The genetic diversity of North American vertebrates in protected areas

Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in
the Graduate School of The Ohio State University

By

Coleen Elizabeth Paige Thompson, B.S.

Graduate Program in Evolution, Ecology & Organismal Biology

The Ohio State University

2019

Thesis Committee

Bryan C. Carstens, Advisor

Lisle Gibbs

Steve Hovick

Andreas Chavez

Copyrighted by
Coleen Elizabeth Paige Thompson
2019

Abstract

Protected areas play a crucial role in the conservation of biodiversity, but it is unclear if these areas have an influence on genetic diversity. Since genetic diversity is a crucial component of a species ability to adapt and persist in an environment over long periods of time, its assessment is valuable when designating areas for conservation. As a first step towards addressing this issue, we compare genetic diversity inside and outside of protected areas in North America using repurposed data. We tested the null hypothesis that there is no difference between genetic diversity inside compared to outside of protected areas in 44 vertebrate species. A substantial portion of vertebrate species exhibit significant differences in the amount of intraspecific genetic diversity in a comparison between protected and unprotected areas. While our simulation testing suggests that this result is not an artifact of sampling, it is unclear what factors influence the relative amount of genetic diversity inside and outside of protected areas across species.

Acknowledgments

My thesis work would not be possible without the guidance and time put forth by my collaborators, committee, and colleagues. Tara Pelletier and Bryan C. Carstens both provided insight on my thesis proposals and drafts. I would also like to thank Bryan for all of the time he set aside to discuss this project, his help with writing, and for pushing me to have confidence in my research. A huge thanks to Tara as well for teaching me coding skills and how to efficiently repurpose data. My committee, Lisle Gibbs, Andreas Chavez, and Steve Hovick, all provided useful comments and thoughts on how to improve my research and different ways to understand my results and for this I am grateful. I thank my colleagues in my lab who all provided feedback and encouragement throughout my two years spent obtaining my degree in the form of proposal and presentation recommendations. Additionally, they helped me understand and write code in various programming languages. Lastly, I want to thank all the scientists who uploaded their genetic and locality data to open access repositories, this research would not be possible without them.

Vita

2013 East Clinton High School, Lees Creek, OH
2017 B.S., The Ohio State University
2015-2017 Undergraduate Research Assistant- Wolfe
Lab, The Ohio State University
2017 Laboratory Technician- Carstens Lab,
The Ohio State University
2017- present Graduate Research and Teaching Assistant,
The Ohio State University

Fields of Study

Major Field: Evolution, Ecology & Organismal Biology

Table of Contents

Abstract.....	ii
Acknowledgments.....	iii
Vita.....	iiv
List of Tables	vi
List of Figures	vii
Chapter 1. The genetic diversity of North American vertebrates in protected areas	1
Bibliography	27
Appendix A. Additional Tables	31

List of Tables

Table 1 Species, gene, p-value (Π), and measures of genetic diversity both inside and outside of protected areas. The average of the PI values from the null distribution are italicized. The other PI values are the test statistic. The “*” denotes significance.....	17
Table 2 Test study species (not in study) and pvalue.	31
Table 3 Random test study species and pvalues.	32
Table 4 Original papers that generated the data used in this study were examined to see if protected areas included in our investigations overlap with a climate refuge during the Pleistocene glaciations. * mentioned that current range is mainly in protected area, but nothing about overlap	33

List of Figures

Figure 1 Map of <i>Gulo Gulo</i> distributed throughout NA. The shaded areas indicate the boundaries of protected areas and the yellow dots indicate the individual(s) inside of protected areas while the blue dots indicate those which are outside.	20
Figure 2 Histograms of the null distribution of genetic diversity in three species with contrasting results. The red line indicates the averaged amount of genetic diversity for the population opposite of the distribution. a) Genetic diversity in <i>Coluber constrictor</i> is significantly greater outside of protected areas, b) There is no significant difference in genetic diversity for <i>Gulo gulo</i> , c) Genetic diversity in <i>Marmota caligata</i> is significantly greater inside of protected areas.	21
Figure 3 Regressions of body mass ($r^2 = 0.331$), home range size ($r^2 = 0.0000541$), and maximum latitude ($r^2 = 0.00321$) on the difference in nucleotide diversity (inside of protected areas – outside of protected areas).	23
Figure 4 The p-values for all 44 vertebrate species.	26

Chapter 1. The genetic diversity of North American vertebrates in protected areas.

Coleen Thompson, Tara Pelletier, and Bryan Carstens

Introduction

Genetic diversity is an important component of biodiversity, and one thought to be essential to long-term species persistence and the ability of a species to adapt to environmental change over time (Frankel 1974; Spielman et al 2004; Hoffmann & Sgro 2011). Populations with high genetic diversity are likely to be more robust in the face of anthropogenic forces and natural disasters (Barabás & D’Andrea, 2016; Reed et al., 2002) because they contain more variation for natural selection to act on (Lacy, 1987). The Convention on Biological Diversity (CBD) began including genetic diversity among its 2010 biodiversity targets because the conservation of genetic diversity is believed to be a vital component in the conservation of threatened species (Laikre et al., 2010). Despite the clear theoretical importance of genetic diversity, this parameter remains unaccounted for in many investigations into biodiversity.

While most protected areas were originally created for recreational use and to preserve scenic vistas, they have potential to act as reservoirs of biodiversity (Chape, Harrison, Spalding, & Lysenko, 2010). Watson et al. (2014) argued that protected areas are better at maintaining populations of species than other approaches, and thus should be the top consideration for threatened species. However, other studies suggest that protected areas do not encompass all of the biodiversity in the area; for example,

protected areas in California do not encompass all of the hotspots of genetic diversity (Vandergast, Bohonak, Hathaway, Boys, & Fisher, 2008). To complicate matters, the effectiveness of protected areas may decrease since these regions are environmentally unstable (Sgrò, Lowe, & Hoffmann, 2011). Aside from indirect anthropogenic forces, such as climate change, human population density alone also negatively effects genetic diversity and in protected areas with less humans there is more genetic diversity (Cahill et al., 2017; Carvalho et al., 2017). A crucial component for evaluating the importance of protected areas for biodiversity conservation is to understand if these areas contain different amounts of genetic diversity than surrounding regions across a wide range of species.

This effort is complicated by numerous other factors that can influence intraspecific genetic diversity. Populations and species at lower latitudes typically exhibit more genetic diversity than their counterparts at higher latitudes (Adams & Hadly, 2013). Species life history traits are also important, particularly those related to population size (Hague & Routman, 2016) and reproductive strategy (Valiente, Juanes, & Garcia-Vazquez, 2005) since these traits largely determine the extent to which genetic drift and inbreeding depression will remove variation from the population. Across broader geographic scales and deeper into evolutionary time, historical demographic processes such as demographic bottlenecks (Hoelzel et al., 2002) and range expansions (Hewitt, 2004) also influence intraspecific genetic diversity. Populations in temperate regions sought refuge from the glaciers and many experienced bottlenecks, founder effects, and genetic drift when they expanded after the retreat of the glaciers (Nicola M. Anthony, Mireille Johnson-Bawe, Kathryn Jeffery, Stephen L. Clifford, Kate A. Abernethy,

Caroline E. Tutin, Sally A. Lahm, Lee J. T. White, John F. Utey, E. Jean Wickings, 2007; Petit et al., 2003). Even tropical regions experienced habitat fluctuation during the Pleistocene, but populations in regions with stable habitat over time contain more genetic diversity (Carnaval, Hickerson, Haddad, Rodrigues, & Moritz, 2009; Hugall, Moritz, Moussalli, & Stanislav, 2002). In temperate regions at the last glacial maxima (LGM) some areas were stable climatically and served as refugia while others were covered in ice or otherwise modified. Decades of phylogeographic work have demonstrated that these refugial areas contain more genetic diversity than other regions (Gavin et al., 2014) but since refugia tend to be taxon-specific it is difficult to generalize across species.

Experimental design can also influence the estimation of genetic diversity. Since variation in the amount of intraspecific genetic diversity within populations is correlated with the physical distance between individuals (Wright, 1943), the environmental conditions across the range (Wang 2013), and the size of the range (Pelletier & Carstens, 2018), sampling of individuals will have a large effect on estimates of genetic diversity. Sample size (Pruett & Winker, 2008) is positively correlated with estimates of genetic diversity. Taken in their entirety, this combination of unique species traits and sampling factors complicates the comparison of genetic diversity across species and make it particularly challenging to identify factors that influence the relative amounts of intraspecific genetic diversity within the geographic range of a given species.

The literature on the conservation implications of intraspecific diversity is limited in both taxonomic and geographic scope, but as it becomes easier to collect and analyze genetic data (Garrick et al., 2015), conservationists are beginning to explore intraspecific genetic diversity in protected areas. For example, Guzman et. al. (2015) found that the

amount of genetic diversity exhibited by the holm oak in Spain did not differ between protected and non-protected areas. In contrast, Miller and Ayre (2008) found that marine protected areas contain nearly all of the genetic diversity in coral reefs (Guzmán, Rodríguez López, Forrest, Cano, & Vargas, 2015; Miller & Ayre, 2008). These contrasting findings highlight the importance of looking at the conservation of genetic diversity within protected areas across broad geographical regions. However, information on genetic diversity in protected areas in North America is lacking. Traditionally, answering questions about biodiversity patterns in multiple species has been challenging due to intensive field and laboratory work, and thus, can be near impossible for one person to complete. However, by re-purposing data and automating pipelines, larger genetic studies are now plausible. By examining genetic diversity, we can begin to find areas that will be more robust in the face of climate change, and potentially provide source populations in the future. Source populations contain majority of the genetic diversity in the population and they influence the diversity of future populations (Roman & Darling, 2007). The main objectives of this study are to 1) identify if there is a difference in genetic diversity of vertebrates in North America inside compared to outside of protected areas, and 2) examine the influence of life history traits, if any, on the differences of genetic diversity.

Methods

Online data repositories enable scientists to ask broad-scale questions about multiple species using data that have been collected to answer a wide array of questions. To assess the potential influence of protected areas on intraspecific genetic diversity, we utilized

GenBank and the Global Biodiversity Information Facility (GBIF) as sources of genetic and collection locality data, respectively. We then built computational pipelines in R to repurpose these data in order to calculate genetic diversity for all 44 species in the study. We focused on mitochondrial DNA (mtDNA) because it has traditionally been used for intraspecific genetic investigations and is easy to collect in vertebrates, and thus has been used in many intraspecific genetic studies (Avise, 1987).

Genetic and Geographic Data

All occurrence data for North American mammals, reptiles, and amphibians were downloaded from Global Biodiversity Information Facility (GBIF) by specifying “Mammal, Reptile, Amphibia” under Scientific Name, “Observation, Human Observation, Material Sample, and Preserved Specimen” under Basis of Record, and “North America” under Location. Only those with GPS coordinates were retained. The associated genetic data were downloaded from GenBank when available using a custom Perl script to extract individual identifications from the GBIF download, and then manually read into GenBank using the Batch Entrez function (Pelletier & Carstens, 2018). We then downloaded genetic data by selecting the Darwin Core Archive as the GBIF format to download the data as a Genbank file. Python scripts were used to rename sequences in a consistent manner (e.g., abbreviated names in Genbank might include *cytb*, *cytochrome-b*, *CYTB*, *CytB*, etc.) and concatenate all sequences on a locus-by-locus and species-by-species basis. Multiple sequence alignments were conducted for each species and locus using only sequences that were >200bp using MUSCLE v. 3.8.31

(Edgar, 2010), with the unrefined settings. All following analyses were conducted using R v. 3.3.3 (RStudio Team, 2015).

Protected Areas

GIS shape files of protected areas in Canada, the United States, and Mexico were downloaded from the World Database on Protected Areas (WDPA), Protected Areas Database of the United States (PAD-US), and Protected Planet, respectively. The files contain information on protected areas at the national, state, and local level. These include parks, forests and refuges to name a few, as each country has 10+ types of protected areas. These were used to identify individuals from our dataset as either inside or outside of protected areas. Localities were plotted for each species in our dataset (Figure 1).

National Parks vs. Forests

National parks and forests are protected differently in that forests allow for multiple human uses such as lumber and grazing (doi.gov). To see if the level of protection influenced the genetic diversity in any way, we calculated the number of species that had significant results (greater diversity inside or outside of protected areas). Shapefiles for the National Park Service unit boundaries in the U.S. were obtained from Data.gov, and the National Forests shapefile was downloaded from the US Forest Service. The analysis was ran as mentioned below, but with using only the National Parks shapefile for the park data and likewise for the forest data. It does not include the other types of protected areas.

Given the well-known effect of geographic distance on genetic diversity (Wright 1943; Pelletier and Carstens 2018) we controlled for geographic distance using the following steps: 1) calculate the geographic distance between all individuals outside of protected areas, 2) count how many individuals are within half of the maximum distance inside of protected areas from the other individuals, 3) use the individual with the largest count from step 2 to randomly select other individuals to calculate distance. This assures that no two individuals outside of protected areas had a greater distance than any of the individuals inside of protected areas, assuming outside of protected areas had the potential to include individuals at a much greater distance than inside. These datasets were then used to estimate genetic diversity.

Genetic Diversity

Nucleotide diversity (π), Watterson's θ , and haplotype diversity (H) were all used to assess genetic diversity. π was the main measure used as it incorporates parts of Watterson's θ and H. These other measures were used as metrics to support the inferences made from the π calculations, and thus significance was not calculated for these measures.

Nucleotide Diversity

Nucleotide diversity (π) is a useful measure of intraspecific genetic diversity using mtDNA, which quantifies the degree of polymorphisms within a nucleotide sequence (Nei, 1978). π was calculated in the package “pegas” and the function “nuc.div” (Paradis,

2010) for individuals inside of protected areas (π_p). For each species we created a null distribution that could be used to calculate the critical value of the empirical data, or the value of the empirical data that would indicate significance. A parametric simulation test was used to assess the significance of the value of π inside compared to outside of protected areas. We developed our test as follows: A null distribution was generated for each species by bootstrap resampling by 1) assessing if there were more individuals inside or outside of protected areas, 2) building a distribution based off the population with the largest sample size and randomly selecting the number of individuals of the smaller sample size with replacement and then calculating π for this group (this was repeated 1000 times to build a distribution), and 3) calculating π once for the smaller sample size and plotting this on the distribution from the larger population (Figure 2). For all comparisons, an alpha level of 0.05 was used to assess significance.

Watterson's Theta and Haplotype Diversity

Watterson's θ (Watterson, 1975) is a measure of the count of segregating sites in relation to the number of expected sites for a neutral gene. We calculated θ using the package "PopGenome" and the function "neutrality stats" (Pfeifer, Wittelsbrger, Ramos-Onsins, & Lercher, 2014). We also calculated haplotype diversity (H), which is the measure of haplotypes, a group of SNPs commonly inherited together, compared within each population. H was calculated using the package "pegas" and the function "haplotype" (Paradis, 2010). The larger population was downsized to the sample size of the smaller population and there was one calculation made for haplotypes in each population.

Influencing Factors

Trait data

To see if life-history traits were influencing our results, we used mammal trait data, since most of the species in the study were mammals and there is a database compiled with trait data for most species, PanTHERIA (Jones et al., 2009). Adult body size (g), home range size (km²), and maximum latitude (dd) were regressed against the difference between the average nucleotide diversity for each species ($\pi_{in} - \pi_{out}$).

Taxonomy

To test the influence of taxonomy on genetic diversity, we conducted a t-test on the biggest representatives of the groups sampled. Using the “t.test” function in R, a Welch two sample t-test (Welch, 1938) was conducted for mammal v. non-mammal species and rodent v. non-rodent species in relation to their p-value.

Results

Genetic and Geographic Data/Protected areas.

Data from some ~5,000 species of reptile, amphibian, and mammal from North America were available on GBIF, but only ~500 of these had associated genetic data on GenBank. After filtering the data such that species were required to have at least two individuals inside and outside of protected areas, contain sequence reads >200 bp, and corrected for geographic distance, we were left with 44 vertebrate species (33 mammals, 2 amphibians, 9 squamates). The amphibian data averaged 120 individuals inside of protected areas and

23 outside of protected areas, with post-filter data sample sizes inside or outside of protected areas at ~23 individuals. Similarly, mammals had an average of 24 individuals inside and outside of protected areas, and squamates had an average of 9 individuals.

Analysis/ Genetic Diversity.

Of the 44 species in the study, 21 species (48%) showed a significant difference in nucleotide diversity inside and outside of protected areas (Figure 4). Of these, 12 species exhibited greater nucleotide diversity inside of protected areas, while 9 exhibited greater nucleotide diversity outside of the protected areas. Within mammals, 30% had more diversity inside of protected areas, and 18% had more diversity outside of protected areas. For squamates 11% of species had greater diversity inside protected areas and 33% had greater diversity outside. There were only 2 amphibian species in the study and only one showed a significant difference in nucleotide diversity inside and outside of protected areas; for this species, nucleotide diversity was greater inside of protected areas.

15 of 41 species had greater haplotype diversity inside of protected areas, while 5 had greater diversity outside, and the rest of the species had identical haplotype diversity regardless of location. More species inside of protected areas had a higher theta (56%) than those outside (39%). For each individual species, haplotype diversity and Watterson's theta both reflected what was concluded by the measure of nucleotide diversity.

There was no difference in the number of species with greater diversity inside or outside of protected areas specifically examining National Forests and National Parks. The National Parks contained 26 species with data both inside and outside, and 4 species had significantly greater diversity inside, while 4 have significantly greater diversity outside. The National Forests also contained 26 species with data and 7 had significant diversity inside, while 5 had significant diversity outside.

Influencing Factors

Neither life-history traits nor taxonomic rank were significantly related to the amounts of genetic diversity in the differing populations. There was no significant correlation between body mass ($r^2 = 0.331$), home range size ($r^2 = .0000541$), or maximum latitude ($r^2 = .00321$) and a significant amount of genetic diversity either in or outside of protected areas for mammals (Figure 3). Additionally, whether the species was of the order rodentia or a mammal did not have significance (p-values = 0.311 and 0.6824, respectively).

Random Test

With a sample size of 44 species, and an α value of 0.05 we would expect ~2 species to show significance by chance. To test for potential bias in our approach, we chose 10 of the species with the greatest sample size that did not occur in protected areas, assigned samples to populations at random, and repeated our non-parametric resampling approach using the average maximum distance inside of protected areas to correct for distance (1832360.81 km). Three of the ten species had a population with significantly greater

nucleotide diversity (Table 2). To investigate this further, we selected 10 of these species with cyt-b data and large sample sizes that had data inside and outside of protected areas and then randomly assigned all samples to two populations (Table 3). We then repeated our non-parametric resampling approach. One species had a significantly differing amount of genetic diversity, however when correctly sorted into populations inside and outside of protected areas, seven individuals were significant. These tests were used to assess the amounts of significance we would expect to see at random.

Discussion

Evaluation of statistical signal

Our results suggest that there are a substantial number of species that contain differing amounts of genetic diversity between protected and unprotected regions. In nearly half of the examined species, the genetic diversity that was estimated from samples that were partitioned inside/outside of protected areas were significantly different then the null expectations which were generated using non-parametric resampling. We designed this approach to correct for geographic distance by limiting the physical separation of the samples in the replicated datasets to a distance that was less than or equal to the greatest geographic distance in the actual samples. Sample size was also held consistent between the actual and replicated data. Since (to our knowledge) this approach has not been applied to this question by other researchers, we designed two simulation tests to assess its statistical power.

For our first test the non-parametric resampling was repeated using data from 10 species with the geographic locality data omitted. By randomly assigning samples to two

groups, we approximated the division of samples in the actual analysis into protected v. unprotected sets. We then calculated π from each of the two sets and compared this value to the null distribution. None of these tests identified a significant difference in π in this simulation, despite the fact that 7/10 of the species were significant when they were analyzed with the geographic locality data. This indicates that the signal we have observed in our actual tests is derived in large part from the spatial distribution of samples.

In our second test, we analyzed data from the group of species that had available locality data but that did not have samples from inside and outside of protected areas. We repeated our non-parametric resampling by randomly selecting samples from sets that approximated those from the protected areas in both their size and the distance of the samples to one another. These pseudo-protected area sets were then used in the non-parametric resampling approach. Results indicate that ~30% of the pseudo-protected area sets of samples were significant. This indicates that the non-parametric approach has a substantial level of error, albeit one that likely does not explain all of the significant tests that we observed in our empirical data. This makes assessment of our results difficult because we expect that some, but not all, of the significant test results may represent false positives.

Interactions between life history traits and genetic diversity

Species-specific traits related to factors such as dispersal ability or reproductive system might influence the degree of gene flow and thus interact with protected areas to influence genetic diversity. Other traits such as body mass, maximum latitude, and home

range size are known to influence the degree of population genetic structure within species (Ellegren & Galtier, 2016; Pelletier & Carstens, 2018). In addition, the diet type (omnivore, herbivore, carnivore) could be linked to a relationship between home range size and body mass (Ottaviani, Cairns, Oliverio, & Boitani, 2006). Exploring these traits, mechanisms, and links could provide information on why we see differences between the amounts of genetic diversity in populations of the same species. In order to explore these effects on a broad level, we tested for a correlation between taxonomic rank and the p -value from our empirical test. However, we identified no significant difference between species with significant results and those that had no difference in genetic diversity. We also conducted several linear regressions but found no relationship between genetic diversity and body mass, home range size, or maximum latitude. Clearly, more work is needed to understand if and how organismal traits interact with protected areas to influence intraspecific genetic diversity.

Edge effects that may occur at the boundaries between protected and non-protected areas are another factor that may influence the results of our study. While it is likely that all species are influenced by edge effects (Muposhi, Chademana, Gandiwa, & Muboko, 2016), these effects are likely to interact with organismal traits to affect species in different ways. Additionally, habitat fragmentation may influence our results by exacerbating edge effects (Wilson et al. 2016), and while the (presumed) lack of development in protected areas may lead species that inhabit to experience fewer edge effects, these likely will occur in other parts of the species range. Machine learning approaches such as random forest analysis have been able to identify key features in other investigations into the factors that influence intraspecific genetic diversity (e.g., Pelletier

& Carstens 2018), but the low number of species (~50) that we identified here preclude us from conducting these analyses.

The demographic history of the species is a final factor that might influence our results because it is also possible that protected areas served as climate refugia in some species during the Pleistocene glaciations. If this were the case, individuals sampled from protected areas might contain more genetic diversity due to these regions acting as a reservoir for genetic diversity (e.g., Hewitt 2004). To explore this question, we reviewed the investigations that generated the data used in this study, which were largely phylogeographic investigations and as such potentially concerned with the identification of climate refugia, in order to assess the degree to which protected areas might overlap with putative climate refugia. Of the 31 papers reviewed, none proposed that the regions that were classified as protected areas in our investigations served as climate refuge during the Pleistocene glaciations. However, for some species most of the species range is currently in protected area (Supplementary table 3).

Repurposed data for conservation biology

This study is one example of how repurposed data can be used to explore questions that are relevant to conservation biology. Similar analyses have been used to explore related questions, for example, these methods have been used to map genetic diversity on a global scale (Miraldo et al., 2016), identify the factors that predict whether genetic variation is structured within species (Pelletier & Carstens, 2018), and explore the factors that influence the response of bat species to the end-Pleistocene climate change (Carstens, Morales, Field, & Pelletier, 2018). The repurposing of data available on GBIF

and GenBank has tangible benefits for future studies concerning the geographic distribution of genetic variation, now that millions of sequence reads can be easily collected from any species. Due to large open access databases, novel questions in the fields of conservation and evolutionary biology can be addressed with the methods used for this study. We encourage all researches that have collected georeferenced and/or genetic data to upload their data to an open-database repository where other scientists can use this data to answer large scale questions. While our work suggest that protected areas may influence intraspecific genetic diversity, it is hampered by the low percentage of studies that provide the necessary metadata to connect genetic sequences with the geographic localities from which they originate. However, these tools developed in this study for the analysis of genetic diversity of multiple species will allow for time-efficient and cost-effective methods for determining the conservation value of land areas by maximizing the utility of existing data.

Tables

Species	Gene	Π_{in}	Π_{out}	Pvalue (π)	Θ_{in}	Θ_{out}	Hap. Div _{in}	Hap. Div _{out}
<u>Amphibians</u>								
<i>Pseudacris regilla</i>	cytb	<i>0.019</i>	0.023	0.681	13.44	12	5	5
<i>Plethodon ouachitae</i>	cytb	* <i>0.026</i>	0.014	0	12.93	11.52	27	18
<u>Mammals</u>								
<i>Antrozous pallidus</i>	cytb	<i>0.026</i>	0.030	0.282	14.67	12	3	3
<i>Corynorhinus townsendii</i>	cytb	<i>0.003</i>	*0.031	1	3.33	32.67	3	3
<i>Sturnira parvidens</i>	cytb	<i>0.007</i>	0.011	0.771	13	13	2	2
<i>Ursus americanus</i>	cytb	0.034	<i>0.016</i>	0.768	39	1	2	2
<i>Glaucomys sabrinus</i>	cytb	*0.019	<i>0.001</i>	1	23.18	1.47	3	3
<i>Lepus othus</i>	cytb	0.000	<i>0.001</i>	0.496	-	-	1	2
<i>Dicrostonyx groenlandicus</i>	cytb	*0.010	<i>0.005</i>	1	8.48	5.01	7	8
<i>Lemmus trimucronatus</i>	cytb	<i>0.029</i>	0.043	0.973	-	-	5	5
<i>Marmota broweri</i>	cytb	<i>0.003</i>	0.000	0.054	0.96	0.48	1	2
<i>Marmota caligata</i>	cytb	* <i>0.013</i>	0.009	0	9.74	7.91	20	11
<i>Martes americana</i>	cytb	0.011	<i>0.012</i>	0.177	4.12	3.04	9	8
<i>Chaetodipus hispidus</i>	cytb	0.018	<i>0.014</i>	0.865	8.76	7.45	6	6
<i>Perognathus flavus</i>	COIII	*0.005	<i>0.002</i>	1	4.38	2.19	6	5
<i>Peromyscus keeni</i>	cytb	<i>0.006</i>	*0.008	1	6.97	10.83	57	57
<i>Peromyscus maniculatus</i>	cytb	<i>0.023</i>	0.019	0.123	24.70	24.46	25	28

Table 1 Species, gene, p-value (Π), and measures of genetic diversity both inside and outside of protected areas. The average of the Π values from the null distribution are italicized. The other Π values are the test statistic. The “*” denotes significance.

Continued

Species	Gene	Π_{in}	Π_{out}	Pvalue (π)	Θ_{in}	Θ_{out}	Hap. Div _{in}	Hap. Div _{out}
<u>Mammals</u>								
<i>Reithrodontomys megalotis</i>	cytb	0.003	0.005	0.938	2.73	4.36	4	4
<i>Ochotona collaris</i>	cytb	0.007	0.006	0.112	11.98	7.67	23	15
<i>Neotoma fuscipes</i>	COII	0.005	0.002	0.343	-	-	2	2
<i>Neotoma cinerea</i>	cytb	*0.038	0.001	0.011	47.33	0.67	3	2
<i>Sorex cinereus</i>	cytb	0.010	*0.015	0.998	13.40	21.36	3	3
<i>Sorex hoyi</i>	cytb	0.001	*0.004	1	2	4	3	3
<i>Sorex monticolus</i>	cytb	*0.029	0.023	0.033	6.51	13.63	-	-
<i>Sorex tundrensis</i>	cytb	0.004	0.005	0.945	5.79	4.24	6	6
<i>Sorex ugunak</i>	cytb	0.004	*0.055	0.994	2	57	-	-
<i>Sorex yukonicus</i>	cytb	0.002	0.001	0.295	1	1	2	2
<i>Tamiasciurus hudsonicus</i>	cytb	*0.017	0.006	1	12.88	11.04	9	8
<i>Uroditellus parryi</i>	cytb	*0.025	0.021	0.002	22.34	18.21	17	13
<i>Microtus longicaudus</i>	cytb	*0.021	0.013	1	9.69	9.88	81	76
<i>Microtus pennsylvanicus</i>	cytb	0.004	0.006	0.112	2.88	4.32	5	5
<i>Microtus miurus</i>	cytb	*0.013	0.004	0	16.61	3.89	10	4
<i>Microtus oeconomus</i>	cytb	0.007	0.007	0.214	9.36	11.76	18	28
<i>Mustela frenata</i>	cytb	0.005	*0.008	1	7.2	10.08	5	5
<i>Gulo gulo</i>	cytb	0.001	0.001	0.522	2.19	0.88	4	3
<u>Squamates</u>								
<i>Phrynosoma mcallii</i>	ND4	*0.010	0.004	1	7.86	5.97	9	6
<i>Phrynosoma platyrhinos</i>	ND4	0.028	0.019	0.084	22.27	14.49	8	6
<i>Sceloporus graciosus</i>	cytb	0.048	0.089	0.765	55	100	2	2
<i>Sceloporus zosteromus</i>	ND4	0.006	*0.089	1	0	61	-	-

Table 1 Continued

Species	Gene	Π_{in}	Π_{out}	Pvalue (π)	Θ_{in}	Θ_{out}	Hap. Div _{in}	Hap. Div _{out}
<i>Uta stansburiana</i>	cytb	0.014	*0.039	1	4.8	8.64	2	2
<i>Coluber constrictor</i>	cytb	0.014	*0.056	1	37.37	36.47	12	12
<i>Contia longicaudae</i>	cytb	0.001	0.002	0.544	1	4	2	1
<i>Contia tenuis</i>	cytb	0.013	0.007	0.131	7.64	8.73	4	4
<i>Diadophis punctatus</i>	cytb	0.045	0.037	0.773	49.93	48.28	19	17

Table 1 Continued

Figures

Figure 1 Map of Gulo Gulo distributed throughout NA. The shaded areas indicate the boundaries of protected areas and the yellow dots indicate the individual(s) inside of protected areas while the blue dots indicate those which are outside.

a)

b)

Figure 2 Histograms of the null distribution of genetic diversity in three species with contrasting results. The red line indicates the averaged amount of genetic diversity for the population opposite of the distribution. a) Genetic diversity in *Coluber constrictor* is significantly greater outside of protected areas, b) There is no significant difference in genetic diversity for *Gulo gulo*, c) Genetic diversity in *Marmota caligata* is significantly greater inside of protected areas.

Continued

Figure 2 Continued

Figure 3 Regressions of body mass ($r^2 = 0.331$), home range size ($r^2 = 0.0000541$), and maximum latitude ($r^2 = 0.00321$) on the difference in nucleotide diversity (inside of protected areas – outside of protected areas).

Continued

Figure 3 Continued

Figure 4 The p-values for all 44 vertebrate species.

Bibliography

- Adams, R. I., & Hadly, E. A. (2013). Genetic diversity within vertebrate species is greater at lower latitudes. *Evolutionary Ecology*, 27(1), 133–143. <https://doi.org/10.1007/s10682-012-9587-x>
- Avise, J. C. (1987). Intraspecific Phylogeography: *Annual Review of Ecology, Evolution, and Systematics*. <https://doi.org/10.1146/annurev.es.18.110187.002421>
- Barabás, G., & D’Andrea, R. (2016). The effect of intraspecific variation and heritability on community pattern and robustness. *Ecology Letters*, 19(8), 977–986. <https://doi.org/10.1111/ele.12636>
- Cahill, A. E., De Jode, A., Dubois, S., Bouzaza, Z., Aurelle, D., Boissin, E., ... Chenuil, A. (2017). A multispecies approach reveals hot spots and cold spots of diversity and connectivity in invertebrate species with contrasting dispersal modes. *Molecular Ecology*, 26(23), 6563–6577. <https://doi.org/10.1111/mec.14389>
- Carnaval, A. C., Hickerson, M. J., Haddad, C. F. B., Rodrigues, M. T., & Moritz, C. (2009). Stability Predicts Genetic Diversity in the Brazilian Atlantic Forest Hotspot. *Science*, 323(February), 785–789.
- Carstens, B. C., Morales, A. E., Field, K., & Pelletier, T. A. (2018). A global analysis of bats using automated comparative phylogeography uncovers a surprising impact of Pleistocene glaciation. *Journal of Biogeography*, 45(8), 1795–1805. <https://doi.org/10.1111/jbi.13382>
- Carvalho, C. da S., Ballesteros-Mejia, L., Ribeiro, M. C., Côrtes, M. C., Santos, A. S., & Collevatti, R. G. (2017). Climatic stability and contemporary human impacts affect the genetic diversity and conservation status of a tropical palm in the Atlantic Forest of Brazil. *Conservation Genetics*, 18(2), 467–478. <https://doi.org/10.1007/s10592-016-0921-7>
- Chape, S., Harrison, J., Spalding, M., & Lysenko, I. (2010). Measuring the extent and effectiveness of protected areas as an indicator for meeting global biodiversity targets, (February 2005), 443–455. <https://doi.org/10.1098/rstb.2004.1592>
- Edgar, R. C. (2010). Quality measures for protein alignment benchmarks. *Nucleic Acids Research*, 38(7), 2145–2153. <https://doi.org/10.1093/nar/gkp1196>

- Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, 17(7), 422–433. <https://doi.org/10.1038/nrg.2016.58>
- Garrick, R. C., Isabel, A. S., Pelletier, T. A., Perez, F., Rice, E., Jordan, D., ... Carstens, B. . (2015). The evolution of phylogeographic data sets. *Molecular Ecology*, 24(6), 1164–1171.
- Gavin, D. G., Stigall, A. L., Tsai, Y.-H. E., Bartlein, P. J., Blois, J. L., Dobrowski, S. Z., ... Rodríguez-Sánchez, F. (2014). Climate refugia: joint inference from fossil records, species distribution models and phylogeography. *New Phytologist*, 204(1), 37–54. <https://doi.org/10.1111/nph.12929>
- Guzmán, B., Rodríguez López, C. M., Forrest, A., Cano, E., & Vargas, P. (2015). Protected areas of Spain preserve the neutral genetic diversity of *Quercus ilex* L. irrespective of glacial refugia. *Tree Genetics and Genomes*, 11(6). <https://doi.org/10.1007/s11295-015-0950-2>
- Hague, M. T. J., & Routman, E. J. (2016). Does population size affect genetic diversity? A test with sympatric lizard species. *Heredity*, 116(1), 92–98. <https://doi.org/10.1038/hdy.2015.76>
- Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359(1442), 183–195. <https://doi.org/10.1098/rstb.2003.1388>
- Hoelzel, A. R., Natoli, A., Dahlheim, M. E., Olavarria, C., Baird, R. W., & Black, N. A. (2002). Low worldwide genetic diversity in the killer whale (*Orcinus orca*): Implications for demographic history. *Proceedings of the Royal Society B: Biological Sciences*, 269(1499), 1467–1473. <https://doi.org/10.1098/rspb.2002.2033>
- Hugall, A., Moritz, C., Moussalli, A., & Stanisic, J. (2002). Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosiphia bellendenkerensis* (Brazier 1875). *Proceedings of the National Academy of Sciences*, 99(9), 6112–6117. <https://doi.org/10.1073/pnas.092538699>
- Jones, K. E., Habib, M., Bielby, J., Boakes, E. H., Gittleman, J. L., Carbone, C., ... Rigby, E. A. (2009). PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, 90(9), 2648–2648. <https://doi.org/10.1890/08-1494.1>
- Lacy, R. C. (1987). Loss of genetic diversity from managed populations: Interacting effects of drift, mutation, immigration, selection, and population subdivision. *Conservation Biology*, 1(2), 143–158. Retrieved from <http://www.jstor.org/stable/2385830>
- Laikre, L., Allendorf, F. W., Aroner, L. C., Baker, C. S., Gregovich, D. P., Hansen, M.

- M., ... Waples, R. S. (2010). Neglect of genetic diversity in implementation of the convention on biological diversity: Conservation in practice and policy. *Conservation Biology*, 24(1), 86–88. <https://doi.org/10.1111/j.1523-1739.2009.01425.x>
- Miller, K. J., & Ayre, D. J. (2008). Protection of genetic diversity and maintenance of connectivity among reef corals within marine protected areas. *Conservation Biology*, 22(5), 1245–1254. <https://doi.org/10.1111/j.1523-1739.2008.00985.x>
- Miraldo, A., Li, S., Borregaard, M. K., Flórez-rodríguez, A., Gopalakrishnan, S., Rizvanovic, M., ... Nogués-bravo, D. (2016). Genetic Diversity. *Science*, 353(6307).
- Muposhi, V. K., Chademana, T. C., Gandiwa, E., & Muboko, N. (2016). Edge effects: impact of anthropogenic activities on vegetation structure and diversity in western Umfurudzi Park, Zimbabwe. *African Journal of Ecology*, 54(4), 450–459. <https://doi.org/10.1111/aje.12300>
- Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89, 583–590. <https://doi.org/10.3390/ijms15010277>
- Nicola M. Anthony, Mireille Johnson-Bawe, Kathryn Jeffery, Stephen L. Clifford, Kate A. Abernethy, Caroline E. Tutin, Sally A. Lahm, Lee J. T. White, John F. Utley, E. Jean Wickings, and M. W. B. (2007). The role of Pleistocene refugia and rivers in shaping gorilla genetic diversity in central Africa. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.0704816105>
- Ottaviani, D., Cairns, S. C., Oliverio, M., & Boitani, L. (2006). Body mass as a predictive variable of home-range size among Italian mammals and birds. *Journal of Zoology*, 269(3), 317–330. <https://doi.org/10.1111/j.1469-7998.2006.00060.x>
- Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, 26(3), 419–420. <https://doi.org/10.1093/bioinformatics/btp696>
- Pelletier, T. A., & Carstens, B. C. (2018). Geographical range size and latitude predict population genetic structure in a global survey. *Biology Letters*, 14(1), 20170566. <https://doi.org/10.1098/rsbl.2017.0566>
- Petit, R. J., Aguinagalde, I., Beaulieu, J.-L. De, Bittkau, C., Brewer, S., Cheddadi, R., ... Martín, J. P. (2003). Glacial Refugia : Hotspots But Not. *Science*, 300(6), 1563–1565. <https://doi.org/10.1038/nature05144>
- Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: An efficient swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, 31(7), 1929–1936.

<https://doi.org/10.1093/molbev/msu136>

- Pruett, C. L., & Winker, K. (2008). The Effects of Sample Size on Population Genetic Diversity Estimates in Song Sparrows *Melospiza melodia*. *Journal of Avian Biology*, 39(2), 252–256. <https://doi.org/10.1111/j.2008.0908-8857.04094.x>
- Reed, D. H., Briscoe, D. A., Frankham, R., Ballou, J., O’Grady, J., Kalinowski, S., ... Lowe, E. (2002). Inbreeding and extinction: The effect of environmental stress and lineage. *Conservation Genetics*, 3(3), 301–307. <https://doi.org/10.1023/A:1019948130263>
- Roman, J., & Darling, J. A. (2007). Paradox lost: genetic diversity and the success of aquatic invasions. *Trends in Ecology and Evolution*, 22(9), 454–464. <https://doi.org/10.1016/j.tree.2007.07.002>
- Sgrò, C. M., Lowe, A. J., & Hoffmann, A. A. (2011). Building evolutionary resilience for conserving biodiversity under climate change. *Evolutionary Applications*, 4(2), 326–337. <https://doi.org/10.1111/j.1752-4571.2010.00157.x>
- Valiente, A. G., Juanes, F., & Garcia-Vazquez, E. (2005). Reproductive strategies explain genetic diversity in Atlantic salmon, *Salmo salar*. *Environmental Biology of Fishes*, 74(3–4), 323–334. <https://doi.org/10.1007/s10641-005-1331-8>
- Vandergast, A. G., Bohonak, A. J., Hathaway, S. A., Boys, J., & Fisher, R. N. (2008). Are hotspots of evolutionary potential adequately protected in southern California? *Biological Conservation*, 141(6), 1648–1664. <https://doi.org/10.1016/j.biocon.2008.04.009>
- Watterson, G. A. (1975). 7to recombination between sites. Of course, this assumption is meaningful only if the gametes making up our population go through a diploid phase, as they do in Burrows and Cockerham’s model to be described below. For 256. *Theoretical Population Biology*, 276(7), 256–276. <https://doi.org/10.1039/b316709g>
- Welch, B. L. (1938). The Significance of the Difference Between Two Means when the Population Variances are Unequal. *Biometrika Trust*, 29(3), 350–362.
- Wright, S. (1943). Isolation by distance. *Genetics*, 28(114), 114–138. <https://doi.org/10.5194/isprs-Archives-XLII-5-W1-419-2017>

Appendix A. Additional Tables

Genus	Species	p.-value
<i>Batrachoseps</i>	<i>attenuatus</i>	0.722
<i>Microtus</i>	<i>abbreviatus</i>	0.028
<i>Plethodon</i>	<i>serratus</i>	0.034
<i>Plethodon</i>	<i>fourchensis</i>	0.138
<i>Plethodon</i>	<i>caddoensis</i>	0.244
<i>Rattus</i>	<i>rattus</i>	0.775
<i>Sorex</i>	<i>tundrensis</i>	0.421
<i>Sorex</i>	<i>jacksoni</i>	1
<i>Sorex</i>	<i>preblei</i>	1
<i>Tamias</i>	<i>rufus</i>	1

Table 2 Test study species (not in study) and pvalues.

Genus	Species	p-value
<i>Coluber</i>	<i>constrictor</i>	0.904
<i>Dicrostonyx</i>	<i>groenlandicus</i>	0.986
<i>Lemmus</i>	<i>trimucronatus</i>	0.581
<i>Marmota</i>	<i>broweri</i>	0.826
<i>Neotoma</i>	<i>cinerea</i>	0.311
<i>Ochotona</i>	<i>collaris</i>	0.774
<i>Peromyscus</i>	<i>keeni</i>	0.937
<i>Plethodon</i>	<i>ouachitae</i>	0.694
<i>Sorex</i>	<i>cinereus</i>	0.817
<i>Sorex</i>	<i>hoyi</i>	0.707

Table 3 Random test study species and pvalues.

Citation	Refugia	Overlap w/ Protected Areas
Andersen & Light 2012	no	no
Bidlack & Cook 2001	no	no *
Brunhoff et al, 2003	yes	no
Burbrink et al 2008	yes	no
Conroy & Cook 2000	yes	no
Cook et al 2001	yes	no
Cook et al 2010	yes	no
Demboski & Cook 2001	yes	no
Demboski & Cook 2003	yes	no
Dragoo et al 2006	yes	no
Fedorov & Stenseth 2002	yes	no
Fedorov et al 2003	yes	no
Feldman & Hoyer 2010	no	no
Fontanella et al 2008	yes	no
Gunderson et al 2009	no	no
Harding & Dragoo 2012	yes	no
Hope et al 2010	yes	no
Hope et al 2012	yes	no
Hope et al 2013	yes	no *
Hope et al 2014	no	no
Kerhoulas et al 2015	yes	no
Mahoney et al 2003	no	no
Melo-Ferreira et al 2011	yes	no
Mulcahy et al 2006	no	no
Recuero et al, 2006	yes	no*
Shepard & Burbrink 2008	yes	no*
Stone & Cook 2000	yes	no
Tomasik & Cook 2005	yes	no
Velazco & Patterson 2013	no	no
Weksler et al 2010	yes	no
Weyandt & Van Den Bussche 2007	yes	no

Table 4 Original papers that generated the data used in this study were examined to see if protected areas included in our investigations overlap with a climate refuge during the Pleistocene glaciations. * mentioned that current range is mainly in protected area, but nothing about overlap.