmental change (Daszak *et al.* 2005). Genetic samples have also been collected over this time period, and offer a unique opportunity to examine temporal genetic diversity. Previous work at RB employing temporal genetic sampling and microsatellite genotyping to estimate effective population size (N_e) and potential demographic changes for these two species failed to detect any correlation between genetic summary statistics and CMR-based census-size estimates (Nunziata *et al.* 2015). The increased power offered by large numbers of single nucleotide polymorphism (SNP) loci coupled with recently developed demographic inference methods may yield greater power and precision to detect these population trends at RB (Morin *et al.* 2009; Smouse 2010).

The goal of this study was to compare observed climate-driven demographic trends to those inferred from analyses of genome-scale SNP data. I leveraged the multi-decade CMR data at RB and genetic samples from two temporally spaced sampling years for both *A. opacum* and *A. talpoideum* to address the following questions: (1) can recent population growth or decline be

55

detected using genome-wide SNP data and coalescent-based demographic inference, and (2) do non-temporally sampled data have similar power as temporal samples to detect recent population trends. My study complements a larger body of work at this long-term study site, which has provided unique insight into amphibian population dynamics and their response to a changing climate (Daszak *et al.* 2005; Pechmann *et al.* 1991; Todd *et al.* 2010). Results of this study are informative in elucidating experimental design and considerations when applying demographic inference to non-model species of conservation concern, which may show rapid population-level impacts from climate or environmental change.

Methods

Population sampling and molecular methods

Rainbow Bay is a seasonal wetland on the US Department of Energy's 780-km^2 Savannah River Site on the Upper Coastal Plain of South Carolina (Figure 3.1). It is completely encircled by a drift fence with pitfall traps, and has been censused daily for amphibians entering and leaving since the fall of 1978 (Gibbons & Semlitsch 1981; Pechmann *et al.* 1991). Ambystomatid salamanders exhibit geographic structuring associated with wetlands, with breeding site fidelity (Gamble *et al.* 2007). This, coupled with a lack of detected population structure for both species at RB using genetic data (see Appendix B), and detection of strong genetic structure between *A. opacum* at RB with a neighboring wetland (Nunziata *et al.* 2015), leads us to consider each species as a single breeding population.

I sampled tissue from twelve *A. opacum* in both 1993 and 2013, and 24 *A. talpoideum* in both 1984 and 2011. Although sample size is lower for *A. opacum*, a large number of SNP markers can provide insight into population history with limited individuals (Brumfield *et al.*

56

2003; Lanier *et al.* 2015; Moura *et al.* 2014). All sampled individuals were metamorphic

juveniles collected at the drift fence as they exited the wetland. For both species, tissues from the

first sampling years were taken from individuals found dead in pitfall traps that were preserved

as whole animals at -70°C. For both species, samples from the later sampling years were taken

from live individuals collected in traps and tail-clipped prior to release.

Whole genomic DNA was extracted using a Qiagen® DNEasy Blood and Tissue Kit,

following protocols recommended by the manufacturers. We followed the ddRAD library

preparation protocol outlined in (Peterson *et al.* 2012), a method that has not yet been performed

in large-genome species. Ambystomatid salamanders have genomes ~10x larger than the human

genome, with recent estimates at ~32 Gb (Keinath *et al.* 2015). To account for the large genome

size, the ddRAD protocol was modified from the suggested initial amount of 200-500 ng of

genomic DNA per individual to instead begin with 3 µg of genomic DNA per individual. Briefly,

individual in-line and indexing Illumina sequencing oligonucleotides were ligated onto DNA

fragments following restriction digest with *Eco*RI and *Sph*I (Peterson *et al.* 2012). After

individual library preparation, all *A. opacum* samples from both years were pooled, while two

library pools were prepared for *A. talpoideum*, one for each sampling year. All three library pools

were size-selected between 338 and 414 base pairs using a Pippin Prep (Sage Science) machine,

and then amplified using High-Fidelity DNA Polymerase (Bio-Rad) with 12 cycles. The three

pooled ddRAD libraries were sequenced separately on three lanes of an Illumina HiSeq 2500 at

Florida State University's College of Medicine using 150 base-pair paired-end reads. My first

round of sequencing was performed on the *A. opacum* libraries and to increase sequence

diversity in the initial portions of sequence reads we used an additional 30% spike in of PhiX

control library. Subsequent to the *A. opacum* sequencing, a 1% PhiX spike in was found to be sufficient, and this level was used for both lanes of *A. talpoideum* sequencing.

Quality filtering and variant detection

　　　Sequences were subjected to standard Illumina chastity filtering and then assigned to individuals based on index sequences using STACKS v1.21(Catchen *et al.* 2013; Catchen *et al.* 2011). Reads were trimmed to remove adaptors and restriction enzyme recognition sites. STACKS was used to identify RAD loci and call SNPs in forward reads only. I set the minimum depth of identical reads to 4 (m = 4), with additional aligned reads having a maximum number of 4 mismatches (M = 4), and we allowed for a maximum of 15% missing data (across individuals) per locus. To account for potential paralogs, I filtered for highly repetitive stacks that exceeded the expected number of reads given the average depth of coverage (-t option in STACKS). To avoid inclusion of linked non-independent SNPs, only a single SNP was allowed per RAD locus.

　　　Exclusion of missing data can potentially bias demographic inference (Huang & Knowles 2014); however, alleles present at very low frequencies could be the result of sequencing error. Therefore, I analyzed nuclear diversity under different filtering conditions to investigate the sensitivity of our results using an approach similar to (Lozier 2014). The following data filtering scenarios were performed: (1) all loci with a maximum of 15% missing data per locus (as discussed above), (2) all loci with a maximum of 5% missing data per locus, (3) all loci with a minor allele frequency cutoff of 0.05 and a maximum of 5% missing data per locus, and (4) exclusion of the two individuals from each year with the lowest mean coverage per RAD locus, a minor allele frequency cutoff of 0.05, and a maximum of 5% missing data per locus (Table 1).

For all filtering scenarios, I ensured that the two sample years were directly comparable by retaining only sites that passed all filter parameters for both sampled years.

Genotype-based population genetic analyses

For both species, nucleotide diversity estimates were calculated separately for each sampled year and were based on individual SNPs. I estimated observed heterozygosity (Het_{obs}) and Wright's F-statistic (F_{IS}), as calculated in the POPULATIONS program in STACKS. To examine differentiation among sampled years, I also used POPULATIONS to calculate pairwise F_{ST} between years. Nucleotide diversity (π) estimates were generated using the program vcftools v0.1.12b (Danecek *et al.* 2011), and confidence intervals (CIs) were generated by bootstrapping per-SNP π estimates 10,000 times in the R package boot v1.3 (Canty & Ripley 2012). To test for a genomic signal of population expansion or decline, I calculated Tajima's D in the program $\partial a \partial i$ v1.6.3 (Gutenkunst *et al.* 2009). Estimates of N_e were generated for each year using the linkage disequilibrium method implemented in N_eEstimator v2.01(Do *et al.* 2014). This method measures the deviation from the expected genotype frequency based on allele frequencies in the population, which increases in small populations due to drift (Hill 1981).

SNP markers were screened for temporal outliers between the two sampling years for each species using the method of (Beaumont & Balding 2004) implemented in BayeScan v2.1 (Foll & Gaggiotti 2008). This method measures the discord between global and population-specific, or in this case, year-specific, allele frequencies. Although this method was designed to identify outlier loci between populations, the method has proved useful for detecting outliers between temporal samples (Therkildsen *et al.* 2013) (See Methods and Results B.1 for detailed Methods and Results).

59

Genotype-free population genetic analyses

Sequencing error and low-coverage sequences can cause incorrect genotyping when heterozygotes are incorrectly inferred as homozygotes, and vice-versa (Johnson & Slatkin 2008). Genotyping error causes biased estimates of allele frequencies, and therefore a misrepresentation of the site frequency spectrum (SFS). To minimize these biases and serve as a complementary analysis to STACKS genotyping, I employed a likelihood-based variant detection method in ANGSD v0.700 (Korneliussen *et al.* 2014; Nielsen *et al.* 2012). The variant detection method estimates genotype likelihoods — which account for sequencing error, coverage, and alignment quality — and can be used directly in analyses instead of SNP calls. Briefly, I used the STACKS-based cstacks program to generate a consensus sequence from the highest coverage individual for both species (*A. opacum:* 1993_3 = 37.18x mean coverage; *A. talpoideum:* 2011_7 = 49.27x mean coverage). These were used as 'RAD reference' genomes in BWA v0.7.10 (Li & Durbin 2009) to create alignments for other individuals. Year-specific maximum-likelihood SFSs were estimated for both species using genotype-likelihood estimation (-GL 1) in ANGSD, excluding loci missing from more than one individual and allowing a minimum base quality score of 20. The genotype-free SFS was optimized with an Expectation Maximization (EM) approach, and included only sites that were present in both sampled years so that the data sets were comparable. I calculated genotype-free genetic diversity statistics directly using the likelihood-estimated SFS, including Watterson's theta (θ), π, and Tajima's D.

Demographic modeling

The influence of recent population size changes on SNP diversity was investigated with a simulation-based approach using the genotype-based SFS, implemented in fastsimcoal2 (fsc2) (Excoffier *et al.* 2013). For both species, I generated the observed folded joint-SFS using SNP data assembled with data filter 1 discussed above, and excluding monomorphic sites (see "*removeZeroSFS*" option in fsc2) and any outlier loci detected with BayeScan. In order to reduce bias with allele frequency estimates, I removed all missing data and included loci found in a minimum of 10 individuals per year for *A. opacum* and 18 individuals per year for *A. talpoideum* to maximize the number of SNPs and individuals included in SFS with complete data. I simulated SFS for six basic demographic models (Figure 3.2), including a model based on species-specific demographic history as estimated through CMR data (Model 1b for both species), and fit these simulated SFS to our observed SFS to generate the likelihood of the data under each demographic model. The models tested included three major models, each with two submodels: (1) RB was founded by a source population and subsequently underwent exponential population growth (positive or negative) between founding and the first sampling event (N_e^{1984} or N_e^{1993}), and between both sampling events using the following equation, where N is current N_e, N_0 is historical N_e, r is growth rate, and t is time in generations:

$$N = N_0 e^{rt}$$

with (1b) or without (1a) migration with the source population, (2) an ancestral population founded RB and another source population with no subsequent exponential growth with (2b) or without (2a) migration with the source population, and (3) RB is a long-standing population that underwent a bottleneck in the past and has subsequently experienced population growth as estimated by the equation above (3a) or did not experience exponential growth (3b). I include an

unsampled source population as a source of migrations with RB, which if unaccounted for can lead to spurious signatures of demographic events (Excoffier *et al.* 2013; Malaspinas *et al.* 2016). Although these models do not encompass all possible evolutionary scenarios, they were chosen to represent the range of likely alternate evolutionary models that potentially underlie the history of these two species at RB.

Parameters estimated from models included N_e of the sister population to RB (N_e Source), N_e at the time of sampling (*A. opacum*: N_e^{1993} and N_e^{2013}; *A. talpoideum:* N_e^{1984} and N_e^{2011}), immigration rates from the sister population (MIG), and the time of initiation of population size change (T_{DIV}). I assumed a generation time of two years for both species when converting estimates to years, based on age at first reproduction (Scott 1994). Defined parameter ranges were uniformly distributed with N_e ranging from 10 to 10,000 and T from 10 to 10,000. A total of 100,000 simulations were performed to estimate the SFS, with a minimum and maximum of 10 and 100 loops (ECM cycles), respectively. The stopping criterion, defined as the minimum relative difference in parameters between two iterations, was set to 0.001. A total of 50 replicate runs were performed per model and the overall maximum likelihood (ML) was retained. The relative likelihood across compared models was generated using Akaike Information Criteria (AIC) as outlined in (Excoffier *et al.* 2013). Confidence intervals for parameters in the best-supported model were obtained with a parametric bootstrap approach by simulating 100 SFS from the ML point estimates.

For each model, because I did not include monomorphic sites in the SFS or mutation rate, I fixed the time of historical sampling with other parameters estimated directly from the SFS. To assess bias in scaling estimates from this parameter, I re-estimated parameters for the ML model by fixing the current N_e (*A. opacum*: N_e^{2013}; *A. talpoideum:* N_e^{2011}) using estimates from

extensive CMR data detailed in (Nunziata *et al.* 2015). Briefly, the effective number of breeders

(N_b) was calculated for each year using the equation $N_b = 4N_mN_f/(N_m+N_f)$, where N_m and N_f are

the annual census estimates of adult males and females, respectively (Hedrick 2011). The

contemporary N_e was estimated as the harmonic mean of the single-generation N_e estimates from

each year since 1978 (estimated as 117 for *A.opacum* in 2013, 154 for *A. talpoideum* in 2011).

To assess the impact of non-temporal sampling on model choice and parameter estimation, I also

performed demographic inference for *A. talpoideum* using an SFS constructed exclusively from

2011 SNP data with N_e^{2011} fixed from CMR data. I performed initial demographic inference with

single-season 2013 data for *A. opacum,* but results were largely inconsistent between model runs,

and I do not present these results.

 To test the impact of reduced sample size in *A. opacum* on parameter estimates from the

SFS, I constructed the SFS for *A. talpoideum* using 12 randomly chosen individuals from each

year using criteria outlined above for *A. opacum*. I then re-estimated parameters for the ML

model using this reduced SFS. To estimate recent population size trends, I calculated lambda (λ)

as the slope of a linear regression of the log transformation of the current and historical N_e

estimates from the best-fitting demographic model. I compared these to empirical estimates of λ,

calculated using estimates of N_b from census data for genetic sampling years, as calculated in

Nunziata *et al.* (2015).

Results

ddRADseq data

 Total reads generated for each of the three library pools ranged from 180 to 270 million

reads (Table B.1). Initial processing of all individuals resulted in 538,628 ddRAD loci with

230,986 SNPs for *A. opacum*, and 845,433 ddRAD loci with 585,025 SNPs for *A. talpoideum*. Using default filtering (filter 1), the mean depth of sequencing coverage was 20.11 (5.67 SD) for *A. opacum*, and 28.48 (8.27 SD) for *A. talpoideum* (Table B.1). No outlier SNPs were detected for either species consistently across runs using BayeScan (Methods and Results B.1).

Genotype-based diversity statistics

Genetic diversity estimates were similar between years for both species (Table 3.1). Estimates of π in *A. opacum* had CIs that were largely overlapping between years and were similar for all filtering scenarios (Table 3.1). Estimates of π in *A. talpoideum* were slightly higher in 2011 than 1984 (Table 3.1). Between species, π was slightly higher in *A. opacum* for all filtering scenarios and in both sampled years, which was driven by a larger proportion of intermediate frequency SNPs in *A. opacum* (Figure B.1). When a minor allele frequency cutoff was applied (Filter 3), π estimates increased for both species due to the exclusion of rare alleles from calculations. Both species had minimal changes in genomic diversity over the studied time spans, as revealed by Tajima's D estimates (Table B.3) and between-year F_{ST} (0.029 and 0.036 for *A. opacum* and *A. talpoideum,* respectively). For both species, estimates of N_e using NeEstimator reached infinity for each year, indicating that the signal of drift cannot be distinguished from sampling error to accurately generate N_e estimates (Waples & Do 2010).

Genotype-free diversity statistics

After quality filtering, *A. opacum* had a total of 236,048 ddRAD loci and *A. talpoideum* had a total of 224,757 ddRAD loci. Site frequency spectra were correlated between years, and were similar in appearance to the genotype-based SFS, although they included a lower number of

rare alleles (Figure B.2). For *A. opacum,* there was a broader distribution of per site θ estimates in 1993 than in 2013 (Figure 3.3a). Mean π was 2.51E-07 and 1.19E-05 in 1993 and 2013, and the distribution of per site π estimates were largely overlapping between years (Figure 3.3a). For *A. talpoideum,* mean π was 4.82E-14 and 7.46E-15 in 1984 and 2011, respectively, and distributions of per site θ and π were largely overlapping between years (Figure 3.3b). These results also indicated slightly higher nucleotide diversity in *A. opacum* than in *A. talpoideum.* These and Tajima's D estimates (Table B.3) were similar to genotype-based analyses and collectively indicate minimal changes in genomic diversity over these time spans. Differences in Tajima's D observed between genotype-based and genotype-free calculations were likely due to differences in filtering between datasets, with no individual calculation indicating a significant departure from neutral expectations (i.e., < -2 or > 2).

Demographic modeling

For *A. opacum,* analyses of temporally-sampled data supported Model 1a (Figure 3.2) as the best fitting model with a relative likelihood of 0.961 (Table B.4), which is consistent with the interpretation that RB was founded by a source population, and has undergone expansion until the present (Tables 3.2 and B.5). However, confidence intervals for estimates of N_e^{1993} and N_e^{2013} were broad (Table 3.2), and the range of parametric bootstraps were overlapping between years (Figure 3.4). When N_e^{2013} was fixed using a CMR-based estimate, ML estimates of N_e differed by an order of magnitude from unconstrained estimates and were not within the 95% CIs (Table 3.2).

For *A. talpoideum,* analyses of temporally-sampled data supported Model 1b (Figure 3.2) as the best fitting model with a relative likelihood of 1.0 (Table B.6), consistent with the

interpretation that RB was founded by a source population and has undergone population decline, with asymmetrical migration with the source population (Tables 3.2 and B.7). Confidence intervals for all N_e estimates encompassed point estimates, and alternate analyses using either a fixed N_e^{2011}, or a reduced-sample SFS (n = 12 individuals per year) provided N_e estimates on the same order of magnitude as other estimates (Table 3.2). The range of N_e estimates from parametric bootstraps was non-overlapping between sample years (Figure 3.4). Confidence intervals for migration estimates were broad and varied among the alternate methods of parameter estimation (i.e., using a fixed N_e^{2011} and using a reduced-size SFS; Table 2.2). The best fitting model for *A. talpoideum* using the single-year SFS from 2011 supported RB being founded by a source population with subsequent population decline (Model 1a; Tables B.8 and B.9).

In both species, estimates of λ using SFS-based genetic estimates and CMR estimates showed similar population size trajectories (Figure 3.5). *Ambystoma opacum* had genetic and CMR λ estimates of 0.0417 and 0.0277, respectively. *Ambystoma talpoideum* had genetic and CMR λ estimates of -0.0637 and -0.0481, respectively.

Discussion

I leveraged temporal genomic data from a well-studied natural wetland community with 37 years of CMR data to perform demographic inference on species with documented population declines (*A. talpoideum*) and expansions (*A. opacum*) in response to changes in wetland hydroperiod. My results provide one of the first demonstrations that coalescent-based demographic model inference based on genome-wide SNP data can potentially be informative in detecting trends in population size change on an ecological time scale (McCoy *et al.* 2014). For

both species, model selection strongly supported histories of recent exponential population size change, with N_e estimates in *A. talpoideum* producing tight confidence intervals in both sampling years that identified a clear decrease in population size, consistent with well-documented trends from CMR data. In *A. opacum*, while the SNP-based point estimates of N_e produced a pattern of population growth that matched the known population expansion at RB, confidence intervals for N_e estimates were broad and overlapping, tempering our confidence in detecting a demographic trend in this species. Overall, I believe the strength of my model selection results, coupled with robust parameter estimation in *A. talpoideum*, point to the potential utility of demographic inference for examining population responses on an ecological scale. I caution, however, that this may not serve as a panacea for the study of all populations, as particular evolutionary scenarios and unknown factors may challenge the recovery of robust results. I further elaborate on these issues and provide basic pragmatic advice in the following sections.

Demographic modeling

For both species, SFS-based demographic inference strongly supported models of population size change, and produced similar population trajectories to census-based estimates (Figure 3.5). In general, a high degree of certainty was found in the N_e parameter estimates from *A. talpoideum* based on bootstrap CIs and these results were robust to the use of alternate inference procedures (models using fixed N_e vs. free N_e), all suggesting that the SNP data were informative in elucidating population trends on an ecological time scale. The robust results in *A. talpoideum* may reflect its lower genetic diversity and the progressively increasing bottleneck (based on field data), which may have led to more pronounced signals of recent allele frequency shifts, and more precise parameter estimates. In the only other study that we are aware of to

67

employ SNP data and coalescent-based demographic inference on an ecological time scale, McCoy et al. (2014) were able to detect a known bottleneck event in an introduced insect population, further suggesting that histories featuring a decrease in N_e may be particularly well recovered using these methods.

However, while my results point to the potential for genetically-based detection of general population trends on an ecological scale, I also note a number of challenges in the recovery of robust results. Most importantly, I highlight the difficulty here with parameter estimation. This was particularly true for estimates of migration in *A. talpoideum*, all of which produced high levels of uncertainty. It is also noteworthy that the best-fitting model for *A. opacum* did not reflect migration, which is apparent from CMR data and metapopulation structuring (Kinkead *et al.* 2007). This may be the result of sampling that was limited to a single population, without incorporating genetic data from potential source populations. Parameter estimation in *A. opacum* further demonstrates the difficulty in demographic inference under some evolutionary scenarios. Bootstrap CIs for *A. opacum* strongly suggested that all parameter estimates carry uncertainty from the SFS-based inference procedure. Across the six tested demographic models for *A. opacum,* N_e-based population trajectories alternated between expansion and decline. In currently large populations that have undergone a recent expansion, like *A. opacum,* large sample sizes are likely needed to observe the rare recent variants that will be needed for SFS-based signatures of population expansion (Fu 1997; Gravel *et al.* 2011; Robinson *et al.* 2014). This will be an issue with any demographic inference procedure employing coalescent modeling, where expansions cause an increase in deeper coalescent events in a genealogy (Kim *et al.* 2015). Ultimately, my detection of a population expansion in *A. opacum* may represent a chance correlation of genetic and census population size trajectory. It is

also possible the uncertainty in parameter estimates in *A. opacum* may be a result of poor model fit (Thomé & Carstens 2016), or that parameter estimates are reflecting patterns of an overall large and stable metapopulation.

Because a nuclear mutation rate is not known for our species (or salamanders in general), I did not include the mutation rate in models, and instead scaled free parameters by fixing the time of historical sampling, which may create a bias as a result of the small time frame between sampling events. To address this possible bias, I re-estimated parameters using the ML model and fixed N_e of the most recent sample year using estimates from CMR data. Scaling parameters from demographically estimated N_e also carries bias, as I am uncertain of the true N_e incorporating population structure, reproductive success, and clutch survival; but, this has been used in a similar demographic inference study without apparent bias (McCoy *et al.* 2014). I found largely different parameter estimates between alternate scaling procedures for *A. opacum,* but similar N_e parameter estimates for *A. talpoideum,* further suggesting uncertainty in *A. opacum* parameter estimates and relatively robust estimates in *A. talpoideum.* This indicates that scaling parameter estimates with temporal sampling may be a feasible option for non-model species to approximate population trajectory, but should be tested against alternate scaling strategies for robustness.

In most cases, researchers will have no *a priori* knowledge of the demographic history of a population to guide sampling designs, scale parameters, or determine the overall applicability of coalescent-based demographic inference on an ecological time scale. However, this procedure may still be of broad use in conservation studies. First, I note that many population genetic studies are focused on species of conservation concern, where populations are typically small and recently bottlenecked, or are known to be decreasing in size. These situations are the most

likely to create apparent signatures in the SFS that can be elucidated at an ecological time scale (as shown here in *A. talpoideum* and in McCoy *et al*., 2014). Ultimately, the application of coalescent modeling to detect recent population declines will depend on the length and severity of the bottleneck, as the rate of coalescence will vary in turn, with smaller population having more coalescent events in the recent past (Hein *et al.* 2004; Johnson *et al.* 2007). Second, researchers should be aware of the difficulties associated with these procedures, particularly when underlying evolutionary scenarios are likely to challenge their implementation. Demographic inference analyses should incorporate thorough attempts at parameter estimation, including the assessment of confidence intervals, using the robustness of their results as a proxy for the informativeness of their data. Finally, as many researchers will not have access to temporally sampled populations, I evaluated model choice and parameter estimates using only contemporary samples for *A. talpoideum.* This analysis resulted in an alternate choice of the best model (model 1a) than when using the joint-SFS, but still supported a decline in N_e from 1984 to 2011. However, relative likelihoods of tested models were much more evenly distributed across models using the single-SFS than when using the joint-SFS from both years, where likelihoods of poorer-fitting models were almost zero. These results demonstrate that the addition of temporal samples, even over a short number of generations, can greatly improve model choice, providing direct information on the magnitude of allele frequency shifts over the sampling time frame (Ramakrishnan *et al.* 2005). Museum tissue collections and personal collections may serve as unique resources for estimating current responses of species to climate change and reconstructing population history (Bi *et al.* 2013).

Genetic diversity

      Ambystoma opacum generally had higher genetic diversity than *A. talpoideum* using both genotype-based and genotype-free estimates, with non-overlapping CIs for estimates of nucleotide diversity for all filtering scenarios. Similar results were seen in the previous study employing microsatellite markers, providing additional support that these differences are real (Nunziata *et al.* 2015). Diversity differences might be due to a long-term decline of the *A. talpoideum* population at RB, as detected from field data and demographic inference. However, even historical samples of *A. talpoideum*, when population size was large, had lower genetic diversity than *A. opacum*. Overall genetic diversity differences between these species may be the result of lineage history; *Ambystoma opacum* is more broadly distributed than *A. talpoideum* (Petranka 1998), and may exhibit higher genetic diversity due to this broader range or differences in life-history characteristics such as metapopulation structure (Leffler *et al.* 2012; Romiguier *et al.* 2014).

      I observed minimal changes in temporal genomic diversity for both species with no correlation with demographic size change, highlighting the utility of demographic inference procedures to reveal patterns in genomic diversity that may not be apparent from summary statistics. This result is not surprising given the time-frame of our sampling, which was likely too short to produce large fluctuations in nucleotide diversity (Bi *et al.* 2013; Lozier 2014). Diversity results were largely similar in genotype-free analyses and across filtering scenarios suggesting biases in data filtering were minor. Similar to my previous study involving microsatellites (Nunziata *et al.* 2015), N_e estimates from linkage disequilibrium estimates reached infinity. Both of my study species likely exist in metapopulations with migration, which is not accounted for in the linkage disequilibrium model of N_e and may confound results (Waples & Do 2010).

Sequence-based demographic inference estimates parameters under more realistic population models, incorporating migration and variable population size, and are more likely to capture important information about population history relative to summary statistics (Drummond *et al.* 2005).

Conservation concerns

The N_e estimates for *A. talpoideum* using both demographic inference and CMR data are at the lower bound of suggested N_e for maintenance of evolutionary potential of a population (Franklin & Frankham 1998). However, RB likely exists in a larger metapopulation characterized by local extinction of some breeding wetlands and continued occupancy and population stability of others (Marsh & Trenham 2001). Because this study focused at a single breeding wetland, I cannot infer if the N_e of the entire metapopulation is in decline or if the observed N_e decline at RB is part of a source-sink dynamic that is typical of amphibian populations. Increased droughts at RB from 1978-2004 were negatively correlated with census size in *A. talpoideum,* a species adapted to wetlands with long hydroperiods (Daszak *et al.* 2005). Drought is possibly impacting surrounding wetlands similarly, causing additional local extinctions of *A. talpoideum* populations, and negatively impacting metapopulation persistence. (Walls *et al.* 2013b) found an increase in local extinction rate of *A. talpoideum* at several wetlands within the southeastern Coastal Plain, likely due to increased intensity of drought. The demographic inference procedure employed in fsc2 allowed me to model an unsampled source population representing the metapopulation of wetlands RB exists within. The results indicate an overall large and stable source population; however, this has not been validated with census or

genetic data. Future work will focus on sampling wetlands at multiple spatial scales surrounding RB and performing similar demographic inference.

ddRAD in large genome species

The generation of large-scale SNP datasets from ddRAD sequencing is relatively inexpensive, straight-forward, and has been validated in numerous studies on small-genome (<20 Gb) taxa (Lanier *et al.* 2015; Lozier 2014; Papadopoulou & Knowles 2015). However, this method has yet to be validated in species with larger genomes. In this study, I demonstrate the use of this procedure in two salamander species with ~32 Gb genomes, and show the scalability of the method with little knowledge of the genomic architecture of the study taxa. For both *A. opacum* and *A. talpoideum* I was able to generate thousands of SNP loci that passed filtering parameters, even with a stringent missing-data threshold of 95% of individuals across loci. The development of RADseq methods and their applicability to species with large and complex genomes is an important step forward in marker development that carries the advantages of *de novo* marker discovery and the removal of ascertainment bias in large-genome species where marker development has been formidable in the past (Garner 2002; Poland *et al.* 2012).

Table 3.1. Filter properties and genetic diversity statistics calculated for *A. opacum* and *A. talpoideum* (*A. talpoide*) juveniles at Rainbow Bay. Included are the total number of SNPs (N), the average observed heterozygosity per locus (Het$_{obs}$), Wright's inbreeding coefficient (F$_{IS}$), and the average nucleotide diversity (π) with 95% confidence intervals obtained by bootstrapping 10,000 times.

Species	Filter Set	Missing Data Allowed	Minimum MAF	Year	# of individuals	N	Het$_{obs}$	F$_{IS}$	π	π lower 95% CI	π upper 95% CI
A. opacum	1	15%	> 0	1993	12	40326	0.2458	0.0111	0.2492	0.2475	0.2509
A. opacum				2013	12	40326	0.2429	0.0172	0.2479	0.2462	0.2496
A. opacum	2	5%	> 0	1993	12	15740	0.2121	0.002	0.2117	0.2090	0.2144
A. opacum				2013	12	15740	0.2066	0.0083	0.2079	0.2052	0.2105
A. opacum	3	5%	> 0.05	1993	12	10555	0.2979	0.0045	0.2972	0.2943	0.3001
A. opacum				2013	12	10555	0.2926	0.0142	0.2945	0.2916	0.2974
A. opacum	4	5%	> 0	1993	10	24793	0.2373	-0.007	0.2339	0.2317	0.2361
A. opacum				2013	10	24793	0.2366	-0.004	0.2328	0.2306	0.2351
A. talpoide	1	15%	> 0	1984	24	45027	0.1718	0.0615	0.1905	0.1889	0.1921
A. talpoide				2011	24	45027	0.1886	0.0181	0.1949	0.1934	0.1965
A. talpoide	2	5%	> 0	1984	24	28644	0.1649	0.0306	0.1737	0.1717	0.1756
A. talpoide				2011	24	28644	0.1805	-0.002	0.1800	0.1781	0.1819
A. talpoide	3	5%	> 0.05	1984	24	23132	0.2497	0.0465	0.2633	0.2612	0.2654
A. talpoide				2011	24	23132	0.2693	-0.005	0.2676	0.2656	0.2696
A. talpoide	4	5%	> 0	1984	22	31497	0.1713	0.0396	0.1835	0.1816	0.1854
A. talpoide				2011	22	31497	0.1876	-0.003	0.1871	0.1853	0.1889

Table 3.2. Maximum-likelihood (ML) demographic parameter estimates and confidence intervals (CIs) for A.

talpoideum and A. opacum under the best-fitting model (*A. talpoideum:* Model1b; *A. opacum*: Model1a) from

fastsimcoal2. Parameter estimates from an alternate analysis that fixed N_e^{2011} and N_e^{2013} (in bold) and with *A.*

talpoideum SFS constructed from 12 individuals for each year, are also presented, but CIs are based off of

original ML estimates. CIs were obtained by parametric bootstrapping.

Species	Parameter	ML Estimate	95% CI-Lower	95% CI-Upper	ML N_e Fixed	ML n = 12	Unit
A. opacum	N_eSource	5,613	2,363	28,491	22,937	-	Individuals
A. opacum	N_e^{2013}	17,740	11,419	41,135	**117**	-	Individuals
A. opacum	N_e^{1993}	2,596	451	44,786	46,882	-	Individuals
A. opacum	N_e_Out	228	50	1,077	77	-	Individuals
A. opacum	T_{DIV}	416	262	610	412	-	Years before 2013
A. talpoideum	N_eSource	3,742	1,174	10,840	1,540	4,397	Individuals
A. talpoideum	N_e^{2011}	26	17	35	**154**	25	Individuals
A. talpoideum	N_e^{1984}	508	187	1,742	421	995	Individuals
A. talpoideum	N_e_Out	34,310	5,114	39,935	18,690	31,056	Individuals
A. talpoideum	T_{DIV}	4,806	2,206	5,724	4,212	4,734	Years before 2011
A. talpoideum	MIG into RB2011 from Source	2.54	0.00	281.34	0.03	0.66	Individuals/Generation
A. talpoideum	MIG into RB1984 from Source	1.28	0.00	23.77	0.35	0.31	Individuals/Generation
A. talpoideum	MIG into Source from RB2011	5.91	1.19	8.32	1.35	9.02	Individuals/Generation
A. talpoideum	MIG into Source from RB1984	2.51	0.04	15.00	0.68	0.86	Individuals/Generation
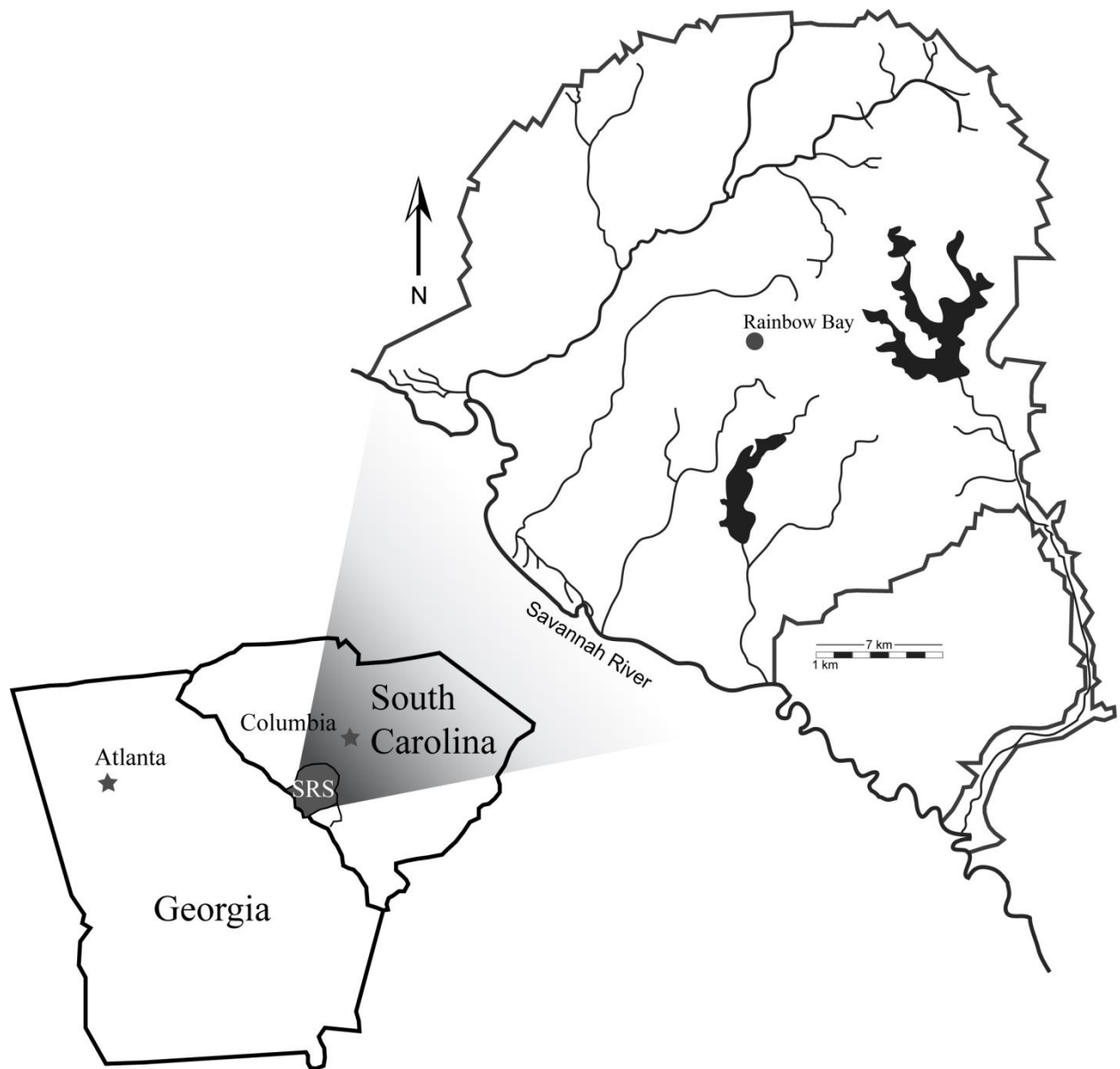
Figure 3.1. Map of the Rainbow Bay study site on the Savannah River Site (SRS).
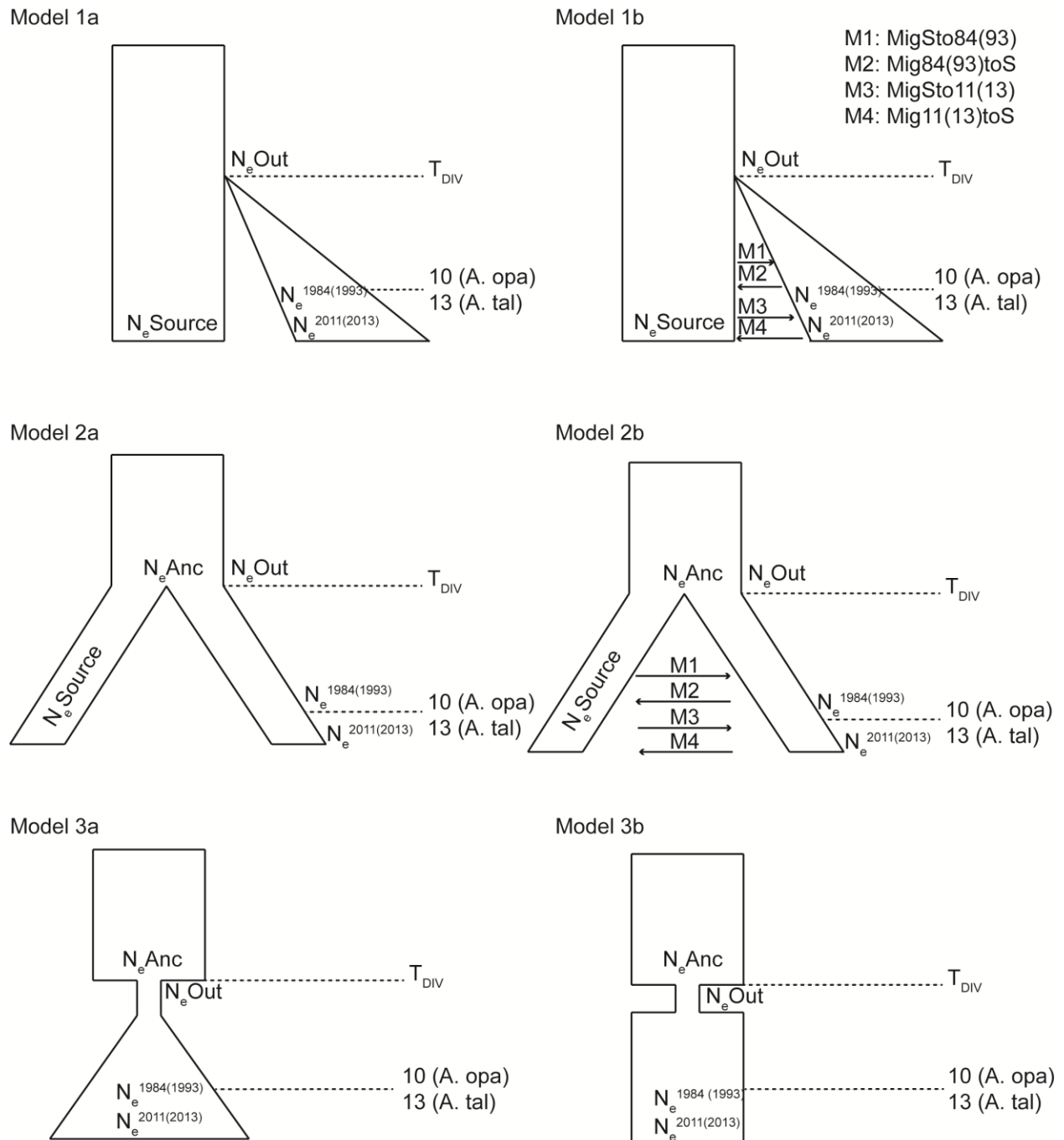
Figure 3.2. All tested demographic models for A. opacum (A. opa) and A. talpoideum (A. tal). (1a) Rainbow Bay (RB) was founded by a source population and subsequently underwent exponential population growth. (1b) RB was founded by a source population and subsequently underwent exponential population growth, with asymmetrical migration with its source. (2a) An ancestral population founded RB and a source population with population size remaining

constant. (2b) An ancestral population founded RB and a source population, with population size remaining constant, and asymmetrical migration with source. (3a) RB is a long-standing population that underwent a bottleneck in the past and has subsequently experienced population growth. (3b) RB is a long-standing population that underwent a bottleneck in the past and has subsequently maintained a stable population size. Parameters estimated from models included N_e of the sister population to RB (N_eSource), N_e at the time of sampling (*A. opacum*: N_e^{1993} and N_e^{2013}; *A. talpoideum:* N_e^{1984} and N_e^{2011}), migration rates (M1-M4), and the time of initiation of population size change (T_{DIV}).
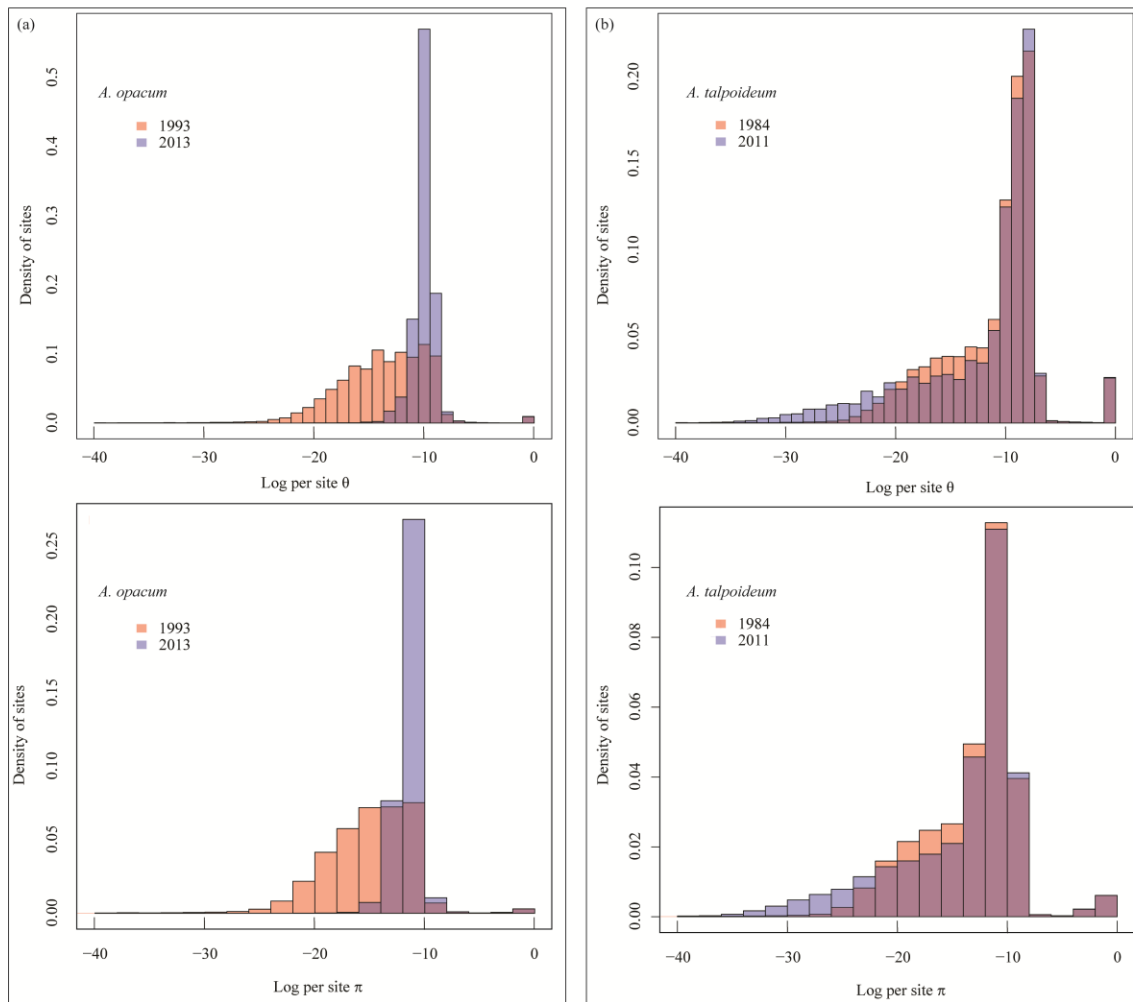
Figure 3.3. Distribution of genome-wide per-site average pairwise nucleotide diversity (π) and Watterson's theta (θ) for populations of (a) *A. opacum* from 1993 and 2013, and (b) *A. talpoideum* from 1984 and 2011 at Rainbow Bay. Historic samples are represented in pink, contemporary samples in blue, and the overlap between historic and contemporary samples in purple.
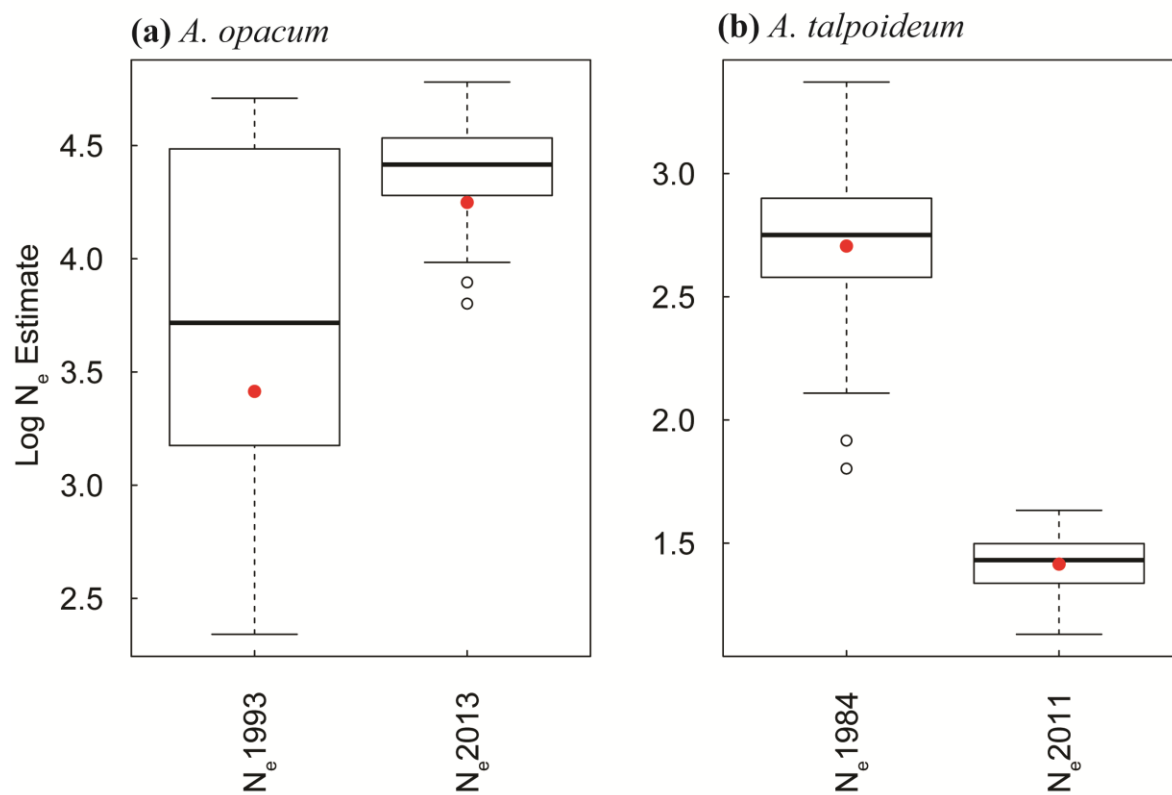
Figure 3.4. Boxplots showing the median, interquartile ranges, and overall ranges of log N_e estimates for 100 parametric bootstraps for (a) *A. opacum* and (b) *A. talpoideum*. Maximum likelihood point estimates under the best-fitting model (*A. talpoideum:* Model1b; *A. opacum*: Model1a) from fastsimcoal2 are shown as red dots.
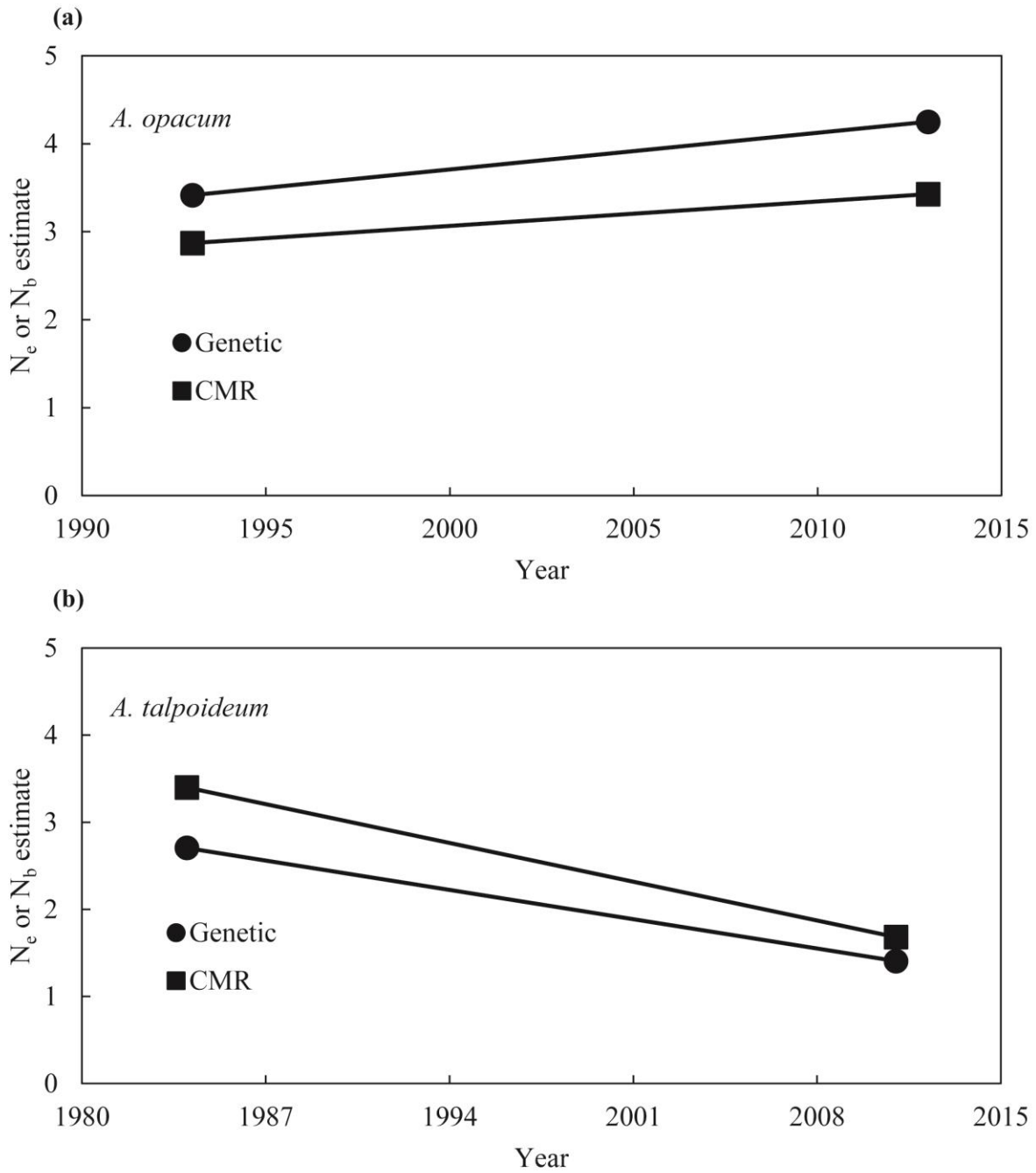
Figure 3.5. Comparisons of estimates of the log of effective number of breeder (N_b) as calculated from capture-mark-recapture (CMR) data (squares), and the log of N_e for the maximum likelihood model from fastsimcoal2 (circles) for (a) *A. opacum* and (b) *A. talpoideum* at Rainbow Bay. The slopes of these linear regressions were used as estimates of λ.

CHAPTER FOUR

USING GENTIC MONITORING TO UNDERSTAND EXTINCTION RISK OF POND-
BREEDING AMPHIBIANS IN THE FACE OF CLIMATE CHANGE

Abstract

Genetic approaches of estimating effective population size (N_e) and effective breeding

number (N_b) are useful in conservation, ecology, and evolution, providing information on

population abundance and demographic history, and insight into the evolutionary history and

genetic viability of populations. Although widely used in amphibian population genetic studies,

there is uncertainty regarding the relationship between N_b, N_e, and census size (N_C) in

iteroparous species (repeat breeding) with complex life-histories. In particular, little is known

about how catastrophic reproductive failure (CRF, i.e., complete mortality of all larval

individuals), a common life-history characteristic in many pond-breeding amphibians, impacts

the stability and level of genetic diversity, and the relation of effective sizes to census size. I used

individual-based population models of the pond-breeding salamander, *Ambystoma opacum,* with

life-history parameters estimated from a long-term dataset, and multiple levels of CRF (a proxy

for climate change) over a 50 year projection, to quantify annual $\widehat{N}_b$ estimated with both genetic

and demographic data, $\widehat{N}_e$, and N_C, and generate predictions about future genetic viability and

extinction risk. My results demonstrate that genetically estimated $\widehat{N}_b$ is positively correlated

with N_C across multiple levels of CRF in isolated and subdivided *A. opacum* populations, and

has the potential to be informative in detecting population declines in amphibians at a

contemporary time-scale. I found few correlations between $\widehat{N}_b$ and $\widehat{N}_e$, or between $\widehat{N}_b$ and N_C,

indicating that N_e while providing important information on genetic viability and evolutionary

potential of populations, may not be the best tool if quick detection of population size change is

the study goal. My models suggest that CRF causes significant extinction risk and reduced genetic diversity at 70% probability of CRF and above, and that including gene flow in models prevented extinctions and maintained genetic variability of populations across all levels of CRF. These results highlight the importance of metapopulation scale conservation of amphibian populations, and the potential of monitoring N_b for inferring demography of populations of conservation concern.

Introduction

Amphibian declines and extinctions have been a global concern over the past several decades, with many factors to blame, including disease, climate change, and habitat degradation/alteration (Blaustein *et al.* 1994; Collins & Storfer 2003). Declines continue to occur even in protected areas, indicating habitat protection is not sufficient to halt many of these declines (McCartney-Melstad & Shaffer 2015). It is therefore critical to understand the processes regulating amphibian populations, and identify factors important in the design of conservation plans (Semlitsch 2002).

Population models serve as excellent tools to investigate amphibian population dynamics and how they regulate population processes, including census size, population stability, genetic diversity, and extinction risk (Taylor *et al.* 2006; Weir *et al.* 2016). One factor that has received considerable attention in constructing amphibian population models is catastrophic reproductive failure (CRF), the complete mortality of larval individuals in a reproductive season. High variance in reproductive success between years is a common life-history characteristic across many pond-breeding amphibians, where juvenile recruitment alternates between years when tens of thousands of juveniles are produced and others when no juveniles are produced. Multiple factors can cause CRF, including predation by fish or other amphibians, and durations of hydroperiod, the length of time a wetland holds water (Alford & Richards 1999; Semlitsch *et al.* 1996). For this study, I focus on CRF as single season reproductive failure due to shortened hydroperiod. Multiple studies have found climatic changes and increased droughts correlated with shortened hydroperiods, increased probability of CRF, and declines in amphibian populations (Daszak *et al.* 2005; McMenamin *et al.* 2008; Walls *et al.* 2013a). As climate change

projections predict altered precipitation patterns across the United States (US), it is important to understand potential impacts on amphibian populations and associated extinction risk.

While simulation studies have examined the impact of CRF on amphibian extinction risk and population size, none have evaluated the impact on genetic diversity and effective population size (N_e). Effective population size is important in assessing genetic variability, and can possibly serve as an indicator of population size trends (Charlesworth 2009; Schwartz *et al.* 2007; Tallmon *et al.* 2010). It may also serve as an important indicator of population viability, reflecting the evolutionary potential and risk of inbreeding depression (Frankham 1995). A recent review article highlighted the utility of molecular ecology to gain insight into population size trends in amphibians (McCartney-Melstad & Shaffer 2015). Natural high variance in population size and recruitment success combined with often cryptic life-history characteristics in amphibians may make population declines difficult to detect without intense field studies (Pechmann & Scott 1991). Genomic tools may therefore serve as an important tool in amphibian monitoring and conservation biology.

It is unclear what impact temporally variable recruitment has on N_e and genetic diversity, but it has been hypothesized to lead to a naturally low N_e/N ratio (Funk *et al.* 1999; Rowe & Beebee 2004; Schmeller & Merilä 2007). Genetic estimates of N_e for amphibians are typically low (< 100), even in large, stable populations with N_e/N often below 0.1 (Phillipsen *et al.* 2011; Schmeller & Merilä 2007). This leads to a number of questions. What mechanisms are regulating genetic variation in amphibians? Perhaps more importantly, is genetic variation stable over time? Before genetic monitoring can be applied to estimate population size and time-course of decline in amphibians it is necessary to address these questions and evaluate if detection of overall

population growth trends is achievable with the complex demography of pond breeding amphibians.

Effective population size is notoriously difficult to estimate in natural iteroparous populations (repeat breeding), leading to much interest in estimating the effective number of breeders per generation (N_b)(Waples & Antao 2014). While N_e is dependent on the lifetime variance in reproductive success among all adults in a population, N_b is focused only on breeding adults in a single season. Estimating and interpreting effective size for a single season rather than the entire adult lifetime is understandably much more straightforward. Estimates of N_b may be of use in pond-breeding amphibians where collection of single-cohort DNA samples from larvae or metamorphs can be collected with minimal effort at breeding sites (Polich *et al.* 2013). However, the relationship between N_e and N_b is unclear for amphibians, as is their correlation with census size (N_c) (Waples *et al.* 2013).

In this study, I use simulated data for populations of a pond-breeding salamander, *Ambystoma opacum,* to examine the impacts of amphibian life-history characteristics on N_e and N_b, including increasing probability of CRF (a proxy for climate change). I use these models to understand null expectations of genetic diversity in isolated and subdivided populations and to provide insight into the usefulness of genetic monitoring to detect population size declines in amphibians. Specifically, I quantified correlations between effective sizes and census size, and estimated population viability under CRF over a 50 year projection.

Methods

Study organism

The marbled salamander, *Ambystoma opacum*, served as my study species because of the amount of long-term field and experimental data available, as well as previous modeling studies done on the species (Taylor & Scott 1997; Taylor *et al.* 2006). Marbled salamanders are a common species throughout the eastern US and like most other *Ambystoma* species they are explosive breeders that migrate from terrestrial overwintering habitat to wetland breeding sites (Petranka 1998). *Ambystoma opacum* females migrate to breeding wetlands in late summer or autumn before wetlands fill and deposit eggs terrestrially in the dried beds of wetlands or along the edges of partially filled wetlands. Females often remain with nests until wetlands are inundated with rain and eggs hatch (Petranka 1998). After a larval stage *A. opacum* metamorphose into terrestrial juveniles, with individuals mating annually after reproductive age/size is reached.

Population models

I conducted simulations of *A. opacum* populations using the Python program simuPOP v1.1.4 (Peng & Kimmel 2005), a forward-time and individual-based population genetic modeling program. Age-structured population models were constructed using age- and sex-specific birth and death rates estimated from published and unpublished demographic data from long-term study sites at the Savannah River Site in South Carolina, and from additional published studies on *A. opacum* (Table C.1). The model included 11 life-history stages: pre-adult (egg, larval, and juvenile stages), and ages 1 through 10 years. Survival to age 1 was based on a constant combined survival rate of egg and larval stages, and a function based on larval density (Lar_{surv})

outlined below, and as used in (Taylor *et al.* 2006). All individuals over age 1 were considered adults, however age at first reproduction was variable between individuals. Mating was random, with each female only having the opportunity to mate once per reproductive season, whereas males could mate multiple times and fertilize multiple clutches. Populations were constructed with an initial size of 60 adult individuals with an average sex ratio of 1. Initial diploid genotypes for each individual were simulated for 100 "microsatellite-like" markers, each with 10 possible alleles. Allele frequencies were generated from a Dirichelet distribution with a mean of eight alleles and were randomly assigned to individuals, with no mutation or selection. All simulated populations went through an equilibrium phase of 30 generations to establish the base population and reach Hardy-Weinberg equilibrium (Antao *et al.* 2011; Tallmon *et al.* 2010; Waples 2006), after which each replicate population evolved for an additional 50 years with a set probability of CRF. A total of 100 replicate populations were modeled for each demographic scenario.

I modeled the *A. opacum* population as a single isolated deme, and also as a subpopulation with the possibility for migration. Many amphibians, including *A. opacum*, form metapopulations across the landscape, with low levels of migration between subpopulations (Marsh & Trenham 2001). My migration model consisted of one additional breeding population initialized with the same population size and vital rates as the focal population, no CRF, and with equal probability of immigration and emigration between it and the focal populations. In *A. opacum* juveniles are an important dispersal stage, so I set this as the dispersing age with 5% probability of dispersal (Gamble *et al.* 2007).

Larval density dependence

Density-dependent effects from the larval environment are common in pond-breeding amphibians, and contribute to population dynamics well into adulthood (Taylor & Scott 1997). In *Ambystoma* species larval density has been correlated with larval survival, body size at metamorphosis, survival to first reproduction, and age at first reproduction (Scott 1990, 1994; Taylor & Scott 1997). I used the following model of density-dependent larval survival based on (Getz 1996):

$$Lar_{surv} = \frac{Lar_{max}\,\gamma^{\delta}}{\gamma^{\delta} + \left[\frac{Lar_{init}}{R}\right]^{\delta}},$$

where Lar_{surv} is estimated larval survival, Lar_{max} is maximum larval survival (Lar_{max} =0.8), Lar_{init} is the initial number of larvae, γ is the density where larval survival is half of maximum survival (γ =11), R is available resources (R=2), and δ is the shape parameter and defines the strength of density-dependent effects at increasing larval density (δ =1.8; Code supplied by Scott Weir). If larval density was high ($\gamma > 11$) minimum age at first reproduction was 2 years, at low larval density ($\gamma <= 11$) minimum age at first reproduction was 1 year, however age at first reproduction is variable between individuals (Table C.1). Because I am unsure of fecundity estimates for year one breeders from low larval densities, I also investigate how eliminating density-dependent age at first reproduction affected final census size and extinction risk.

Catastrophic reproductive failure

Environmental stochasticity was incorporated into models to mimic the relationship between spring rains and survival of larvae to metamorphosis. Wetland hydroperiod is important

for larval recruitment and can lead to CRF if wetlands do not fill or dry before metamorphosis occurs. The amount of time needed for metamorphosis is species-specific and will impact the probability of metamorphosis before pond drying (time needed for metamorphosis: *A. opacum*: 54 days; (Daszak *et al.* 2005)). I ran simulations for a range of CRF probabilities from 0-100%, in increments of 10%. This range encompasses realistic probabilities of CRF for ephemeral ponds across the southeastern US, as well as projected CRF probabilities into the future under increased drought.

Data analysis

I estimated genetic diversity each generation, using all loci and individuals to calculate expected heterozygosity (H_e) and allelic diversity (AD). To estimate extinction probability, defined as when the total adult population (N_C) reached zero individuals, I recorded N_C at the end of each breeding cycle. To estimate population size trends, I calculated lambda (λ), from the time when CRF was initiated to the end of the simulation ($\lambda = N_C\text{Gen}80/N_C\text{Gen}30$).

Estimating N_b and N_e. I estimated N_b per generation using two methods: (1) estimation of linkage disequilibrium (LD; $N_b\text{Gen}$), and (2) direct estimates with demographic data ($N_b\text{Dem}$). For LD-based estimation, I used the program NeEstimator v2.01 with random samples of 500 offspring every five generations without replacement; if 500 individuals were not available, the largest number of individuals available was collected (Do *et al.* 2014). Results are presented using P_{crit} of 0.05, as it has been found to provide balance between bias and precision with large samples sizes (Waples & Do 2010). I also used demographic data to directly calculated N_b per generation using the inbreeding effective size formula for species with two sexes (Crow & Denniston 1988):

$$N_b = \frac{\bar{k}N - 2}{\bar{k} - 1 + \frac{V_k}{\bar{k}}}$$

where $\bar{k}$ was the mean number of offspring per parent, N was the total number of adults, and V_k was the variance in reproductive success among adults.

I estimated contemporary N_e using the LD-based method (N_eGen) in the program NeEstimator v. 2.01 with a random sample of 500 adults every five generations without replacement; if 500 individuals were not available, the largest number of individuals available was collected. Linkage-based N_e estimation has been found to produce accurate estimates of N_e given large samples of loci and individuals, and given that samples represent a range of age classes equal to the generation span (Luikart *et al.* 2010; Robinson & Moyer 2013; Waples & Do 2010). I looked for associations between genetic estimates of N_b and N_e using Spearman correlation (ρ) in python. Genetic estimates of N_e and N_b were also used to estimate population size trends, with λ estimated from the time when CRF was initiated to the end of the simulation ($\lambda = N_e$Gen80/N_eGen30 and $\lambda = N_b$Gen80/N_bGen30).

Impacts of demography – Associations between different estimates of effective sizes with census size and demographic parameters were assessed using Spearman correlation in python. Demographic parameters explored included the number of breeding females per reproductive cycle, and annual standardized variance (SDV) in reproductive success (V_k/k^2) for both males and females. I used estimates of effective size and census size to calculate N_e to N_C, and N_b to N_C

ratios. Ratios were calculated for each replicate in each generation, and then averaged over all replicates for each demographic scenario modeled.

Impacts of sample size – To assess the influence of number of sampled individuals and loci on estimates I also estimated N_bGen and N_eGen, as outlined above, using a random sample of 30 microsatellite markers, and 60 individuals without replacement; if 60 individuals were not available, the largest number of individuals available was collected. These numbers were chosen to represent realistic sampling conditions in population genetic studies, and have been suggested as target numbers of samples and markers for accuracy in N_e estimation when using microsatellites (Tallmon et al. 2010). Associations between N_bGen and N_eGen with census size were assessed using Spearman correlation in python.

Sensitivity analysis

To explore which parameters have the greatest impact on extinction risk and census size at the final generation of simulations, I conducted a relative sensitivity analysis. I investigated the following parameters: adult survival, juvenile survival, and the proportion of females that breed per year. Specifically, I reduced each of these parameters by 10% while keeping all other values unchanged and evaluated proportional changes in extinction probability and adult census size from the base model.

Results

Population genetic variability

The level of genetic diversity remaining in populations at the end of simulations was highly dependent on the probability of CRF imposed and the presence of migration in models. In models without migration, genetic variability was maintained at over 90% of initial estimates up to a CRF probability of 70% (Table 4.1). When CRF probability was over 70%, genetic diversity was rapidly lost from populations. In models with migration, genetic diversity was maintained at 99% to 100% of initial estimates, over all probabilities of CRF.

Annual population demographic data

Demographic parameters including N_C, number of breeding females, and SDV for males and females were estimated each reproductive cycle to characterize relationships with N_e and N_b. The initial census size across simulations stabilized at a mean of 3,651 for models with and without migration, after 30 generations of burn-in. A full summary of initial and final census sizes with confidence intervals across replicates can be found in Tables C.2 and C.3. Final census size at generation 80 ranged from $0 - 3,651$ without migration, and $185 - 3,652$ with migration, over all probabilities of CRF. Genetic estimates of N_e were generally higher than N_b estimates, while demographic estimates of N_b were consistently higher than genetic estimates of N_b (Figure 4.1). Few significant correlations were found between N_C and $\widehat{N}_e$Gen (hat denotes parameter estimates; Table 4.2 and Tables C.4 and C.5). Census size showed significant correlations with $\widehat{N}_b$ across many simulations, with the probability of CRF impacting which was correlated more often, $\widehat{N}_b$Demo or $\widehat{N}_b$Gen (Table 4.2 and Tables C.4 to C.5). When probability of CRF was low ($\leq 20\%$) $\widehat{N}_b$Demo was significantly correlated with N_C across almost all replicates, while at higher CRF probabilities ($\geq 30\%$) $\widehat{N}_b$Gen was correlated with N_C most often (Table 4.2). Genetic

estimates of N_bGen showed few correlations with N_eGen across simulations. Neither $\widehat{N}_b$ nor $\widehat{N}_e$ was significantly correlated with number of breeding mothers, or SDV of males or females, across most replicate simulations.

CRF

Models with and without migration both showed strong effects of CRF, with sharp declines in census size observed across all levels of CRF (Figure 4.1). However, after ~10 generations of initiating CRF all estimators oscillated around a stable size, except for at high CRF ($\geq$ 70%) where populations declined towards extinction. For models without migration, extinction risk increased with increasing CRF probabilities over 40%, with extinction probability > 50% when CRF was above 70% (Table 4.3). There was no extinction when including migration in the model. Increasing probability of CRF led to decreased lambda estimates across all estimators (N_C, N_eGen, and N_bGen; Table 3), except when CRF was $\geq$ 70% in models including migration where $\widehat{N}_e$Gen increased with increasing CRF.

The majority of effective size to census size ratios at the end of simulations (generation 80) were similar across low to median probabilities of CRF (CRF $\leq$ 50%), and became variable at higher CRF probabilities (Figure 4.2). In particular, $\widehat{N}_b$Gen /N_c ratios were remarkably stable across most levels of CRF, up to $\geq$70% CRF in models without migration, and across all levels of CRF when including migration. Similar patterns were apparent when looking at ratios temporally across generations, $\widehat{N}_b$Gen /N_C and $\widehat{N}_e$Gen / N_C ratios were stable at low to median levels of CRF (CRF $\leq$50%), and became variable at higher CRF probabilities (Figure 4.3). Again, $\widehat{N}_b$Gen /N_C ratios were more stable than $\widehat{N}_e$Gen / N_C. All ratios showed greater temporal stability when including migration in models.

Sample size and sensitivity analysis

The subsampled dataset of 30 microsatellite markers and 60 sampled individuals yielded slightly negatively biased estimates of effective sizes (Figure C.1), and fewer correlations of $\widehat{N}_b$Gen with census size relative to the complete dataset (Table C.6). Sensitivity analysis revealed that changes in juvenile survival had the largest influence on adult abundance (Table 4.5, for full sensitivity analysis see Table C.7). The proportion of breeding adults impacted extinction risk with CRF $\geq$70% for models with no migration. Eliminating density-dependent age to first breeding had little impacts on results, except at 60-70% CRF where including it maintained slightly larger population sizes.

Discussion

Understanding the relationship between N_b, N_e, and N_C is critical when applying genetic monitoring for biological conservation, yet remains difficult in iteroparous species with complex life-histories. My simulation study incorporated life-history characteristics common in pond-breeding amphibians allowing me to quantify annual $\widehat{N}_b$ estimated with both genetic and demographic data, $\widehat{N}_e$, and N_C, and generate predictions about future genetic viability and extinction risk under multiple levels of CRF, a proxy for climate change. I highlight several important findings. First, my results demonstrate that genetically estimated $\widehat{N}_b$ is positively correlated with N_C across multiple levels of CRF in isolated and subdivided amphibian populations, and can potentially be informative in inferring abundance and detecting population declines in amphibians at a contemporary scale. Second, I provide evidence of the consequences

of CRF on population viability of *A. opacum* populations. In the following sections, I discuss these findings in the context of genetic monitoring and conservation of amphibian populations.

Effective size relation to census size

Using genetic monitoring to infer changes in population abundance requires a clear and predictable relationship to infer N_C from $\widehat{N}_e$ or $\widehat{N}_b$, and therefore stability of N_e/N_C or N_b/N_C ratios across time and demographic conditions (Luikart *et al.* 2010; Palstra & Ruzzante 2008). I found stability of effective size ratios temporally and across low to moderate levels of CRF, but saw ratios increase when CRF was high. This environmentally-driven variation in ratios implies interactive effects with life-history, and possible density-dependent processes at decreased population sizes, i.e. 'genetic compensation' (Palstra & Ruzzante 2008). I modeled increased survival and breeding probability with low larval densities, which could maintain population size and buffer genetic diversity loss when populations decline. I also observed short-term fluctuations in N_C across all simulations, yet most models showed overall long-term stability of population size. This stability was observed in both N_e and N_b estimates, and suggests genetic monitoring is reflective of long-term trends, and not short-term fluctuations common to amphibians (Alford & Richards 1999; Pechmann & Wilbur 1994). Given these findings, genetic monitoring for inferring contemporary abundance trends has potential in amphibian populations, but is generally not advisable in populations where CRF is known to be high.

In long-lived species with overlapping generations, using large samples of mixed-age cohorts for N_e estimation approximates per-generation N_e (Robinson *et al.* 2014; Waples & Do 2010). However, these models assume a constant population size, and when applied in declining or variable sized populations the relationship between N_C and N_e is difficult to interpret, and the

exact time period to which $\widehat{N}_e$ applies is difficult to determine. Maybe not surprisingly then, I saw few correlations between N_e and N_C because the estimate of N_e is reflecting the entire generation where N_C was variable, and not the single year from which I calculated N_C. This also complicates the relationship between N_b and N_e, with N_b tracking evolutionary factors at an ecological scale in a single season, with N_e relating to evolutionary processes acting across overlapping generations (Waples & Antao 2014). Given the complexity of properly linking per generation N_e with N_C in variable sized populations, N_e may not be the best tool for short term detection of population declines, but is itself an important indicator of population viability and evolutionary potential.

Sample size limitations

Overall, I found census size to be positively correlated with $\widehat{N}_b Gen$ across the majority of our modeled scenarios. Several empirical studies support this finding and have found positive correlations between $\widehat{N}_b Gen$ and $\widehat{N}_C$ in other species (Charlier *et al.* 2012; Osborne *et al.* 2010). However, other studies have failed to detect any correlation, highlighting that genetic monitoring using N_b estimation may not be a panacea across systems (Duong *et al.* 2013; Whiteley *et al.* 2015). My simulation based study did not have the constraint of limited numbers of sampled individuals or genetic markers, which may have large influences on accuracy and precision of $\widehat{N}_e$ and $\widehat{N}_b$ in empirical studies that often go unnoticed (Tallmon *et al.* 2010). When analyzing the subsampled dataset with numbers of microsatellite markers and sampled individuals commonly used in population genetic studies, I found fewer correlations between $\widehat{N}_b Gen$ and $\widehat{N}_C$. This was most evident when population size was large, resulting in large variance in effective size estimates, and a negative bias. This result demonstrates the potential benefit of genetic

monitoring using N_b measured from single-aged cohorts over time, but cautions that sufficient sampling is critical for accuracy.

Catastrophic reproductive failure

I tested a large range of CRF probabilities to evaluate their impact on genetic diversity, N_C, and extinction risk. Over the course of the models, most replicate populations experienced declines in census size from increasing levels of CRF, but no losses of genetic diversity until CRF was 80% and above. Fluctuations in population size were also apparent when imposing any level of CRF, and although this is suggested to be one of the largest variables reducing N_e/N_C, ratios remained stable across most models (Frankham 1995). Maintenance of N_e and genetic diversity with moderate levels of CRF is encouraging for population and evolutionary viability of *A. opacum* populations, and is likely the result of a long life span allowing for high lifetime fitness and cumulative fecundity despite missed breeding in some years due to environmental stochasticity.

Significant extinction risk became apparent at 70% probability of CRF, yet probabilities this high are unlikely across most small ephemeral ponds in the southeastern US, with probabilities of 10-40% being more in-line with current conditions (Daszak *et al.* 2005; Semlitsch *et al.* 1996). However, extreme climatic conditions are predicted to increase into the future, with the potential to increase CRF probabilities and extinction risk to levels observed in my models. Increased drought conditions have already been observed throughout the southeastern US with correlated declines in many amphibian populations (Walls *et al.* 2013b). Sensitivity analysis revealed juvenile and adult survival had large influences on census size. This suggests that improving conditions at the terrestrial level to increase survival would increase

persistence probability of *A. opacum* populations. This result emphasizes the importance of terrestrial buffer zones adjacent to breeding wetlands, which should be much larger than previously thought to adequately protect *A. opacum* populations (Scott *et al.* 2013).

The final N_e over my simulations ranged from 163 to 1485 (750-2099 with migration), falling slightly above the N_e estimates from a previous empirical study of *A. opacum* at an ephemeral wetland in South Carolina, USA ($N_e = 152$) (Nunziata *et al.* 2015). Estimates from my empirical study were likely affected by individual sample size and the amount of marker information (number of loci and alleles). Genetically estimated N_e from other amphibians are often much lower than observed in my models, often falling below 100 (Funk *et al.* 1999; Jehle & Arntzen 2002; Savage *et al.* 2010; Wang *et al.* 2011). My results suggest that CRF alone does not act to reduce N_e to such low levels, with reasons likely being population specific, such as historical demography and habitat quality.

Gene flow

Amphibian populations often exhibit metapopulation structure across landscapes, with gene flow acting to maintain genetic variation and buffer extinction risk. I found that including a moderate level of migration in models had large impacts on maintenance of genetic diversity and overall extinction risk by allowing for the rescue of declining populations. This source-sink dynamic is well documented in amphibian populations, where sink populations can be maintained through migrants from stable populations. I observed high genetic diversity and $\widehat{N}_e$ in migration models when probability of CRF was 80-100% (i.e. a sink population), as the majority of sampled individuals were migrants. This highlights the dangers of limiting genetic sampling to single ponds, where one could misidentify a declining sink population as stable

(Griffiths *et al.* 2010). Preserving metapopulations as a whole and maintaining connectivity of habitat patches may allow *A. opacum* and other amphibian species to cope with environmental stochasticity and drought conditions and retain genetic diversity and census size. Management at the pond level should be focused on source populations, with special attention to ensure that they have been properly identified and not in fact population sinks.

Conclusions

This study highlights the utility of genetic monitoring to track contemporary population size trends in amphibians. My findings are relevant to other *Ambystoma* species of conservation concern that have similar but less well documented life-history characteristics, such as the endangered flatwoods salamander (*A. cingulatum*). Optimistically, my models illustrate that fluctuations in population size from CRF do not cause large decreases in genetic diversity or N_e, at least at low to moderate levels of CRF. I emphasize that population management and monitoring of pond-breeding amphibians should be focused at the landscape scale, keeping in mind that local populations acting as sinks could be misidentified as stable if metapopulation structure is not accounted for. One important aspect of species conservation is that multiple stressors are acting on populations at any one time, and can be interactive and/or additive (Weir *et al.* 2016). Here, I focus on CRF and provide guidelines for persistence in the face of this risk, but this does not mean pond-breeding amphibians do not face other important threats. Understanding the impacts of multiple concurrent stressors, as well as interactive effects of co-occurring species will be critical for designing conservation management plans and will be the focus of future modeling efforts.

Table 4.1. Average proportion of allelic diversity (AD) and expected heterozygosity (He) remaining in *Ambystoma opacum* models at the end of simulations (generation 80 relative to generation 30) over 100 replicate simulations, or of those replicates not extinct at end of simulation.

Probability of CRF (%)	Base Model-No Migration		Base Model-With Migration	
	AD	H_e	AD	H_e
0	0.98	1.00	1.00	0.99
10	1.00	1.00	1.00	1.00
20	0.99	1.00	0.99	1.00
30	0.99	1.00	1.00	1.00
40	0.97	0.99	1.00	1.00
50	0.96	0.99	0.99	1.00
60	0.95	0.98	1.00	1.00
70	0.92	0.97	0.99	1.00
80	0.37	0.62	1.00	1.00
90	0.20	0.40	0.99	0.99
100	-	-	0.99	0.99

Table 4.2. Number of times Spearman rank correlations were significant ($p < 0.05$) for *Ambystoma opacum* models with and without migration (mig.) over 100 replicate simulations under different probabilities of catastrophic reproductive failure (CRF). Mean Spearman rank correlation coefficients across replicates are in parentheses.

	No CRF		20% CRF		40% CRF		60% CRF	
	no mig.	with mig.	no mig	with mig.	no mig.	with mig.	no mig.	with mig.
Census v. N_bDemo	100 (0.59)	99 (0.58)	100 (0.63)	100 (0.65)	81 (0.40)	72 (0.36)	28 (0.17)	28 (0.17)
Census v. N_bGen	5 (0.20)	11 (0.17)	83 (0.75)	86 (0.75)	99 (0.86)	99 (0.88)	96 (0.77)	100 (0.92)
Census v. N_eGen	4 (0.00)	2 (0.02)	11 (0.27)	9 (0.08)	8 (0.30)	4 (0.04)	8 (0.20)	2 (-0.01)
Moms v. N_bGen	7 (0.16)	11 (0.18)	43 (0.51)	35 (0.51)	18 (0.30)	18 (0.27)	7 (0.07)	5 (0.11)
Moms v. N_eGen	2 (0.01)	5 (0.04)	7 (0.15)	7 (0.02)	8 (0.11)	4 (-0.01)	7 (0.05)	2 (-0.05)
N_bGen v. N_eGen	2 (-0.01)	3 (0.06)	32 (0.46)	13 (0.20)	37 (0.53)	11 (0.22)	14 (0.21)	6 (0.12)
SDVmale v. N_bGen	8 (-0.09)	8 (-0.13)	7 (0.05)	5 (0.04)	8 (0.11)	5 (0.14)	5 (0.05)	5 (-0.05)
SDVmale v. N_eGen	6 (0.00)	4 (-0.02)	10 (-0.04)	8 (-0.13)	3 (-0.04)	4 (-0.12)	11 (-0.10)	6 (-0.15)
SDVfemale v. N_bGen	11 (-0.06)	8 (-0.08)	8 (-0.01)	8 (0.00)	10 (0.10)	6 (0.14)	7 (0.04)	5 (-0.05)
SDVfemale v. N_eGen	7 (-0.01)	3 (0.01)	7 (-0.1)	8 (-0.16)	5 (-0.04)	5 (-0.11)	11 (-0.10)	7 (-0.14)

Table 4.3. The probability of extinction and average $\hat{\lambda}$ (λ = NGen80/NGen30) for *Ambystoma opacum* under different probabilities of catastrophic reproductive failure (CRF) at the end of simulations (Generation 80) over 100 replicate simulations, or of those replicates not extinct at end of simulation.

	Base Model-No Migration				Base Model-With Migration			
CRF (%)	Extinction	$\hat{\lambda}$ N$_c$	$\hat{\lambda}$ N$_e$Gen	$\hat{\lambda}$ N$_b$Gen	Extinction	$\hat{\lambda}$ N$_c$	$\hat{\lambda}$ N$_e$Gen	$\hat{\lambda}$ N$_b$Gen
0	0	1.00	0.99	1.01	0	1.00	1.08	1.04
10	0	0.93	0.90	0.93	0	0.91	0.98	0.93
20	0	0.88	0.77	0.84	0	0.90	0.84	0.90
30	0	0.76	0.66	0.74	0	0.78	0.76	0.76
40	0	0.72	0.56	0.66	0	0.67	0.68	0.66
50	4	0.57	0.42	0.51	0	0.58	0.55	0.54
60	6	0.47	0.29	0.37	0	0.42	0.50	0.40
70	40	0.21	0.18	0.27	0	0.34	0.55	0.32
80	75	0.04	0.11	0.10	0	0.20	0.64	0.20
90	99	0.00	0.00	0.00	0	0.09	0.92	0.11
100	100	0.00	0.00	0.00	0	0.05	1.37	0.06

Table 4.4. Summary of sensitivity analysis under different probabilities of catastrophic reproductive failure (CRF) for *Ambystoma*

opacum models with and without migration (mig.).

Parameter, by endpoint	No CRF		20% CRF		40% CRF		60% CRF		80% CRF	
	no mig.	with mig.	no mig.	with mig.	no mig.	with mig.	no mig.	with mig.	no mig.	with mig.
Adult pop Size										
Base Model	3651	3652	3221	3282	2628	2437	1718	1515	136	738
Juvenile Survival	0.61	0.60	0.56	0.55	0.48	0.54	0.37	0.44	0.01	0.38
Adult Survival	0.78	0.78	0.78	0.75	0.66	0.73	0.61	0.64	0.08	0.58
Proportion Breeding	1.06	1.05	1.00	1.00	0.97	1.08	0.76	1.27	1.09	1.01
No Density-Dependent Breeding*	1.00	1.00	0.97	0.99	1.00	0.96	0.82	0.99	1.05	0.99
Extinction risk										
Base Model	0	0	0	0	0	0	0.06	0	0.75	0
Juvenile Survival	0	0	0	0	0	0	0.15	0	0.94	0
Adult Survival	0	0	0	0	0	0	0.12	0	0.91	0
Proportion Breeding	0	0	0	0	0.01	0	0.15	0	0.71	0
No Density-Dependent Breeding*	0	0	0	0	0	0	0.12	0	0.70	0

Notes: Each parameter was decreased by 10%. Adult population size is in proportion to the base model. *This model does not include density-dependent age at first reproduction, i.e. no individuals breed at age one.
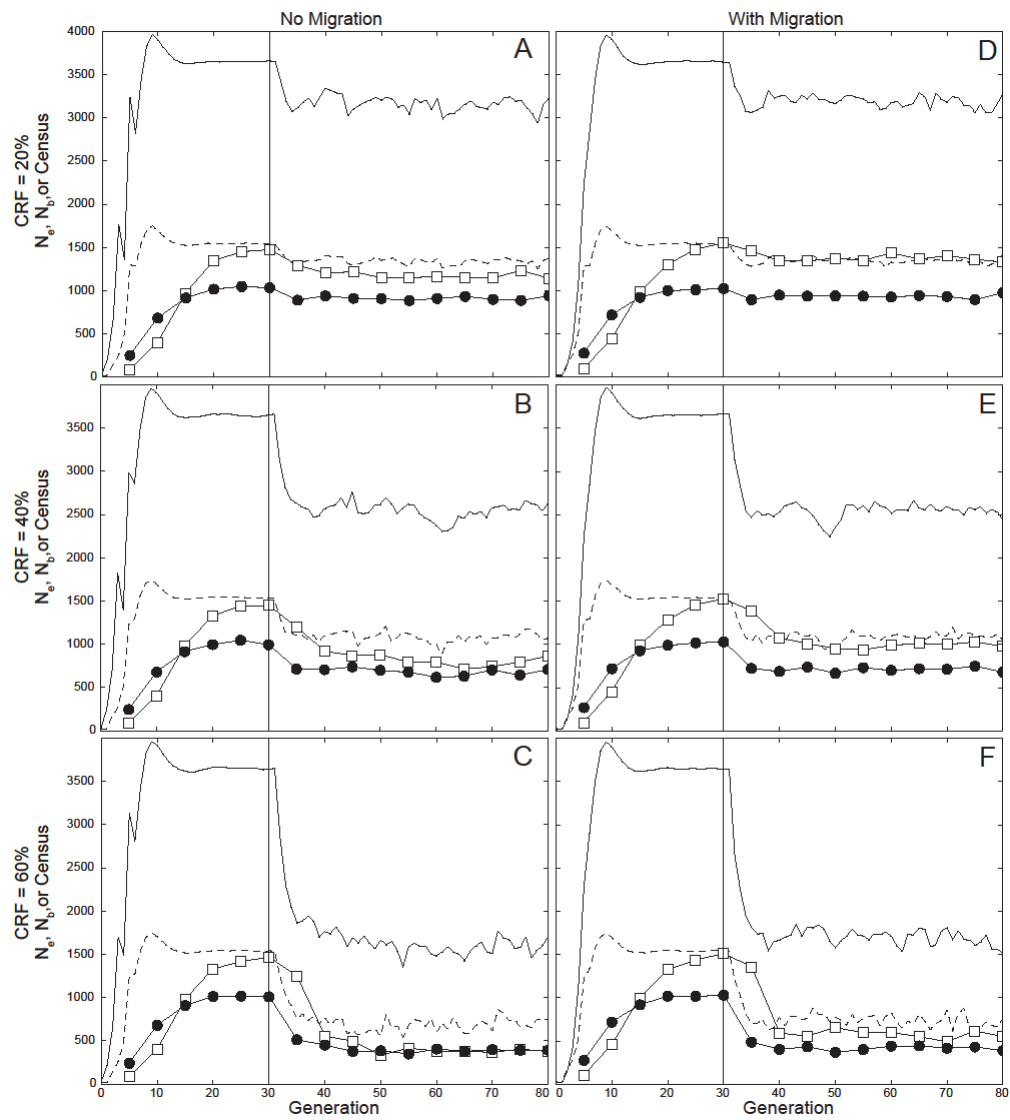
Figure 4.1. *Ambystoma opacum* mean census size (black line) and N_bDemo (dashed line) each generation, with N_bGen (filled circles), and N_eGen(open squares) every 5 generations, over 80 years of projections. Catastrophic reproductive failure (CRF) was imposed after 30 year burn-in (vertical line). Top panels represent 20% probability of CRF (A&D), middle 40% (B&E), and bottom 60% (C&F). Left panels (A-C) represent models with no migration, and right panels (D-F) are models with migration.
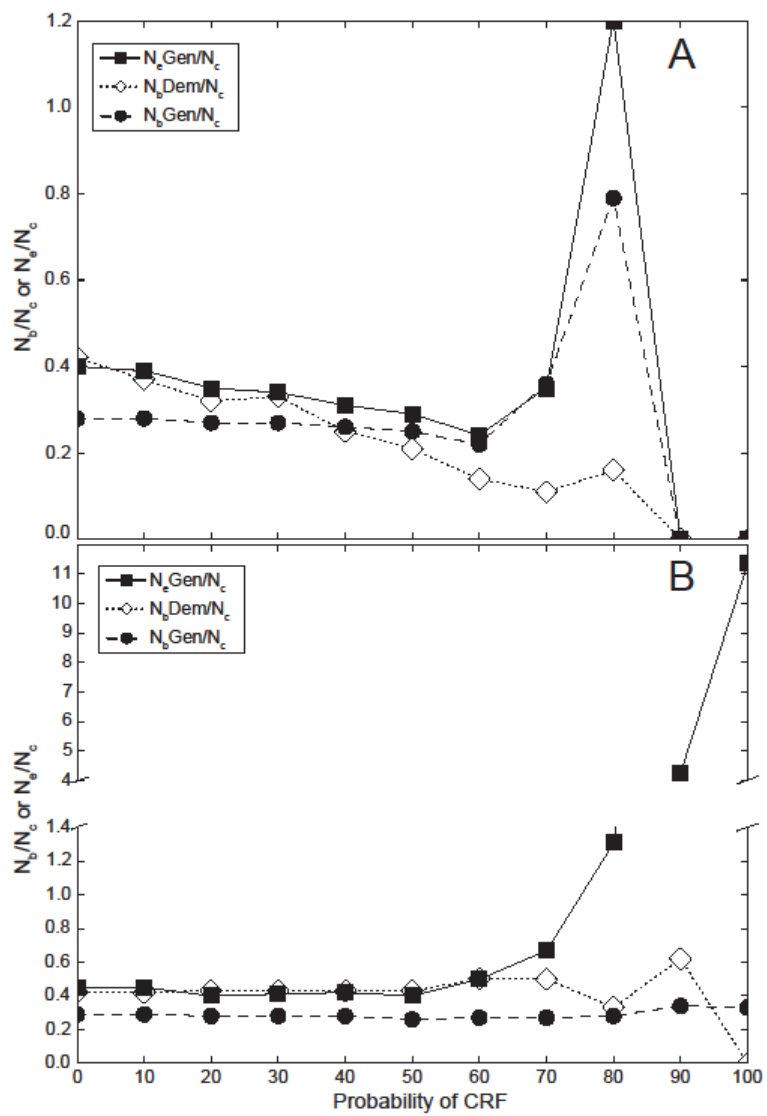
Figure 4.2. *Ambystoma opacum* mean effective size to census size ratios at generation 80 across multiple probabilities of catastrophic reproductive failure (CRF), represented are models without migration (A), and including migration (B).
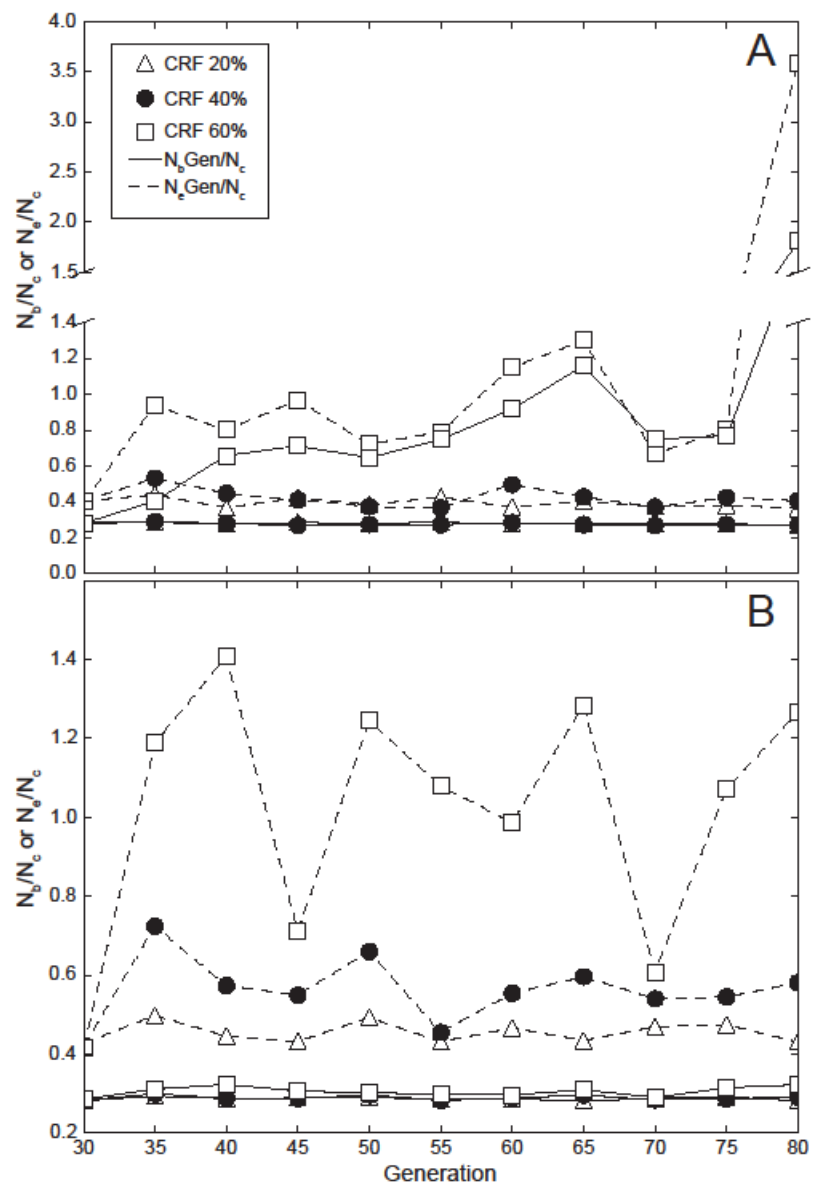
Figure 4.3. *Ambystoma opacum* mean effective size to census size ratios across generations at multiple levels of catastrophic reproductive failure (CRF), represented are models without migration (A), and including migration (B).

CHAPTER FIVE

TEMPORAL CHANGES IN POPULATION CONNECTIVITY AND EFFECTIVE
POPULATION SIZES ACROSS TWO SALAMANDER METAPOPULATIONS

Abstract

Environmental shifts can impact population demography across a landscape, with influences on census size, N_e, and population connectivity. In a previous study focused at a single wetland (Rainbow Bay), I found genetic-based N_e estimates between two temporally sampled time points supported population size trends that corroborated mark-recapture data in two species of salamanders (*Ambystoma opacum* expansion, *A. talpoideum* decline). The goal of this study was to estimate the geographic scale reflected by these demographic inference results and to characterize how migration may have changed across the landscape. I generated ddRAD data from 7 additional wetlands surrounding the focal Rainbow Bay wetland, estimated N_e at each wetland, as well as population structure and migration paths between wetlands. For both species I found significant differences between N_e estimates at the subpopulation (wetland) level, despite the appearance of panmictic population structure within a neighboring cluster of wetlands. Population structure across the landscape differed between species, and showed temporal stability in *A. opacum* and temporal changes in *A. talpoideum.* These preliminary results suggest that demographic history can be detected at the subpopulation (wetland) level using genomic datasets, even over recent time and fine geographic scales with the presence of migration. This study is ongoing, and I am currently modeling alternative patterns of population structure, and evaluating temporal changes in connectivity.

Introduction

Understanding the demographic response of species to environmental disturbance such as habitat loss, modification, and climate change is critical in conservation management. Estimates of effective population (N_e) provide information on population abundance and demographic history, offer insight into the evolutionary history and genetic viability of populations, and are useful for monitoring the status of populations of conservation concern (Schwartz *et al.* 2007). However, the continuous distribution of individuals across landscapes creates complications in the estimation and interpretation of N_e, with a major question being what geographic scale is reflected in contemporary N_e estimates (Neel *et al.* 2013). Most species do not occur as independent and isolated populations, but instead occur in networks that vary in their degree of connectivity spatially and temporally.

In disturbed habitats, migration and dispersal are important for population viability, and allow for maintenance of population size, genetic diversity, and recolonization after local extinction events (Hanski 1998). Estimating and understanding N_e simultaneously with migration is critical when applying estimates within conservation and monitoring frameworks. If migration is left unaccounted for in N_e estimation models, its influences can be attributed to drift, creating biased N_e estimates (Gilbert & Whitlock 2015; Wang & Whitlock 2003; Waples & England 2011). However, disentangling the contributions of population growth trajectory and migration to N_e is difficult when population size is variable and unlikely to be at drift-migration equilibrium (Leblois *et al.* 2006). Additionally, when continuously distributed populations face disturbance, N_e and migration may both be impacted simultaneously. Comparative studies of N_e and migration that incorporate temporal samples from populations undergoing population size

112

changes may provide important insights into population dynamics within structured populations, and the geographic scale of disturbance reflected in N_e estimates.

In a previous study, I studied the utility of genetically-based demographic inference and N_e estimation on salamander species with documented population declines (*Ambystoma talpoideum*) and expansions (*A. opacum*) (Nunziata *et al.* 2016); patterns that have been shown to be correlated with changes in wetland hydroperiod (Daszak *et al.* 2005). For both species, I found genetic-based N_e estimates between two temporally sampled time points supported population size trends that corroborated mark-recapture data. However, sampling was limited to a single wetland (Rainbow Bay) in both species. Both species exist in metapopulations across the landscape with breeding wetlands acting as subpopulations connected by migration, as revealed through mark-recapture data and a previous landscape genetic study (Kinkead *et al.* 2007). It is unclear how the genetic interaction between individual subpopulations across the landscape influenced the observed population size trends at Rainbow Bay, if N_e estimates were actually reflective of a larger geographic scale beyond Rainbow Bay, or how migration itself may have responded as population size at the wetland level changed.

The goal of this study was to characterize the metapopulation structure within which Rainbow Bay exists and estimate how migration may have changed across the landscape as populations of *A. opacum* and *A. talpoideum* at Rainbow Bay have changed in size. I sampled from 7 additional wetlands surrounding Rainbow Bay, one of which was sampled temporally, to address the following questions: (1) at what geographic scale does population structure exist, (2) are differences in N_e and demographic history apparent between wetlands, and (3) has the connectivity between wetlands changed over time. The results of this study will elucidate if Rainbow Bay exists as an isolated or structured population, and reveal if demographic history

and N_e estimates can be distinguished between neighboring wetlands that exist within a larger metapopulation. This also has broader implications for scale of sampling in investigating population demography using genetics in continuously distributed populations.

Methods

Population sampling

      The Savannah River Site (SRS) is a US Department of Energy facility on the Upper Coastal Plain of South Carolina (Figure 5.1). I sampled tissue from 8 wetlands for *A. opacum*, and 7 wetlands for *A. talpoideum,* including samples from Rainbow Bay which were generated for a previous study (Nunziata *et al.* 2016*).* Samples were analyzed temporally for both Rainbow Bay and Gingers Bay. From Ginger's Bay, I pooled samples collected prior to 2000 (Pre-2000), and after 2000 (Post-2000) for temporal comparison. Rainbow Bay sampling is described in my previous study (Nunziata *et al.* 2016) and includes samples from twelve *A. opacum* in both 1993 and 2013, and 24 *A. talpoideum* in both 1984 and 2011. All other wetlands were only sampled from contemporary time points with sample collection taking place between 2014 and 2016. A total of 120 *A. opacum* and 117 *A. talpoideum* individuals were sequenced, with a mean of 12 and 13 individuals per population for each species, respectively. Most individuals sampled were larvae or metamorphs collected at breeding wetlands. Adults were collected from two wetlands in each species when larvae were not available temporally, or where larvae could not be collected (Table D.1).

ddRAD sequencing

Whole genomic DNA was extracted using a Qiagen® DNEasy Blood and Tissue Kit, following protocols recommended by the manufacturers. I followed the ddRAD library preparation protocol used for Rainbow Bay samples as outlined in my previous study (Nunziata *et al.* 2016), modified slightly from Peterson *et al*. (2012). Four library pools were prepared for *A. opacum*, each containing 24 samples mixed between sampling sites. Three library pools were prepared for *A. talpoideum*, two containing 24 samples and one with 21 samples, all mixed between sampling sites. The pooled ddRAD libraries were sequenced separately on seven lanes of an Illumina HiSeq 2500 at Florida State University's College of Medicine using 150 base-pair paired-end reads. To increase sequence diversity in the initial portions of sequence reads I used an additional 1% spike in of PhiX control library.

Quality filtering and variant detection

Sequences were subjected to standard Illumina chastity filtering and then demultiplexed, quality filtered, and SNPs called using STACKS v.1.37 (Catchen *et al.* 2013; Catchen *et al.* 2011). To avoid the inclusion of linked sites only first reads were retained. Demultiplexing and quality-filtering were done with the *process_radtags* program in STACKS v1.37. The *denovo_map.pl* script in Stacks was used to identify unique RAD loci and call SNPS, with default settings except minimum coverage (m=3), number of mismatches allowed between RAD loci for an individual (M=3), and number of mismatches allowed when building the catalog (n=2). SNP genotypes were called using a multinominal-based likelihood model accounting for sequencing errors using a conservative upper bound error rate of 0.2.

The *populations* script in Stacks was used to generate population datasets with loci that occurred in at least 4 sampling sites (p = 4), in at least 50% of individuals at each site (r = 0.50) and with a minimum sequencing coverage of 10x (m = 10). To avoid the inclusion of linked, non-independent SNPs, only a single SNP was allowed per RAD locus (--write_single_snp). After an initial run of the stacks pipeline I excluded 18 *A. opacum* individuals, and 4 *A. talpoideum* individuals due to excessive missing data (Table D.1). This brought the total number of wetlands sampled down to 7 for *A. opacum* as all individuals from Linda's had excessive missing data. This resulted in 104,623 SNPs in *A. talpoideum* and 75, 805 SNPs in *A. opacum*. Further filtering was done in vcftools to retain SNPs only genotyped across 50% of all individuals. This resulted in 38,126 SNPs for *A. talpoideum* and 42,879 SNPs for *A. opacum*.

Population genomics

Nucleotide diversity estimates were calculated separately for each sampled wetland and were based on individual SNPs. I estimated observed heterozygosity (Het_{obs}), nucleotide diversity (π), and Wright's F-statistic (F_{IS}), as calculated using the *populations* script in Stacks. To examine differentiation among wetlands, I also calculated pairwise F_{ST} between all wetland pairs using the *populations* script and separately in Arlequin v.3.5.2.2 (Excoffier & Lischer 2010). To test for isolation by distance (IBD) I compared genetic distance to natural logarithm of geographic distance, and performed Mantel tests in R using the ADE4 package with 10,000 permutations. Genetic distance was calculated as the ratio $F_{ST}/(1-F_{ST})$. Geographic distances (meters) between wetlands were calculated using ArcGIS v10.3 and a least cost path analysis. Least cost path analysis is a distance analysis tool that calculates the least accumulative cost route between a source and destination over a cost surface. A cost distance surface was created

by using a slope raster generated by a digital elevation model. Steeper slopes were assigned

higher costs when reclassifying this dataset. Using the slope cost path dataset least cost routes

were drawn and measured from a source wetland to a destination wetland.

Estimates of N_e were generated for each wetland and temporal sample pool using the

linkage disequilibrium method implemented in N_eEstimator v2.01 (Do *et al.* 2014). This method

measures the deviation from the expected genotype frequency based on allele frequencies in the

population, which increases in small populations due to drift (Hill 1981). Using the complete

SNP dataset exceeded memory requirements for NeEstimator. Therefore, I filtered the SNP

dataset to retain SNPs only genotyped across 75% of all individuals resulting in a dataset of

13,585 SNPs in *A. opacum* and 15,124 SNPs in *A. talpoideum*.

Genetic structure and population differentiation

To infer genetic clusters of individuals and the degree of admixture in individuals I used

two Bayesian clustering programs: ADMIXTURE v.1.3.0 (Alexander *et al.* 2009) and

fastSTRUCTURE v.1.0 (Raj *et al.* 2014). Data was filtered to exclude SNPs with minor allele

frequency (MAF) less than 0.05, which have been shown to create biases in quantifying genetic

connectivity (Benestan *et al.* 2015; Roesti *et al.* 2012), resulting in 30,187 SNPs for *A. opacum*

and 22,781 SNPs in *A. talpoideum*. ADMIXTURE was run using default parameters and models

were fit with K ranging from 1 to 10, with the optimal K chosen as having the lowest cross-

validation error. fastSTRUCTURE uses a similar algorithm as the program STUCTURE

(Pritchard *et al.* 2000) to estimate global ancestry, but runs faster on large genome wide SNP

datasets by employing a variational inference scheme. I conducted analyses with K ranging from

1 to 10 using a beta flat prior, or "simple prior". The most likely value of K was identified by

calculating heuristic scores using the *chooseK.py* script in fastSTRUCTURE. After identifying

the most likely values of K, I reran fastSTRUCTURE using those K values and a logistic prior,

which is more computationally demanding but better at resolving subtle structure. For both

ADMIXTURE and fastSTRUCTURE, I used the web server CLUMPAK (Cluster Markov

Packager Across K) to visualize bar plots across multiple values of K (Kopelman *et al.* 2015).

Landscape genomics

I used a Bayesian assignment test in BAYESASS v.3.04 (Wilson & Rannala 2003) to

estimate gene flow between wetlands, and to look for changes in migration rate over time.

BAYESASS was run for 100,000,000 Markov chain Monte Carlo (MCMC) iterations, discarding

the first 10,000,000 iteration as burnin and sampling every 1000 iterations. A migration rate

mixing parameter (m) of 0.3 was used to optimize the acceptance rate. Convergence of runs was

assessed in Tracer v 1.5 (Rambaut & Drummond 2012). To balance adequate sampling and

computational effort, I used the dataset generated for NeEstimator, outlined above, with 13,585

SNPs in *A. opacum* and 15,124 SNPs in *A. talpoideum*.

I estimated mutation-scaled migration (M) and N_e (θ), and looked for changes between

historical and contemporary populations using MIGRATE-n v.3.6.11 (Beerli 2006). Since

MIGRATE-n is better suited to DNA sequences than SNP data, I used the entire 146-bp

sequences in these analyses. To balance sample size and computational effort, subsamples of

250, 500, and 1000 loci were randomly chosen and used as input. Initial runs were used to find

priors for M and θ, after which each analysis was run twice to ensure consistency with the

following settings: a burnin of 20 million steps, and then 10 million generations with metropolis

sampling every 1000 steps.

I tested 3 models of population connectivity for each species, using two different datasets to investigate if N_e and connectivity have changed over time (Figure 5.2). My sampling scheme did not allow me to estimate gene flow historically between all sampled wetlands. Instead, my estimates of historical connectivity were based on a dataset that included historical samples from Rainbow and Gingers Bay and contemporary samples from other wetlands. I make the assumption that because all sampled wetlands except Rainbow Bay showed large and temporally stable census-size over this short time span, the genetic composition of individual wetlands should be comparable over the sampling time span. This allows me to investigate if connectivity of Rainbow Bay with other wetlands has changed in conjunction with observed population size trends. For comparison, I estimated contemporary connectivity using a dataset that included contemporary samples from Rainbow and Gingers Bay and contemporary samples from other wetlands. The following three models were evaluated independently for each dataset: (i) panmixia of all wetlands, (ii) full migration between all wetlands, and (iii) the Rainbow Bay metapopulation is one panmictic population with migration paths with other wetlands (see Figure 5.2 for model schematics). Model choice was compared by ranking marginal likelihoods between different models and calculating their probabilities based on the Bayes factor (Beerli & Palczewski 2010).

Results

Population genomics

Genetic diversity estimates were similar between all wetlands for *A.talpoideum,* while Flamingo Bay showed lower genetic diversity compared to other wetlands for *A. opacum* (Table 5.1). Between species, genetic diversity was slightly higher in *A. opacum*; this pattern was

119

apparent across all sampled wetlands. Estimates of N_e using NeEstimator reached infinity for most wetlands (Table 5.2), indicating that the signal of drift cannot be distinguished from sampling error to accurately generate N_e estimates (Waples & Do 2010). However, both sampled years for *A. talpoideum* generated finite estimates of N_e, and showed a clear and significant signal of decline.

Genetic structure and population differentiation

Bayesian clustering tests revealed no population structuring for the wetlands immediately surrounding Rainbow Bay (NPR, Pickerel, Lindas, Bullfrog, and NPR) for *A. opacum* or *A. talpoideum*, hereafter referred to as the Rainbow Bay cluster. I reran fastSTRUCTURE and ADMIXTURE for the Rainbow Bay cluster, and found no further sub-structuring in either species. All other more distant wetlands grouped into separate clusters for *A. opacum,* with an optimal K of 4 using fastSTRUCTURE and ADMIXTURE (Figure 5.3). Population structure for *A. talpoideum* also distinguished these wetlands as largely differentiated, but recognized a large number of individuals with admixed ancestry, with an optimal K of 6 for both fastSRUCTURE and ADMIXTURE. Pairwise F_{ST} values were low to moderate (Tables 5.3 and 5.4) and were generally similar between species (Stacks estimates: *A. opacum:* 0.027-0.101; *A. talpoideum:* 0.029-0.128). Isolation by distance was obvious for the contemporary population of *A. opacum,* with borderline significance ($R^2 = 0.932$, p = 0.079) and showed a similar pattern when using historical samples from Rainbow and Gingers Bay ($R^2 = 0.870$, p = 0.095); while IBD was absent for *A. talpoideum* in both contemporary and historical samples (contemporary: $R^2 = -0.126$, p = 0.850; historical: $R^2 = -0.150$, p = 0.906).

Landscape genomics

For *A.opacum*, the BAYESASS estimates of migration rates between wetlands were low (mean values of 0.0157-0.0186), with the exception of the proportion of Bullfrog Bay migrants to other wetlands within the Rainbow Bay cluster (mean values 0.1795-0.2064) (Table 5.5). These results suggest that Bullfrog Bay supplies a greater proportion of migrants to wetlands within the Rainbow Bay cluster, relative to other wetlands. There were no significant difference in migration rates between temporal samples and other sampled wetlands. Estimates of migration were similar for *A. talpoideum* (mean values 0.0100-0.1025); with the only significant migration routes being from NPR to Linda's and Pickerel (Table 5.5).

MIGRATE-n runs are computationally demanding and are still in progress.

Discussion

Environmental shifts can impact population demography across a landscape, with influences on census size, N_e, and population connectivity. Life-history characteristics will dictate the response of individual species to disturbance, and even species with similar life-history can have differing responses. In this study, I examined two co-occurring species of *Ambystoma* salamanders with observed population trends at a single wetland (*A. opacum* expansion, *A. talpoideum* decline) from a landscape perspective and observed clear differences in the genetic structure and migration patterns between species. For *A. opacum* I observed an isolation-by-distance pattern, limited gene flow between geographically distant sites, and the appearance of panmictic population structure within the Rainbow Bay wetland cluster. For *A. talpoideum,* there was no isolation by distance pattern, admixture of individuals was apparent in geographically distant wetlands, and the Rainbow Bay wetland cluster appeared to be composed

of individuals with largely admixed ancestry. In both species there were differences in N_e at the subpopulation (wetland) level, with preliminary evidence of asymmetric gene flow between wetlands in *A. opacum,* and changes in population structure over time in *A. talpoideum.* This study is currently ongoing, and it will be exciting to test further models of migration and N_e patterns from a temporal perspective and compare capture-mark-recapture estimates of census size and migration to N_e and gene flow. For now, I discuss the results to date below.

Effective population size

Interestingly, I found significant differences in N_e at neighboring wetlands within the *A. opacum* Rainbow Bay cluster, despite the appearance of panmictic population structuring in Bayesian clustering assignments and non-significant pairwise F_{ST} between most wetlands. In the four wetlands sampled within the Rainbow Bay cluster, three had finite estimates of N_e with 95% CIs suggesting significant differences between wetlands. Bullfrog Bay was the only wetland where N_e could not be precisely estimated and was estimated as infinite; presumably the result of a large N_e. The Bullfrog Bay wetland also appeared to supply a disproportionate number of migrants to the other wetlands as revealed through BAYESASS analyses. This wetland is located centrally to the other wetlands and had a higher density of larvae than other wetlands at the sampling time. However, F_{ST} between the wetlands of the Rainbow Bay cluster was very low ($F_{ST} < 0.01$) and may be leading to biased estimates at this low level of differentiation (Faubet *et al.* 2007). Further exploration with migrate-n may help elucidate if these patterns are real, or the result of low genetic differentiation between wetlands.

For *A. talpoideum,* the only wetland yielding finite estimates of N_e was Rainbow Bay, which showed a decrease in N_e over time. I failed to detect this pattern in an earlier analysis of

this data using a different subset of SNPs (Nunziata *et al.* 2016). It is possible that the large number of SNPs (45,027 SNPs) used for N_e estimation in (Nunziata *et al.* 2016) produced a greater noise-to-signal ratio that dissipated when a more stringently filtered and smaller SNP dataset was used (15,124 SNPs used here). I am currently subsampling the dataset to assess the cause of this discrepancy. Overall, these results indicate that the entire metapopulation of *A. talpoideum* across the SRS is not undergoing the population decline I detected previously at the Rainbow Bay wetland, and that source-sink dynamics can be inferred with genomic data. This knowledge is useful from a conservation and management context where efforts should be focused on source populations. Demographic inference methods can serve as potential tools to distinguish source from sink populations, and to estimate the spatial extent of population declines where long-term field data is absent.

Landscape genomics

Species of *Ambystoma* salamanders have long been thought to have strong natal philopatry, breeding site fidelity, and be poor dispersers (Gamble *et al.* 2007). In several population genetic studies of *A. opacum* strong population structure was found at a localized scale, providing support for these hypotheses (Greenwald *et al.* 2009; Whiteley *et al.* 2015). I found evidence for IBD in *A. opacum*, which is commonly observed in amphibian populations and reflects that gene flow is geographically restricted (Funk *et al.* 2005; Spear *et al.* 2005; Wang & Summers 2010). Isolation-by-distance was not detected in *A. talpoideum,* suggesting individuals may travel and disperse further than *A. opacum.* This result is supported by a study of the terrestrial distribution of these salamanders surrounding Ginger's Bay. *Ambystoma talpoideum* individuals were found at greater distances from the wetland than *A. opacum* (Scott

et al. 2013) and although there was no distinguishing between emigrating and dispersing individuals, it implies differences in the average movement distances between the two species. Both of these species may be predominantly philopatric, but even a small amount of dispersal may equate to functional gene flow that homogenizes genetic diversity in neighboring breeding wetlands, as evident in wetlands within the Rainbow Bay cluster.

Estimates of migration from Bayesass indicate no changes in migration temporally as population size has changed in both species at Rainbow Bay. For *A. opacum* estimates of pairwise F_{ST} were similar over time for both Rainbow and Gingers Bay, and Bayesian clustering plots revealed similar genetic structure between sampling time points. However, for *A. talpoideum* distinct genetic changes are obvious at Rainbow Bay, while pairwise F_{ST} and genetic structure appear stable over time at Ginger Bay. The contemporary Rainbow Bay population appears to be derived from individual with mixed ancestry versus the historical sample which appears to be a homogeneous genetic cluster, additionally, pairwise F_{ST} between Rainbow Bay and other wetlands appeared to decline over time. One possible explanation is that as the population has declined at Rainbow Bay, a greater proportion of the breeders as Rainbow Bay originated from the surrounding wetlands. It will be interesting to compare capture-mark-recapture data to these results to see if indeed there is a change in proportion of migrants breeding at Rainbow Bay over time.

These results have clear conservation implications on the extent of population structure and potential gene flow, and highlight the utility of a genomic perspective on fine-scale landscape genetics. In a study published a decade ago assessing population genetic structure in these two species across the SRS with AFLP markers, there was no structure detected (Kinkead *et al.* 2007). It is likely that the level of resolution offered by AFLP markers does not provide

insight at such a localized scale. The contrasting results between Kinkead et al. (2007) and the results presented here highlight the increased resolution into fine-scale population processes a genomic perspective can offer.

However, while genomic datasets may provide increased resolution into fine-scale processes they create additional computational problems in their analyses. As outlined in the introduction, accurate genetically based estimates of N_e within local populations should account for population structure. Here, I estimate local N_e and migration between subpopulations separately and speculate on their relationship, but have faced challenges while attempting to simultaneously estimate N_e and migration. In my previous study, I employ the program fastsimcoal2 to infer recent demography at Rainbow Bay; however the program assumes a tree-like structure of populations, a framework that is likely violated when applied to a metapopulation. I am currently estimating N_e and migration in migrate-n, but when trying to take full advantage of the thousands of loci and landscape level sampling, the computational effort needed is staggering and impractical. This highlights the limitations that currently exist within the field, where data generation is quickly out pacing the ability to analyze these datasets.

Genetic diversity

Evolutionary processes operating at both contemporary and historical time scales have likely both acted to structure the genetic diversity of these breeding populations across the SRS. My results indicate genetic diversity was significantly higher in *A. opacum* than in *A. talpoiduem,* across all sampled wetlands. Genetic diversity in wetland-breeding salamanders has been correlated with population connectivity and the size or condition of the wetlands (Cosentino *et al.* 2012; Wang *et al.* 2011). Because I sampled wetlands where both species co-occur and

migration rates appear similar between both species, observed differences in genetic diversity between species are likely the result of long-term population processes and historical demography. To answer the question of why genetic diversity differs between the two species, it will be necessary to take a comparative approach from a phylogeographic perspective.

Table 5.3. Genetic diversity statistics calculated for *Ambystoma opacum* and *A. talpoideum* at wetlands across the Savannah River Site. Included are average observed heterozygosity per locus (Het$_{obs}$), Wright's inbreeding coefficient (F$_{IS}$), and the average nucleotide diversity (π) with standard errors for each parameter estimate.

Species	Wetland	Year	# of individuals	Lifestage	Het$_{obs}$	StdErr	π	StdErr	F$_{IS}$	StdErr
A talpoideum	Rainbow	1984	24	M	0.1455	0.0012	0.1771	0.0013	0.098	0.0229
A talpoideum	Rainbow	2011	24	M	0.1538	0.0012	0.1824	0.0012	0.0964	0.0211
A talpoideum	Gingers	Pre-2000	13	A/M	0.1569	0.0013	0.178	0.0013	0.0607	0.0125
A talpoideum	Gingers	Post-2000	11	M	0.1573	0.0014	0.176	0.0014	0.0495	0.0105
A talpoideum	Bay 92	2014/2015	12	A/M	0.1578	0.0009	0.1912	0.0009	0.0943	0.0088
A talpoideum	Flamingo	2015	12	L	0.1502	0.0009	0.1847	0.0009	0.0983	0.0074
A talpoideum	NPR	2016	15	L	0.1555	0.0009	0.1872	0.001	0.0969	0.0102
A talpoideum	Pickerel	2016	2	L	0.1523	0.0015	0.1723	0.0015	0.0299	0.002
A talpoideum	Lindas	2016	4	L	0.1546	0.0012	0.1857	0.0012	0.0602	0.0037
A opacum	Rainbow	1993	12	M	0.2178	0.0012	0.235	0.0011	0.0459	0.0094
A opacum	Rainbow	2013	12	M	0.228	0.0013	0.2341	0.0011	0.0197	0.0085
A opacum	Gingers	Pre-2000	13	A/M	0.2293	0.0015	0.2413	0.0013	0.0355	0.0121
A opacum	Gingers	Post-2000	11	A/M	0.2431	0.0018	0.2359	0.0015	-0.0131	0.0101
A opacum	Bay_92	2015	12	L	0.2406	0.0011	0.2443	0.001	0.0136	0.0064
A opacum	Flamingo	2014	12	A	0.199	0.0013	0.1912	0.0011	-0.0155	0.0084
A opacum	Pickerel	2016	12	L	0.2286	0.0011	0.2398	0.001	0.0338	0.0069
A opacum	Bullfrog	2016	13	L	0.2292	0.0011	0.2403	0.001	0.0344	0.0059
A opacum	NPR	2016	11	L	0.231	0.0012	0.2382	0.001	0.0217	0.0066

Table 5.2. Effective population size estimates for *Ambystoma opacum* and *A. talpoideum* at wetlands across the Savannah River Site.

Species	Wetland	Year	N_e (95% CI)
A talpoideum	Rainbow	1984	2011.1 (1555.1-2843.4)
A talpoideum	Rainbow	2011	12.5 (12.4-12.5)
A talpoideum	Gingers	Pre-2000	Infinite
A talpoideum	Gingers	Post-2000	Infinite
A talpoideum	Bay 92	2014/2015	Infinite
A talpoideum	Flamingo	2015	Infinite
A talpoideum	NPR	2016	Infinite
A talpoideum	Pickerel	2016	Infinite
A talpoideum	Lindas	2016	Infinite
A opacum	Rainbow	1993	Infinite
A opacum	Rainbow	2013	1000.7 (814.0-1298.0)
A opacum	Gingers	Pre-2000	Infinite
A opacum	Gingers	Post-2000	Infinite
A opacum	Bay_92	2015	Infinite
A opacum	Flamingo	2014	Infinite
A opacum	Pickerel	2016	178.5 (172.6-184.9)
A opacum	Bullfrog	2016	Infinite
A opacum	NPR	2016	11.7 (11.6-11.7)

Table 5.3. Pairwise F_{ST} values as calculated by Stacks (above the diagonal) and Arlequin (below the diagonal) for *Ambystoma*

talpoideum. Absolute values differ because of differences in how the programs treat missing data and nonpolymorphic sites.

Significance was determined with 10,000 permutations in Arlequin, values in bold are significant at the 0.05 level.

	RB_1984	RB_2011	GB_Pre-2000	GB_Post-2000	Bay92	Flamingo	NPR	Pickerel	Lindas
RB_1984	-	0.042	0.055	0.069	0.045	0.046	0.046	0.059	0.057
RB_2011	**0.132**	-	0.057	0.064	0.035	0.037	0.029	0.039	0.039
GB_Pre-2000	-0.050	**0.140**	-	0.030	0.045	0.043	0.040	0.067	0.059
GB_Post-2000	-0.010	**0.210**	-0.017	-	0.049	0.047	0.043	0.080	0.069
Bay92	**0.464**	**0.129**	**0.504**	**0.558**	-	0.041	0.033	0.056	0.050
Flamingo	0.024	0.003	0.036	0.116	**0.267**	-	0.036	0.059	0.053
NPR	**0.103**	-0.026	**0.138**	**0.218**	**0.181**	-0.023	-	0.043	0.040
Pickerel	**0.752**	**0.423**	**0.869**	**0.921**	0.111	**0.597**	**0.524**	-	0.128
Lindas	0.130	-0.045	0.109	0.218	0.094	-0.019	-0.047	0.355	-

Table 5.4. Pairwise F_{ST} values as calculated by Stacks (above the diagonal) and Arlequin (below the diagonal) for *Ambystoma*

opacum. Absolute values differ because of differences in how the programs treat missing data and nonpolymorphic sites. Significance

was determined with 10,000 permutations in Arlequin, values in bold are significant at the 0.05 level.

	RB_1993	RB_2013	GB_Pre-2000	GB_Post-2000	Bay_92	Flamingo	Pickerel	Bullfrog	NPR
RB_1993	-	0.029	0.055	0.062	0.070	0.074	0.036	0.031	0.039
RB_2013	0.000	-	0.055	0.063	0.070	0.077	0.037	0.033	0.039
GB_Pre2000	**0.058**	**0.054**	-	0.029	0.066	0.086	0.063	0.058	0.064
GB_Post2000	**0.070**	**0.068**	-0.006	-	0.071	0.098	0.071	0.066	0.073
Bay_92	**0.091**	**0.091**	**0.075**	**0.076**	-	0.101	0.067	0.066	0.069
Flamingo	**0.094**	**0.094**	**0.109**	**0.129**	**0.165**	-	0.072	0.067	0.081
Pickerel	**0.004**	**0.005**	**0.061**	**0.072**	**0.092**	**0.097**	-	0.027	0.033
Bullfrog	-0.003	-0.002	**0.057**	**0.065**	**0.090**	**0.080**	0.001	-	0.031
NPR	**0.009**	**0.007**	**0.057**	**0.073**	**0.083**	**0.114**	0.005	-0.001	-

130

Table 5.5. BAYESASS estimates of the mean posterior distribution for contemporary migration rates from a source wetland to a destination wetland, with 95% HPD in parentheses. Nonsignificant dispersal rates are not shown.

Species	Source → Destination	Dispersal Rate
A. opacum	Bullfrog→RB_1993	0.206(0.139-0.271)
A. opacum	Bullfrog→ Pickerel	0.206(0.138-0.271)
A. opacum	Bullfrog → RB_2013	0.200(0.128-0.265)
A. opacum	Bullfrog → NPR	0.180(0.113-0.254)
A. talpoideum	NPR → Linda's	0.103(0.029-0.184)
A. talpoideum	NPR → Pickerel	0.091(0.016-0.174)
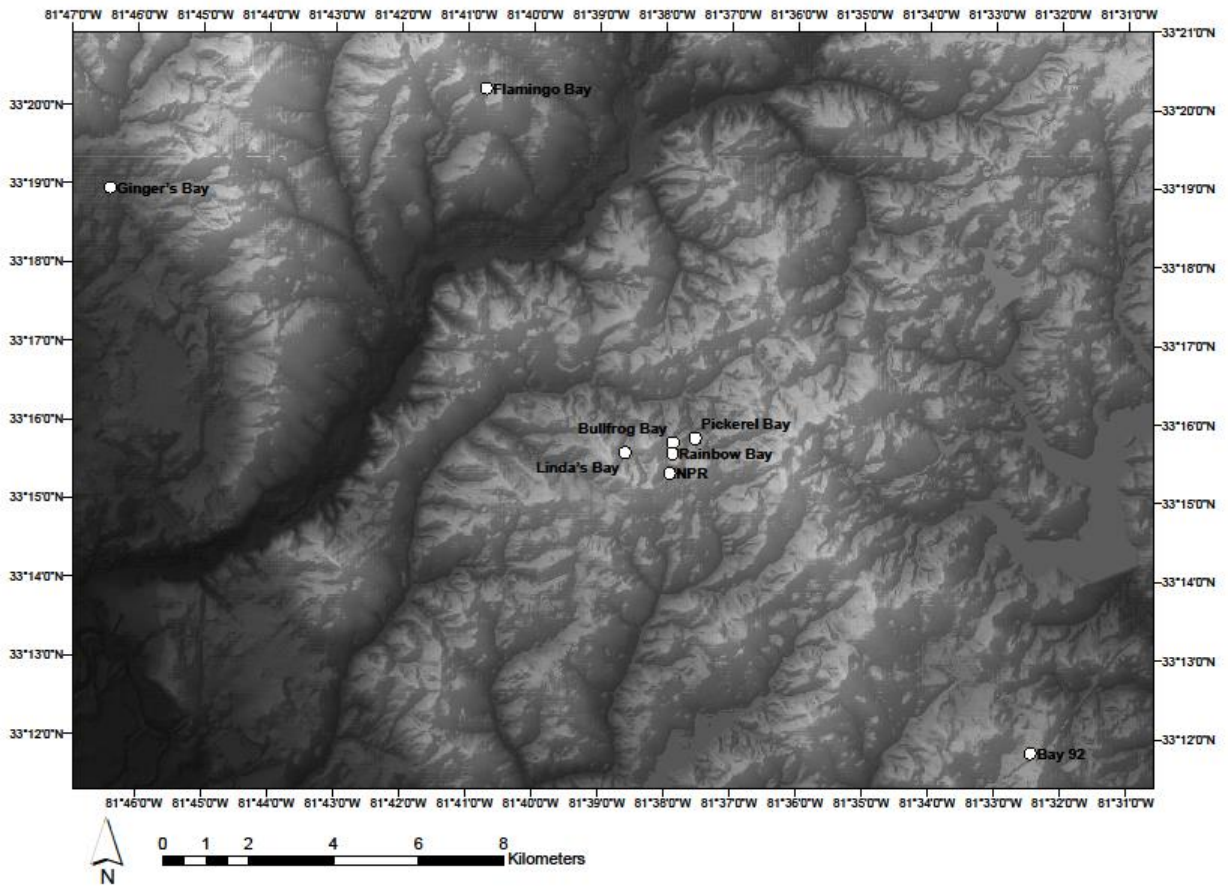
Figure 5.1. Sampling locations of *Ambystoma opacum* and *A. talpoideum* across the Savannah

River Site (SRS).
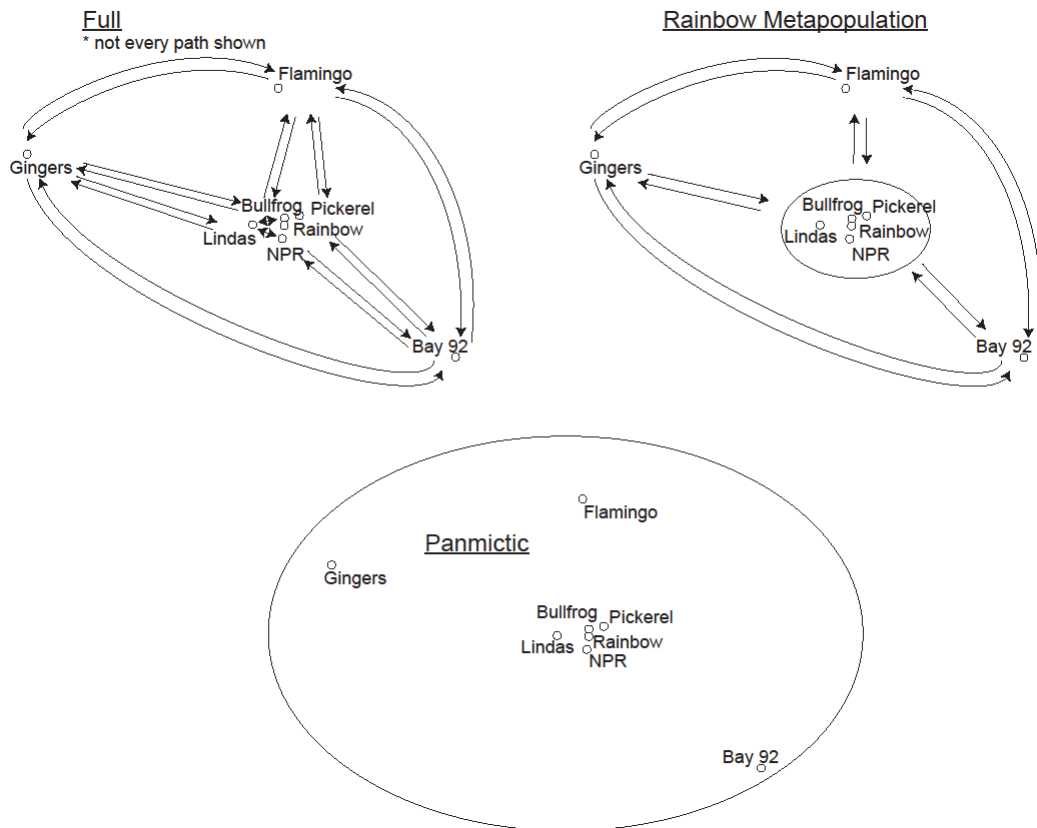
Figure 5.2. Geneflow models compared in MIGRATE-n. Models were compared for both *Ambystoma opacum* and *A. talpoideum*, and using a historic and contemporary dataset. See text for detailed description of datasets and methods used.
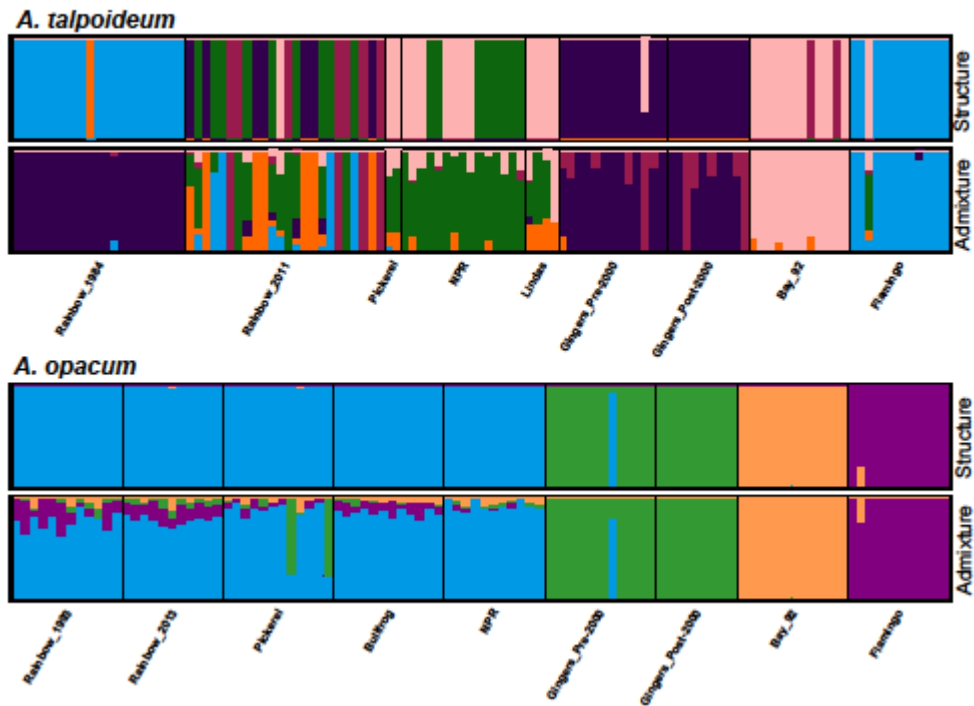
Figure 5.3. Bayesian clustering of *Ambystoma opacum* and *A. talpoideum* individuals for the most likely number of clusters (*A. opacum* K =4; *A. talpoideum* K=6), identified by the Bayesian clustering programs fastSTRUCTURE and ADMIXTURE.

CHAPTER SIX

CONCLUSIONS AND FUTURE DIRECTIONS

This dissertation provides a framework for assessing the application of genomic datasets to infer recent population demography. I focus on a well-studied salamander community, however my results are not limited to amphibians but have implications for a range of species of conservation concern. I demonstrate that both linkage-disequilibrium and coalescent approaches can be successfully applied to infer recent population demographic history at an ecological time scale (i.e., tens of generations), both in the simulation of ideal populations (Chapter 2) and in real populations (Chapters 3). These methods offer the potential to monitor the demographic response of populations facing potential threats from climate change, or other environmental disturbance, and appear to track contemporary demography even in long-lived species with complex life-history, such as amphibians (Chapter 4). This work highlights the amazing resolution possible from genome-scale data, providing insight into short term demographic processes, even within a single subpopulation of a large and interconnected metapopulation (Chapter 5). This dissertation integrates long-term mark-recapture data, simulation studies, and population genomics from multiple spatial and temporal scales. Below I discuss how my work provides new information on monitoring contemporary demographic history of populations with genomic data, limitations, and future directions.

Simulation studies

Characterizing the demographic history of populations, both on contemporary and historical time scales, is a long and active area of research in evolution and ecology. The

estimation of contemporary N_e is complicated by multiple eco-evolutionary processes acting simultaneously within populations, as well as the high numbers of individuals and genetic markers needed for fine-scale resolution. These complications may be overcome with the growth of affordable and scalable next-generation sequencing technologies and development of computationally efficient coalescent modeling software capable of evaluating complex models of population history. However, with any method, it is important to evaluate potential biases and limitations. In Chapter 2, I developed a simulation framework to evaluate the application of restriction-associated DNA sequencing (RADseq) in N_e estimation and the characterization of population declines. Two pieces of important knowledge were found in this chapter. First, I found that RADseq datasets, even with the inclusion of possible null alleles, have minimal biases when used for N_e estimation. Second, I found that LD-based methods generally outperformed coalescent methods in quick detection of population declines, and accuracy of N_e estimates. One major limitation of this study is the computational effort required for coalescent simulations precluded the inclusion of a limitation analysis to evaluate the number of SNPs needed to produce accurate N_e estimates. In my future research, I will subsample my simulated SNP dataset to try to find a minimum threshold of SNPs needed for accurate N_e estimation using coalescent methods.

The results in Chapter 2 point to the potential of genomics to yield insight into recent population demography; however few populations in nature will meet the criteria of an idealized population that was modeled (equal sex ratio, non-overlapping generations, random mating, no mutation or selection). The complex and dynamic nature of natural populations, such as population size fluctuations and the presence of migration, are complicating factors in the estimation and interpretation of N_e. Pond-breeding amphibians have complex life-histories

136

where catastrophic reproductive failure (CRF) is a common property of populations, which may lead to difficulty distinguishing long-term population trends from short term fluctuations. In Chapter 4, I developed a simulation framework to model *Ambystoma opacum* populations under multiple levels of CRF, and examine if correlations exist between N_e and census size. Two pieces of important knowledge were found in this chapter. First, the effective number of breeders (N_b) was found to be correlated with census size more often than N_e, suggesting that targeted sampling of single-cohort newborns in long-lived species may be preferred over samples of adults for monitoring recent population size trends. Second, N_b tracked the overall long-term census size and not short term fluctuation across most levels of CRF, helping to validate the potential of these methods for genomic monitoring of populations. A limitation of this study is the memory requirements of the modeling software precluded the use of large SNP datasets. I have currently developed a new more computationally efficient model incorporating SNP markers, and in the future will use these new datasets to test alternate models of demographic history within a coalescent model-testing framework.

Validation in empirical systems

　　While the above simulation studies suggest the potential of genomic datasets to offer insight into recent population history, validation in empirical systems is critical. In Chapter 3, I apply the LD-based and coalescent-based N_e estimation methods tested in Chapters 2 and 4 in an empirical system with two species of salamanders with known population decline (*A. talpoideum*) and expansion (*A. opacum*). In both species, the correct population size trend was detected using coalescent-based demographic inference, but results were only robust in *A. talpoideum.* I found estimates of LD-based N_e reached infinity for both sampled years in each

species, but further filtering of the dataset in Chapter 5 produced LD-based estimates correlated with census-size for *A. talpoideum.* This discrepancy in results highlights the need for stringent data filtering when using large genomic-datasets, and I am currently further investigating how filtering conditions impact LD-based estimates. Overall, the results in Chapter 4 help to confirm that even very recent population size changes create population genomic signatures that can be inferred with coalescent modeling techniques and LD-based N_e estimates. The biggest advantage of coalescent-based demographic inference is the ability to test complex models of population history, including migration, population size changes, and temporal sampling. However, in this chapter sampling was restricted to a single wetland, Rainbow Bay, and by only sampling from a single subpopulation it was impossible to determine whether the results were reflecting subpopulation processes, or processes of the overall metapopulation.

The continuous distribution of species across landscapes not only complicates the interpretation of geographic scale reflected in N_e estimates, but also N_e estimation itself when migration is not accounted for and its influence is attributed to drift. In Chapter 5, I extend sampling of *A. opacum* and *A. talpoideum* to 7 additional wetlands across the landscape and estimate population structuring and N_e at these individual wetlands. Estimates of N_e at wetlands surrounding Rainbow Bay suggest large and stable populations of both *A. opacum* and *A. talpoideum*, helping to confirm that the decline observed in *A. talpoideum* was a subpopulation process at the Rainbow Bay wetland and the entire metapopulation does not appear to be in decline. Preliminary results suggest that the population decline of *A. talpoiduem* at Rainbow Bay has increased the proportion of migrants to resident breeders at the wetland and created a shift in population structuring, while preliminary results for *A. opacum* suggest temporal stability of population structure. Ongoing work in this system includes testing alternate models of migration

in the program migrate-n and examining if migration patterns have changed over time. While this study is currently ongoing, preliminary results suggest that demographic history can be detected at the subpopulation (wetland) level using genomic datasets, even over recent time and fine geographic scales with the presence of migration.

Conclusions

To conclude, in this dissertation I aimed to assess whether genomic methods could be applied to accurately estimate demographic processes taking place at an ecological scale, from a theoretical standpoint and in natural populations. I was able to take advantage of the amazing wealth of a long-term field dataset on an amphibian community and the explosion of next-generation-sequencing to develop genomic markers, and integrate information from life-history, mark-recapture, and genomics. My findings suggest the ability of genomic data to reconstruct recent demographic changes in complex and dynamic populations, which can have important applications to conservation biology, and ultimately can help us elucidate the effects of environmental disturbances in the demography of endangered and invasive species. Although I am not advocating that molecular methods can replace long-term field studies, I feel that they are a great complement to gain insight into population processes. I believe that the studies in this dissertation are merely the forefront of an explosion of studies using a temporal samples and a genome-scale approach to infer recent population demography, which may be an important step forward in conservation biology.

Table A.1. Number of times $\widehat{N}_e$ was estimated as "Infinite" out of 100 replicate runs in NeEstimator using the LD-based RAD mutation dataset simulated under a stable population size. Numbers are provided from the five different time sampling points (t_0 - t_{20}), and are broken down across the different initial population sizes (N), individual sample size (n), and minimum allele frequency cutoff (MAF_{cut}).

		t_0 (& MAF_{cut})			t_5 (& MAF_{cut})			t_{10} (& MAF_{cut})			t_{15} (& MAF_{cut})			t_{20} (& MAF_{cut})		
		0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05
N = 250	n = 15	21	21	0	22	22	0	23	23	0	16	16	0	19	19	0
	n = 30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	n = 60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N = 500	n = 15	79	86	7	66	74	5	66	73	6	56	65	5	55	60	2
	n = 30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	n = 60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N = 1000	n = 15	89	90	29	94	96	28	90	92	20	92	98	13	86	88	19
	n = 30	7	0	0	11	2	0	5	1	0	8	1	0	7	1	0
	n = 60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table A.2. Number of times $\widehat{N}_e$ was estimated as "Infinite" out of 100 replicate runs in NeEstimator using the 10% missing dataset simulated under a stable population size. Numbers are provided from the five different time sampling points (t_0 - t_{20}), and are broken down across the different initial population sizes (N), individual sample size (n), and minimum allele frequency cutoff (MAF_{cut}).

		t_0 (& MAF_{cut})			t_5 (& MAF_{cut})			t_{10} (& MAF_{cut})			t_{15} (& MAF_{cut})			t_{20} (& MAF_{cut})		
		0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05
N = 250	n = 15	24	24	0	21	21	1	21	21	0	17	17	1	19	19	0
	n = 30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	n = 60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N = 500	n = 15	87	87	10	77	77	5	78	78	13	68	68	7	63	63	6
	n = 30	1	0	0	0	0	0	1	0	0	2	0	0	0	0	0
	n = 60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N = 1000	n = 15	90	90	34	96	96	37	93	93	26	97	97	23	88	88	24
	n = 30	30	0	0	20	1	2	19	1	1	13	1	1	24	0	0
	n = 60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table A.3. Number of times $\widehat{N}_e$ was estimated as "Infinite" out of 100 replicate runs in NeEstimator using the 50% missing dataset simulated under a stable population size. Numbers are provided from the five different time sampling points (t_0 - t_{20}), and are broken down across the different initial population sizes (N), individual sample size (n), and minimum allele frequency cutoff (MAF_{cut}).

		t_0 (& MAF_{cut})			t_5 (& MAF_{cut})			t_{10} (& MAF_{cut})			t_{15} (& MAF_{cut})			t_{20} (& MAF_{cut})		
		0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05
N = 250	n = 15	35	35	0	32	32	3	24	24	1	19	19	0	19	19	0
	n = 30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	n = 60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N = 500	n = 15	92	92	23	80	80	14	82	82	18	72	72	14	75	75	20
	n = 30	2	0	0	0	0	0	1	0	0	2	0	0	0	0	0
	n = 60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N = 1000	n = 15	93	93	49	97	97	49	94	94	49	98	98	43	89	89	49
	n = 30	30	1	0	21	3	1	20	2	1	19	0	0	25	0	0
	n = 60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table A.4. Number of times that a declining population trend was correctly identified out of 100 replicate runs for LD-based analysis in NeEstimator under a declining population model ($\lambda = 0.9$). Results are presented for increasing intervals of time and by the combination of population size (N), individual sampling level (n), and minor allele frequency cutoff (MAF_{cut}). Results are based on the analysis of the LD RAD mutation datasets.

		$t_0 - t_5$			$t_0 - t_{10}$			$t_0 - t_{15}$			$t_0 - t_{20}$		
		(& MAF_{cut} level)			(& MAF_{cut} level)			(& MAF_{cut} level)			(& MAF_{cut} level)		
		0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05
N = 250	n = 15	51	51	69	65	65	82	72	72	94	73	73	100
	n = 30	86	81	80	99	95	95	100	100	100	-	-	-
	n = 60	94	94	94	100	100	100	-	-	-	-	-	-
N = 500	n = 15	16	11	53	28	19	73	47	27	83	40	30	89
	n = 30	79	72	69	91	85	84	100	98	98	100	99	99
	n = 60	79	82	85	97	98	98	100	100	100	-	-	-
N = 1000	n = 15	0	0	42	0	0	49	1	2	62	2	2	72
	n = 30	66	66	67	82	80	82	89	93	96	91	97	100
	n = 60	71	73	72	96	95	96	100	100	100	100	100	100

For some parameter combinations, there were insufficient numbers of individuals for target n (-).

Table A.5. Number of times that a declining population trend was correctly identified out of 100 replicate runs for LD-based analysis in NeEstimator under a declining population model ($\lambda = 0.9$). Results are presented for increasing intervals of time and by the combination of population size (N), individual sampling level (n), and minor allele frequency cutoff ($\mathrm{MAF_{cut}}$). Results are based on the analysis of datasets with 50% missing data.

		$t_0 - t_5$			$t_0 - t_{10}$			$t_0 - t_{15}$			$t_0 - t_{20}$		
		(& $\mathrm{MAF_{cut}}$ level)			(& $\mathrm{MAF_{cut}}$ level)			(& $\mathrm{MAF_{cut}}$ level)			(& $\mathrm{MAF_{cut}}$ level)		
		0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05
N = 250	n = 15	45	45	68	58	58	81	64	64	94	65	65	99
	n = 30	86	83	79	99	94	93	100	100	100	-	-	-
	n = 60	93	94	93	100	100	100	-	-	-	-	-	-
N = 500	n = 15	8	8	45	15	15	57	20	20	70	22	22	76
	n = 30	78	75	72	93	86	86	100	99	98	100	99	99
	n = 60	82	80	83	98	98	98	100	100	100	-	-	-
N = 1000	n = 15	0	0	20	0	0	27	1	1	37	1	1	48
	n = 30	55	65	67	71	80	79	74	92	92	76	97	98
	n = 60	70	73	73	94	94	96	100	100	100	100	100	100

For some parameter combinations, there were insufficient numbers of individuals for target n (-).

Table A.6. Number of times $\widehat{N}_e$ was estimated as "Infinite" out of 100 replicate runs using LD-based analysis in NeEstimator under a declining population model ($\lambda = 0.9$). Results are presented for five different sampling points (t_0 through t_{20}) and by the combination of population size (N), individual sampling level (n), and minor allele frequency cutoff (MAF_{cut}). Results are based on the analysis of LD RAD mutation datasets.

		t_0 (& MAF_{cut} level)			t_5 (& MAF_{cut} level)			t_{10} (& MAF_{cut} level)			t_{15} (& MAF_{cut} level)			t_{20} (& MAF_{cut} level)		
		0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05
N = 250	n = 15	27	27	0	1	1	0	0	0	0	0	0	0	0	0	0
	n = 30	0	0	0	0	0	0	0	0	0	0	0	0	-	-	-
	n = 60	0	0	0	0	0	0	0	0	0	-	-	-	-	-	-
N = 500	n = 15	59	69	2	26	43	1	5	6	0	0	0	0	0	0	0
	n = 30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	n = 60	0	0	0	0	0	0	0	0	0	0	0	0	-	-	-
N = 1000	n = 15	98	98	22	79	83	10	43	52	1	12	15	0	1	1	0
	n = 30	9	3	0	1	0	0	0	0	0	0	0	0	0	0	0
	n = 60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

For some parameter combinations, there were insufficient numbers of individuals for target n (-).

Table A.7. Number of times $\widehat{N}_e$ was estimated as "Infinite" out of 100 replicate runs using LD-based analysis in NeEstimator under a declining population model ($\lambda = 0.9$). Results are presented for five different sampling points (t_0 through t_{20}) and by the combination of population size (N), individual sampling level (n), and minor allele frequency cutoff (MAF_{cut}). Results are based on the analysis of datasets with 10% missing data.

		t_0 (& MAF_{cut} level)			t_5 (& MAF_{cut} level)			t_{10} (& MAF_{cut} level)			t_{15} (& MAF_{cut} level)			t_{20} (& MAF_{cut} level)		
		0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05
N = 250	n = 15	28	28	0	1	1	0	0	0	0	0	0	0	0	0	0
	n = 30	0	0	0	0	0	0	0	0	0	0	0	0	-	-	-
	n = 60	0	0	0	0	0	0	0	0	0	-	-	-	-	-	-
N = 500	n = 15	71	71	6	43	43	2	8	8	0	0	0	0	0	0	0
	n = 30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	n = 60	0	0	0	0	0	0	0	0	0	0	0	0	-	-	-
N = 1000	n = 15	98	98	30	87	87	12	52	52	1	18	18	0	1	1	1
	n = 30	22	3	3	2	0	0	0	0	0	0	0	0	0	0	0
	n = 60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

For some parameter combinations, there were insufficient numbers of individuals for target n (-).

Table A.8. Number of times $\widehat{N}_e$ was estimated as "Infinite" out of 100 replicate runs using LD-based analysis in NeEstimator under a declining population model ($\lambda = 0.9$). Results are presented for five different sampling points (t_0 through t_{20}) and by the combination of population size (N), individual sampling level (n), and minor allele frequency cutoff (MAF$_{cut}$). Results are based on the analysis of datasets with 50% missing data.

		t_0 (& MAF$_{cut}$ level)			t_5 (& MAF$_{cut}$ level)			t_{10} (& MAF$_{cut}$ level)			t_{15} (& MAF$_{cut}$ level)			t_{20} (& MAF$_{cut}$ level)		
	MAF$_{cut}$	0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05
N = 250	n = 15	35	35	1	2	2	0	0	0	0	0	0	0	0	0	0
	n = 30	0	0	0	0	0	0	0	0	0	0	0	0	-	-	-
	n = 60	0	0	0	0	0	0	0	0	0	-	-	-	-	-	-
N = 500	n = 15	77	77	15	52	52	4	11	11	0	0	0	0	0	0	0
	n = 30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	n = 60	0	0	0	0	0	0	0	0	0	0	0	0	-	-	-
N = 1000	n = 15	99	99	47	92	92	23	59	59	4	26	26	0	1	1	0
	n = 30	24	3	2	3	0	0	0	0	0	0	0	0	0	0	0
	n = 60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

For some parameter combinations, there were insufficient numbers of individuals for target n (-).

Table A.9. Number of times a stable population was incorrectly identified as declining out of 100 replicate runs for LD-based analysis in NeEstimator under a stable population model ($\lambda = 1.0$). Results are presented for increasing intervals of time and by the combination of population size (N), individual sampling level (n), and minor allele frequency cutoff (MAF_{cut}). Results are based on the analysis of LD RAD mutation datasets.

		$t_0 - t_5$			$t_0 - t_{10}$			$t_0 - t_{15}$			$t_0 - t_{20}$		
		(& MAF_{cut} level)			(& MAF_{cut} level)			(& MAF_{cut} level)			(& MAF_{cut} level)		
		0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05
N = 250	n = 15	27	27	38	25	25	27	24	24	23	19	19	8
	n = 30	30	29	25	18	13	11	15	7	5	3	1	0
	n = 60	8	10	7	1	1	1	1	1	1	0	0	0
N = 500	n = 15	1	1	40	3	1	34	5	3	33	6	6	26
	n = 30	46	42	43	31	22	22	24	17	12	18	9	8
	n = 60	27	25	27	5	6	6	1	2	2	0	0	0
N = 1000	n = 15	1	1	19	0	0	15	0	0	15	0	0	15
	n = 30	32	38	38	28	31	26	28	26	21	20	18	13
	n = 60	21	20	23	14	16	11	3	5	5	1	1	1

Table A.10. Number of times that a population trend was incorrectly identified as declining out of 100 replicate runs for LD-based analysis in NeEstimator under a stable population model ($\lambda = 1.0$). Results are presented for increasing intervals of time and by combination of population size (N), individual sampling level (n), and minor allele frequency cutoff (MAF_{cut}). Results are based on the analysis of datasets with 50% missing data.

		$t_0 - t_5$			$t_0 - t_{10}$			$t_0 - t_{15}$			$t_0 - t_{20}$		
		(& MAF_{cut} level)			(& MAF_{cut} level)			(& MAF_{cut} level)			(& MAF_{cut} level)		
		0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05	0.01	0.02	0.05
N = 250	n = 15	21	21	41	21	21	29	18	18	21	15	15	12
	n = 30	30	28	28	19	14	12	16	9	7	5	2	2
	n = 60	8	12	9	1	1	1	1	1	1	0	0	0
N = 500	n = 15	0	0	28	0	0	25	2	2	22	3	3	18
	n = 30	46	41	45	35	22	24	27	16	15	20	10	9
	n = 60	26	25	31	6	6	5	1	2	1	0	0	0
N = 1000	n = 15	1	1	12	0	0	6	0	0	9	0	0	4
	n = 30	19	34	35	12	28	29	17	28	27	12	16	16
	n = 60	22	19	23	14	13	12	4	3	4	1	1	0
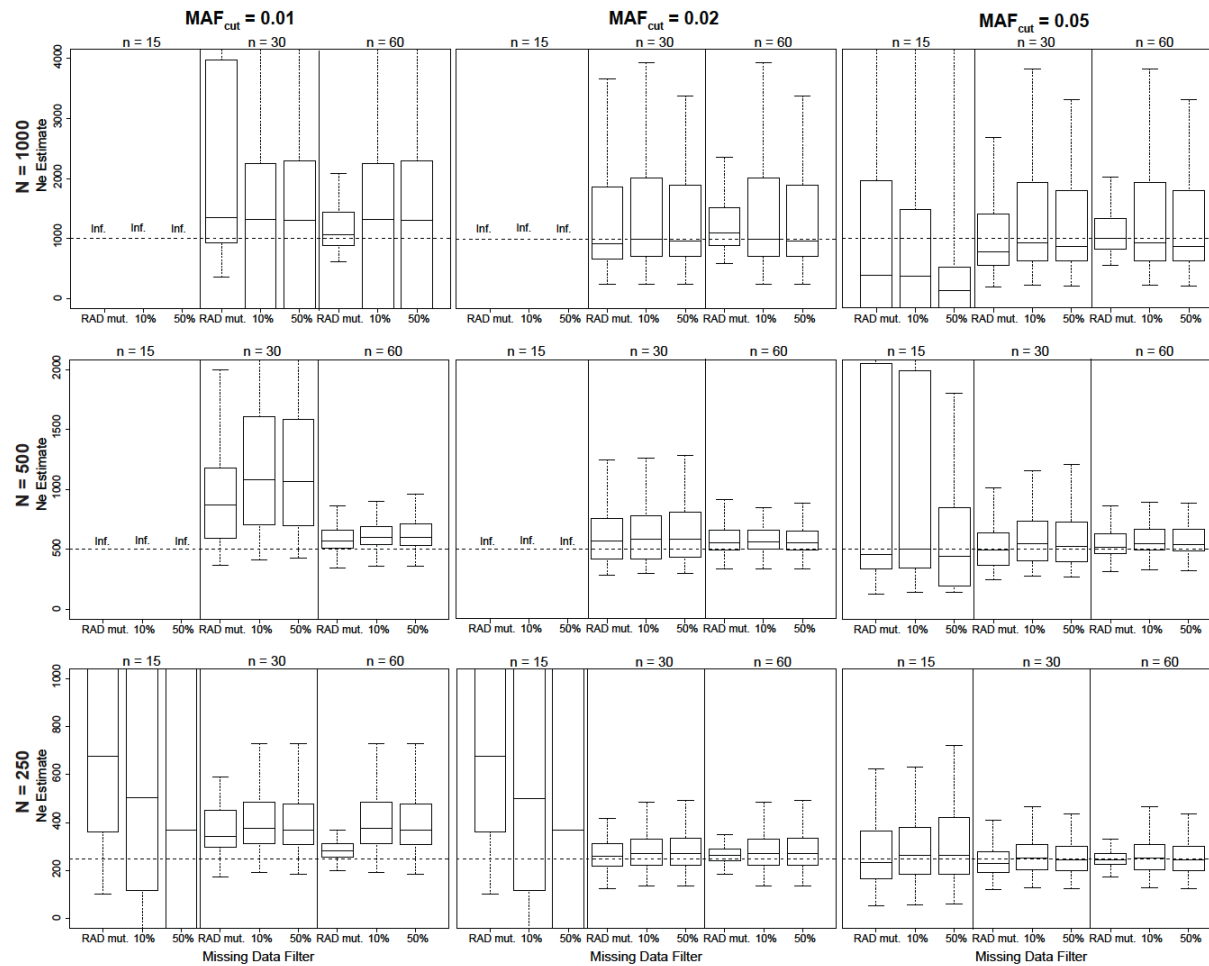
Figure A.1. Boxplots of the distribution of $\widehat{N}_e$ estimates from 100 replicate simulations for LD-based estimation at generation 20 from temporal simulations under stable population sizes ($\lambda = 1.0$). Dashed lines represent true N_e (N = 250, 500, or 1000). The left column represents filtering with MAF_{cut} of 0.01, center MAF_{cut} of 0.02 and the right column with MAF_{cut} of 0.05. Different missing data filtering strategies are shown at the bottom of each panel, with number of individuals sampled at top.
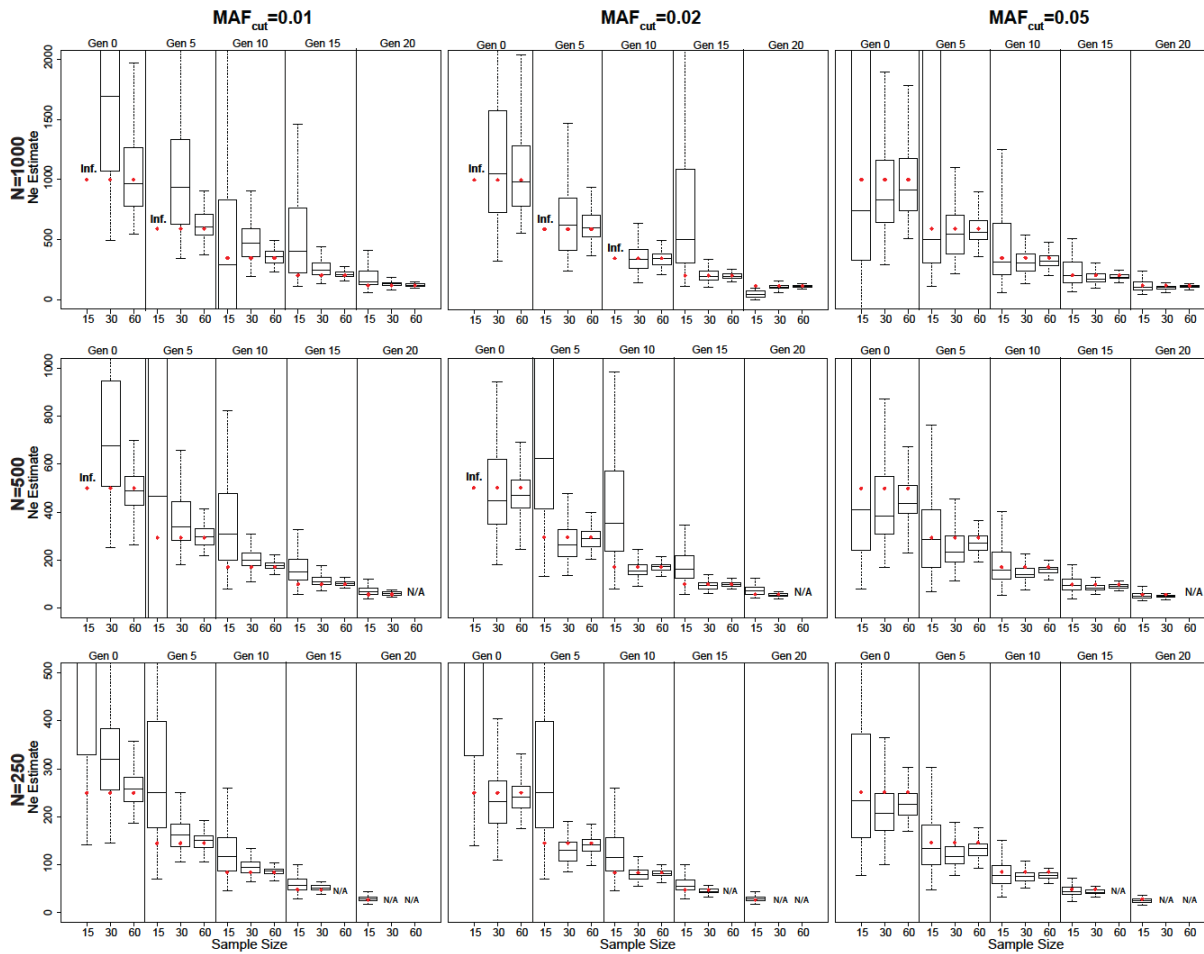
Figure A.2. Boxplots of the distribution of point estimates from 100 replicate simulations for LD-based N_e estimation from five temporal sampling points (t_0 - t_{20}) under a declining population growth model ($\lambda = 0.9$) using the *LD RAD mutation* dataset. Red dots represent true N_e over time, starting from an initial N of 1000 (top), 500 (middle) or 250 (bottom). Results are presented for analyses using different levels of individual sample size (n = 15, 30, or 60) and different MAF_{cut} levels (0.01, 0.02, or 0.05). For some parameter combinations, there were insufficient numbers of individuals for target n.
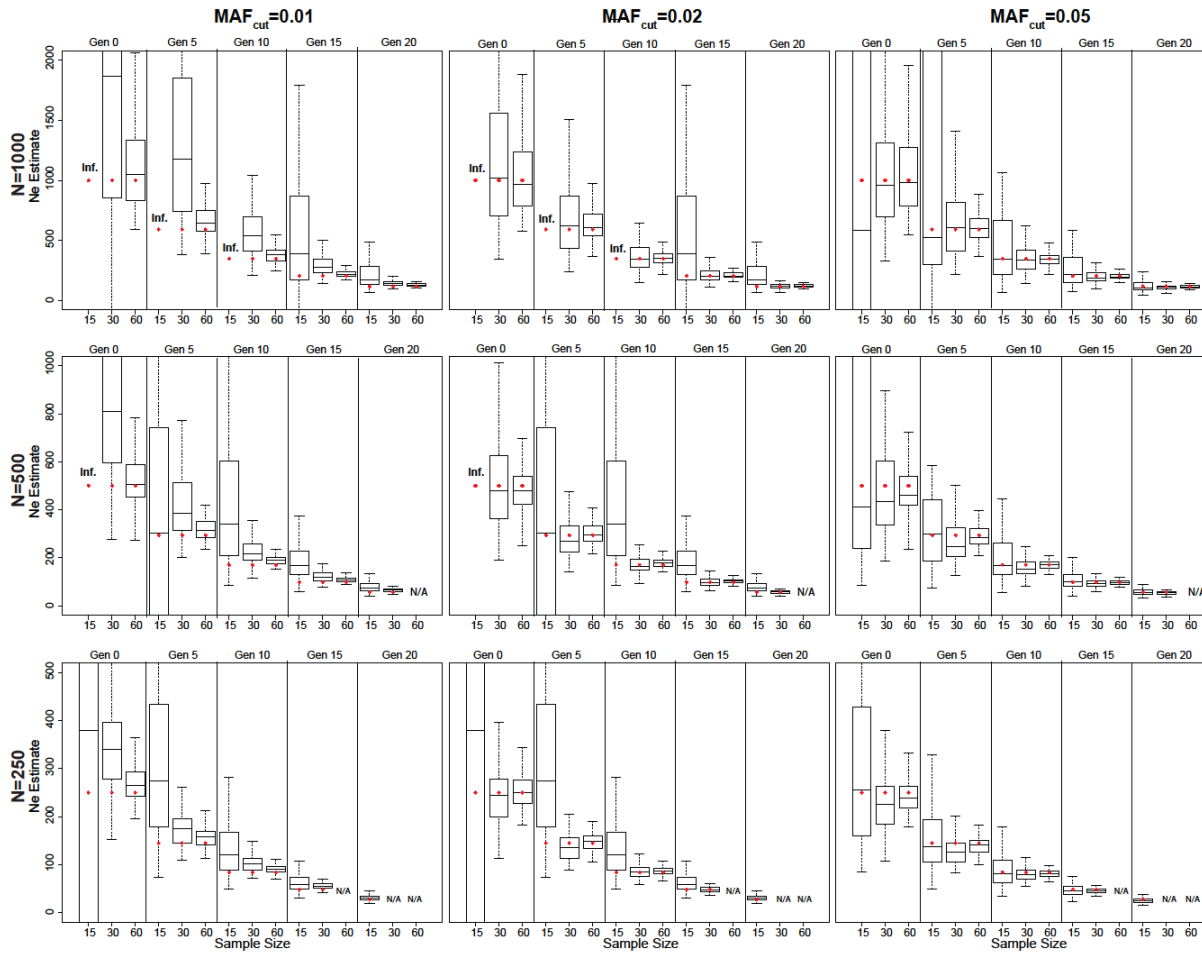
Figure A.3. Boxplots of the distribution of point estimates from 100 replicate simulations for LD-based N_e estimation from five temporal sampling points (t_0 - t_{20}) under a declining population growth model ($\lambda = 0.9$) using the 10% missing dataset. Red dots represent true N_e over time, starting from an initial N of 1000 (top), 500 (middle) or 250 (bottom). Results are presented for analyses using different levels of individual sample size (n = 15, 30, or 60) and different MAF_{cut} levels (0.01, 0.02, or 0.05). For some parameter combinations, there were insufficient numbers of individuals for target n.
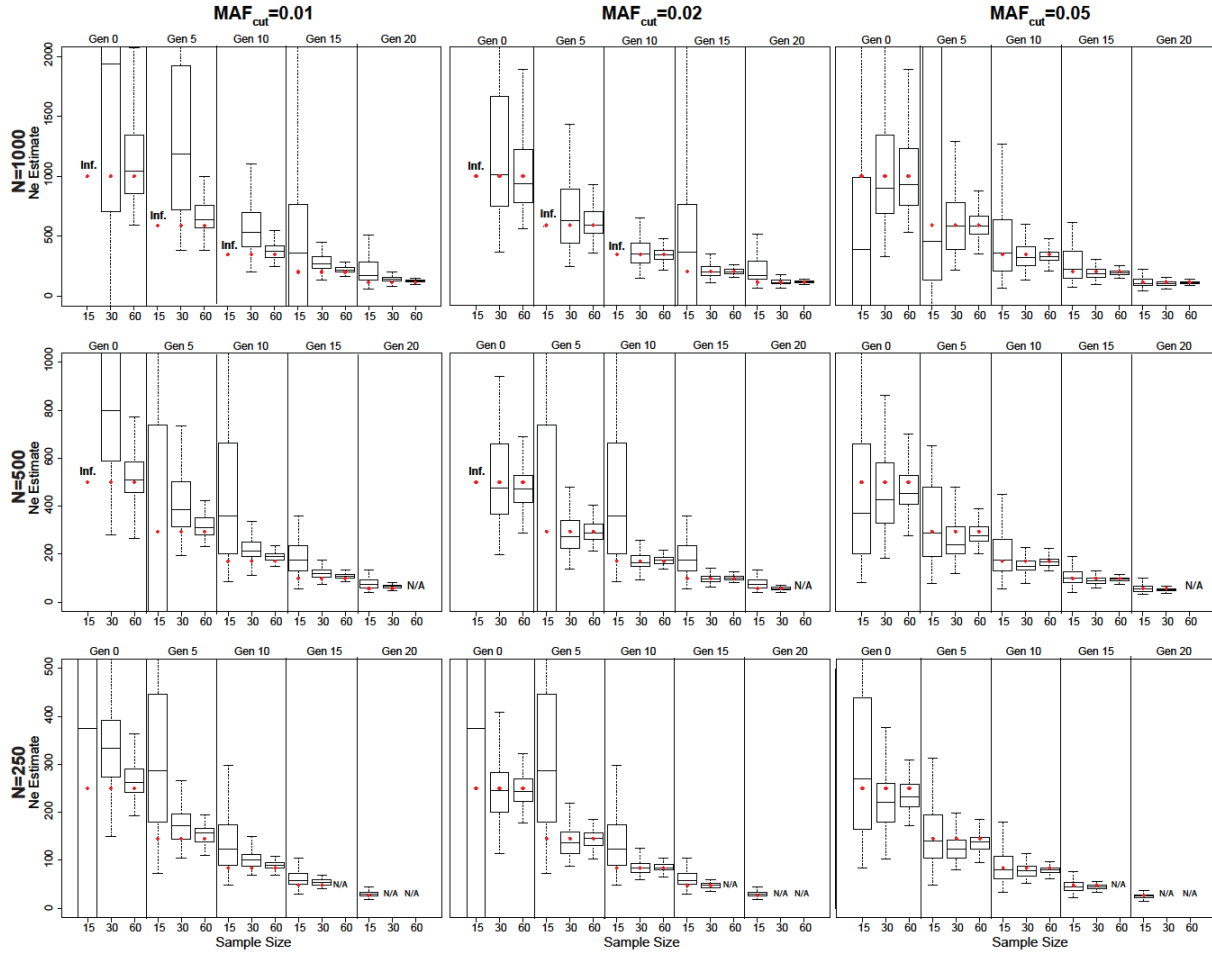
Figure A.4. Boxplots of the distribution of point estimates from 100 replicate simulations for LD-based N_e estimation from five temporal sampling points (t_0 - t_{20}) under a declining population growth model ($\lambda = 0.9$) using the 50% missing dataset. Red dots represent true N_e over time, starting from an initial N of 1000 (top), 500 (middle) or 250 (bottom). Results are presented for analyses using different levels of individual sample size (n = 15, 30, or 60) and different MAF_{cut} levels (0.01, 0.02, or 0.05). For some parameter combinations, there were insufficient numbers of individuals for target n.

APPENDIX B

Methods and Results B.1.

Outlier Detection

BayeScan was run using default settings, as well as using prior probabilities of 100 and 1000 in addition to the default 10, with an exploration of false discovery rates (FDR) of 0.05 and 0.01. Any outliers detected were then subjected to a BLASTN search of all sequences in the NCBI database, as well as against the *Ambystoma mexicanum* transcriptome v3.0 (www.ambystoma.org). BLASTN searches of both data sets used an e-value threshold of 1^{-10} with at least 85% sequence identity.

No outlier SNPs were detected for *A. opacum* with BayeScan. For *A. talpoideum,* no outliers were detected at an FDR of 0.01, but a total of nine outliers were detected at an FDR of 0.05. The same nine outliers were detected in three independent runs of BayeScan with prior probability of 10, but no outliers were detected at prior probabilities of 100 or 1000. After aligning outliers to the *A. mexicanum* transcriptome, 5 of 9 had significant alignment to expressed contigs (Table S2). None of the outlier ddRAD loci aligned to contigs corresponding to mapped markers in the *Ambystoma* linkage map. An NCBI database blastn search of these outlier ddRAD loci produced no significant matches.

Population Structure

To investigate if either species displayed subpopulation structure or evidence of admixture, we used a Bayesian clustering approach in the program fastSTRUCTURE on each sampling time point with K ranging from 1 to 3 using a "simple prior" (Raj et al. 2014). The most likely value of K was identified by calculating heuristic scores using the *chooseK.py* script in fastSTRUCTURE, and identified K=1 for each species and time point. Therefore, we considered the RB wetland to be a single breeding population for both *A. opacum* and *A. talpoideum*.

Literature Cited
Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics, **197**, 573–589.

Table B.1. Summary of Illumina reads collected for each individual. Shown are the total number of reads obtained for each individual, retained reads after filtering out low quality scores, and the mean depth of coverage after default filtering with filter set 1 (See Table 2-1 for conditions).

Species	Year	Individual ID	Total Reads	Retained Reads	Mean Depth Coverage
A. opacum	1993	Amop_RB_1993_1	8235678	5734074	17.93
A. opacum	1993	Amop_RB_1993_2	10000486	6450689	20.83
A. opacum	1993	Amop_RB_1993_3	14291878	11495878	37.18
A. opacum	1993	Amop_RB_1993_4	7369872	5755929	17.99
A. opacum	1993	Amop_RB_1993_5	8244798	6072107	18.14
A. opacum	1993	Amop_RB_1993_6	11757322	8085035	23.80
A. opacum	1993	Amop_RB_1993_7	6842698	4895880	13.80
A. opacum	1993	Amop_RB_1993_8	7518160	5440096	17.26
A. opacum	1993	Amop_RB_1993_9	6878838	5003113	15.07
A. opacum	1993	Amop_RB_1993_10	11360396	8585956	26.02
A. opacum	1993	Amop_RB_1993_11	11053706	8029195	24.16
A. opacum	1993	Amop_RB_1993_12	7015048	5207476	16.73
A. opacum	2013	Amop_RB_2013_1	9931710	7557595	24.41
A. opacum	2013	Amop_RB_2013_2	10275598	7606087	22.44
A. opacum	2013	Amop_RB_2013_3	12788398	9813339	28.27
A. opacum	2013	Amop_RB_2013_4	8000074	5732904	17.78
A. opacum	2013	Amop_RB_2013_5	8497784	5950536	18.05
A. opacum	2013	Amop_RB_2013_6	10272342	7256834	23.32
A. opacum	2013	Amop_RB_2013_7	8648848	6433796	21.61
A. opacum	2013	Amop_RB_2013_8	7109458	5142614	16.86
A. opacum	2013	Amop_RB_2013_9	4542706	3040763	8.34
A. opacum	2013	Amop_RB_2013_10	7174202	5458347	16.18
A. opacum	2013	Amop_RB_2013_11	8525550	6304070	18.84
A. opacum	2013	Amop_RB_2013_12	7788168	5858888	17.53
A. talpoideum	1984	Atalp_RB_1984_1	5201034	4970846	17.31
A. talpoideum	1984	Atalp_RB_1984_2	5285492	5045218	17.92
A. talpoideum	1984	Atalp_RB_1984_3	6423864	6170749	21.88
A. talpoideum	1984	Atalp_RB_1984_4	8239206	8005560	28.42
A. talpoideum	1984	Atalp_RB_1984_5	8897272	8616111	29.85
A. talpoideum	1984	Atalp_RB_1984_6	7011954	6694535	22.40
A. talpoideum	1984	Atalp_RB_1984_7	4163104	3972086	14.02
A. talpoideum	1984	Atalp_RB_1984_8	8678484	8370783	28.97
A. talpoideum	1984	Atalp_RB_1984_9	10774404	10456155	37.23
A. talpoideum	1984	Atalp_RB_1984_10	8228498	7963498	27.24
A. talpoideum	1984	Atalp_RB_1984_11	9801470	9438278	32.12
A. talpoideum	1984	Atalp_RB_1984_12	8664668	8417362	29.30
A. talpoideum	1984	Atalp_RB_1984_13	5891260	5610242	19.05
A. talpoideum	1984	Atalp_RB_1984_14	4070148	3908276	13.34

A. talpoideum	1984	Atalp_RB_1984_15	8777026	8479250	29.63
A. talpoideum	1984	Atalp_RB_1984_16	5345142	5107754	17.03
A. talpoideum	1984	Atalp_RB_1984_17	8709914	8415420	28.75
A. talpoideum	1984	Atalp_RB_1984_18	10018056	9711532	33.56
A. talpoideum	1984	Atalp_RB_1984_19	7557396	7315848	25.52
A. talpoideum	1984	Atalp_RB_1984_20	5441638	5264282	18.28
A. talpoideum	1984	Atalp_RB_1984_21	5914320	5756938	20.10
A. talpoideum	1984	Atalp_RB_1984_22	10578608	10290008	36.59
A. talpoideum	1984	Atalp_RB_1984_23	7010640	6815875	24.26
A. talpoideum	1984	Atalp_RB_1984_24	4159234	4024422	13.43
A. talpoideum	2011	Atalp_RB_2011_1	7603190	7405953	28.49
A. talpoideum	2011	Atalp_RB_2011_2	7549346	7327102	27.52
A. talpoideum	2011	Atalp_RB_2011_3	11013480	10711160	39.74
A. talpoideum	2011	Atalp_RB_2011_4	7296906	7158906	26.97
A. talpoideum	2011	Atalp_RB_2011_5	9065096	8910760	33.47
A. talpoideum	2011	Atalp_RB_2011_6	9332472	9020403	33.36
A. talpoideum	2011	Atalp_RB_2011_7	13354566	13071473	49.26
A. talpoideum	2011	Atalp_RB_2011_8	7826798	7619264	28.39
A. talpoideum	2011	Atalp_RB_2011_9	9116490	8939736	33.14
A. talpoideum	2011	Atalp_RB_2011_10	11543414	11335237	41.54
A. talpoideum	2011	Atalp_RB_2011_11	10725524	10471919	39.28
A. talpoideum	2011	Atalp_RB_2011_12	8292034	8145370	30.15
A. talpoideum	2011	Atalp_RB_2011_13	10894828	10621351	38.99
A. talpoideum	2011	Atalp_RB_2011_14	9683536	9507080	35.38
A. talpoideum	2011	Atalp_RB_2011_15	12873498	12619220	46.42
A. talpoideum	2011	Atalp_RB_2011_16	7876102	7677014	26.40
A. talpoideum	2011	Atalp_RB_2011_17	7870228	7688714	29.06
A. talpoideum	2011	Atalp_RB_2011_18	6872292	6719144	25.29
A. talpoideum	2011	Atalp_RB_2011_19	4540526	4434182	15.84
A. talpoideum	2011	Atalp_RB_2011_20	8036542	7878789	29.20
A. talpoideum	2011	Atalp_RB_2011_21	9246970	9090926	32.75
A. talpoideum	2011	Atalp_RB_2011_22	8026760	7879283	29.44
A. talpoideum	2011	Atalp_RB_2011_23	9237782	9046118	33.77
A. talpoideum	2011	Atalp_RB_2011_24	8037642	7903066	27.05

Table B.2. *A. talpoideum* outlier sequences producing significant alignments to the *A. mexicanum* transcriptome.

ddRAD locus ID	*A. mexicanum* V3.0 Contig Name
5519	FUQAVB301DW33B
19677	FSIRIH301BNI9U
23479	FUQAVB302FI8BK
27275	EPTY8IW01ES5SK
37295	contig48253

Table B.3. Tajima's D estimates from ∂a∂i (genotype-based) and ANGSD (genotype-free).

Species	Year	Genotype-based Tajima's D	Genotype-Free Tajima's D
A. opacum	1993	0.4686	0.305
A. opacum	2013	0.3859	0.328
A. talpoideum	1984	0.2204	0.485
A. talpoideum	2011	0.0841	0.434

Table B.4. Relative likelihood of the models described in Figure 1 for *A. opacum.*

Model	Maximum Likelihood	Number of Parameters	AIC	delta	Model normalized relative likelihood
1a	-50845.72	5	234163.20	0.00	9.61E-01
1b	-50847.25	9	234178.22	15.02	5.27E-04
2a	-50850.30	5	234184.29	21.09	2.53E-05
2b	-50848.90	9	234185.82	22.62	1.18E-05
3a	-50849.57	6	234182.91	19.71	5.04E-05
3b	-50846.68	6	234169.63	6.43	3.85E-02

Table B.5. Demographic parameters for *A. opacum* estimated in fastsimcoal2 using models described in Figure 1. Parameter estimates are in haploid units and were obtained from the run with the maximum likelihood.

Model	N_e^{1993}	N_e^{2013}	N_eSource	T_{DIV}	N_eOut	N_eAnc	Mig13toS	MigSto13	Mig93toS	MigSto93
1a*	5192	35479	11226	208	456	N/A	N/A	N/A	N/A	N/A
1b	74363	55017	111924	1572	3182	N/A	1.13E-04	3.28E-02	1.03E-05	6.60E-06
2a	29136	44019	58275	6135	N/A	64283	N/A	N/A	N/A	N/A
2b	59346	25728	161888	13688	N/A	86665	6.82E-06	4.21E-06	8.86E-07	1.06E-07
3a	37657	46047	N/A	12469	97554	69059	N/A	N/A	N/A	N/A
3b	50520	83539	N/A	291	11315	71592	N/A	N/A	N/A	N/A

*Best Model

Table B.6. Relative likelihood of the models described in Figure 1 for *A. talpoideum.*

Model	Max $\log_{10}$ Likelihood	Number of Parameters	AIC	delta	Model normalized relative likelihood
1a	-51857.47	5	238822.49	875.73	6.87E-191
1b	-51665.57	9	237946.76	0.00	1.00E+00
2a	-52217.34	5	240479.75	2532.99	0.00E+00
2b	-51676.25	9	237995.91	49.15	2.13E-11
3a	-52150.15	6	240172.33	2225.57	0.00E+00
3b	-52853.79	6	243412.68	5465.92	0.00E+00

Table B.7. Demographic parameters for *A. talpoideum* estimated in fastsimcoal2 using models as described in Figure 1. Parameter estimates are in haploid units and were obtained from the run with the maximum likelihood.

Model	N_e^{1984}	N_e^{2011}	N_eSource	T_{DIV}	N_eOut	N_eAnc	Mig11toS	MigSto11	Mig84toS	MigSto84
1a	7638	31	161	14	15	N/A	N/A	N/A	N/A	N/A
1b*	1016	51	7484	2403	68619	N/A	9.97E-02	1.58E-03	2.52E-03	6.72E-04
2a	202	127	44459	14511	N/A	27017	N/A	N/A	N/A	N/A
2b	60146	266	109902	10575	N/A	138166	0.14202	3.69E-05	4.86E-05	8.78E-06
3a	174	107	N/A	4301	78210	37694	N/A	N/A	N/A	N/A
3b	10052	176	N/A	879	22016	11230	N/A	N/A	N/A	N/A

*Best Model

Table B.8. Demographic parameter estimates from fastsimcoal2 for *A. talpoideum* analysis using only frequency spectra from 2011 and models as described in Figure 1. Parameter estimates are in haploid units and were obtained from the run with the maximum likelihood.

Model	N_e^{1984}	N_e^{2011}	N_eSource	T_{DIV}	N_eOut	N_eAnc	Mig11toS	MigSto11	Mig84toS	MigSto84
1a*	1477	308	1204	600	88382	N/A	N/A	N/A	N/A	N/A
1b	27276	308	36547	169	47634	N/A	1.77E-07	1.50E-07	2.48E-04	2.58E-04
2a	8240	308	68831	5611	N/A	21	N/A	N/A	N/A	N/A
2b	3056	308	6029	2299	N/A	8907	0.01731	9.98E-05	4.00E-05	3.96E-05
3a	53093	308	N/A	4560	39356	103002	N/A	N/A	N/A	N/A
3b	1870	308	N/A	142	1869	21642	N/A	N/A	N/A	N/A

*Best Model

Table B.9. Relative likelihood of the models described in Figure 1 for *A. talpoideum* using frequency spectra from 2011.

Model	Max log$_{10}$ Likelihood	Number of Parameters	AIC	delta	Model normalized relative likelihood
1a	-20484.59	4	94343.04	0.00	0.37
1b	-20484.49	8	94350.54	7.50	0.01
2a	-20485.12	4	94345.44	2.40	0.11
2b	-20484.32	8	94349.79	6.75	0.01
3a	-20484.40	5	94344.16	1.12	0.21
3b	-20484.28	5	94343.60	0.55	0.28
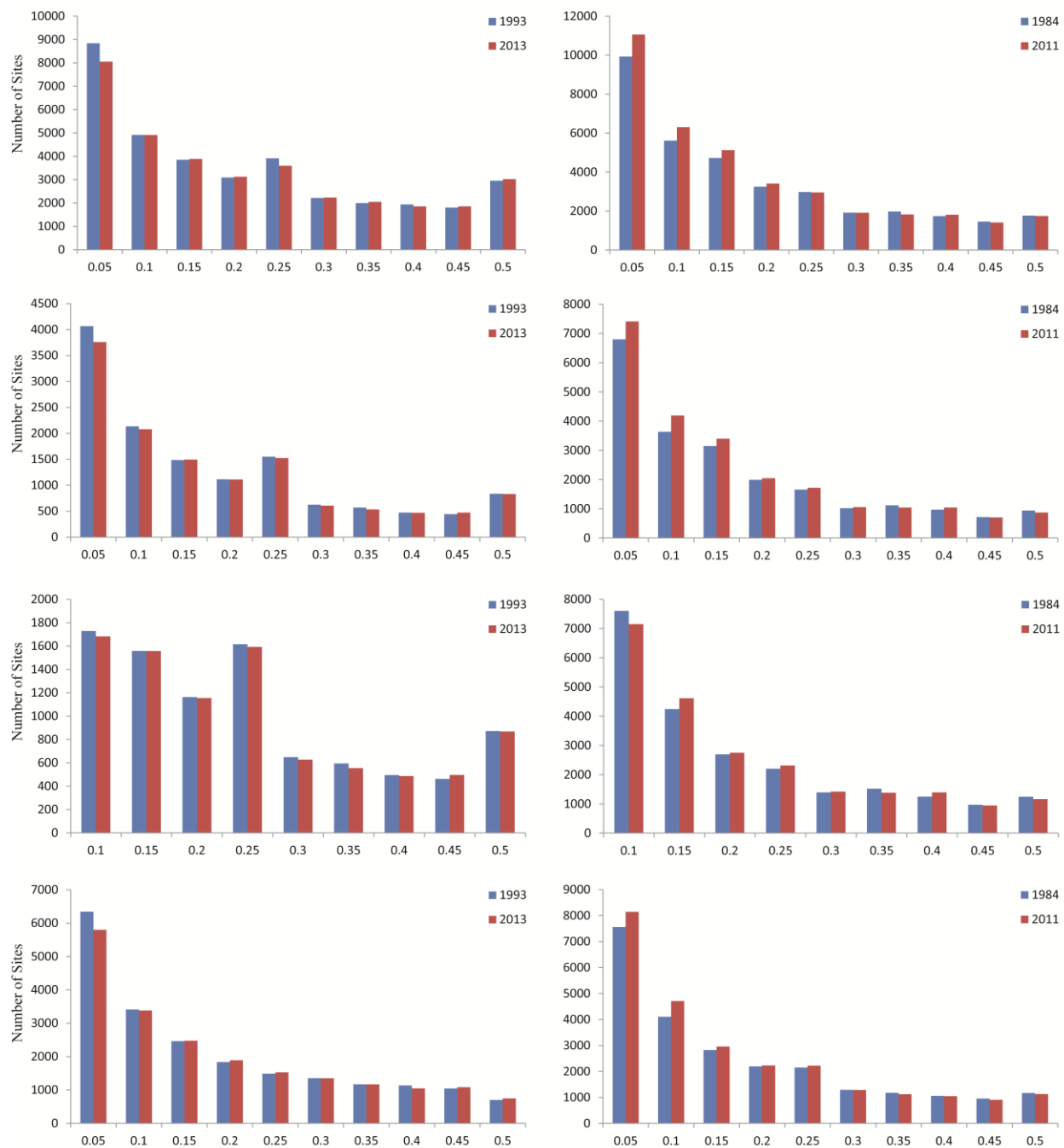
Figure B.1. Density histograms of the minor allele frequencies for *A. opacum* (left) and *A. talpoideum* (right) using four filtering scenarios as described in Table 1.
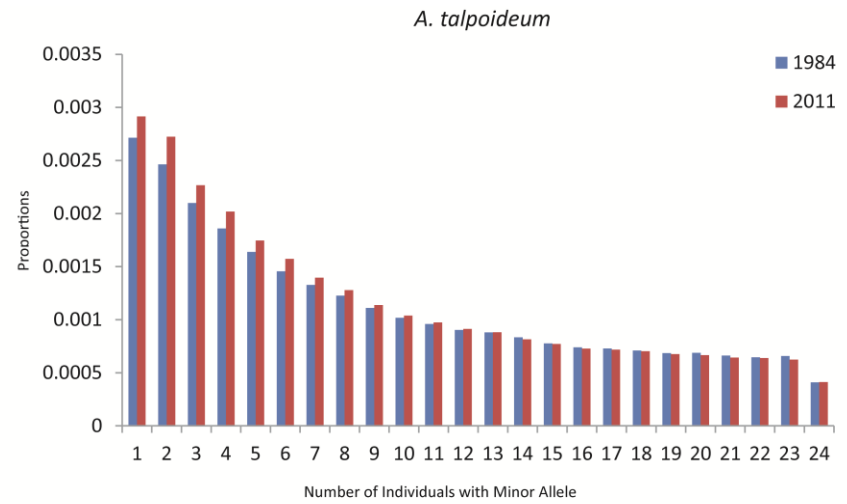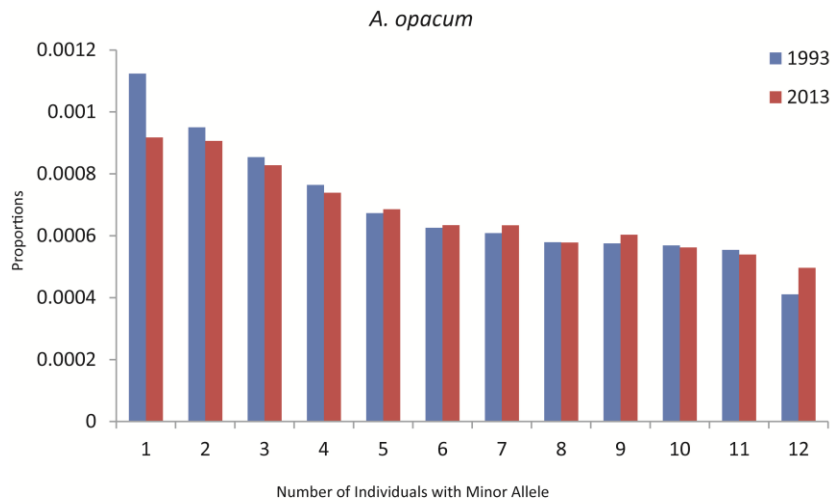
Figure B.2. One-dimensional folded site frequency spectrum (1dSFS) generated in ANGSD illustrating the probability of sampling 1 to 12 of the minor allele for *A. opacum,* and of sampling 1 to 24 of the minor allele for *A. talpoideum*

Table C.4. Summary of population parameters for *Ambystoma opacum* models.

Parameter	Value	Source
Initial population size	60	
Maximum clutch size (low larval density)	age < 1 : 0 1 <= age < 3 : 40 3 <=age < 5: 80 5 <=age < 7: 120 7 <=age < 9: 160 9 <=age >= 10: 200	1
Maximum clutch size (high larval density)	0 <= age < 2 : 0 2 <= age < 3: 40 3 <=age < 5: 80 5 <=age < 7: 120 7 <=age < 9: 160 9 <=age >= 10: 200	1
Male fecundity (low larval density)	age < 1 : 0 1 <= age >= 10: 0.96	2
Male fecundity (high larval density)	age < 2 : 0 2 <= age >= 10: 0.96	2
Female fecundity (low larval density)	age < 1 : 0 1 <= age >= 10: 0.67	2
Female fecundity (high larval density)	age < 2 : 0 1 <= age >= 10: 0.67	2
Annual Male Survival Rate	age < 1 : 0.05 1 <=age <10: 0.6 age >= 10: 0	2 3 4
Annual Female Survival Rate	age < 1 : 0.05 1 <=age <10: 0.6 age >= 10: 0	2 3 4

Sources: (1) Scott (1994), (2) Gamble (2009), (3) Scott (1990), (4) Taylor and Scott (1997)

Table C.2. Average adult census size (N_c), effective number of breeders (N_bGen and N_bDem), and effective population size (N_eGen), for *Ambystoma opacum* models with no migration after 30 generation burn-in and at generation 80, over 100 replicate simulations, with 95% confidence intervals in parenthesis.

Probability of CRF (%)	Generation	N_c	N_eGen	N_bDem	N_bGen
0	30	3651(3512-3776)	1498(1304-1758)	1543(1457-1624)	1037(876-1210)
0	80	3651(3522-3755)	1465(1249-1710)	1550(1461-1639)	1038(885-1203)
10	30	3647(3519-3772)	1466(1256-1684)	1418(0-1632)	1030(881-1210)
10	80	3381(1537-3858)	1308(897-1636)	1267(0-1666)	952(516-1174)
20	30	3656(3536-3788)	1483(1239-1754)	1284(0-1621)	1035(902-1177)
20	80	3221(1620-3875)	1132(475-1638)	1045(0-1711)	868(493-1110)
30	30	3652(3515-3769)	1464(1258-1730)	1040(0-1630)	1034(881-1231)
30	80	2790(1157-3886)	959(390-1470)	933(0-1663)	760(356-1119)
40	30	3653(3548-3777)	1475(1263-1773)	968(0-1593)	1015(890-1180)
40	80	2628(761-3882)	827(166-1412)	652(0-1656)	672(220-1052)
50	30	3652(3542-3807)	1451(1235-1787)	753(0-1609)	1026(858-1171)
50	80	2083(108-3887)	607(66-1114)	429(0-1612)	518(66-988)
60	30	3645(3502-3765)	1464(1262-1692)	694(0-1601)	1030(887-1250)
60	80	1718(0-3810)	419(67-930)	234(0-1614)	384(38-798)
70	30	3642(3518-3790)	1487(1254-1752)	429(0-1597)	1036(883-1229)
70	80	771(0-3321)	271(23-875)	85(0-1004)	281(18-869)
80	30	3651(3525-3764)	1479(1265-1755)	248(0-1586)	1042(878-1215)
80	80	136(0-1713)	163(10-722)	21(0-134)	107(6-474)
90	30	3654(3533-3766)	1481(1269-1778)	186(0-1573)	1034(891-1198)
90	80	0(0-0)	0(0-0)	0(0-0)	0(0-0)
100	30	3659(3521-3781)	1474(0-0)	0(0-0)	1035(838-1267)
100	80	0(0-0)	0(0-0)	0(0-0)	0(0-0)

Table C.3. Average adult census size (N_c), effective number of breeders (N_bGen and N_bDem), and effective population size (N_eGen), for *Ambystoma opacum* models with migration after 30 generation burn-in and at generation 80, over 100 replicate simulations, with 95% confidence intervals in parenthesis.

Probability of CRF (%)	Generation	N_c	N_eGen	N_bDem	N_bGen
0	30	3641(3504-3753)	1548(1299-1864)	1538(1443-1644)	1027(838-1235)
0	80	3652(3554-3760)	1658(1356-2071)	1543(1499-1627)	1061(904-1221)
10	30	3653(3550-3754)	1560(1295-1828)	1543(1473-1618)	1047(868-1255)
10	80	3340(2164-3846)	1512(999-1968)	1411(962-1666)	963(649-1221)
20	30	3640(3501-3764)	1554(1323-1805)	1533(1456-1610)	1026(866-1198)
20	80	3282(1459-3888)	1298(690-1806)	1416(681-1704)	913(479-1203)
30	30	3650(3524-3758)	1536(1303-1819)	1545(1464-1630)	1036(889-1214)
30	80	2829(1096-3971)	1154(520-1647)	1219(477-1727)	780(380-1133)
40	30	3664(3538-3806)	1532(1306-1803)	1541(1460-1627)	1035(883-1194)
40	80	2437(673-3854)	1024(330-1751)	1047(293-1681)	680(247-1104)
50	30	3642(3537-3742)	1557(1258-1859)	1539(1419-1619)	1043(903-1203)
50	80	2109(394-3844)	848(311-1862)	901(235-1617)	557(186-1049)
60	30	3642(3530-3773)	1515(1242-1829)	1545(1459-1650)	1039(901-1170)
60	80	1515(231-3459)	750(197-2472)	756(144-1525)	416(106-861)
70	30	3654(3536-3775)	1529(1282-1795)	1551(1483-1630)	1041(859-1243)
70	80	1227(184-3389)	821(147-2904)	607(85-1621)	329(68-819)
80	30	3657(3537-3772)	1540(1252-1867)	1538(1453-1606)	1031(889-1178)
80	80	738(188-2718)	964(111-2951)	243(88-487)	203(57-577)
90	30	3645(3508-3757)	1532(1297-1872)	1543(1468-1605)	1039(879-1260)
90	80	333(164-1028)	1418(112-3458)	206(81-423)	112(50-309)
100	30	3655(3544-3758)	1540(1342-1861)	0(0-0)	1054(901-1244)
100	80	185(159-207)	2099(1260-3269)	0(0-0)	62(47-79)

Table C.4. Number of times Spearman rank correlations were significant ($p < 0.05$) for *Ambystoma opacum* models without migration over 100 replicate simulations under different probabilities of catastrophic reproductive failure (CRF). Mean Spearman rank correlation coefficients across replicates are in parentheses.

	No CRF	10% CRF	20% CRF	30% CRF	40% CRF	50% CRF	60% CRF	70% CRF	80% CRF	90% CRF	100% CRF
Census v. N_bDemo	100 (0.59)	100(0.72)	100 (0.63)	92 (0.51)	81 (0.40)	59 (0.30)	28 (0.17)	19 (0.14)	28 (0.17)	23 (0.16)	0 (0.07)
Census v. N_bGen	5 (0.20)	44 (0.50)	83 (0.75)	95 (0.82)	99 (0.86)	98 (0.85)	96 (0.77)	82 (0.33)	76 (-0.18)	93 (-0.65)	100 (-0.66)
Census v. N_eGen	4 (0.00)	8 (0.04)	11 (0.27)	7 (0.22)	8 (0.30)	11 (0.25)	8 (0.20)	29 (-0.03)	63 (-0.36)	93 (-0.69)	100 (-0.67)
Moms v. N_bGen	7 (0.16)	30 (0.46)	43 (0.51)	28 (0.39)	18 (0.30)	15 (0.15)	7 (0.07)	2 (-0.08)	0 (-0.10)	0 (-0.09)	0 (0.00)
Moms v. N_eGen	2 (0.01)	4 (0.05)	7 (0.15)	6 (0.04)	8 (0.11)	4 (0.05)	7 (0.05)	3 (-0.12)	0 (-0.10)	0 (-0.10)	0 (0.00)
N_bGen v. N_eGen	2 (-0.01)	10 (0.20)	32 (0.46)	33 (0.45)	37 (0.53)	31 (0.48)	14 (0.21)	75 (0.71)	85 (0.85)	99 (0.97)	100 (0.99)
SDVmale v. N_bGen	8 (-0.09)	5 (-0.01)	7 (0.05)	7 (0.12)	8 (0.11)	6 (0.14)	5 (0.05)	26 (0.24)	51 (0.48)	93 (0.87)	100 (0.99)
SDVmale v. N_eGen	6 (0.00)	4 (-0.06)	10 (-0.04)	6 (0.00)	3 (-0.04)	4 (-0.05)	11 (-0.10)	32 (0.19)	57 (0.48)	94 (0.88)	100 (1.00)
SDVfemale v. N_bGen	11 (-0.06)	7 (-0.05)	8 (-0.01)	8 (0.11)	10 (0.10)	4 (0.12)	7 (0.04)	26 (0.24)	51 (0.48)	93 (0.87)	100 (0.99)
SDVfemale v. N_eGen	7 (-0.01)	1 (-0.07)	7 (-0.10)	9 (-0.01)	5 (-0.04)	8 (-0.07)	11 (-0.10)	34 (0.19)	57 (0.48)	94 (0.88)	100 (1.00)

Table C.5. Number of times Spearman rank correlations were significant ($p < 0.05$) for *Ambystoma opacum* models with migration over 100 replicate simulations under different probabilities of catastrophic reproductive failure (CRF). Mean Spearman rank correlation coefficients across replicates are in parentheses.

	No CRF	10% CRF	20% CRF	30% CRF	40% CRF	50% CRF	60% CRF	70% CRF	80% CRF	90% CRF	100% CRF
Census v. N_bDemo	99 (0.58)	100 (0.69)	100 (0.65)	94 (0.50)	72 (0.36)	48 (0.26)	28 (0.17)	13 (0.10)	19 (0.05)	27(0.01)	19 (0.05)
Census v. N_bGen	11 (0.17)	57 (0.58)	86 (0.75)	98 (0.83)	99 (0.88)	100 (0.91)	100 (0.92)	100 (0.92)	98 (0.88)	97 (0.85)	84 (0.73)
Census v. N_eGen	2 (0.02)	10 (0.14)	9 (0.08)	9 (0.08)	4 (0.04)	6 (0.07)	2 (-0.01)	2 (-0.10)	8 (-0.35)	39 (-0.54)	9 (-0.30)
Moms v. N_bGen	11 (0.18)	44 (0.49)	35 (0.51)	27 (0.35)	18 (0.27)	9 (0.26)	5 (0.11)	5 (0.09)	1 (0.04)	2 (0.00)	0 (0.00)
Moms v. N_eGen	5 (0.04)	5 (0.08)	7 (0.02)	6 (0.01)	4 (-0.01)	5 (0.01)	2 (-0.05)	6 (-0.01)	3 (-0.02)	0 (-0.01)	0 (0.00)
N_bGen v. N_eGen	3 (0.06)	10 (0.24)	13 (0.20)	20 (0.26)	11 (0.22)	16 (0.25)	6 (0.12)	5 (0.06)	3 (-0.19)	17 (-0.42)	9 (-0.26)
SDVmale v. N_bGen	8 (-0.13)	3 (0.04)	5 (0.04)	12 (0.17)	5 (0.14)	6 (0.01)	5 (-0.05)	9 (-0.16)	9 (-0.23)	22 (-0.36)	24 (-0.44)
SDVmale v. N_eGen	4 (-0.02)	7 (-0.08)	8 (-0.13)	7 (-0.11)	4 (-0.12)	2 (-0.09)	6 (-0.15)	11 (-0.17)	7 (-0.18)	5 (-0.05)	6 (0.20)
SDVfemale v. N_bGen	8 (-0.08)	7 (-0.05)	8 (0.00)	8 (0.15)	6 (0.14)	9 (0.03)	5 (-0.05)	10 (-0.16)	9 (-0.23)	22 (-0.36)	24 (-0.44)
SDVfemale v. N_eGen	3 (0.01)	13 (-0.16)	8 (-0.16)	6 (-0.13)	5 (-0.11)	1 (-0.09)	7 (-0.14)	8 (-0.17)	7 (-0.17)	5 (-0.05)	6 (0.20)

171

Table C.6. Number of times Spearman rank correlations were significant ($p < 0.05$) for *Ambystoma opacum* models without migration using the subsampled dataset of 30 microsatellite markers and 60 individuals, over 100 replicate simulations under different probabilities of catastrophic reproductive failure (CRF). Mean Spearman rank correlation coefficients across replicates are in parentheses.

	Census v. N_bGen	Census v. N_eGen
No CRF	4(0)	6(-0.02)
10% CRF	7(0.12)	3(-0.01)
20% CRF	11(0.19)	3(0.04)
30% CRF	21(0.35)	11(0.11)
40% CRF	19(0.38)	11(0.19)
50% CRF	46(0.48)	11(0.15)
60% CRF	69(0.6)	18(0.2)
70% CRF	68(0.25)	27(-0.02)
80% CRF	75(-0.18)	59(-0.36)
90% CRF	93(-0.65)	93(-0.68)
100% CRF	100(-0.66)	100(-0.67)

Table C.7. Summary of sensitivity analysis under different probabilities of catastrophic reproductive failure (CRF) for *Ambystoma opacum* models with and without migration (mig).

Parameter	No CRF		10% CRF		20% CRF		30% CRF		40% CRF		50% CRF		60% CRF		70% CRF		80% CRF		90% CRF		100% CRF	
	no mig	with mig	no mig	with mig	no mig	with mig	no mig	with mig	no mig	with mig	no mig	with mig	no mig	with mig	no mig	with mig	no mig	with mig	no mig	with mig	no mig	with mig
Adult pop Size																						
Base Model	3651	3652	3381	3340	3221	3282	2790	2829	2628	2437	2083	2109	1718	1515	771	1227	136	738	0	333	0	185
Juvenile Survival	0.6	0.6	0.6	0.6	0.6	0.6	0.5	0.6	0.5	0.5	0.5	0.5	0.4	0.4	0.2	0.4	0.0	0.4	-	0.5	-	0.6
Adult Survival	0.8	0.8	0.8	0.8	0.8	0.8	0.7	0.8	0.7	0.7	0.7	0.7	0.6	0.6	0.4	0.6	0.1	0.6	-	0.7	-	0.8
Proportion Breeding	1.1	1.1	1.1	1.1	1	1	1	1.1	1	1.1	1	1	0.8	1.3	1	1	1.1	1	-	1.1	-	1.1
No Density-Dependent Breeding*	1	1	1	1	1	1	1	1	1	1	0.9	1	0.8	1	0.7	0.9	1.1	1	-	1	-	1
Extinction risk																						
Base Model	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0.4	0	0.8	0	1	0	1	0
Juvenile Survival	0	0	0	0	0	0	0	0	0	0	0	0	0.2	0	0.6	0	0.9	0	1	0	1	0
Adult Survival	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0.5	0	0.9	0	1	0	1	0
Proportion Breeding	0	0	0	0	0	0	0	0	0	0	0	0	0.2	0	0.3	0	0.7	0	1	0	1	0
No Density-Dependent Breeding*	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0.4	0	0.7	0	1	0	1	0

Notes: Each parameter was decreased by 10%. Adult population size is in proportion of the base model. *This model does not include density-dependent age at first reproduction, i.e. no individuals breed at age one.
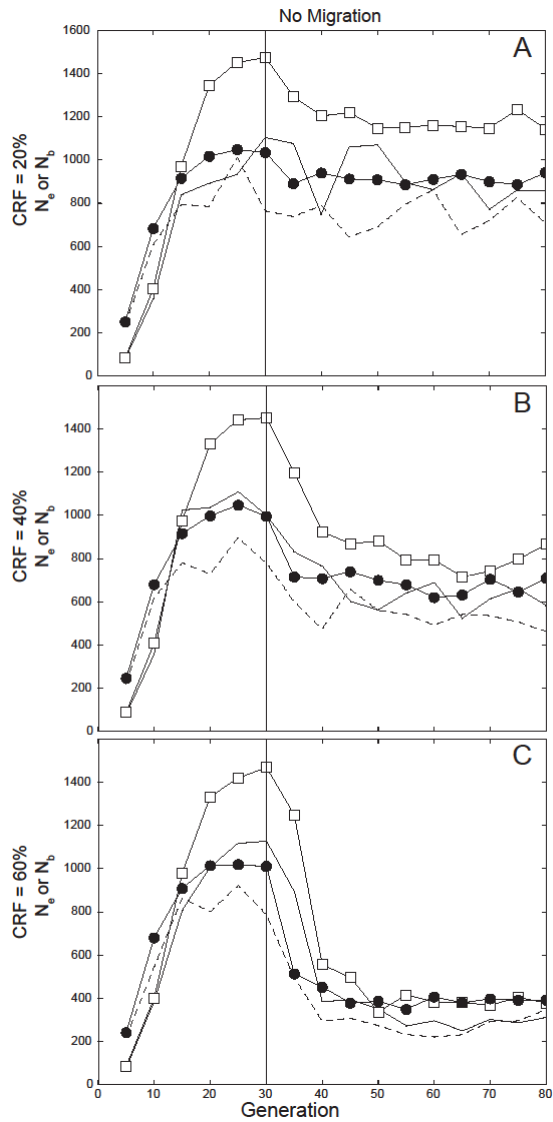
Figure C.2. *Ambystoma opacum* mean N_eGen (black line) and N_bGen (dashed line) using subsampled dataset (30 microsatellites and 60 individuals), and mean N_eGen(open squares) and N_bGen (filled circles) using the complete dataset (100 microsatellites and 500 individuals) every 5 generations, over 80 years of projections. Catastrophic reproductive failure (CRF) was imposed after 30 year burn-in (vertical line). Top panel represent 20% probability of CRF (A), middle 40% (B), and bottom 60% (C), all with no migration.

Table D.5. Individual sample information for *Ambystoma opacum* and *A. talpoideum* across the SRS.

Species	Wetland	Year	Season	Stage	Individual ID	Total Reads	Retained Reads
A. opacum	Rainbow	2013	N/A	metamorph	Amop_RB_2013_9	4542706	3041263
A. opacum	Rainbow	1993	N/A	metamorph	Amop_RB_1993_7	6842698	4896551
A. opacum	Rainbow	2013	N/A	metamorph	Amop_RB_2013_10	7174202	5459819
A. opacum	Rainbow	1993	N/A	metamorph	Amop_RB_1993_9	6878838	5004389
A. opacum	Rainbow	2013	N/A	metamorph	Amop_RB_2013_8	7109458	5143848
A. opacum	Rainbow	2013	N/A	metamorph	Amop_RB_2013_4	8000074	5733841
A. opacum	Rainbow	2013	N/A	metamorph	Amop_RB_2013_12	7788168	5860264
A. opacum	Rainbow	2013	N/A	metamorph	Amop_RB_2013_5	8497784	5952381
A. opacum	Rainbow	1993	N/A	metamorph	Amop_RB_1993_8	7518160	5440838
A. opacum	Rainbow	2013	N/A	metamorph	Amop_RB_2013_11	8525550	6305570
A. opacum	Rainbow	1993	N/A	metamorph	Amop_RB_1993_12	7015048	5208398
A. opacum	Rainbow	1993	N/A	metamorph	Amop_RB_1993_1	8235678	5735247
A. opacum	Rainbow	1993	N/A	metamorph	Amop_RB_1993_4	7369872	5757267
A. opacum	Rainbow	1993	N/A	metamorph	Amop_RB_1993_5	8244798	6073797
A. opacum	Rainbow	2013	N/A	metamorph	Amop_RB_2013_7	8648848	6435261
A. opacum	Rainbow	2013	N/A	metamorph	Amop_RB_2013_2	10275598	7608513
A. opacum	Rainbow	2013	N/A	metamorph	Amop_RB_2013_6	10272342	7258797
A. opacum	Rainbow	1993	N/A	metamorph	Amop_RB_1993_2	10000486	6452557
A. opacum	Rainbow	2013	N/A	metamorph	Amop_RB_2013_1	9931710	7559485
A. opacum	Rainbow	1993	N/A	metamorph	Amop_RB_1993_11	11053706	8031302
A. opacum	Rainbow	1993	N/A	metamorph	Amop_RB_1993_6	11757322	8087569
A. opacum	Rainbow	2013	N/A	metamorph	Amop_RB_2013_3	12788398	9815887
A. opacum	Rainbow	1993	N/A	metamorph	Amop_RB_1993_10	11360396	8588603

A. opacum	Rainbow	1993	N/A	metamorph	Amop_RB_1993_3	14291878	11498891
A. opacum	Pickerel	2016	Spring	Larvae	Pick_12	10146856	9888793
A. opacum	Pickerel	2016	Spring	Larvae	Pick_1	11999476	11686973
A. opacum	Pickerel	2016	Spring	Larvae	Pick_4	12777616	12424949
A. opacum	Pickerel	2016	Spring	Larvae	Pick_14	14246732	13899688
A. opacum	Pickerel	2016	Spring	Larvae	Pick_7	13907220	13616733
A. opacum	Pickerel	2016	Spring	Larvae	Pick_15	13669678	13371081
A. opacum	Pickerel	2016	Spring	Larvae	Pick_11	15042514	14725041
A. opacum	Pickerel	2016	Spring	Larvae	Pick_9	14063816	13785235
A. opacum	Pickerel	2016	Spring	Larvae	Pick_2	15945830	15502016
A. opacum	Pickerel	2016	Spring	Larvae	Pick_10	17641280	17270482
A. opacum	Pickerel	2016	Spring	Larvae	Pick_8	18863050	18394261
A. opacum	Pickerel	2016	Spring	Larvae	Pick_6	18591472	18259999
A. opacum	NPR	2016	Spring	Larvae	NPR_11	2266558	2131803
A. opacum	NPR	2016	Spring	Larvae	NPR_2	2797184	2664963
A. opacum	NPR	2016	Spring	Larvae	NPR_10	4786440	4612368
A. opacum	NPR	2016	Spring	Larvae	NPR_5	6615602	6429776
A. opacum	NPR	2016	Spring	Larvae	NPR_9	6921788	6734074
A. opacum	NPR	2016	Spring	Larvae	NPR_1	8563430	8341924
A. opacum	NPR	2016	Spring	Larvae	NPR_4	10881400	10635508
A. opacum	NPR	2016	Spring	Larvae	NPR_8	10863850	10609312
A. opacum	NPR	2016	Spring	Larvae	NPR_6	12144204	11846519
A. opacum	NPR	2016	Spring	Larvae	NPR_3	12842564	12522246
A. opacum	NPR	2016	Spring	Larvae	NPR_7	13761456	13440479
A. opacum	Lindas	2016	Spring	Larvae	Lind_330	244280	132486
A. opacum	Lindas	2016	Spring	Larvae	Lind_331	731344	581495
A. opacum	Lindas	2016	Spring	Larvae	Lind_332	1346110	1236551
A. opacum	Lindas	2016	Spring	Larvae	Lind_355	6127328	5961033
A. opacum	Lindas	2016	Spring	Larvae	Lind_351	4455380	4242267
A. opacum	Lindas	2016	Spring	Larvae	Lind_329	6588888	6362862

A. opacum	Lindas	2016	Spring	Larvae	Lind_369	7117446	6866311
A. opacum	Lindas	2016	Spring	Larvae	Lind_350	8175632	7808174
A. opacum	Lindas	2016	Spring	Larvae	Lind_356	6057520	5830345
A. opacum	Lindas	2016	Spring	Larvae	Lind_354	8108892	7854548
A. opacum	Lindas	2016	Spring	Larvae	Lind_357	8365316	8098889
A. opacum	Lindas	2016	Spring	Larvae	Lind_353	7346320	7121386
A. opacum	Gingers	2007	N/A	Adult	GB_77	116798	59833
A. opacum	Gingers	2010	N/A	metamorph	GB_221	1263648	1012012
A. opacum	Gingers	1996	N/A	Adult	GB_56	2392222	2234887
A. opacum	Gingers	2005	N/A	metamorph	GB_207	4580066	4287860
A. opacum	Gingers	2007	N/A	Adult	GB_43	5200034	4938330
A. opacum	Gingers	2005	N/A	metamorph	GB_204	5306622	5017471
A. opacum	Gingers	2005	N/A	metamorph	GB_205	5860300	5551885
A. opacum	Gingers	2010	N/A	metamorph	GB_220	6604018	6245418
A. opacum	Gingers	2007	N/A	Adult	GB_42	6339330	5957179
A. opacum	Gingers	2007	N/A	Adult	GB_75	6165042	5880159
A. opacum	Gingers	1996	N/A	metamorph	GB_261	5451970	5034133
A. opacum	Gingers	2005	N/A	metamorph	GB_208	7573718	7138228
A. opacum	Gingers	1990	N/A	Adult	GB_46	4775600	4619821
A. opacum	Gingers	1996	N/A	metamorph	GB_256	7698812	7231713
A. opacum	Gingers	1996	N/A	Adult	GB_30	8270544	7884628
A. opacum	Gingers	1996	N/A	Adult	GB_34	8585316	8105599
A. opacum	Gingers	1990	N/A	Adult	GB_48	8966340	8711190
A. opacum	Gingers	1996	N/A	Adult	GB_36	11632988	11214656
A. opacum	Gingers	1990	N/A	Adult	GB_44	12365740	11998394
A. opacum	Gingers	2007	N/A	Adult	GB_76	20948404	20066442
A. opacum	Gingers	1996	N/A	Adult	GB_35	14222508	13702065
A. opacum	Gingers	1990	N/A	Adult	GB_66	14538168	14102012
A. opacum	Gingers	1990	N/A	Adult	GB_47	16515292	16061686
A. opacum	Gingers	1990	N/A	Adult	GB_45	17506972	17057456

A. opacum	Flamingo	2014	Fall	Adult	Flam_34	1829540	1772068
A. opacum	Flamingo	2014	Fall	Adult	Flam_30	2467098	2383206
A. opacum	Flamingo	2014	Fall	Adult	Flam_37	2613110	2521231
A. opacum	Flamingo	2014	Fall	Adult	Flam_27	3150326	3027810
A. opacum	Flamingo	2014	Fall	Adult	Flam_45	3226988	3119680
A. opacum	Flamingo	2014	Fall	Adult	Flam_29	3826756	3731008
A. opacum	Flamingo	2014	Fall	Adult	Flam_36	4121162	4026593
A. opacum	Flamingo	2014	Fall	Adult	Flam_17	4154388	4029694
A. opacum	Flamingo	2014	Fall	Adult	Flam_33	6627474	6503456
A. opacum	Flamingo	2014	Fall	Adult	Flam_35	9886348	9697063
A. opacum	Flamingo	2014	Fall	Adult	Flam_16	13247416	13002789
A. opacum	Flamingo	2014	Fall	Adult	Flam_18	16127744	15826428
A. opacum	Bullfrog	2016	Spring	Larvae	Bull_15	92250	37541
A. opacum	Bullfrog	2016	Spring	Larvae	Bull_11	6237608	6025818
A. opacum	Bullfrog	2016	Spring	Larvae	Bull_2	9783670	9468381
A. opacum	Bullfrog	2016	Spring	Larvae	Bull_8	10534678	10247831
A. opacum	Bullfrog	2016	Spring	Larvae	Bull_1	11696882	11403884
A. opacum	Bullfrog	2016	Spring	Larvae	Bull_12	12554862	12258737
A. opacum	Bullfrog	2016	Spring	Larvae	Bull_13	12042368	11737644
A. opacum	Bullfrog	2016	Spring	Larvae	Bull_7	12711628	12320330
A. opacum	Bullfrog	2016	Spring	Larvae	Bull_10	12603398	12306094
A. opacum	Bullfrog	2016	Spring	Larvae	Bull_4	12515576	12186467
A. opacum	Bullfrog	2016	Spring	Larvae	Bull_6	13544238	13240127
A. opacum	Bullfrog	2016	Spring	Larvae	Bull_9	13009686	12660884
A. opacum	Bullfrog	2016	Spring	Larvae	Bull_5	14134860	13698028
A. opacum	Bay 92	2015	Spring	Larvae	B92_67	5327288	5206438
A. opacum	Bay 92	2015	Spring	Larvae or Paedomorph	B92_90	7614066	7445288
A. opacum	Bay 92	2015	Spring	Larvae	B92_68	9850996	9594929
A. opacum	Bay 92	2015	Spring	Larvae	B92_75	11162350	10933274
A. opacum	Bay 92	2015	Spring	Larvae	B92_87	12672132	12400265

A. opacum	Bay 92	2015	Spring	Larvae	B92_89	12798682	12554439
A. opacum	Bay 92	2015	Spring	Larvae	B92_69	12571880	12350189
A. opacum	Bay 92	2015	Spring	Larvae	B92_72	13908840	13567549
A. opacum	Bay 92	2015	Spring	Larvae	B92_76	14544360	14224693
A. opacum	Bay 92	2015	Spring	Larvae	B92_70	14958314	14557504
A. opacum	Bay 92	2015	Spring	Larvae	B92_71	14206726	13980090
A. opacum	Bay 92	2015	Spring	Larvae	B92_73	19431690	19066645
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_24	4159234	4024422
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_14	4070148	3908276
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_7	4163104	3972086
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_16	5345142	5107754
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_1	5201034	4970846
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_20	5441638	5264282
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_2	5285492	5045218
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_13	5891260	5610242
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_19	4540526	4434552
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_21	5914320	5756938
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_3	6423864	6170749
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_6	7011954	6694535
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_23	7010640	6815875
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_19	7557396	7315848
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_10	8228498	7963498
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_17	8709914	8415420
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_8	8678484	8370783
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_4	8239206	8005560
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_12	8664668	8417362
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_15	8777026	8479250
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_5	8897272	8616111
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_11	9801470	9438278
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_16	7876102	7677552

A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_18	6872292	6719639
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_24	8037642	7903674
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_18	10018056	9711532
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_4	7296906	7159411
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_8	7826798	7619862
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_2	7549346	7327713
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_1	7603190	7406548
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_22	10578608	10290008
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_17	7870228	7689319
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_22	8026760	7879994
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_20	8036542	7879535
A. talpoideum	Rainbow	1984	N/A	metamorph	Atalp_RB_1984_9	10774404	10456155
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_12	8292034	8145999
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_21	9246970	9091769
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_23	9237782	9046954
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_6	9332472	9021178
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_9	9116490	8940450
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_5	9065096	8911475
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_14	9683536	9507894
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_13	10894828	10622200
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_3	11013480	10712127
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_11	10725524	10472765
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_10	11543414	11336249
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_15	12873498	12620284
A. talpoideum	Rainbow	2011	N/A	metamorph	Atalp_RB_2011_7	13354566	13072499
A. talpoideum	Pickerel	2016	Spring	Larvae	Pick_1	15014970	14793260
A. talpoideum	Pickerel	2016	Spring	Larvae	Pick_2	11365460	11159401
A. talpoideum	NPR	2016	Spring	Larvae	NPR_1	13836438	13610294
A. talpoideum	NPR	2016	Spring	Larvae	NPR_10	15762060	15531976
A. talpoideum	NPR	2016	Spring	Larvae	NPR_11	7132162	6999295

A. talpoideum	NPR	2016	Spring	Larvae	NPR_12	16716496	16446810
A. talpoideum	NPR	2016	Spring	Larvae	NPR_13	14417742	14186447
A. talpoideum	NPR	2016	Spring	Larvae	NPR_14	15000134	14756997
A. talpoideum	NPR	2016	Spring	Larvae	NPR_15	16948084	16690685
A. talpoideum	NPR	2016	Spring	Larvae	NPR_2	12228934	11997922
A. talpoideum	NPR	2016	Spring	Larvae	NPR_3	9923346	9682321
A. talpoideum	NPR	2016	Spring	Larvae	NPR_4	11621874	11443957
A. talpoideum	NPR	2016	Spring	Larvae	NPR_5	16497824	16224796
A. talpoideum	NPR	2016	Spring	Larvae	NPR_6	13911100	13653397
A. talpoideum	NPR	2016	Spring	Larvae	NPR_7	13878306	13662487
A. talpoideum	NPR	2016	Spring	Larvae	NPR_8	8710786	8568969
A. talpoideum	NPR	2016	Spring	Larvae	NPR_9	14976910	14754834
A. talpoideum	Lindas	2016	Spring	Larvae	Lindas_3824	13666664	13456339
A. talpoideum	Lindas	2016	Spring	Larvae	Lindas_3827	11473646	11270382
A. talpoideum	Lindas	2016	Spring	Larvae	Lindas_3828	14743622	14510584
A. talpoideum	Lindas	2016	Spring	Larvae	Lindas_4403	6483964	6380846
A. talpoideum	Gingers	2005	N/A	metamorph	GB_92	2043730	1909850
A. talpoideum	Gingers	1996	N/A	Paedomorph	GB_106	3810090	3554181
A. talpoideum	Gingers	1995	N/A	metamorph	GB_145	4112064	3844492
A. talpoideum	Gingers	1991	N/A	Adult	GB_102	4619530	4313896
A. talpoideum	Gingers	2005	N/A	metamorph	GB_93	4286712	4087270
A. talpoideum	Gingers	1991	N/A	Adult	GB_101	8286436	7866091
A. talpoideum	Gingers	2005	N/A	metamorph	GB_88	8432494	7981875
A. talpoideum	Gingers	1996	N/A	Paedomorph	GB_109	8781850	8515789
A. talpoideum	Gingers	2010	N/A	metamorph	GB_158	8185252	7919276
A. talpoideum	Gingers	1991	N/A	Adult	GB_115	10168720	9555829
A. talpoideum	Gingers	1995	N/A	metamorph	GB_146	10368144	9790601
A. talpoideum	Gingers	2005	N/A	metamorph	GB_85	8979882	8704549
A. talpoideum	Gingers	2005	N/A	metamorph	GB_90	9571296	9167578
A. talpoideum	Gingers	1996	N/A	Paedomorph	GB_134	12332634	11960088

A. talpoideum	Gingers	2010	N/A	metamorph	GB_147	12791138	12404859
A. talpoideum	Gingers	2010	N/A	metamorph	GB_167	12288548	11937295
A. talpoideum	Gingers	2010	N/A	metamorph	GB_148	13051454	12721298
A. talpoideum	Gingers	1995	N/A	metamorph	GB_144	15632300	14883092
A. talpoideum	Gingers	1996	N/A	Paedomorph	GB_105	14529738	14100985
A. talpoideum	Gingers	1991	N/A	Adult	GB_116	15388400	14907570
A. talpoideum	Gingers	1995	N/A	Paedomorph	GB_138	16805364	16321405
A. talpoideum	Gingers	2010	N/A	metamorph	GB_163	15628882	15240835
A. talpoideum	Gingers	1996	N/A	Paedomorph	GB_104	18665640	18143056
A. talpoideum	Gingers	2010	N/A	metamorph	GB_166	18899056	18547806
A. talpoideum	Flamingo	2015	Spring	Larvae or Paedomorph	Flam_128	8203072	8018228
A. talpoideum	Flamingo	2015	Spring	Larvae or Paedomorph	Flam_132	8307224	8139126
A. talpoideum	Flamingo	2015	Spring	Larvae or Paedomorph	Flam_131	9962006	9717343
A. talpoideum	Flamingo	2015	Spring	Larvae or Paedomorph	Flam_186	9035296	8885002
A. talpoideum	Flamingo	2015	Spring	Larvae or Paedomorph	Flam_59	10842030	10601268
A. talpoideum	Flamingo	2015	Spring	Larvae or Paedomorph	Flam_188	11749056	11549733
A. talpoideum	Flamingo	2015	Spring	Larvae or Paedomorph	Flam_129	14498666	14192594
A. talpoideum	Flamingo	2015	Spring	Larvae or Paedomorph	Flam_62	13790640	13526238
A. talpoideum	Flamingo	2015	Spring	Larvae or Paedomorph	Flam_189	12935194	12758588
A. talpoideum	Flamingo	2015	Spring	Larvae or Paedomorph	Flam_63	14397856	14141961
A. talpoideum	Flamingo	2015	Spring	Larvae or Paedomorph	Flam_187	15869332	15664554
A. talpoideum	Flamingo	2015	Spring	Larvae or Paedomorph	Flam_190	15969440	15758013
A. talpoideum	Bay 92	2015	Spring	Larvae or Paedomorph	B92_597	5410416	5187954
A. talpoideum	Bay 92	2015	Spring	Larvae or Paedomorph	B92_596	6324632	6157327
A. talpoideum	Bay 92	2015	Spring	Larvae or Paedomorph	B92_600	6754656	6578531
A. talpoideum	Bay 92	2015	Spring	Larvae or Paedomorph	B92_601	6761934	6510032
A. talpoideum	Bay 92	2015	Spring	Larvae or Paedomorph	B92_602	6230980	6090257
A. talpoideum	Bay 92	2015	Spring	Larvae or Paedomorph	B92_598	7062726	6805635
A. talpoideum	Bay 92	2015	Spring	Larvae or Paedomorph	B92_599	7722130	7560243
A. talpoideum	Bay 92	2015	Spring	Larvae or Paedomorph	B92_603	11488226	11282316

A. talpoideum	Bay 92	2015	Spring	Larvae or Paedomorph	B92_618	12213488	12017048
A. talpoideum	Bay 92	2015	Spring	Larvae or Paedomorph	B92_620	13220120	12993084
A. talpoideum	Bay 92	2015	Spring	Larvae or Paedomorph	B92_619	14107556	13884742
A. talpoideum	Bay 92	2014	Fall	Adult	B92_31	17295944	17029435

REFERENCES

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in
unrelated individuals. *Genome research* **19**, 1655-1664.

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

Allendorf FW, Hard JJ (2009) Human-induced evolution caused by unnatural selection through
harvest of wild animals. *Proceedings of the National Academy of Sciences* **106**, 9987-
9994.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation
genetics. *Nature Reviews Genetics* **11**, 697-709.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of
RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* **17**, 81-92.

Antao T, Pérez-Figueroa A, Luikart G (2011) Early detection of population declines: high power
of genetic monitoring using effective population size estimators. *Evolutionary
Applications* **4**, 144-154.

Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RADseq underestimates diversity and
introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*
22, 3179-3190.

Baird NA, Etter PD, Atwood TS*, et al.* (2008) Rapid SNP discovery and genetic mapping using
sequenced RAD markers. *PloS one* **3**, e3376.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations
from genome scans. *Molecular Ecology* **13**, 969-980.

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025-2035.

Beerli P (2006) Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* **22**, 341-345.

Beerli P, Palczewski M (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics* **185**, 313-326.

Benestan L, Gosselin T, Perrier C, *et al.* (2015) RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (Homarus americanus). *Molecular Ecology* **24**, 3299-3315.

Bi K, Linderoth T, Vanderpool D, *et al.* (2013) Unlocking the vault: next-generation museum population genomics. *Molecular Ecology* **22**, 6018-6032.

Blaustein AR, Wake DB, Sousa WP (1994) Amphibian declines: judging stability, persistence, and susceptibility of populations to local and global extinctions. *Conservation Biology* **8**, 60-71.

Blois JL, Zarnetske PL, Fitzpatrick MC, Finnegan S (2013) Climate Change and the Past, Present, and Future of Biotic Interactions. *Science* **341**, 499-504.

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution* **18**, 249-256.

Canty A, Ripley B (2012) boot: Bootstrap R (S-Plus) functions. *R package version* **1**.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**, 3124-3140.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* **1**, 171-182.

Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* **10**, 195-205.

Charlier J, Laikre L, Ryman N (2012) Genetic monitoring reveals temporal stability over 30 years in a small, lake-resident brown trout population. *Heredity* **109**, 246-253.

Collins JP, Storfer A (2003) Global amphibian declines: sorting the hypotheses. *Diversity and distributions* **9**, 89-98.

Cosentino BJ, Phillips CA, Schooley RL, Lowe WH, Douglas MR (2012) Linking extinction–colonization dynamics to genetic structure in a salamander metapopulation. *Proceedings of the Royal Society of London B: Biological Sciences* **279**, 1575-1582.

Crow JF, Denniston C (1988) Inbreeding and variance effective population numbers. *Evolution*, 482-495.

Danecek P, Auton A, Abecasis G*, et al.* (2011) The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158.

Daszak P, Scott D, Kilpatrick A*, et al.* (2005) Amphibian population declines at Savannah River site are linked to climate, not chytridiomycosis. *Ecology* **86**, 3232-3237.

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

Do C, Waples RS, Peel D*, et al.* (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources* **14**, 209-214.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular biology and evolution* **22**, 1185-1192.

Duong TY, Scribner KT, Forsythe PS, Crossman JA, Baker EA (2013) Interannual variation in effective number of breeders and estimation of effective population size in long-lived iteroparous lake sturgeon (Acipenser fulvescens). *Molecular Ecology* **22**, 1282-1294.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLoS Genet* **9**, e1003905.

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

Faubet P, Waples RS, Gaggiotti OE (2007) Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. *Molecular Ecology* **16**, 1149-1166.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977-993.

Frankham R (1995) Inbreeding and extinction: a threshold effect. *Conservation Biology* **9**, 792-799.

Frankham R, Bradshaw CJ, Brook BW (2014) Genetics in conservation management: revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation* **170**, 56-63.

Franklin I, Frankham R (1998) How large must populations be to retain evolutionary potential? *Animal Conservation* **1**, 69-70.

Franklin IR (1980) Evolutionary change in small populations. *Conservation biology: an evolutionary-ecological perspective*, 135-149.

Fraser BA, Künstner A, Reznick DN, Dreyer C, Weigel D (2015) Population genomics of natural and experimental populations of guppies (Poecilia reticulata). *Molecular Ecology* **24**, 389-408.

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

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

Funk WC, Tallmon DA, Allendorf FW (1999) Small effective population size in the long-toed salamander. *Molecular Ecology* **8**, 1633-1640.

Gamble LR, McGarigal K, Compton BW (2007) Fidelity and dispersal in the pond-breeding amphibian, Ambystoma opacum: implications for spatio-temporal population dynamics and conservation. *Biological Conservation* **139**, 247-257.

Garner TW (2002) Genome size and microsatellites: the effect of nuclear size on amplification potential. *Genome* **45**, 212-215.

Gattepaille LM, Jakobsson M, Blum MG (2013) Inferring population size changes with sequence and SNP data: lessons from human bottlenecks. *Heredity* **110**, 409-419.

Gautier M, Gharbi K, Cezard T *, et al.* (2013) The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology* **22**, 3165-3178.

Getz WM (1996) A hypothesis regarding the abruptness of density dependence and the growth rate of populations. *Ecology* **77**, 2014-2026.

Gibbons JW, Semlitsch RD (1981) Terrestrial drift fences with pitfall traps: an effective technique for quantitative sampling of animal populations. *Brimleyana* **7**, 1-16.

Gilbert KJ, Whitlock MC (2015) Evaluating methods for estimating local effective population size with and without migration. *Evolution* **69**, 2154-2166.

Glenn TC, Stephan W, Braun MJ (1999) Effects of a population bottleneck on whooping crane mitochondrial DNA variation. *Conservation Biology* **13**, 1097-1107.

Gravel S, Henn BM, Gutenkunst RN, *et al.* (2011) Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences* **108**, 11983-11988.

Greenwald KR, Gibbs HL, Waite TA (2009) Efficacy of Land-Cover Models in Predicting Isolation of Marbled Salamander Populations in a Fragmented Landscape. *Conservation Biology* **23**, 1232-1241.

Griffiths RA, Sewell D, McCrea RS (2010) Dynamics of a declining amphibian metapopulation: survival, dispersal and the impact of climate. *Biological Conservation* **143**, 485-491.

Grover CE, Salmon A, Wendel JF (2012) Targeted sequence capture as a powerful tool for evolutionary analysis1. *American journal of botany* **99**, 312-319.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* **5**, e1000695.

Hanski I (1998) Metapopulation dynamics. *Nature* **396**, 41-49.

Hedrick PW (2011) *Genetics of populations* Jones & Bartlett Learning.

Hein J, Schierup M, Wiuf C (2004) *Gene genealogies, variation and evolution: a primer in coalescent theory* Oxford University Press, USA.

Hill WG (1981) Estimation of effective population size from data on linkage disequilibrium. *Genetical Research* **38**, 209-216.

Hollenbeck C, Portnoy D, Gold J (2016) A method for detecting recent changes in contemporary effective population size from linkage disequilibrium at linked and unlinked loci. *Heredity* **117**, 207-216.

Huang H, Knowles LL (2014) Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. *Systematic Biology*, syu046.

Jehle R, Arntzen J (2002) Review: microsatellite markers in amphibian conservation genetics. *Herpetological journal* **12**, 1-9.

Johnson JA, Dunn PO, Bouzat JL (2007) Effects of recent population bottlenecks on reconstructing the demographic history of prairie-chickens. *Molecular Ecology* **16**, 2203-2222.

Johnson PL, Slatkin M (2008) Accounting for bias from sequencing error in population genetic estimates. *Molecular biology and evolution* **25**, 199-206.

Keinath MC, Timoshevskiy VA, Timoshevskaya NY*, et al.* (2015) Initial characterization of the large genome of the salamander Ambystoma mexicanum using shotgun and laser capture chromosome sequencing. *Scientific reports* **5**, 16413.

Kim J, Mossel E, Rácz MZ, Ross N (2015) Can one hear the shape of a population history? *Theoretical population biology* **100**, 26-38.

Kingman JFC (1982) The coalescent. *Stochastic processes and their applications* **13**, 235-248.

Kinkead K, Abbott A, Otis D (2007) Genetic variation among Ambystoma breeding populations on the Savannah River Site. *Conservation Genetics* **8**, 281-292.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* **15**, 1179-1191.

Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD: analysis of next generation sequencing data. *BMC bioinformatics* **15**, 356.

Lanier HC, Massatti R, He Q, Olson LE, Knowles LL (2015) Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (Ochotona collaris). *Molecular Ecology* **24**, 3688-3705.

Leblois R, Estoup A, Streiff R (2006) Genetics of recent habitat contraction and reduction in population size: does isolation by distance matter? *Molecular Ecology* **15**, 3601-3615.

Leffler EM, Bullaughey K, Matute DR*, et al.* (2012) Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol* **10**, e1001388.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760.

Lischer H, Excoffier L (2012) PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28**, 298-299.

Lozier J (2014) Revisiting comparisons of genetic diversity in stable and declining species: assessing genome-wide polymorphism in North American bumble bees using RAD sequencing. *Molecular Ecology* **23**, 788-801.

Luikart G, Ryman N, Tallmon DA, Schwartz MK, Allendorf FW (2010) Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conservation Genetics* **11**, 355-373.

Mace GM, Collar NJ, Gaston KJ*, et al.* (2008) Quantification of extinction risk: IUCN's system for classifying threatened species. *Conservation Biology* **22**, 1424-1442.

Malaspinas A-S, Westaway MC, Muller C*, et al.* (2016) A genomic history of Aboriginal Australia. *Nature*.

Marsh DM, Trenham PC (2001) Metapopulation dynamics and amphibian conservation. *Conservation Biology* **15**, 40-49.

McCartney-Melstad E, Shaffer HB (2015) Amphibian molecular ecology and how it has informed conservation. *Molecular Ecology* **24**, 5084-5109.

McCoy RC, Garud NR, Kelley JL, Boggs CL, Petrov DA (2014) Genomic inference accurately predicts the timing and severity of a recent bottleneck in a nonmodel insect population. *Molecular Ecology* **23**, 136-150.

McMenamin SK, Hadly EA, Wright CK (2008) Climatic change and wetland desiccation cause amphibian decline in Yellowstone National Park. *Proceedings of the National Academy of Sciences* **105**, 16988-16993.

Morin PA, Martien KK, Taylor BL (2009) Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources* **9**, 66-73.

Moura AE, Kenny JG, Chaudhuri R*, et al.* (2014) Population genomics of the killer whale indicates ecotype evolution in sympatry involving both selection and drift. *Molecular Ecology* **23**, 5179-5192.

Nachman MW, Crowell SL (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**, 297-304.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology* **22**, 2841-2847.

Neel MC, McKelvey K, Ryman N, *et al.* (2013) Estimation of effective population size in

    continuously distributed populations: there goes the neighborhood. *Heredity* **111**, 189-

    199.

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in

    populations. *Evolution*, 1-10.

Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J (2012) SNP calling, genotype calling,

    and sample allele frequency estimation from new-generation sequencing data. *PloS one*

    7, e37558.

Nordborg M, Krone SM (2002) Separation of time scales and convergence to the coalescent in

    structured populations. *Modern Developments in Theoretical Population Genetics: The*

    Legacy of Gustave Malécot, 194-232.

Nunziata SO, Lance SL, Scott DE, Lemmon EM, Weisrock DW (2016) Genomic data detect

    corresponding signatures of population size change on an ecological time scale in two

    salamander species. *Molecular Ecology*.

Nunziata SO, Scott DE, Lance SL (2015) Temporal genetic and demographic monitoring of

    pond-breeding amphibians in three contrasting population systems. *Conservation*

    Genetics **16**, 1335-1344.

Osborne MJ, Davenport SR, Hoagstrom CW, Turner TF (2010) Genetic effective size, Ne, tracks

    density in a small freshwater cyprinid, Pecos bluntnose shiner (Notropis simus

    pecosensis). *Molecular Ecology* **19**, 2832-2844.

Palstra FP, Ruzzante DE (2008) Genetic estimates of contemporary effective population size:

    what can they tell us about the importance of genetic stochasticity for wild population

    persistence? *Molecular Ecology* **17**, 3428-3447.

Palut MPJ, Canziani OF (2007) Contribution of working group II to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press.

Papadopoulou A, Knowles LL (2015) Genomic tests of the species-pump hypothesis: Recent island connectivity cycles drive population divergence but not speciation in Caribbean crickets across the Virgin Islands. *Evolution* **69**, 1501-1517.

Pechmann JH, Scott DE (1991) Declining amphibian populations: the problem of separating human impacts from natural fluctuations. *Science* **253**, 892.

Pechmann JH, Wilbur HM (1994) Putting declining amphibian populations in perspective: natural fluctuations and human impacts. *Herpetologica*, 65-84.

Pechmann JHK, Scott DE, Semlitsch RD*, et al.* (1991) Declining Amphibian Populations: The Problem of Separating Human Impacts from Natural Fluctuations. *Science* **253**, 892-895.

Peng B, Kimmel M (2005) simuPOP: a forward-time population genetics simulation environment. *Bioinformatics* **21**, 3686-3687.

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

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

Phillipsen IC, Funk WC, Hoffman EA, Monsen KJ, Blouin MS (2011) Comparative analyses of effective population size within and among species: ranid frogs as a case study. *Evolution* **65**, 2927-2945.

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PloS one* **7**, e32253.

Polich RL, Searcy CA, Shaffer HB (2013) Effects of tail-clipping on survivorship and growth of larval salamanders. *The Journal of Wildlife Management* **77**, 1420-1425.

Pool JE, Hellmann I, Jensen JD, Nielsen R (2010) Population genetic inference from genomic sequence variation. *Genome research* **20**, 291-300.

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

Pujolar J, Jacobsen M, Frydenberg J*, et al.* (2013) A resource of genome-wide single-nucleotide polymorphisms generated by RAD tag sequencing in the critically endangered European eel. *Molecular Ecology Resources* **13**, 706-714.

Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* **197**, 573-589.

Ramakrishnan U, Hadly EA, Mountain JL (2005) Detecting past population bottlenecks using temporal genetic data. *Molecular Ecology* **14**, 2915-2922.

Rambaut A, Drummond A (2012) Tracer v1. 4. 2007.

Robinson JD, Coffman AJ, Hickerson MJ, Gutenkunst RN (2014) Sampling strategies for frequency spectrum-based population genomic inference. *BMC evolutionary biology* **14**, 254.

Robinson JD, Moyer GR (2013) Linkage disequilibrium and effective population size when generations overlap. *Evolutionary Applications* **6**, 290-302.

Roesti M, Salzburger W, Berner D (2012) Uninformative polymorphisms bias genome scans for signatures of selection. *BMC evolutionary biology* **12**, 94.

Rokas A, Abbot P (2009) Harnessing genomics for evolutionary insights. *Trends in Ecology & Evolution* **24**, 192-200.

Romiguier J, Gayral P, Ballenghien M, *et al.* (2014) Comparative population genomics in

    animals uncovers the determinants of genetic diversity. *Nature* **515**, 261-263.

Rowe G, Beebee T (2004) Reconciling genetic and demographic estimators of effective

    population size in the anuran amphibian Bufo calamita. *Conservation Genetics* **5**, 287.

Saura M, Tenesa A, Woolliams JA, Fernández A, Villanueva B (2015) Evaluation of the linkage-

    disequilibrium method for the estimation of effective population size when generations

    overlap: an empirical case. *BMC genomics* **16**, 922.

Savage WK, Fremier AK, Bradley Shaffer H (2010) Landscape genetics of alpine Sierra Nevada

    salamanders reveal extreme population subdivision in space and time. *Molecular Ecology*

    19, 3301-3314.

Schmeller DS, Merilä J (2007) Demographic and genetic estimates of effective population and

    breeding size in the amphibian Rana temporaria. *Conservation Biology* **21**, 142-151.

Schoener TW (2011) The newest synthesis: understanding the interplay of evolutionary and

    ecological dynamics. *Science* **331**, 426-429.

Schwartz MK, Luikart G, Waples RS (2007) Genetic monitoring as a promising tool for

    conservation and management. *Trends in Ecology & Evolution* **22**, 25-33.

Scott DE (1990) Effects of larval density in Ambystoma opacum: an experiment large-scale field

    enclosures. *Ecology* **71**, 296-306.

Scott DE (1994) The effect of larval density on adult demographic traits in Ambystoma opacum.

    Ecology **75**, 1383-1396.

Scott DE, Komoroski MJ, Croshaw DA, Dixon PM (2013) Terrestrial distribution of pond-

    breeding salamanders around an isolated wetland. *Ecology* **94**, 2537-2546.

Semlitsch R, Scott D, Pechmann J, Gibbons J (1996) Structure and dynamics of an amphibian community: evidence from a 16-year study of a natural pond. *Long-term studies of vertebrate communities. Academic Press, San Diego, California, USA*, 217-248.

Semlitsch RD (2002) Critical elements for biologically based recovery plans of aquatic-breeding amphibians. *Conservation Biology* **16**, 619-629.

Shafer AB, Wolf JB, Alves PC*, et al.* (2015) Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution* **30**, 78-87.

Shultz AJ, Baker AJ, Hill GE, Nolan PM, Edwards SV (2016) SNPs across time and space: population genomic signatures of founder events and epizootics in the House Finch (Haemorhous mexicanus). *Ecology and evolution* **6**, 7475-7489.

Slatkin M (2008) Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics* **9**, 477-485.

Smouse PE (2010) How many SNPs are enough? *Molecular Ecology* **19**, 1265-1266.

Soule ME, Wilcox BA (1980) *Conservation biology: an evolutionary-ecological perspective* Sinauer Associates.

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

Stapley J, Reger J, Feulner PG*, et al.* (2010) Adaptation genomics: the next generation. *Trends in Ecology & Evolution* **25**, 705-712.

Steiner CC, Putnam AS, Hoeck PE, Ryder OA (2013) Conservation genomics of threatened animal species. *Annu. Rev. Anim. Biosci.* **1**, 261-281.

Tallmon DA, Gregovich D, Waples RS*, et al.* (2010) When are genetic methods useful for

    estimating contemporary abundance and detecting population trends? *Molecular Ecology*

    Resources **10**, 684-692.

Taylor BE, Scott DE (1997) Effects of larval density dependence on population dynamics of

    Ambystoma opacum. *Herpetologica*, 132-145.

Taylor BE, Scott DE, Gibbons JW (2006) Catastrophic reproductive failure, terrestrial survival,

    and persistence of the marbled salamander. *Conservation Biology* **20**, 792-801.

Therkildsen NO, Hemmer-Hansen J, Als TD*, et al.* (2013) Microevolution in time and space:

    SNP analysis of historical DNA reveals dynamic signatures of selection in Atlantic cod.

    Molecular Ecology **22**, 2424-2440.

Thomé MTC, Carstens BC (2016) Phylogeographic model selection leads to insight into the

    evolutionary history of four-eyed frogs. *Proceedings of the National Academy of Sciences*

    113, 8010-8017.

Todd BD, Scott DE, Pechmann JH, Gibbons JW (2010) Climate change correlates with rapid

    delays and advancements in reproductive timing in an amphibian community.

    Proceedings of the Royal Society of London B: Biological Sciences, rspb20101768.

Wakeley J, Sargsyan O (2009) Extensions of the coalescent effective population size. *Genetics*

    181, 341-345.

Walls SC, Barichivich WJ, Brown ME (2013a) Drought, deluge and declines: the impact of

    precipitation extremes on amphibians in a changing climate. *Biology* **2**, 399-418.

Walls SC, Barichivich WJ, Brown ME, Scott DE, Hossack BR (2013b) Influence of drought on

    salamander occupancy of isolated wetlands on the southeastern coastal plain of the

    United States. *Wetlands* **33**, 345-354.

Wang IJ, Johnson JR, Johnson BB, Shaffer HB (2011) Effective population size is strongly correlated with breeding pond size in the endangered California tiger salamander, Ambystoma californiense. *Conservation Genetics* **12**, 911-920.

Wang IJ, Summers K (2010) Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Molecular Ecology* **19**, 447-458.

Wang J (2005) Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **360**, 1395-1409.

Wang J, Whitlock MC (2003) Estimating effective population size and migration rates from genetic samples over space and time. *Genetics* **163**, 429-446.

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57-63.

Waples RK, Larson WA, Waples RS (2016) Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. *Heredity*.

Waples RS (1991) Genetic methods for estimating the effective size of cetacean populations. *Report of the International Whaling Commission (special issue)* **13**, 279-300.

Waples RS (2005) Genetic estimates of contemporary effective population size: to what time periods do the estimates apply? *Molecular Ecology* **14**, 3335-3352.

Waples RS (2006) A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conservation Genetics* **7**, 167-184.

Waples RS (2016) Making sense of genetic estimates of effective population size. *Molecular Ecology* **25**, 4689-4691.

Waples RS, Antao T (2014) Intermittent breeding and constraints on litter size: consequences for effective population size per generation (Ne) and per reproductive cycle (Nb). *Evolution* **68**, 1722-1734.

Waples RS, Do C (2010) Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Applications* **3**, 244-262.

Waples RS, England PR (2011) Estimating contemporary effective population size on the basis of linkage disequilibrium in the face of migration. *Genetics* **189**, 633-644.

Waples RS, Luikart G, Faulkner JR, Tallmon DA (2013) Simple life-history traits explain key effective population size ratios across diverse taxa. *Proceedings of the Royal Society of London B: Biological Sciences* **280**, 20131339.

Weber D, Stewart BS, Garza JC, Lehman N (2000) An empirical genetic assessment of the severity of the northern elephant seal population bottleneck. *Current Biology* **10**, 1287-1290.

Weir SM, Scott DE, Salice CJ, Lance SL (2016) Integrating copper toxicity and climate change to understand extinction risk to two species of pond-breeding anurans. *Ecological Applications* **26**, 1721-1732.

Whiteley AR, Coombs JA, Cembrola M*, et al.* (2015) Effective number of breeders provides a link between interannual variation in stream flow and individual reproductive contribution in a stream salmonid. *Molecular Ecology* **24**, 3585-3602.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**, 1177-1191.

Education

Eastern Kentucky University	**M.S.**, Biological Sciences	2011
	Advisor: Dr. Stephen Richter	
Towson University	**B.S.**, Biology and Environmental Science	2009
	summa cum laude with honors	
	Advisor: Dr. Joel Snodgrass	

Publications

PUBLISHED AND IN PRESS

Nunziata SO, EM Lemmon, DE Scott, SL Lance, & DW Weisrock. 2017. Genomic data detect corresponding signatures of population size change on an ecological time scale in two salamander species. Molecular Ecology, 26(4), 1060–1074.

Nunziata SO, P Wallenhorst, MA Barrett, RE Junge, AD Yoder, & DW Weisrock. 2016. Population and conservation genetics in an endangered lemur, *Indri indri,* across three forest reserves in Madagascar. International Journal of Primatology, 37(6), 688–702. (Cover Image)

Love CN, ME Winzeler, RR Beasley, DE Scott, **SO Nunziata,** & SL Lance. 2016. Patterns of amphibian infection prevalence across wetlands on the Savannah River Site, South Carolina, USA. Diseases of Aquatic Organisms, 121(1), 1–14.

Nunziata SO, DE Scott, & SL Lance. 2015. Temporal genetic and demographic monitoring of pond- breeding amphibians in three contrasting population systems. Conservation Genetics, 16(6), 1335–1344.

Ramos AG, **SO Nunziata**, SL Lance, C Rodríguez, BC Faircloth, PA Gowaty, & H Drummond. 2014. Interactive effects of male and female age on extra-pair paternity in a socially monogamous seabird. Behavioral Ecology and Sociobiology, 68(10), 16031–1609.

Ramos AG, **SO Nunziata**, SL Lance, C Rodríguez, BC Faircloth, PA Gowaty, & H Drummond. 2014. Habitat structure and colony structure constrain extrapair paternity in a colonial bird. Animal Behaviour, 95, 1211–127.

Richter SC, EM O'Neil, **SO Nunziata**, A Rumments, ES Gustin, J Young, & BI Crother. 2014. Cryptic diversity and conservation of gopher frogs across the Southeastern United States. Copeia, 2, 231–237.

Lance SL, CN Love, **SO Nunziata**, J O'Bryhim, DE Scott, RW Flynn, & K Jones. 2013. 32 species validation of a new Illumina paired-end approach for the development of microsatellites. PlosOne, 8(11), e81853.

Richer SC, & **SO Nunziata**. 2013. Survival to metamorphosis is positively related to genetic variability in a critically endangered amphibian species. Animal Conservation, 17(3), 265-274.

Nunziata SO, MJ Lannoo, JR Robb, DR Karns, SL Lance, & SC Richter. 2013. Population and Conservation Genetics of Crawfish Frogs, *Lithobates areolatus*, at Their Northeastern Range Limit. Journal of Herpetology, 47(2), 361–368.

Gray EP, **S Nunziata**, JW Snodgrass, DR Ownby, & JE Havel. 2010. Predation on Green Frog Eggs (*Rana clamitans*) by Ostracoda. Copeia, 2010(3), 452-456.

PRIMER NOTES

Ricono A, G Bupp, C Peterson, **SO Nunziata**, SL Lance, & CL Pruett. 2015. Development and characterization of microsatellite loci for the endangered scrub lupine, *Lupinus aridorum* (*Fabaceae*). Applications in Plant Sciences, 3(4).

Love CN, RW Flynn, **SO Nunziata**, KL Jones, & SL Lance. 2013. Development of 31 polymorphic microsatellite markers for the mole salamander (*Ambystoma talpoideum*) using Illumina paired-end sequencing. Conservation Genetic Resources, 5,951–954.

Jenkins TM, TD Eaton, **SO Nunziata**, & SL Lance. 2013. Paired-end illumina shotgun sequencing used to develop the first microsatellite primers for *Megacopta cribraria* (F.)(*Hemiptera: Heteroptera: Plataspidae).* Journal of Entomological Science, 48(4), 345–351.

Nunziata SO, SL Lance, KL Jones, SA Nerkowski, & AE Metcalf. 2013. Development and characterization of twenty-three microsatellite markers for the freshwater minnow Santa Ana speckled dace (*Rhinichthys osculus* spp., Cyprinidae) using paired-end Illumina shotgun sequencing. Conservation Genetics Resources, 5(1), 145–148.

Nunziata SO, JD Karron, RJ Mitchell, SL Lance, KL Jones, & DW Trapnell. 2012. Characterization of 42 polymorphic microsatellite loci in *Mimulus ringens* (Phrymaceae) using Illumina sequencing. American Journal of Botany, 99(12), e477–80.

Allen JM, SG Obae, MH Brand, JA Silander, KL Jones, **SO Nunziata**, & SL Lance. 2012. Development and characterization of microsatellite markers for *Berberis thunbergii* (Berberidaceae). American Journal of Botany, 99(5), e220–2.

Nunziata SO, SC Richter, RD Denton, JM Yeiser, DE Wells, KL Jones, C Hagen, & SL Lance. 2011. Fourteen novel microsatellite markers for the gopher frog, *Lithobates capito* (Amphibia: Ranidae). Conservation Genetics Resources, 4(1), 201–203.

Nunziata SO, DE Scott, KL Jones, C Hagen, & SL Lance. 2011. Twelve novel microsatellite markers for the marbled salamander, *Ambystoma opacum*. Conservation Genetics Resources 3:773-775.

Presentations

ORAL PRESENTATIONS (*Denotes Presenter)

Nunziata SO*, EM Lemmon, DE Scott, SL Lance, and DW Weisrock. "Genomic data detect corresponding signatures of population size change on an ecological time scale in two salamander species"
2016 *Ecological Society of America Meeting* – Ft Lauderdale, FL

SO Nunziata, SL Lance, DE Scott, and DW Weisrock* "Genomic data detect corresponding signatures of very recent population expansion in marbled salamanders"
2015 *Society for the Study of Amphibians and Reptiles Annual Meeting* – Lawrence, KS

ME Winzeler*, CN Love, DE Scott, R Beasley, **SO Nunziata**, and SL Lance "Prevalence of two amphibian diseases at contaminated wetlands of the Savannah River Site, SC"
2014 *Joint Meeting of Ichthyologists and Herpetologists* – Chattanooga, TN

SO Nunziata*, SL Lance, DE Scott, and DW Weisrock "Temporal genetic and demographic monitoring of a pond breeding salamander"
2014 *Evolution Meeting* – Raleigh, NC

SL Lance*, CN Love, **SO Nunziata**, and DE Scott "Prevalence of two amphibian diseases, ranavirus and chytridiomycosis, in contaminated and uncontaminated sites on the Savannah River Site"
2013 *Wildlife Disease Association Meeting* – Knoxville, TN

SO Nunziata*, P Wallenhorst, M Barrett, RE Junge, AD Yoder, and DW Weisrock "Population and conservation genetics of an endangered lemur species across protected forest reserves in Madagascar"
2013 *UK Center for Ecology Evolution and Behavior's Spring Symposium* – Lexington, KY

SO Nunziata*, DE Scott, and SL Lance "Genetic and demographic patterns of an expanding population of *Ambystoma opacum"*
2012 Society for Integrative and Comparative Biology Meeting – Charleston, SC

SO Nunziata*, MJ Lannoo, JR Robb, DR Karns, SL Lance, and SC Richter "Population and conservation genetics of crawfish frogs in Indiana"
2011 *Southeastern Population Ecology & Evolutionary Genetics Meeting* – Reidsville, NC

POSTER PRESENTATIONS (*Denotes Presenter)
CN Love*, ME Winzeler, R Beasley, **SO Nunziata**, DE Scott, and SL Lance "Prevalence of Ranavirus at contaminated and uncontaminated sites"
2014 *Society for Integrative and Comparative Biology Meeting* – Austin, TX

CN Love*, ME Winzeler, **SO Nunziata**, DE Scott, and SL Lance "Effects of contaminants on amphibian disease prevalence"
2013 *Georgia Chapter of the Wildlife Society Meeting* – Athens, GA

SO Nunziata*, P Wallenhorst, M Barrett, RE Junge, AD Yoder, and DW Weisrock "Population and conservation genetics of an endangered lemur species across protected forest reserves in Madagascar" 2013 *Evolution Meeting* – Snowbird, UT, June 2013.

SL Lance, **SO Nunziata**, CN Love*****, W Flynn, and KL Jones "A new and better method for genetic microsatellite marker development validated in four amphibian species"
2012 *Southeastern Partners in Amphibian and Reptile Conservation Meeting* – Fall Creek Falls, TN

SO Nunziata*, MJ Lannoo, JR Robb, DR Karns, SL Lance, and SC Richter "Population and Conservation Genetics of Crawfish Frogs, *Lithobates areolatus*, at their Northeastern Range Limit"
2012 *Southeastern Partners in Amphibian and Reptile Conservation Meeting* – Fall Creek Falls, TN

SO Nunziata*, DE Scott, and SL Lance "Genetic and demographic patterns of an expanding population of *Ambystoma opacum* on the Savannah River Site"
2011 *Joint Meeting of Ichthyologists and Herpetologists* – Minneapolis, MN

SC Richter*, and **SO Nunziata** "Relationship between genetic variability and survival in a critically imperiled amphibian, *Lithobates sevosus"*
2011 *Joint Meeting of Ichthyologists and Herpetologists* – Minneapolis, MN

SO Nunziata*, and SC Richter "Distribution and Abundance of the Southern Two-lined Salamander (*Eurycea cirrigera*) in a Central Kentucky Stream System"
2010 *Annual Meeting of the Kentucky Academy of Sciences* – Bowling Green, KY

SO Nunziata* and DC Forester "Nest-site Selection by Northern Two-lined Salamanders (*Eurycea bislineata*)"
2009 *Towson University Student Research Expo* – Towson, MD

SO Nunziata*, and JW Snodgrass "Impacts of Road Salt on Stream Salamanders"
2008 *Biogeochemistry REU Poster Presentation* – Towson, MD

Grants and Funding

NSF Doctoral Dissertation Improvement Grant - $18,946	2016
University of Kentucky, Ribble mini-grant - $700	2015
University of Kentucky, Ribble Travel Grant - $500	2014
University of Kentucky, Graduate School Travel Grant - $400	2014
University of Kentucky, Ribble mini-grant - $700	2014
University of Kentucky, Ribble Travel Grant - $500	2013
University of Kentucky, Graduate School Travel Grant - $400	2013
Society for the Study of Amphibians and Reptiles - $500	2013
University of Kentucky, Ribble mini-grant - $700	2013
Kentucky Academy of Sciences, Marcia Athey Grant - $992	2012
Kentucky NSF EPSCoR Research Scholars Program - $5,000	2012
Kentucky NSF EPSCoR Research Scholars Program - $5,000	2010
Undergraduate Research Grant, Fisher College of Science and Mathematics - $462	2008
Undergraduate Research Grant, Towson University - $462	2008