Synthesizing vertebrate population richness and genetic diversity across the American continents

Elizabeth Rachel Lawrence

A thesis
In
The Department
of
Biology

Presented in Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy (Biology) at
Concordia University
Montreal, Quebec,Canada

September 2020

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Elizabeth Rachel Lawrence

Entitled: Synthesizing vertebrate population richness and genetic diversity across the American continents

and submitted in partial fulfillment of the requirements for the degree of

Doctor Of Philosophy (Biology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. Kristen Dunfield

_____ External Examiner
Dr. Janna Willoughby

_____ External to Program
Dr. Pascale Biron

_____ Examiner
Dr. James Grant

_____ Examiner
Dr. Jean-Philippe Lessard

_____ Thesis Supervisor
Dr. Dylan Fraser

Approved by _____
Dr. Robert Weladji, Graduate Program Director

September 11, 2020

_____
Dr. André Roy, Dean
Faculty of Arts and Science

Abstract

Synthesizing vertebrate population richness and genetic diversity across the American continents

Elizabeth Rachel Lawrence, Ph.D.
Concordia University, 2020

Intraspecific diversity is an important facet of biodiversity, both for the understanding of broad-scale biodiversity distribution and for the prioritization of conservation hotspots below the species level. It is the level of biodiversity that responds first to environmental change, yet few studies have assessed its broad-scale distribution. By constructing and analyzing an extensive population-genetics database, my thesis aims to demonstrate both the links and differences between species richness, population richness, and population-specific genetic diversity (PGD). Chapter 1 details the database and provides an exploration of population genetic data across five vertebrate taxonomic groups. The database collated geo-referenced information from 895 vertebrate species, 1308 studies, and 9090 genetically distinct populations. I found that anadromous species tended to be both the most population rich and genetically diverse, while mammals had lower levels of genetic diversity. In Chapter 2, I synthesized the conceptual foundation for broad-scale expectations of genetic and population diversity patterns by drawing from theories in the species diversity literature. I also tested the relationship between range size and population richness or PGD, finding a positive and a non-significant relationship for population richness and PGD, respectively. For Chapter 3, I assessed the latitudinal gradient in vertebrate PGD and assessed how environmental variables and variation among genera may mediate patterns in PGD. I found minimal evidence for a latitudinal gradient in PGD, a weak influence of environmental variables, and strong evidence for genera-specific patterns. In Chapter 4, I evaluated the influence of anthropogenic impacts (namely human population density and heterogeneity in land use intensity) on metrics of PGD across broad-scales. I found inconsistent support for the expected negative impacts, instead finding that human impact varies both between and within taxonomic groups. Collectively, my thesis demonstrates the difficulties in applying species theories to intraspecific diversity and that species-centric views overlook important variation below the species level. Taxonomic-dependent responses are common, and

there is no "broad brush" for biodiversity – considering differences among taxa, even down to the genus-level, can be vital for biodiversity conservation. Intraspecific diversity does not have the same distribution as species diversity, and more extensive sampling would be needed to investigate these patterns more clearly.

Acknowledgements

"At top and bottom of big rock, there are not many kinds of animals, but the animals are more changey" – Jonathan D. Lawrence

I first would like to dedicate this thesis to my twin brother, Jonathan D. Lawrence, who summarized my work with the quote above. You have always been there for me, whether through emotional support, editorial means, and even helping me extract data from papers. I could not have gotten through this degree or indeed this life without you. I hope you know how much you mean to me, and you always will.

I have many other people to thank for supporting me throughout this thesis. To my doctoral supervisor, Dylan Fraser, you have been an incredible source of inspiration and motivation. I am deeply thankful for your continued support during my degree and feel extremely lucky to have had you as a supervisor. Your encouragement and emphasis for mental and physical health is so refreshing in a field where burn out can be so high. You have been a continual pool of knowledge and I appreciate all the time you have spent providing me with valuable feedback.

Fellow lab member Dr. Jean-Michel Matte I want to thank you for your statistical support, feedback on written drafts, and general discussion as it has been instrumental to my understanding of the work throughout this thesis. As well, to Dr. Matthew Yates who has been there for many discussions regarding statistical approaches. To Dr. Ella Bowles, your friendship and mentorship has been and is incredibly valued.

I am grateful to all the co-authors and volunteers who assisted in the construction of the database this thesis is based on: Kia Marin, Javiera Benavente Paredes, Nia Krasteva, Thaïs Bernos, Zachery Wells, Andrew Habrich, Ramela Arax Koumrouyan, Gabrielle Nessel, Katherine Levasseur-Bhatia, Astrid de Jaham, Danielle Mac Rae, Jonathan Lawrence, and Charlotte Clegg. Additionally, the many colleagues who have provided feedback on my drafts at some point or another. I would also like to thank all of my committee members who have supported this work and provided insightful criticism that has improved the development of the project.

Thank you to my committee members, Dr. James Grant and Dr. Jean-Philippe Lessard, as well as my external examiners, Dr. Pascale Biron and Dr. Janna Willoughby. I also thank all the

authors from various studies who have provided me with data that has supported this thesis, inclding Erle Ellis (anthropogenic data), Kirsten J Monsen (population data), Janna Willoughby (bird data), the IUCN, and BirdLife International (bird data).

Finally, I would like to thank my funding sources: Concordia University, NSERC, and the Quebec Centre for Biodiversity Science (QCBS) for their instrumental financial support. In particular the QCBS has not only supported the construction of the database itself in the form of a seed grant that allowed us to hire help, but, along with Concordia, also supported my growth as a person and a scientist by supplying financial aid to attend international conferences in Colombia and Malaysia.

Contribution of Authors

As primary author I contributed to the conception, planning, data collection/extraction, data analyses, and writing for all four thesis chapters. Similarly, Dr. D. Fraser contributed to the conception, planning, and writing/editing of all chapters. For chapter 1, co-authorship contribution is as follows:

E.R.L. initiated the database construction, collected data on North American fish, mammals, amphibians, reptiles, and South American mammals; she also combined all files from other contributors, checked data for consistency, ran statistical analyses, built the reference file, and wrote the paper.

J.N.B. collected population data on North American fish, and South American fish, mammals, amphibians, and reptiles, assisted with collating reference information, and provided feedback on manuscript revisions.

J.-M.M. completed statistical modeling, focusing on analyses for beta distributions and coefficients of variation; he also provided feedback on manuscript revisions.

K.M. collected population data on North American mammals, fish, reptiles, and amphibians, and provided feedback on manuscript revisions.

Z.W. collected population data on North American fish and provided feedback on manuscript revisions.

T.B. and N.K. collected population data on bird species within the American continents and provided feedback on manuscript revisions.

A.H. collected population data on North American mammal species and provided feedback on manuscript revisions.

G.N. collected F_{ST} values for bird species and assisted with collating reference information.

R.A.K. collected F_{ST} values for bird species and provided feedback on manuscript revisions.

D.J.F. conceived the general idea of the study, collected population data on North American fish, and provided feedback on manuscript revisions.

Chapters 1 and 2 have been published in *Scientific Data* and *Global Ecology and Biogeography*, respectively, as:

Lawrence, E.R., Javiera N. Benavente, Jean-Michel Matte, Kia Marin, Zachery Wells, Thaïs A. Bernos, Nia Krasteva, Andrew Habrich, Gabrielle Nessel, Ramela Arax Avedis Koumrouyan, & Dylan J. Fraser. (2019). Geo-referenced population-specific microsatellite data across the American continents: the MacroPopGen Database. *Scientific Data* 6(1), 14.

Lawrence, E.R. & Fraser, D.J. (2020). Latitudinal biodiversity gradients at three levels: linking species richness, population richness, and genetic diversity. *Global Ecology and Biogeography 29*(5), 770-788.

Chapter 3 has been submitted to *Nature Communications* as:

Lawrence, E.R. & Fraser, D.J. (*in review*). Weak latitudinal and environmental influences on vertebrate population genetic diversity across the Americas. *Nature Communications.*

Chapter 4 is in preparation to be submitted to *Biological Conservation* as:

Lawrence, E.R. & Fraser. D.J. Variability in anthropogenic impacts on nuclear vertebrate population genetic diversity across the Americas. *Biological Conservation*.

Table of Contents

List of Figures

summed across all the species and populations in an area (TotGenDiv, c), this value masks the nuances of genetic diversity of the species present (GenPerSpp). When each individual species' total genetic diversity is considered, nuances of the genetic diversity in an area are more apparent.

Figure 2.2. A) Number of vertebrate species sampled in each 500 x 500km^2 grid cell. B) Number of genetically distinct populations across vertebrate species in each grid cell. Data obtained from *MacroPopGen* database (Lawrence et al., 2019) and projected with the World Behrmann projection.

Figure 2.3. Results for testing the Geographic Distribution Hypothesis (A, B), the Overlapping Range Hypothesis (C, D), and the Range Restricted Gene Hypothesis (E, F). A) Log of species range size and the number of genetically distinct populations within a species (PopPerSpp) for all taxonomic groups. B) Linear prediction estimates from a GLMM for the relationship between range size and number of populations for each taxonomic group. Error bars represent upper and lower confidence intervals. C-D) The number of unique species within grid cells (n=250) of an area of 500km^2 (x axis) and C) the total number of populations (Tot .Pop. Richness, R^2=0.75, $p<0.001$) or D) the average number of populations within each species (PopPerSpp; $R^2$0.22, $p<0.001$) for each grid cell. Solid line represents linear regression between the two variables. E-F) Linear prediction estimates from a GLMM of the relationship between E) log of range size and genetic diversity, measured as mean number of alleles (MNA) or F) observed heterozygosity (Ho) and MNA. Error bars represent upper and lower confidence intervals. Results for the other genetic diversity metric, observed heterozygosity (Ho), not shown as relationships were very similar as MNA.

Figure 3.1. Summary of the three predictions for a latitudinal gradient in population genetic diversity, indicating which variables are likely to contribute to expectations for a) positive, b) negative, or c) no latitudinal gradient. All y axes represent population genetic diversity, indicating the generally expected trend for genetic diversity with each variable.

Figure 3.2. The predicted effect of latitude and environmental variables on vertebrate population genetic diversity (mean number of alleles) for vertebrate species across the Americas. Predictors from the selected generalized additive mixed model were fitted by smoothers (s; tensor products (te) for interaction) and include a) Degrees Latitude, b) Elevation (m), c) MAT = mean annual temperature (°C), d) AP = annual precipitation (mm/year), e) the interaction between Elevation

nearest Urban biome (Urban, km), (b-g) the interaction between human population density (HPD, persons km^{-2}) and Originating Biome (Croplands, Rangelands, Semi-Natural, Urban, Village, Wild), (h) Proportion of Biome (POB), and (i) distance to nearest Natural biome (Natural, km). Confidence intervals represent standard error.

Figure 4.5. The predicted effect of anthropogenic variables selected through model selection for observed heterozygosity. Variables were fitted by smoothers (s) and include: (a) distance to nearest Urban biome (Urban, km), (b) Proportion of Biome (POB), (c) distance to nearest Natural biome (Natural, km), and (d) Originating Biome (CL = Croplands; FW = Freshwater; OC = Ocean; RL = Rangelands; SN = Semi-Natural; UR = Urban; VI =Villages; WI = Wild). Confidence intervals represent the standard error.

List of Tables

Table 4.2. AIC comparison and model fit of select Generalized Additive Mixed Models GAMMs during model selection. Ho = observed heterozygosity; MNA = mean number of alleles; HPD = human population density (persons km^{-2}), anthrome, Natural = distance to Natural biomes (km), Urban = distance to Urban biomes (km), POB = weighted metric of proportion of biomes within 100km of a population; Anthrome = the anthrome that a population was located in. Variables are fit with a smoother (s) and denoted as a random effect by bs="re".

Table 4.3. Summary of final GAMMs selected through model selection for either observed heterozygosity (Ho) or mean number of alleles (MNA). HPD = human population density (persons km^{-2}), anthrome, Natural = distance to Natural biomes (km), Urban = distance to Urban biomes (km), POB = weighted metric of proportion of biomes within 100km of a population; Anthrome = the anthrome that a population was located in (i.e. Originating Biome). Variables are fit with a smoother s() and denoted as a random effect by bs="re". Bold values indicate statistical significance.

General Introduction

Biodiversity is the variety of life and while this variation can be assessed at multiple levels, researchers have historically focused on one level at a time (most prominently species richness) due to technological and theoretical constraints. As these constraints have become less limiting due to technological advances, there has been a shift towards assessing intraspecific diversity and its role and distribution in large-scale biodiversity (Millette et al., 2019; Miraldo et al., 2016; Willoughby et al., 2015). Intraspecific diversity can determine the maintenance and/or establishment of species within ecosystems, and how a species will respond to future biotic/abiotic change (Barrett & Schluter, 2008; Bernatchez, 2016; Ghalambor et al., 2007; Jump et al., 2009; Willoughby et al., 2018). Overlapping and examining broad-scale patterns of inter- and intraspecific diversity together can provide insights into biodiversity distribution (e.g. latitudinal gradient in species richness). It can even have conservation management implications by identifying hotspots (Marchese, 2015; Paz-Vinas et al., 2018). This thesis aimed to integrate species-level biodiversity research with two facets of intraspecific diversity – population richness and genetic diversity – and then investigated how these levels may or may not be linked together at broad-scales.

While there are different aspects of intraspecific diversity, such as phylogenetic diversity, functional diversity, population richness, and genetic diversity, this thesis focused on the latter two. Population richness is a key component between species richness and genetic diversity. As genetic divergence increases, population differentiation leads to eventual speciation (Schluter, 2016; Schluter & Pennell, 2017; Taylor, 1999; Wiens & Donoghue, 2004). Thus, I defined population richness as the number of genetically distinct populations in a region. This metric can provide insight into the state of an ecosystem or an individual species. For example, comparing the species and population richness within an area could indicate the age and speciation potential of that community, where high species richness but low population richness could be indicative of an older community (i.e. millions or tens of millions years old) with lower rates of speciation since all populations have likely diverged (Kennedy et al., 2018; Schluter, 2016; Schluter & Pennell, 2017). Conversely, an individual species with many genetically distinct populations could indicate higher potential for local adaptation and persistence throughout the species range as environmental change occurs. Populations will experience variable environments throughout

the range, thus some may be better adapted for different conditions. The differences in local adaptation may enable future persistence compared to an endemic species with one population, which may not have the variability required to adapt (Ghalambor et al., 2007). As such, population richness is one of two important facets of intraspecific diversity I focused on here and is intrinsically linked to genetic diversity.

Genetic diversity can be either neutral or adaptive. Neutral genetic diversity represents standing genetic diversity that selection is currently not acting upon, whereas adaptive genetic diversity represents the parts of genomes associated with active selection, adapting to future change, and can have definitive effects on fitness (Gebremedhin et al., 2009; Jarne & Lagoda, 1996; Kirk & Freeland, 2011; Mittell et al., 2015; Selkoe & Toonen, 2006). Although adaptive genetic diversity may seem to provide more direct information on a population's fitness, neutral genetic diversity can provide important insights into a population's potential for cryptic genetic diversity that could allow future adaptation to novel environments (Brennan et al., 2019; Kirk & Freeland, 2011; Paaby & Rockman, 2014; Selkoe & Toonen, 2006). Neutral genetic diversity is also often used for delineating populations due to higher mutation rates resulting from lower selection pressure (Jarne & Lagoda, 1996). This makes neutral genetic diversity a useful aspect of diversity to assess at broad scales, and to compare across many taxonomic groups. Additionally, largescale syntheses of adaptive genetic diversity are at present restrained by a lack of data, whereas there are two decades of data on neutral genetic diversity (Jarne & Lagoda, 1996; Selkoe & Toonen, 2006). As such, I concentrated on neutral genetic diversity – explicitly population-specific neutral genetic diversity (PGD) – for this thesis.

Neutral PGD can be assessed with different molecular markers (e.g. mitochondrial DNA (mtDNA), allozymes, Random Amplified Polymorphic DNA, Single Nucleotide Polymorphisms (SNPs), etc.) but microsatellites were chosen here over other markers. Microsatellites have elevated mutation rates, a polymorphic nature, an ability to represent nuclear genomic diversity, and widespread usage in recent decades (Jarne & Lagoda, 1996; Schlötterer, 2004; Selkoe & Toonen, 2006; Väli et al., 2008). Microsatellites have high polymorphism and elevated mutation rates which allows them to resolve fine-scale population structure, particularly for closely related populations (Angers & Bernatchez, 1998; Jarne & Lagoda, 1996). As microsatellites are obtained from the nuclear genome, they reflect genome-wide nuclear diversity – an integral aspect for future adaptation – providing a distinct advantage over mtDNA markers (e.g. commonly assessed

cytochrome c oxidase subunit I gene; Miraldo et al., 2016; Millette et al., 2019), which do not originate from the nuclear genome. Admittedly, other nuclear markers such as barcoding can assess phylogenetic signals across populations and species. However, microsatellites allow for the comparison of genetic characteristics between populations such as heterozygosity and allelic diversity, which has been noted to indicate levels of inbreeding or adaptive potential (Fraser et al., 2019; Hansson & Westerberg, 2002; Jump et al., 2009; Reed & Frankham, 2003). Finally, microsatellites presently provide the largest abundance of collectable data across taxa relative to more recent molecular developments associated with SNPs or barcoding.

Chapter one formed the basis and foundation of this thesis – the construction of the MacroPopGen database. This database was founded upon two decades of microsatellite-generated PGD data. I focused on five vertebrate classes across the American continents: amphibians, birds, fishes (anadromous and freshwater), mammals, and reptiles. I chose to exclude marine animals, plants, and invertebrates due to difficulties in delineating a genetically distinct population for marine animals and plants, and a paucity of data that suited my criteria for invertebrates. The collation of these data not only standardized decades of population-genetic data but also was used to i) determine any broad-scale patterns with respect to PGD across vertebrate classes, ii) detect differences in population differentiation among classes, and iii) identify any publication bias with respect to marker type, geographic region, or taxonomic groups.

Chapter two constructed a conceptual framework that applied species diversity theories to population richness and genetic diversity. Genetic diversity can have different definitions to different authors, so part of my goal for this chapter was to explicitly delineate population richness and genetic diversity, and the various broad-scale methods of assessing them. In this chapter I first applied theories describing the distribution in species diversity to generate predictions for the latitudinal distribution of population richness and genetic diversity (including but not limited to the time and area hypotheses, and Rapoport's rule; (Mittelbach et al., 2007; Pianka, 1966; Stevens, 1989; Wallace, 1878). After establishing a theoretical framework, I then generated and tested three hypotheses that related species range size to intraspecific diversity to test some of the predictions discussed in the chapter.

For Chapter three, I formally investigated the latitudinal gradient for vertebrate PGD, drawing on the theoretical foundation established in Chapter two. I tested models for two aspects

of neutral PGD (observed heterozygosity and mean number of alleles), including both latitudinal and environmental correlates. Specifically, I assessed the role of mean annual temperature, annual precipitation, total annual range, elevation, and net primary productivity. I also contrasted Last Glacial Maximum data for the first three environmental variables. Lastly, I accounted for variation among taxonomic levels, and tested which level accounted for the most variation.

Finally, in Chapter four I assessed how human impacts influence vertebrate PGD at broad scales. I tested the hypothesis that human impacts should negatively affect PGD. Metrics for human impact included human population density and land usage assessed by anthropogenic biomes (Ellis et al., 2010; Ellis & Ramankutty, 2008). As with Chapter three, a main goal of this chapter was to account for variation among taxonomic groups, but I also wanted to account for heterogeneity in the surrounding habitat of a population that may have buffering effects on a population's PGD (for example a "natural" area nearby an urban region may provide refuge for populations). By assessing multiple aspects of human impact and accounting for taxonomic and impact variability, I was able to more effectively evaluate the human influence on PGD.

My thesis focused on macro-population genetics and has important implications for the related fields, as well as for conservation. It showed the importance of accounting for variability in data that can better explain large-scale patterns. Finally, I demonstrated that there is rarely one comprehensive solution that can account for nuances within and among data.

Chapter 1: Geo-referenced population-specific microsatellite data across American continents, the MacroPopGen Database

Elizabeth R. Lawrence[1]; Javiera N. Benavente [1,2]; Jean-Michel Matte[1]; Kia Marin[1,3]; Zachery Wells[1,4]; Thaïs A. Bernos[1]; Nia Krasteva[1]; Andrew Habrich[1,5]; Gabrielle Nessel[1]; Ramela Arax Avedis Koumrouyan[1]; Dylan J. Fraser[1]

Published as:
Lawrence, E.R., Javiera N. Benavente, Jean-Michel Matte, Kia Marin, Zachery Wells, Thaïs A. Bernos, Nia Krasteva, Andrew Habrich, Gabrielle Nessel, Ramela Arax Avedis Koumrouyan, & Dylan J. Fraser. (2019). Geo-referenced population-specific microsatellite data across the American continents: the MacroPopGen Database. *Scientific Data* 6(1), 14.

Abstract

Population genetic data from nuclear DNA has yet to be synthesized to allow broad scale comparisons of intraspecific diversity versus species diversity. The MacroPopGen database collates and geo-references vertebrate population genetic data across the Americas from 1,308 nuclear microsatellite DNA studies, 895 species, and 9,090 genetically distinct populations where genetic differentiation (F_{ST}) was measured. Caribbean populations were particularly distinguished from North, Central, and South American populations, in having higher differentiation (F_{ST}=0.12 vs. 0.07-0.09) and lower mean numbers of alleles (MNA=4.11 vs. 4.84-5.54). While mammalian populations had lower MNA (4.86) than anadromous fish, reptiles, amphibians, freshwater fish, and birds (5.34-7.81), mean heterozygosity was largely similar across groups (0.57 – 0.63). Mean F_{ST} was consistently lowest in anadromous fishes (0.06) and birds (0.05) relative to all other groups (0.09-0.11). Significant differences in Family/Genera variance among continental regions or taxonomic groups were also observed. MacroPopGen can be used in many future applications including latitudinal analyses, spatial analyses (e.g. central-margin), taxonomic comparisons, regional assessments of anthropogenic impacts on biodiversity, and conservation of wild populations.

Background and Summary

Collating large quantities of data is useful not only for assessing large-scale patterns but also for testing theories, informing conservation initiatives, and providing a valuable resource for future data comparisons. In particular, macro-ecological biodiversity assessments are becoming increasingly popular to identify hotspots of species biodiversity that can inform local management strategies (Abell et al., 2008; Brum et al., 2017; Gaston, 2000; Miraldo et al., 2016; Schluter & Pennell, 2017). However, populations, not species, are generally recognized as the appropriate scale for the management of sustainable harvesting and protection in endangered species legislation (Species at Risk Act, 2002; Endangered Species Act of 1973 As Amended through the 108th Congress, 2003; Stephenson, 1999). Nevertheless, population diversity – the number of genetically distinct populations within species – is typically excluded from most biodiversity syntheses and large-scale conservation planning (e.g. DeWoody & Avise, 2000; Hughes, Daily, & Ehrlich, 1997; Medina, Cooke, & Ord, 2018; Miraldo et al., 2016; Santini et al., 2018; Willoughby et al., 2015). This has consequences when assessing biodiversity loss, as population extinction occurs at a much faster rate than species loss, and as such, a species' vulnerability could be grossly misrepresented (Ceballos et al., 2017).

Molecular markers provide an increasingly effective way to differentiate populations and estimate population diversity (Allendorf, 2017). One example is the global population diversity estimate based on allozymes and restriction fragment length polymorphisms where authors found on average 220 populations per species and estimated annual loss of 16 million populations, a coarse estimate obtained by dividing the number of sampling locations by the sampling area (Hughes et al., 1997). The collated data from this study was not made publicly available for future usage and is outdated following the advancement of genetic tools. No study has formally revisited these concepts since this 1997 study (He & Hubbell, 2011; Costello, May & Stork, 2013; Rybicki and Hanski, 2013, but see Ceballos, 2002; Ceballos, Ehrlich and Dirzo, 2017; World Wildlife Fund, 2017 for exceptions), indicating the need for collating population information.

Population genetic technologies have seen advances in recent years, switching from allozymes to microsatellites to single nucleotide polymorphisms (SNPs), largely due to the better resolution of within-population variation that more recent technologies provide (Allendorf, 2017; Schlötterer, 2004). Population structure studies and vulnerability assessments have used

microsatellites as their molecular marker for the past two decades, yet this wealth of data has not been thoroughly collated, although a few authors have collated related information in the form of microsatellite genetic variation (DeWoody & Avise, 2000; Willoughby et al., 2015), population density estimates (Santini et al., 2018), and pairwise F_{ST} estimates (Medina et al., 2018). Despite the great degree of data collation across these studies, no work has combined the geo-referencing of population-specific genetic variation, F_{ST} measurements, and the number of populations within a species to create a single database across a wide variety of taxa and geographic regions.

Here we provide the first description of the release of the Macro-ecological, Population Genetics Database (MacroPopGen Database) – a database that contains geo-referenced population-specific characteristics based on nuclear DNA microsatellites. It contains information on 895 species from 1,308 studies published between 1994 – 2017, and 9,090 distinct populations of amphibians, birds, fish [anadromous, brackish, catadromous, or freshwater], mammals, and reptiles, totalling 561,605 genotyped individuals. Every population entry is georeferenced to permit large-scale spatial analyses, opening a variety of opportunities for overlaying microsatellite genetic data with environmental, geographic, or anthropogenic variables. It allows for population diversity and F_{ST} to be directly compared to species and genetic diversity (e.g. heterozygosity and mean number of alleles) through mapping applications.

MacroPopGen exemplifies the importance and usefulness of collating population genetic data by standardizing data from >1000 different studies, allowing for large-scale comparisons and many future applications, including latitudinal analyses, spatial or temporal analyses, taxonomic comparisons and regional assessments of genetic diversity across taxa or in relation to anthropogenic effects. Previous works focusing on older markers have already shown incredible usefulness in testing a variety of genetic and ecological theories (Miraldo et al., 2016; Schluter & Pennell, 2017; Willoughby et al., 2015). We provide a baseline database for future works to build from and to compare to, particularly for comparing results to different, newer technologies. We urge future population studies using newer technologies to strive for a similar standardized repository for reporting population-specific statistics.

Methods

Data collection

To collect population-genetic data from vertebrate populations located in the Americas, we first scanned Web of Science and Google Scholar for relevant articles using key search terms

including country of occurrence, species common names, author names, and scientific names in combination with "microsatellite", "distinct population", and/or "F_{ST}". A full list of the 1304 key terms and combinations used can be found online (Appendix 1). We also cross-referenced the list of bird microsatellite papers from Willoughby et al. (2015).

Search results with over 1000 hits would be filtered where if two consecutive pages did not yield a relevant result, further pages would not be considered (on average the first 15 pages on Google Scholar would be filtered for relevant articles). This preliminary screening limited results down to 6,297 peer-reviewed studies, technical reports, dissertations and government documents, of which only 1,308 fulfilled our criteria, including 142 of which were obtained from Willoughby et al.'s (2015) bird reference list. Once a study was selected, we extracted where possible: population locality name, latitude-longitude coordinates, average population-specific F_{ST} (Wright's F_{ST} or Weir and Cockerham's unbiased F_{ST} estimator $\theta_{FST;}$ Wright, 1951; Weir & Cockerham, 1984), population-specific observed and expected heterozygosity averaged across loci (H_O/H_E, respectively), sample size (N), population-specific mean number of alleles per loci (MNA), study-specific corrected allelic richness (AR), and the number of microsatellite loci used in the study. For each population, we also documented the taxonomic group (amphibians, birds, fish [anadromous, brackish, catadromous, or freshwater], mammals, or reptiles), family, genus, species, common name, continent, and country. We chose not to include marine species because microsatellites have typically been unable to detect fine-scale population structure in such species, in contrast to the increased power and resolution of more recent genome-scale analyses for such species (Corander et al., 2013). Instead we focus on terrestrial and aquatic ecosystems.

All populations were georeferenced in decimal degrees as a point estimate; if coordinates were not provided, they were inferred from the text or maps in a study. To calculate a metric of population-specific F_{ST}, we consulted pairwise F_{ST} tables and averaged across values that included the focal population, or population group if there was no significance between one or more population pairs. When only a global or regional F_{ST} was reported then that value would be used for all populations within the study; such F_{ST} values are indicated in the database where applicable.

Inclusion criteria and assumptions

A study was retained if two criteria were met: 1) microsatellites were used as molecular markers and 2) genetic differentiation was measured by Weir and Cockerham's pairwise F_{ST} as

opposed to other differentiation estimators because of its wide usage. Microsatellites were favoured over other molecular markers (e.g. SNPs, mitochondrial DNA, allozymes, RAPD, etc.) because their polymorphic nature allows them to resolve population structure at fine scales, particularly for closely related populations (Angers & Bernatchez, 1998; Jarne & Lagoda, 1996). Additionally, microsatellites have higher mutation rates than other markers (Schlötterer, 2004) and have been one of the most widely used genetic markers in recent decades (Schlötterer, 2004). Therefore, microsatellites presently provide an abundance of collectable data across taxa relative to more recent molecular developments associated with single nucleotide polymorphisms (SNPs) or barcoding. While barcoding can assess phylogenetic signals across populations and species, microsatellites allow for the comparison of genetic characteristics between populations such as heterozygosity and allelic diversity, which has been noted to indicate levels of inbreeding or adaptive potential (Fraser et al., 2019; Hansson & Westerberg, 2002; Jump et al., 2009; Reed & Frankham, 2003).

Studies were assumed to have used selectively neutral nuclear microsatellite loci unless otherwise indicated because microsatellites are located within non-coding regions of the genome (Wiehe, 1998) and have relatively fast mutation rates (Väli et al., 2008; Wiehe, 1998). Microsatellite loci are often selected based on their polymorphism due to these faster mutation rates, causing concern that microsatellites may bias measures of genetic diversity compared to whole DNA sequencing-based measures (Ellegren et al., 1997; Väli et al., 2008). Polymorphism bias has also been recognized in studies using other genetic markers such as SNPs (Clark et al., 2005; Nielsen, 2004; Schlötterer, 2004; Väli et al., 2008), and will continue to present challenges in genetic studies. An inherent assumption of this database is that ascertainment bias is similar across all studies and taxa, and therefore comparable. Additionally, previous work (Willoughby et al., 2015) has concluded that the number of loci and primer type (whether cross-species or focal species) were not important in explaining variability in genetic diversity, an indication that ascertainment bias may not be very significant for large quantities of microsatellite data such as this database. Regardless, we tested ascertainment bias with a subset of the database, as described below.

Demarcating Populations

Populations were considered genetically distinct above a threshold F_{ST} value of 0.02. F_{ST} was used as the statistical measure of differentiation because of its standardized and common use

in the literature for measuring genetic differentiation. The chosen threshold was based on a previous analytical review (Waples & Gaggiotti, 2006), which indicated that genetic differentiation is not negligible if $F_{ST} \geq 0.05$, but an F_{ST} value as low as 0.01 can also denote statistically significant differentiation (Waples & Gaggiotti, 2006). While lower values of F_{ST} (0.02 to 0.01) are sufficient to show significant genetic differentiation, such values are more relevant for distinguishing specific taxonomic groups, such as marine fish populations which exhibit more gene flow (Waples, 1998; Waples & Gaggiotti, 2006). Freshwater and terrestrial species tend to experience lower rates of gene flow than marine species and therefore an F_{ST} threshold above 0.01 is more appropriate (Medina et al., 2018; Waples, 1998). To avoid accepting biologically insignificant population differentiation (type I error) or rejecting biologically significant differentiation (type II error) when demarcating populations, we considered the significance of F_{ST} values where available. We ensured that any pairwise comparisons >0.02 were statistically significant; we also checked significance when F_{ST} was <0.02 and significance implied two separate populations despite a lower F_{ST}. We also accounted for sample sizes with respect to significant F_{ST}. If sample size was five or less (occurring <0.1% of all cases in this study) and populations were found to be significantly different, the populations were instead grouped as one unless an adequate biological explanation was provided (n=5). Likewise, if sample size was very large (e.g. >50) but F_{ST} was <0.02, consideration would be taken to determine if the populations were significantly different given the statistical support large sample sizes provide (usually given by p-values in the specific study, n=63 cases where $n \geq 50$ but $F_{ST} \leq 0.02$). Additionally, if multiple studies were conducted in the same location for the same species, data from the most recent study or the one with the most microsatellite loci was used (n=268 populations were duplicates and removed). When F_{ST} tables were unclear (e.g. many low F_{ST} values and no significance given), we considered results from population structure analyses (e.g. STRUCTURE, BAPS, etc.) to make informed decisions about population structure.

Geographic Breadth

We also report (i) how differentiated each population is in relation to all other populations it was compared to by calculating the average F_{ST} between a focal population and all other populations within that study, and (ii) the number of populations included in the calculation as well as the geographic distance or breadth that they span. For example, low F_{ST} values

resulting from only a few sampling locations (e.g. 5) in a small geographic region (e.g. 10 km) may have a different interpretation than low F_{ST} values across many (e.g. >10) sampling locations in a broad geographic range (e.g. 10,000 km). To estimate the geographic breadth that sampled populations cover, we obtained coordinates for each population including locations that had been combined into one population. These data were put into a separate file that contains 10, 921 sampling localities. Next, we used custom code (Appendix 1) utilizing the R package geosphere to calculate the maximum, minimum, and mean distances in metres between all populations of a study; distances are reported in metres in the database. We additionally note how many sampling localities make up each population in the database and how coordinates were obtained/estimated for populations that encompass multiple sampling localities.

Statistical Analysis

To calculate mean genetic diversity for taxonomic groups and continental regions we used generalized linear mixed models (GLMMs) that accounted for the random effect of study, species, genus, and family. Fixed effects included either the taxonomic group, or the continental region. Beta distributions were used to model H_O and F_{ST} (R package glmmTMB v 0.2.2.0) because both these response variables and distributions are bounded between zero and one with no exact zeros or ones. Gamma distributions were used for MNA (R package lme4 v 1.1-18-1) as MNA follows a positively right skewed distribution characteristic of gamma distributions. We then used the R package and function emmeans (v 1.2.3) to calculate the mean values while accounting for model structure. For the models that used beta distributions, we used the function back.emmeans (R package RVAideMemoire v0.9-69-3) to back transform estimates.

To compare the degree of variation in each taxonomic or continental group, we calculated the coefficient of variation grouped at the species level for H_O, MNA, and F_{ST}. Mixed models using the gamma distribution and random effects of reference, genus, and family were constructed. We then used model selection to see which between taxonomic group or continental region best explained differences between groups.

We assessed trends of ascertainment bias related to microsatellite loci development using a subset focusing on North American mammalian data (n=1579 populations, 73 species; Appendix 1). In addition to the number of microsatellite loci, we obtained from 230 mammalian studies the number of species used to develop those loci (ranged from 1 to 7), and whether the species were focal (n=384), non-focal (n=545), or mixed (n=692), as well as information on the

senior author's country of affiliation. Using IUCN descriptions for each species, we also determined whether the species was harvested and to what extent (no n=317, low n=957, or high n=347), the species' IUCN status (Least Concern n=1335, Near Threatened n=45, Vulnerable n=193, Endangered n=41, Critically Endangered n=7), whether the species was of conservation concern (no n=561, low n=211, or high n=849), charismatic (no n=495, low n=189, or high n=937), or of economic value (no n=602, low n=887, or high n=132). Extent of harvesting was determined by the degree of harvesting described in IUCN's "Use and Trade" category: none ("no"), subsistence or local harvesting ("low"), or substantial commercial harvesting ("high"). Conservation concern was specified to account for species that may have a lower IUCN rank (e.g. Least Concern, LC) but still have populations at risk or aspects of their habitat at risk (e.g. 563 LC species were still of conservation concern and therefore considered as "low"); this was largely described in IUCN's "Threats" and "Conservation Action" categories. Charisma of species was somewhat subjective as it was determined by how generally well-known the species was, and whether the species may be considered a nuisance which would negatively affect their charisma score (e.g. the coyote is well known but can be considered a pest and as such its score was "low"). Economic value of a species was determined by the "Use and Trade" section, where if the species was commercially harvested it would be considered to have economic value ("high"); if the harvest has declined or is relatively low, a species' economic value was considered as "low".

We tested the fixed effects and interactions among these factors for ascertainment bias as well as the random effects of reference, species, genus, and family. We used GLMMs, using a beta distribution for H_O (R package glmmTMB) and a gamma distribution for MNA (R package lme4). Following Zuur et al. (2009) guidelines for forwards and backwards model selection, we used the likelihood ratio test to find significant factors for the H_O and MNA models, respectively.

Code and Data Availability

The data and R code used for the analyses are available from FigShare (Appendix 1).

Data Records

Data from the MacroPopGen database is hosted at Figshare (Appendix 1) and can be downloaded as one XLSX file. It consists of 9,098 rows (distinct populations), and 22 columns. The columns include taxonomic identifiers (family, genus, species, common name), population locality names, and study-specific data (sample size, population-specific F_{ST}, observed and expected heterozygosity, mean number of alleles, standardized allelic richness, latitude and longitude coordinates, reference ID, and year).

An additional XLSX file containing the corresponding references for each reference ID, and the list of key terms used in searches is also available on Figshare (Appendix 1). Most of the references were published in English, although a minority are in Spanish.

Technical Validation

Geographic and taxonomic bias

Between 1994 and 2017, most population microsatellite data came from species studied in North America (85.1%, Table 1.1). Fish species were the most represented taxonomic group, making up 44.8% of the database (Table 1.1). Salmonid species made up 55.9% of fish population data and represented 25.0% of data across all taxa.

When accounting for model structure, mean population genetic diversity differed significantly between some continental regions for H_O and MNA (Figure 1.1). Populations of South American species had the lowest H_O while Caribbean populations showed significantly lower MNA (Table 1.1, Figure 1.1). Despite some significant differences, the range of mean population genetic diversity metrics among continental regions was limited, between 0.57 and 0.61 for H_O, and 4.11 and 5.5 for MNA (Figure 1.1). Continental population differences in F_{ST} were stronger than for genetic diversity metrics, wherein Caribbean populations showed significantly higher population-specific F_{ST}, suggestive of less gene flow overall for these populations. This result follows general island-mainland expectations where island populations tend to be more isolated than mainland populations (Frankham, 1997; Jaenike, 1973).

Among taxonomic groups, populations of anadromous fish had statistically higher mean genetic diversity (MNA = 7.8), and lower average F_{ST} values (0.06) aside from birds (mean F_{ST} =0.05) (Figure 1.1), consistent with previous work (DeWoody & Avise, 2000; Medina et al., 2018). Mammalian populations also had lower mean MNA than all other groups (Figure 1.1).

However, there were no significant differences in mean H_O between taxonomic groups (Figure 1.1).

Variation among taxonomic and continental groups

There were significant differences in the coefficient of variation for H_O among taxonomic groups but not continental regions, with bird species showing the least variation (Figure 1.1). There were no significant differences in the coefficient of variation for species MNA across taxonomic groups or continental regions (Figure 1.1). For F_{ST}, the only statistical difference was for the coefficient of variation to be larger in North American species relative to species in other regions, i.e. no taxonomic group differences in F_{ST} variance were found (Figure 1.1). More variance among taxonomic distinctions was observed when considering within-family and within-genera variance in genetic metrics (Figure 1.2). For example, the mean family H_O ranged between 0.07 – 0.88, while MNA ranged from 1.40 – 24.97, and mean F_{ST} ranged from 0.0008 – 0.72; genera averages had a similar range for both metrics.

Bias with microsatellite loci

We assessed how genetic diversity and the number of microsatellite loci employed in empirical research has changed over time using linear models (Figure 1.3). There has been a significant trend for increasing number of loci per year (R^2=0.07, p<0.001) as well as a weak increase in genetic diversity with year (H_O: R^2=0. 0.009, p<0.001 and MNA: R^2=0.001, p=0.003). Additionally, we evaluated bias with respect to the number of microsatellite loci and the degree of genetic variation in H_O and MNA using funnel plots (Figure 1.4) and linear models. The plots appear to be largely symmetrical and show little bias with respect to number of loci, indicating the data capture a reasonable degree of genetic variation for the number of loci used. Note that we could not use a formal funnel plot test such as the Egger test because we do not have variance for H_O and MNA for each study. However, the number of microsatellite loci was a significant predictor in linear models for both H_O and MNA (p<0.001 for both), although adjusted R^2 values were very small (0.002 and 0.03, respectively).

Ascertainment bias

After model selection testing for ascertainment bias with respect to loci type and origin, only the interaction between level of harvesting and conservation concern as well as the random effects of reference, family, and genus were significant for the Ho model (Table 1.2). For the MNA model, the significant factors only included the interaction between conservation concern

and charisma, as well as the random effects for reference and genus. None of the factors associated with microsatellite bias were retained in model selection (i.e. number of species used to derive loci, whether those species were focal, non-focal, or mixed). These results are consistent with previous assessments (Willoughby et al., 2015) but indicate that microsatellite loci and loci origin do not significantly affect genetic diversity metrics when analyzed across diverse taxa.

Acknowledgements

Much work went into building this database and we would like to thank J. D. Lawrence, K. Levasseur-Bhatia, A. de Jaham, D. MacRae, and C. Clegg, for additional help in building the database, finding population coordinates, and helping collate references; M. Yates for statistical support. Additionally, we would like to thank two anonymous reviewers who gave constructive feedback on the manuscript. We would sincerely like to thank Dr. K. J. Monsen, Dr. T. Beacham, and Dr. J. Willoughby for sharing datasets with us that enabled completion of the database. This work was funded by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant, and a Quebec Centre for Biodiversity Science Seed Grant.

Tables

Table 1.1. Summary statistics for data collected from microsatellite studies published between 1994 and 2017 broken down by taxonomic group. N = sample size; H_O = observed heterozygosity; MNA = mean number of alleles; SD = standard deviation; SE = standard error. Amph = amphibians; Anad = anadromous fishes; FW = freshwater fishes; Mam = mammals; Rep = reptiles; NOR = North America; CEN = Central America; CAR = Caribbean; SOU = South America. Brackish and catadromous fishes are not shown due to their low number of populations (25 and 33, respectively), but their populations are included in the regional summaries.

	Amph	**Bird**	**Anad**	**FW**	**Mam**	**Rep**	**NOR**	**CEN**	**CAR**	**SOU**	**Total**
Unique families	17	66	6	42	37	28	135	31	16	98	**195**
Unique genera	46	170	9	99	93	66	308	40	18	173	**480**
Unique species	104	254	15	231	158	133	578	45	26	282	**895**
Number populations	1117	608	1315	2704	1943	1349	7738	230	107	1015	**9090**
Studies	136	265	72	298	344	203	962	46	32	299	**1308**
Countries	10	28	2	16	19	30	4	6	15	14	**39**
Published year range	2001-2016	1997-2017	1997-2016	1997-2017	1994-2016	1997-2017	1994-2017	2002-2016	2002-2017	1997-2017	**1994-2017**
Mean latitude	32.713	25.923	50.546	37.445	34.188	27.520	43.415	11.643	18.384	-14.585	**35.83**
Total number of loci	10870	6713	18958	28069	23213	13869	88259	2421	1050	10701	**102431**
Mean number loci per study	9.740	10.987	14.439	10.450	11.947	10.273	11.437	10.526	9.813	10.543	**11.29**
SD number loci across studies	3.689	6.711	4.0329	4.465	5.587	4.928	5.161	5.124	6.924	3.975	**5.08**
Total individuals genotyped	46015	48393	181606	140569	91147	50978	507765	8990	3904	40946	**561605**

Median study N	22	34	83	30	25	22	30	28	20	24	**30**
SD N	88.472	126.508	174.205	198.611	96.694	69.460	156.897	54.052	35.703	71.330	**147.43**
Mean H_O	0.596*	0.592*	0.627*	0.566*	0.594*	0.582*	0.596*	0.610*	0.576*	0.567*	**0.59**
SE H_O	0.023*	0.031*	0.014*	0.077*	0.017*	0.019*	0.022*	0.029*	0.009*	0.012*	**0.16**
Mean MNA	5.650*	5.339*	7.807*	5.629*	4.855*	6.077*	4.838*	5.536*	4.110*	5.203*	**7.92**
SE MNA	0.313*	0.189*	0.692*	0.219*	0.140*	0.293*	0.159*	0.383*	0.348*	0.212*	**5.57**
Mean population F_{ST}	0.106*	0.052*	0.062*	0.092*	0.091*	0.086*	0.073*	0.120*	0.079*	0.086*	**0.13**
SE population F_{ST}	0.015*	0.006*	0.011*	0.009*	0.009*	0.009*	0.009*	0.017*	0.005*	0.008*	**0.12**

***** Calculated to account for model structure. See text for details.

Table 1.2. Summary of model selection results for testing ascertainment bias within H_O and MNA. H_O model selection was done in a forwards fashion, while MNA model selection was done in a backwards fashion; see text for details.

Model	**AIC**	**DF**
H_O ~ 1 + (1|Reference) + (1|Species) + (1|Genus) + (1|Family)	-2196.0	6
H_O ~ MsatType + (1|Reference) + (1|Genus) + (1|Family)	-2183.6	7
H_O ~ ConservC + (1|Reference) + (1|Genus) + (1|Family)	-2202.3	7
H_O ~ Harvested + ConservC + (1|Reference) + (1|Genus) + (1|Family)	-2212.2	9
H_O ~ MsatType + ConservC + (1|Reference) + (1|Genus) + (1|Family)	-2198.2	9
H_O ~ Harvested * ConservC + (1|Reference) + (1|Genus) + (1|Family)	-2215.2	13
H_O ~ MsatType + Harvested * ConservC + (1|Reference) + (1|Genus) + (1|Family)	-2214.2	14
H_O ~ NSpp +Harvested * ConservC + (1|Reference) + (1|Genus) + (1|Family)	-2215.3	14
H_O ~ msat + Harvested * ConservC + (1|Reference) + (1|Genus) + (1|Family)	-2212.1	15
MNA ~ ConservC:Charisma + (1 | Reference) + (1 | Genus) + (1|Species)	4015.6	13
MNA ~ NSpp + ConservC: Charisma + (1 | Reference) + (1 | Genus) + (1|Species)	4016.2	14
MNA ~ NSpp + MsatLoci + ConservC + AuthorCountry + ConservC: Charisma + (1 | Reference) + (1 | Genus) + (1|Species)	4021.6	19
MNA ~ NSpp + MsatLoci + MsatType + Harvested + ConservC + Economic + Charisma + AuthorCountry + ConservC: Charisma + (1 | Reference) + (1 | Genus) + (1|Species)	4031.5	25
MNA ~ NSpp + MsatLoci + MsatType + Harvested + ConservC + Economic + Charisma + AuthorCountry + NSpp: MsatLoci + NSpp: MsatType + MsatLoci: MsatType + Harvested: ConservC + Harvested:cmn + ConservC: Charisma + (1 | Reference) + (1 | Genus) + (1|Species)	4043.7	36
MNA ~ NSpp + MsatLoci + MsatType + Harvested + ConservC + Economic + Charisma + AuthorCountry + NSpp:	4050.9	37

MsatLoci + NSpp: MsatType + MsatLoci: MsatType + Harvested: ConservC + Harvested: Charisma + ConservC: Charisma + (1 | Reference) + (1 | Species) + (1 | Genus)

NSpp: number of species used to derive loci; MsatLoci: total number of microsatellite loci; MsatType: microsatellite type (focal, non-native, native); Harvested: level of harvesting; ConservC: degree of conservation concern; Economic: economic value; Charisma: charisma of focal species; AuthorCountry: senior author's country of residence.

Figures

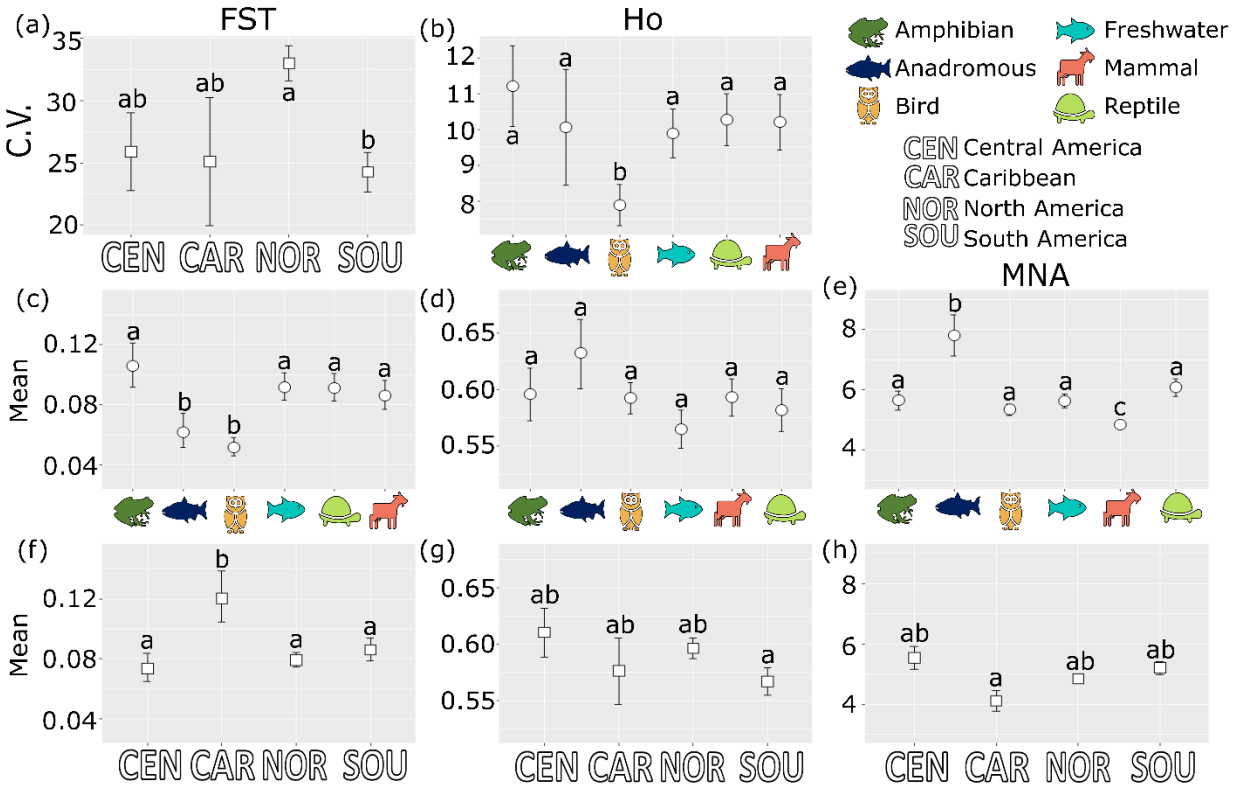

Figure 1.1. Coefficient of variation and mean values for observed heterozygosity (Ho), mean number of alleles (MNA), and population-specific F_{ST} calculated to account for GLMM structure. Error bars represent standard error. Significant differences between groups indicated by letter grouping where groups sharing the same letter(s) are not significantly different from one another. (a, b) Coefficient of variation calculated across (a) taxonomic groups (circles) and (b) between continental regions (squares). (c - e) Mean (c) F_{ST}, (d) Ho, and (e) MNA calculated across taxonomic groups. (f - h) Mean (f) F_{ST}, (g) Ho, and (h) MNA calculated between continental regions.

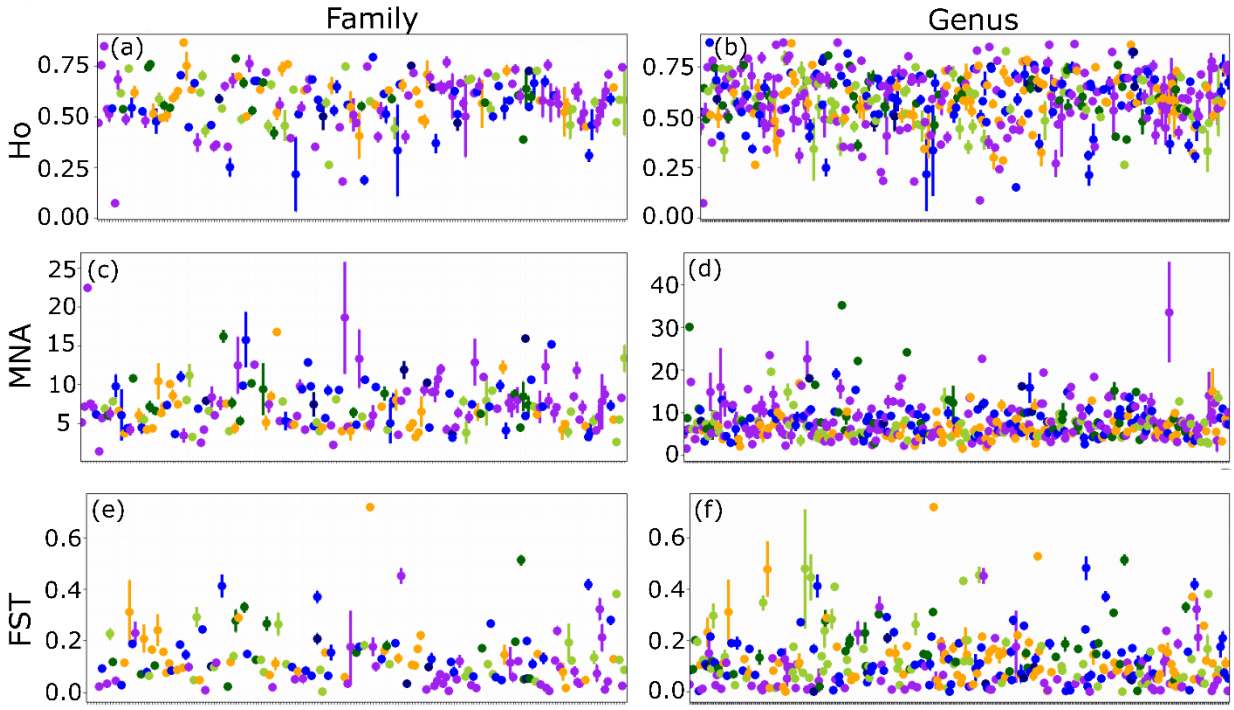

Figure 1.2. Microsatellite observed heterozygosity (H_O), mean number of alleles (MNA), and population-specific F_{ST} averaged across each vertebrate group, according to Family (left column) or Genus (right column), indicated on the x axis. Colours indicate the taxonomic group each family or genus belongs to: dark green = amphibians, purple = birds, blue = fish, orange = mammals, light green = reptiles. Error bars represent standard error. (a, c, e) Ho, MNA, and F_{ST} are averaged across vertebrate families (n=195). (b, d, f) Ho, MNA, and F_{ST} are averaged across vertebrate genera (n=480).

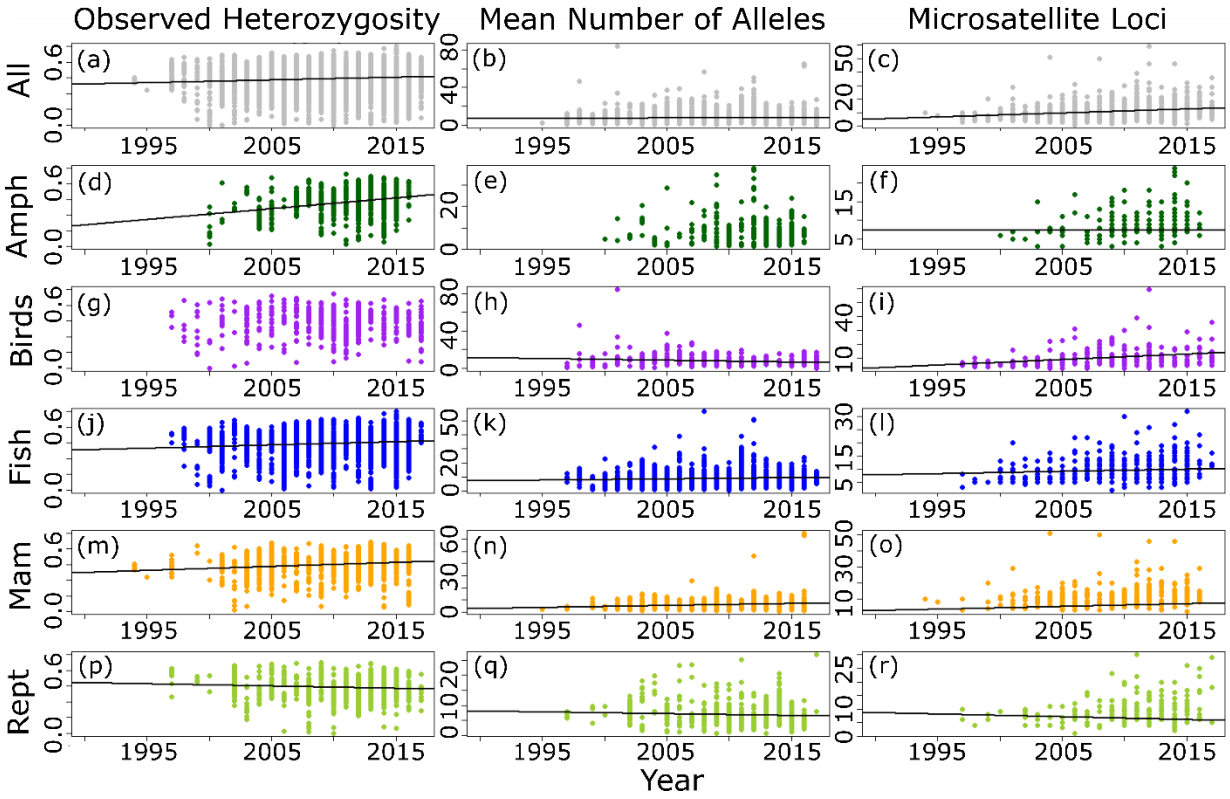

Figure 1.3. Observed heterozygosity, mean number of alleles, and number of microsatellite loci for populations of each taxonomic group sampled between the years 1994 to 2017. (a – c) All vertebrate groups together; (d – f) only amphibian species; (g – i) bird species; (j – l) all fish species; (m – o) mammalian species; (p – r) reptile species. Linear models are indicated for significant relationships.

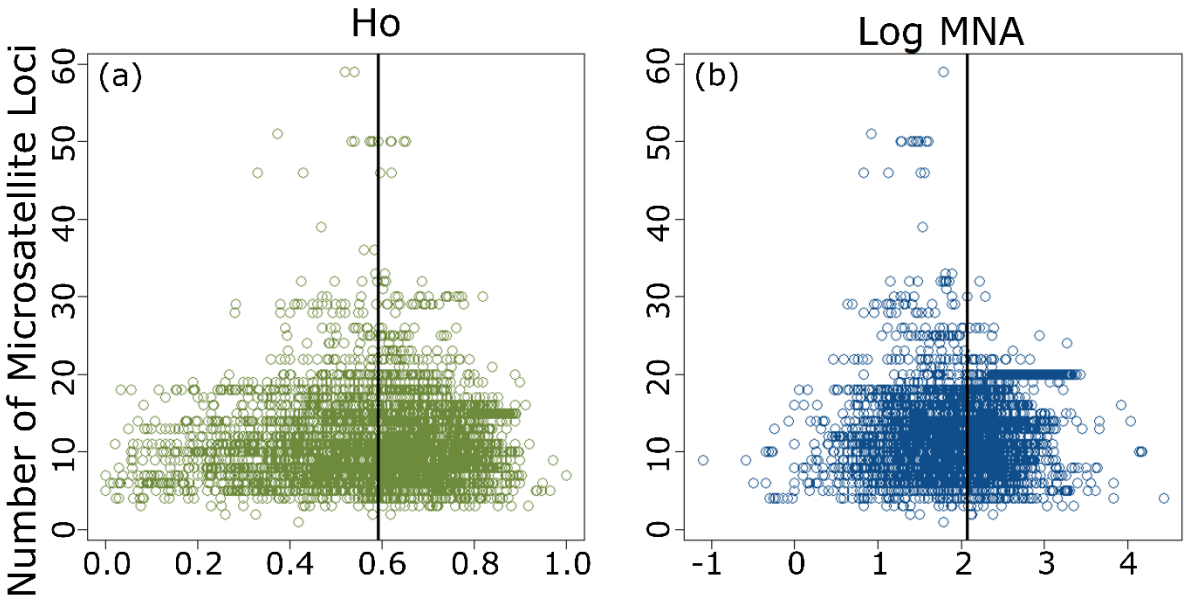

Figure 1.4. Funnel plots for all populations; y axis for both plots is the number of microsatellite loci, and (a) x axis is observed heterozygosity (H_O) or (b) mean number of alleles (MNA). Vertical line represents the mean value.

Chapter 2: Latitudinal biodiversity gradients at three levels: linking species richness, population richness, and genetic diversity

Published as:

Lawrence, E. R., & Fraser, D. J. (2020). Latitudinal biodiversity gradients at three levels: Linking species richness, population richness and genetic diversity. *Global Ecology and Biogeography*, **29**(5), 770-788

Lawrence, Elizabeth; Fraser, Dylan (2020), Data from: Latitudinal biodiversity gradients at three levels: linking species richness, population richness, and genetic diversity. Dryad Digital Repository. https://doi.org/10.5061/dryad.xgxd254ck

Abstract

Motivation: Theory describing biodiversity gradients has focused on species richness with less conceptual synthesis outlining expectations for intraspecific diversity gradients, i.e. broad-scale population richness and genetic diversity. Consequently, there is a need for a diversity-gradient synthesis that complements species richness with population richness and genetic diversity.

Review Methods: Species and population richness are the number of different species or populations in an area, respectively. Population richness can be totalled across species, within a species, or averaged across species. Genetic diversity within populations can be summed or averaged across all species in an area or be averaged across an individual species. Using these definitions, we apply historical, ecological, and evolutionary frameworks of species richness-gradients to formulate predictions for intraspecific diversity gradients.

Review Conclusions: All frameworks suggest higher average population richness at high latitudes, but similar total population richness across latitudes. Predictions for genetic diversity patterns across species are not consistent across frameworks and latitudes.

New Analysis Methods: Species range size tends to increase with latitude, so we used empirical data from ~900 vertebrate species to test hypotheses relating species range size and richness to population richness and genetic diversity.

New Analysis Conclusions: Species range size was positively associated with its population richness but not with species-specific genetic diversity. Furthermore, a positive linear relationship was supported between species richness and total population richness, but only weakly for average population richness.

Overall conclusion: Through the lens of species richness theories, our synthesis identifies an uncoupling between species richness, population richness, and genetic diversity in many instances due to historical and contemporary factors. Range size and taxonomic differences appear to play a large role in moderating intraspecific diversity gradients. We encourage further analyses to jointly assess diversity-gradient theory at species, population, and genetic levels towards better understanding Earth's biodiversity distribution and refining biodiversity conservation.

Keywords:

Genetic diversity; intraspecific diversity; latitudinal gradient; population ecology; population genetics; population richness; species diversity; species richness

Introduction

The latitudinal distribution of species richness is one of the most widely recognized and predictable patterns of biodiversity (Brown, 2014; Costello et al., 2013; Fine, 2015; Mittelbach et al., 2007; Pianka, 1966; Roll et al., 2017; Stork, 1993; Willig et al., 2003). To date, however, the extensive theoretical and empirical attention directed to species richness gradients has not been extended to understand the broad-scale distribution of other important components of biodiversity, namely intraspecific diversity. Levels of intraspecific diversity – whether characterised as functional diversity, phylogenetic diversity, population richness, and/or genetic diversity within and among populations – can influence species geographic distributions, species responses to environmental change (Barrett & Schluter, 2008; Bernatchez, 2016; Jump et al., 2009; Willoughby et al., 2018), community structure, and ecosystem functioning (Des Roches et al., 2018; Raffard et al., 2019). This influence of intraspecific diversity on species themselves could suggest that the distribution of species richness is affected by intraspecific diversity patterns, or vice-versa. Notably lacking in the literature is a foundation for theoretical expectations of intraspecific diversity and its distribution, a gap we aim to resolve in this review.

Large quantities of data on intraspecific diversity have recently become available for broad-scale analyses due to technological advances and the accumulation of smaller-scale empirical works (DeWoody & Avise, 2000; Hughes et al., 1997; Lawrence et al., 2019; Martinez et al., 2018; Medina et al., 2018; Miraldo et al., 2016; Willoughby et al., 2015). When collated, such data allow for the extension of species-centric latitudinal concepts towards understanding how broad-scale intraspecific diversity patterns may better inform, for example, the speciation process (Adams & Hadly, 2013; Schluter & Pennell, 2017; Smith et al., 2017) and biodiversity conservation by revealing hot/cold spots of intraspecific diversity (Marchese, 2015; Paz-Vinas et al., 2018). Herein, we focus our discussion specifically on broad-scale patterns of two metrics of intraspecific diversity: population richness within species and genetic diversity, and how these metrics relate to species richness gradients. Related topics on other components of intraspecific diversity, such as functional and phylogenetic diversity, are discussed in Marske et al. (2013), Economo et al. (2018), and Martinez et al. (2018).

Dynamics between- and within-populations are the steppingstone linking genetic diversity with species richness (Fine, 2015; Marchese, 2015; Marske et al., 2013; Paz-Vinas et al., 2018; Singhal et al., 2018). Increasing genetic differentiation leads to population divergence

and eventually speciation due to isolation and/or selection (Schluter, 2016; Schluter & Pennell, 2017; Taylor, 1999; Wiens & Donoghue, 2004). Therefore, characterizing population richness relative to species richness and genetic diversity is fundamental to refine our understanding of the interrelationships between species richness and genetic diversity. For example, an area's species and population richness could be one indicator of the age and speciation potential of that community: high species richness but low population richness can indicate older communities with lower rates of speciation as all niches may be filled (Kennedy et al., 2018; Schluter, 2016; Schluter & Pennell, 2017). Furthermore, the joint investigation of the distribution of species richness, population richness, and genetic diversity may allow more accurate inferences about ecological history, including glacial refugia, recolonization, and founder effects (Bernatchez & Wilson, 1998; Blanchet et al., 2017; Fedorov & Stenseth, 2002; Galbreath & Cook, 2004; Marske et al., 2013; Tamkee et al., 2010). Some research has bridged how aspects of genetic diversity may relate to and have consequences for communities and species richness (Antonovics, 1976, 2003; Hughes et al., 2008; Lamy et al., 2017; Laroche et al., 2014; Marchesini et al., 2018; Marske et al., 2013; Pfeiffer et al., 2018; Vellend et al., 2014; Vellend, 2005, 2010; Vellend & Geber, 2005). Still lacking, however, is a strong conceptual foundation linking species richness, population richness, and genetic diversity within the framework of the latitudinal gradient. To build this foundation we draw from theories presented in the species richness literature.

Many of the non-mutually exclusive theories and hypotheses proposed to explain the latitudinal gradient in species richness (Fine, 2015; Willig et al., 2003) can be structured into historical, ecological, and evolutionary frameworks (Brown, 2014; Mittelbach et al., 2007; Schemske & Mittelbach, 2017). We begin by elaborating on each of these three broad frameworks and how they relate to species richness, population richness, and genetic diversity. We focus on vertebrate groups across the American continents, as they tend to be more mobile and have a large body of focal research (Bazin et al., 2006). The Americas offer a unique opportunity for discussing latitudinal gradients because the continents are largely arranged in a north to south fashion. To ensure use of standardized terms throughout the review as well as to clarify distinctions among past works, we have broken down population richness and genetic diversity into five categories (see Glossary): 1) total number of populations across species in an area (TotPopR); 2) the number of populations for a given species in an area (PopPerSpp); 3) the

average number of populations per species in an area (AvgPopSpp); 4) total genetic diversity in an area, as the sum of genetic diversity across all species at a population level (TotGenDiv); and 5) average genetic diversity across populations for a given species in an area (GenPerSpp). Note that our definition of genetic diversity refers largely to neutral genetic diversity, not adaptive genetic diversity, as it allows us to make usage of a comprehensive population genetics database that we recently compiled from ~900 vertebrate species spanning the American continents (Lawrence et al., 2019). While data on adaptive genetic diversity are increasing, to date they are insufficiently rich for similar, standardized collation and would likely have different expectations that should be explored in future works.

We structure this paper into two parts. First, we synthesize the general expectations for latitudinal patterns in species richness, population richness, and genetic diversity under historical, ecological and evolutionary frameworks (Table 2.1). Second, using the aforementioned database, we conducted new analyses to test the following hypotheses for population richness and genetic diversity, specifically considering the role of a species' range size: (i) larger range sizes are associated with greater numbers of genetically distinct populations per species (PopPerSpp); (ii) areas with more overlapping species ranges, i.e. higher species richness, have lower PopPerSpp but higher TotPopR; and (iii) larger range sizes have higher levels of genetic diversity (GenPerSpp).

To our knowledge, this is the first conceptual review to link core concepts of population genetics, population ecology, and macroecology towards a better understanding of biodiversity gradients at species and below species levels. We hope the review encourages further interdisciplinary collaboration and continuous integration of such broad-scale concepts.

Review: Understanding the latitudinal gradient of biodiversity via three levels

Historical framework

Historical hypotheses examine Earth's history and how the duration and extent of environments in the past structure current species richness patterns (Brown, 2014; Miller & Román-Palacios, 2019; Mittelbach et al., 2007; Sandel et al., 2011). Two of the most encompassing historical hypotheses are the time and area hypotheses (Table 2.1; Fine, 2015; Li & Wiens, 2019; Mittelbach et al., 2007; Pianka, 1966; Schluter, 2016; Willig et al., 2003). These hypotheses are based largely on the historical extent of the tropical region being larger and older than that of temperate regions (e.g. not covered with glaciers, greater latitudinal extent due to

warming periods in the early Tertiary; Fine, 2015; Mittelbach et al., 2007; Pianka, 1966; Sluijs et al., 2006). Namely, the tropics have had more time and historically more space for organisms to speciate, leading to many species radiating from tropical environments towards temperate ones (Fine, 2015; Stephens & Wiens, 2003; Wiens & Donoghue, 2004). Of course, these expectations may be context-dependent. For example, they may apply less in the Americas since temperate North America is larger than more tropical Central/South America. The area available for speciation in the tropics would then be considerably less than in the relatively species-poor northern hemisphere, with consequences for broad-scale patterns of intraspecific diversity.

Historical framework: Population richness

Extended to the population level, the greater time and area for species to live in the tropics may have also provided ample time for populations to differentiate across a heterogeneous tropical habitat (Mittelbach et al., 2007; Rosenzweig, 1995; Terborgh, 1973). Therefore, across the very large number of species at low latitudes, we might expect high numbers of populations overall (TotPopR; Table 2.1). This high TotPopR might still apply to tropical environments in the Western hemisphere even though contemporarily they have a smaller geographic area relative to North America, because the low latitude environments have been open and inhabited longer by species than habitable area at higher latitudes. On the other hand, the smaller contemporary area of tropical environments also means that there is now less opportunity for new populations to diverge within species (PopPerSpp) compared to within temperate environments (see also related concepts on species range sizes and diversification rates in the ecological and evolutionary frameworks below, respectively). Moreover, the lower number of temperate species in North America have had less time residing in open (connected) habitats, resulting in incomplete speciation but greater population structure across a species' range. Such incomplete speciation in turn would likely lead to higher PopPerSpp than found at low latitudes and reduce or perhaps even eliminate tropical vs. temperate differences in TotPopR. Aside from the ample time that species have had occupying low latitudes in the past, we next consider how historical adaptations can structure future evolution, and the implications for population richness.

The ancestral niche of a species clade determines future regions and habitat that the clade can disperse to and persist in, a phenomenon known as phylogenetic niche conservatism (Ackerly, 2003; Peterson et al., 1999; Ricklefs & Latham, 1992; Wiens, 2004; Wiens & Donoghue, 2004). A species can then only broaden its niche through niche evolution and

dispersal (Sandel et al., 2011; Wiens & Donoghue, 2004), which would influence population differentiation across a species' range. For example, niche evolution may result in an increase in genetic differentiation among populations at range edges due to strong directional selection, similar to expectations from the central-marginal hypothesis (Eckert et al., 2008; Guo, 2012; Hargreaves & Eckert, 2019; Willi et al., 2018). With more differentiated populations at range edges, we might expect high latitudes to have higher PopPerSpp, as species that have expanded their ranges outwards from the tropics would likely have more populations in these high latitude areas. Conversely, low latitude clades would again be expected to have lower PopPerSpp. These clades are typically older, having had more time for speciation of the edge populations to occur, thus resulting in fewer populations in the remaining range.

Overall, when considering historical influences, we would expect higher PopPerSpp at high latitudes relative to low latitudes. Predictions about TotPopR are less clear in part because they are dependent on the magnitude of the difference in species richness and PopPerSpp between tropical and temperate regions. Nonetheless, more similar levels of TotPopR across latitudes seem plausible. For example, while the sheer number of species in the tropics may result in a high TotPopR at low latitudes, multiple factors discussed here may also generate high or higher TotPopR at high latitudes; or at least TotPopR might not be proportional to species richness or PopPerSpp. In short, these points illustrate how species richness and population richness may be uncoupled in many instances.

Historical Framework: Genetic diversity

The historical extent of habitat and its availability can also influence current patterns of genetic diversity, but we would expect an opposite pattern from population richness – higher genetic diversity at low to intermediate latitudes. For example, since more time has passed for evolution to occur at lower latitudes, genetic diversity would accumulate across species (TotGenDiv) (Adams & Hadly, 2013). Average genetic diversity across populations within a tropical species (GenPerSpp) could also follow the same pattern as TotGenDiv, since older clades tend to have more genetic diversity (Willi et al., 2018). Conversely, habitat has been available for less time at high latitudes so organisms may not have had sufficient time to accumulate as much genetic diversity.

In addition to purely just having had more time or not, clade age and history would play a large role in structuring patterns of genetic diversity. For example, many North American clades

have experienced glacial fragmentation across their ranges, followed by founder effects after glacial retreat. Founder effects like these can lead to an overall decrease of genetic diversity within a species (GenPerSpp) (Galbreath & Cook, 2004; Green et al., 1996; Hewitt, 2000; Provan & Bennett, 2008; Stewart et al., 2010; Willi et al., 2018). Additionally, adaptation to colder or new environments (i.e. niche evolution, Weins & Donoghue, 2004) tends to elicit strong directional selection, often leading to losses in GenPerSpp (Eckert et al., 2008; Hargreaves & Eckert, 2019; Pierce et al., 2017).

GenPerSpp and TotGenDiv therefore may be lowest at high latitudes due to the strong influence of historical events and the recent establishment of populations (e.g. after the last glacial maximum 23,000 – 18,000 ybp; Hewitt, 2004), leading to less time for alleles to accumulate in a given population from mutations. Miraldo et al. (2016) may support this expectation in finding higher mitochondrial genetic diversity (TotGenDiv) at low latitudes. However, re-analyses of their data (Gratton et al., 2017a; Schluter & Pennell, 2017) showed a systematic northward bias in spatial autocorrelation and that the pattern was not consistent across species (i.e. GenPerSpp). Despite this, Schluter & Pennell (2017) demonstrated that mammalian and amphibian mitochondrial genetic diversity, equivalent to GenPerSpp here, has a slightly negative slope with latitude. Another study found some evidence for scale- and taxa-dependent latitudinal gradients in genetic diversity (Millette et al., 2019). These results are mostly consistent with the expectations from a historical perspective, wherein species at low latitudes have experienced more time for genetic diversity to accumulate, but some nuances appear to blur gradient patterns.

Historical Framework: Conclusion

In general, historical hypotheses tend to predict higher PopPerSpp at high latitudes, similar levels of TotPopR across latitudes, and more genetic diversity at low latitudes. While the historical time and area available for diversification may form the foundation from which species evolve, current patterns of intraspecific diversity may be the product of both historical and contemporary processes. Changes in population differentiation and genetic diversity can occur relatively quickly throughout time (e.g. tens to hundreds of years instead of thousands or millions) due to stochastic processes as well as increasing anthropogenic impacts (Goossens et al., 2006; Riley et al., 2006; Weider et al., 1997). Human impacts can cause species range barriers more quickly than otherwise expected, greatly reducing connectivity and thus increasing

the likelihood of population differentiation (Ascensão et al., 2016; Cheptou et al., 2017; Meyer et al., 2009; Riley et al., 2006). Alternatively, human influences can homogenize populations by the movement of individuals through introductions, translocations, stocking, and supplementation (Johnson et al., 2010; Tringali & Bert, 1998). The roles and effects of more contemporary ecological hypotheses are discussed in more depth below.

Ecological framework

Ecological hypotheses focus on the mechanisms underlying species coexistence, maintenance, and responses to abiotic elements on Earth (Brown, 2014; Mittelbach et al., 2007), which are more relevant to contemporary timeframes. Most hypotheses falling under this umbrella have been reviewed (Fine, 2015; Willig et al., 2003), including population dynamics (e.g. species range sizes or population sizes), resource availability, local dispersal, spatial heterogeneity, and biotic interactions. Due to the wide range of these hypotheses, we will not review each of them here, although many are presented in Table 2.1. Instead, we consider one of the more relevant hypotheses that applies to population richness and genetic diversity: population dynamics.

Ecological framework: Population richness

A prominent hypothesis of population dynamics is Rapoport's rule (Ruggiero & Werenkraut, 2007; Stevens, 1989), the positive correlation of geographical range size with latitude, focusing on the climatic variation that organisms are exposed to and adapted for. Temperate species tend to populate large geographical ranges, so they experience a wide range of climatic variation; tropical species, conversely, are limited to small geographic ranges due to specialization, with some exceptions rule (Ruggiero & Werenkraut, 2007; Stevens, 1989). If a species has a small range, there is less area for populations to become isolated and differentiated, and the species will likely have fewer populations. Additionally, each population may have smaller population sizes as Currie et al. (2004) noted that both population size and density of individuals decrease towards low latitudes. This could interfere with the ability of populations to become established in a new area if they have small suitable ranges with which to disperse. If consistent across taxa, species with large range sizes may, in general, harbour more PopPerSpp, leading to a reverse latitudinal trend for population richness relative to species richness.

The size of a species' range would also influence gene flow and subsequent population differentiation. Greater distances between populations in large-ranged species would make gene

flow more difficult, all else being equal. Thus, according to Rapoport's rule, we would expect populations at or near the edge of a species range to experience lower gene flow. Range-edge populations are more likely to be geographically isolated and more differentiated from neighbouring populations, particularly for large-ranged species (Eckert et al., 2008; Hargreaves & Eckert, 2019; Pelletier & Carstens, 2018; Stevens, 1989; Willi et al., 2018). A likely consequence across a species' range would be more distinct populations at or near range edges, and fewer within the "core" due to increased gene flow (Pelletier & Carstens, 2018). This could result in higher TotPopR in areas where many species range edges overlap extensively, i.e. at low latitudes. Overall, larger range sizes tend to be associated with increased distances between populations, leading to the expectation of increasing PopPerSpp with range size.

Ecological framework: Genetic diversity

A species' range size could also reflect levels of genetic diversity, where small ranges may have lower genetic diversity per species (GenPerSpp) (Fine, 2015). Small population size or density, typical of tropical species, can lead to increased levels of ecological or genetic drift (Fine, 2015; Mittelbach et al., 2007; Siqueira et al., 2020). While genetic drift has not received much empirical support as an explanation for the species richness gradient, it could be more relevant for intraspecific diversity. As patch size or a species range size is correlated with population size (Bernos & Fraser, 2016; Currie et al., 2004), it is a reasonable assumption that species with small ranges have, on average, smaller population sizes. This would lead to increased levels of genetic drift and an increased possibility of inbreeding, and hence lower GenPerSpp. Alternatively, for species that are population rich, different populations could drift in different directions, and show an inflated GenPerSpp collectively across the species. Herein the combined analysis of genetic diversity metrics such as observed heterozygosity and mean number of alleles (MNA) within species (see GenPerSpp Glossary) could provide a better indication of genetic diversity within a species or taxonomic group. MNA responds to inbreeding, population bottlenecks, and genetic drift more quickly than heterozygosity (Allendorf & Luikart, 2009; Allendorf, 1986; Nei et al., 1975) and could indicate which populations are at risk of low genetic diversity. Additionally, heterozygosity can relate to long-term effective population size (see Glossary) (Bazin et al., 2006; Hansson & Westerberg, 2002) and in some instances, adaptation to environmental change (Fraser et al., 2019; Saccheri et al., 1995). Assessing either MNA or heterozygosity alone might mask some of these potential

patterns and thus it is important to distinguish between the two metrics. However, larger geographic ranges may not necessarily correspond to higher levels of GenPerSpp. As previously discussed, many high latitude organisms have also experienced glacial fragmentation that has resulted in a history of small population size and thus lower overall GenPerSpp and TotGenDiv (Galbreath & Cook, 2004; Green et al., 1996; Hewitt, 2000; Provan & Bennett, 2008; Stewart et al., 2010).

A species' range size and dispersal abilities also influence the extent of gene flow between core and edge populations, affecting the maintenance of genetic diversity (Bohonak, 1999; Martinez et al., 2018; Pelletier & Carstens, 2018; Willoughby et al., 2017). For example, having a large geographic range but being mobility-limited, such as in small rodents, reduces the likelihood of gene flow between northern and southern populations strictly because these would be so far apart. Conversely, species with small geographic ranges may have comparatively more gene flow across their range and have fewer genetically distinct populations. Fishes show interesting patterns: large ranges typical of marine or anadromous fishes tend to have higher genetic diversity and lower population differentiation, whereas freshwater fishes with typically limited dispersal capabilities tend to have lower genetic diversity and higher population differentiation (DeWoody & Avise, 2000; Martinez et al., 2018; Medina et al., 2018). Thus, species that are not as capable of dispersing large distances may show stronger latitudinal patterns for both population and genetic diversity, regardless of range size, due to differences in population differentiation across their range (Bohonak, 1999). Populations may become more easily differentiated in large ranges, and as a result, local adaptation may inflate total genetic diversity across all populations (TotGenDiv) within a species' range. Interestingly, however, non-migratory vertebrate species tend to have more genetic diversity than their migratory counterparts, except for birds which show the opposite pattern (Willoughby et al., 2017). This could be an indication that migratory species more frequently encounter fragmentation that causes reductions in genetic diversity, potentially blurring the otherwise expected pattern of increasing genetic diversity with range size.

Ecological framework: Conclusion

Overall, biological differences between taxonomic groups can play a large role in determining population richness and genetic diversity gradients. In general, we expect large-ranged, limited dispersers to show more intraspecific diversity, and small-ranged, capable

dispersers to show lower intraspecific diversity. If intraspecific diversity generally increases with range size, we would expect both higher population richness and genetic diversity at higher latitudes. However, dynamics within particular taxonomic groups could cause population richness and genetic diversity gradients to be much more idiosyncratic than species gradients.

Evolutionary Framework

To further clarify the expectations for diversity gradients, we next consider hypotheses taking an evolutionary approach that focuses on rates of diversification and how these are affected by abiotic and biotic environmental factors (Mittelbach et al., 2007). The premise is simply that the tropics are older, warmer, and have had historically higher diversification rates along with lower extinction rates than temperate latitudes (Mittelbach et al., 2007; Schluter, 2016; Schluter & Pennell, 2017; Stevens, 1989; Weir & Schluter, 2007). Proposed explanations for higher diversification rates in the tropics include: enhanced tropical genetic drift (Fedorov, 1966; Mittelbach et al., 2007); stronger high latitude climate change cycles (Dynesius & Jansson, 2000; Mittelbach et al., 2007); greater geographic extent allowing for diversification across space (Mittelbach et al., 2007; Terborgh, 1973); narrow physiological tolerances in the tropics (Ghalambor, 2006; Janzen, 1967; Mittelbach et al., 2007; Stevens, 1989); temperature effects on evolutionary speed (Mittelbach et al., 2007; Orton et al., 2019; Rohde, 1992); a stronger influence of biotic over abiotic interactions in the tropics; and greater ecological opportunities (Schluter, 2016). Many of these explanations are outlined in Table 2.1 and overlap with discussions under the other two frameworks. The extent of support for these proposed explanations is variable, but whichever factor(s) caused increased tropical speciation rates in the past appear to be shifting in current times (Orton et al., 2019; Schluter, 2016; Schluter & Pennell, 2017; Weir & Schluter, 2007). This shift has consequences for current patterns of biodiversity; as speciation slows at low latitudes and increases at high latitudes the latitudinal gradient in species richness may dissolve as temperate regions "catch up" in species richness. It is also important to consider rates of extinction along with speciation – low latitude extinction rates could increase if climate changes so drastically that species struggle to keep within suitable habitat (Sandel et al., 2011), potentially changing the gradient more quickly.

Evolutionary Framework: Population richness

Temperate clades are seeing an increase in speciation rates, and this is at least in part due to the opening of available habitat (Schluter, 2016; Schluter & Pennell, 2017). Schluter (2016)

describes this as the ecological opportunity hypothesis, wherein areas having more open niches tend to correspond with faster diversification rates. Low latitudes have historically had a wider range of niches and higher rates of speciation than higher latitudes, giving the tropics a "head start" to accumulate species. If this is the case, the tropics could have "maxed out" on speciation rates and population richness; species are now restricted to small ranges due to specialization of niches and have on average fewer populations (PopPerSpp). Note again it is possible that if every species has at least one population, the tropics could have higher TotPopR than temperate regions simply due to having much higher species richness, although this effect could be mediated if high latitude species have much higher PopPerSpp. This is where comparing PopPerSpp and TotPopR is useful. The tropics might have a higher or similar *absolute* number of populations (TotPopR), but higher latitudes – which tend to have larger species ranges (Stevens, 1989) and now increasing speciation rates – would have a higher PopPerSpp (Figure 2.1).

Increasing diversification rates at high latitudes has consequences for PopPerSpp. As climate shifts open historically inhospitable regions in temperate areas, more diversification is facilitated (Schluter, 2016; Schluter & Pennell, 2017; Weir & Schluter, 2007). This diversification process leads to higher PopPerSpp as species begin moving into novel habitats, without completing speciation due to insufficient time. Faster diversification rates should lead to more populations diverging, thus leading to an increase in PopPerSpp, and TotPopR for a given area over time.

Evolutionary Framework: Genetic diversity

Diversification rates can also play a role in a species' ability to adapt and maintain genetic diversity. For example, if a species experiences faster diversification rates at the edge of its range due to strong directional selection pressures, a given population could become locally adapted and may see a drop in genetic diversity relative to other populations (Eckert et al., 2008; Ellegren & Galtier, 2016; Guo, 2012; Hargreaves & Eckert, 2019; Willi et al., 2018). Higher diversification rates could lead to more rapid population differentiation, leading to decreases in fitness should ongoing gene flow occur (Schluter, 2016; Seidel et al., 2008) and encouraging further speciation. Because this process of diversification has likely already occurred at low latitudes, they may remain a hotspot for genetic diversity across species (TotGenDiv). However, we expect patterns of within-species genetic diversity (GenPerSpp) to be sensitive to taxonomy,

particularly for older clades which may have accumulated more genetic diversity, whether such clades originated at high or low latitudes.

Evolutionary Framework: Conclusion

Evolutionary processes such as diversification rates influence intraspecific diversity expectations by affecting the trajectory of populations within species. While PopPerSpp is expected to be highest at high latitudes due to incomplete diversification within species, TotPopR may have similar levels across latitudes regardless due to many populations across a much larger number of species having already diverged or partially diverged over time. Genetic diversity is predicted to be typically higher at low latitudes, especially for TotGenDiv where diversification across species has led to the accumulation of genetic diversity. On the other hand, trends for GenPerSpp are much more variable across taxonomic groups due to different diversification rates and decreases in genetic diversity from niche specialization and/or novel adaptations.

Review Summary: Latitudinal predictions

Clearly, overlap exists between historical, ecological, and evolutionary hypotheses and how these might influence population richness and genetic diversity across latitudes. In many cases, the effects are associated with the younger age of high latitude clades combined with the increase in diversification rates at high latitudes. Following the conceptual considerations above, we can make the following summary for predictions of latitudinal patterns for species richness, population richness, and genetic diversity:

- Species richness is highest at low latitudes.

Due to several factors including geographic history and past diversification rates, the tropics show higher species richness than temperate latitudes.

- Within-species population richness (PopPerSpp) is highest at high latitudes, but among-species population richness (TotPopR) may be more similar across latitudes.

Many of the expectations for population richness gradients stem from the assumption that species clades at high latitudes are generally younger, likely to experience less gene flow among populations in wide-ranged species and have had less time for speciation to occur throughout their usually larger ranges. Therefore, we expect greater PopPerSpp at higher latitudes where species ranges are larger, allowing for more populations across each individual range. Though less clear, more similar levels of TotPopR across latitudes may be expected overall. TotPopR is influenced by the magnitude of the difference in species richness and PopPerSpp between tropical and temperate regions – for example, depending on the taxonomic group, high latitudes may only need a modestly higher PopPerSpp to effectively equalize TotPopR across latitudes.

- Genetic diversity patterns are more variable and have no clear latitudinal gradient across species.

The predictions for latitudinal genetic diversity patterns are more difficult to untangle due to the combined effects of history and current population size/distribution, and perhaps even the limited range/variability of genetic diversity levels (see Leffler et al., 2012). On one hand, because tropical species tend towards smaller geographic ranges, one could envision lower genetic diversity in these groups. On the other hand, tropical species tend to be older and inhabit more stable environments, and so some authors have suggested that genetic diversity could be maintained/accumulated throughout time (Adams & Hadly, 2013; Smith et al., 2017). Complicating expectations further, temperate species have a longer history of fragmentation,

bottlenecks, and founder effects, which all may contribute to a sharp decline in genetic diversity at high latitudes. This glacial history at high latitudes likely plays a large role in structuring genetic diversity patterns, with greater TotGenDiv at low latitudes but perhaps the highest GenPerSpp at intermediate latitudes. For example, species at intermediate latitudes are likely to have more variable clade ages (Schluter, 2016), to have experienced fewer genetic bottlenecks, to have larger range sizes than tropical species (Stevens, 1989), and to have intermediate levels of gene flow across their range. Complicating expectations even further, anthropogenic impacts are highest at intermediate latitudes where most land conversion for agriculture and human population density exist (Cincotta et al., 2000; Ellis & Ramankutty, 2008; Gibbs et al., 2010; Matthews, 1982). While the broad-scale impacts of humans on species genetic diversity are unclear (Millette et al., 2019), they could blur latitudinal patterns if human activities causing habitat loss reduce genetic diversity in regions where high levels of genetic diversity might be otherwise expected (see Ascensão et al., 2016; Cardillo et al., 2004; Cincotta et al., 2000). Collectively, a number of factors operating differently along the latitudinal gradient appear to have varying consequences for genetic diversity both among and within species. Thus, genetic diversity is not expected to have a clear latitudinal gradient relative to species or population richness.

New Analyses Drawing from Review

Intraspecific Diversity and Range Size: Hypotheses and Predictions

Hypotheses describing latitudinal species richness have direct links to both population richness and genetic diversity. These links form the foundation upon which we further elaborate on population richness and genetic diversity expectations relative to a species' range size. We outline and test three novel hypotheses to explain latitudinal trends in intraspecific diversity. Data used to test the hypotheses below were obtained from the *MacroPopGen* database (Lawrence et al., 2019), a georeferenced dataset of microsatellite genetic diversity for almost 900 vertebrate species and over 9000 genetically-distinct populations across the Americas (see Appendix 2 Supplementary Methods for details). Populations were designated as genetically distinct within *MacroPopGen* using a commonly applied, operational definition of a population (reviewed in Waples & Gaggiotti, 2006); population richness in the database (PopPerSpp or TotPopR) represented only populations that had been sampled with microsatellite loci. As such, some observed patterns may not be as strong as otherwise expected perhaps due to sampling bias

of populations. We strive to acknowledge this in our discussion of results below. Range size data came from IUCN and BirdLife International (BirdLife International, 2017; IUCN, 2016) and Meiri et al. (2017). As an indication of sampling intensity across the Americas, we mapped sampled species richness and population richness as well (Figure 2.2).

H1: Geographic Distribution Hypothesis

We term the first hypothesis the **Geographic Distribution Hypothesis,** which posits that a positive relationship exists between a species' geographic range size and its population richness. PopPerSpp should therefore increase with increasing latitude because temperate species ranges are typically larger than in tropical species. Broadly speaking, we also expect different vertebrate groups to show different strengths for this pattern because of inherent differences between dispersal capabilities and environments inhabited (Sandel et al., 2011). For example, relative to other vertebrates, freshwater and anadromous fish species may show greater TotPopR and/or greater PopPerSpp across their ranges due to the easily fragmented nature of aquatic freshwater habitats through natural barriers (Tatarenkov et al., 2010; Underwood et al., 2016; Wofford et al., 2005), dams (Roberts et al., 2013; Underwood et al., 2016; Wofford et al., 2005), and the connectivity between fluvial environments and lakes (Hébert et al., 2000; Underwood et al., 2016). Amphibians and reptiles (collectively, herptiles) may also show strong patterns between range size and PopPerSpp due to their generally limited ability for dispersal (Araújo et al., 2005; Green et al., 1996; Medina et al., 2018; Sandel et al., 2011) that leads to high subpopulation differentiation across a given species range. Birds and some mammals, conversely, tend to have greater dispersal capabilities than herptiles and some freshwater fishes (Araújo et al., 2005; Medina et al., 2018; Munguía et al., 2008; Servín et al., 2003; Sutherland et al., 2000). Thus, we expect these groups will have a lower TotPopR than fishes and herptiles due to homogenization of population structure, but more variable PopPerSpp depending on the specific species' dispersal ability.

To test the Geographic Distribution Hypothesis, we used a generalized linear model fitted with a gamma distribution where the number of populations for a given species (i.e. PopPerSpp, Table A2.1) was our dependent variable (n=567 species, 5172 populations; see Appendix 2 Supplementary Methods), while the natural logarithm of range size (km^2), latitudinal extent (decimal degrees), and taxonomic class (amphibian, bird, anadromous or freshwater fish, mammal, reptile) were fixed effects. PopPerSpp and range size for each species can be found in

Table A2.1. We also tested the linear relationships between range size and PopPerSpp for each taxonomic group (Figure 2.3a, b). These linear relationships were significant for all taxonomic groups combined ($p=<0.001$, $R^2=0.03$; Figure 2.3A), and fish separately ($p=0.003$, $R^2=0.07$), although they did not explain much variation in the data. There were no significant relationships between range size and PopPerSpp within amphibians, birds, mammals, or reptiles (Figure 2.3b). For the mixed model, both the natural logarithm of range size and the latitudinal extent were significant ($p=0.022$, <0.001 respectively, Figure 2.3a, Table A2.2). The discrepancy across taxonomic groups could be due to a lack of thorough sampling across species ranges. When assessing taxonomic groups separately, amphibians, reptiles, and birds tended to have data that were sparsely sampled across species ranges compared to other species, especially fishes. To account for this, we recommend future studies estimate the area represented by each population so that the percent of the species range that has been sampled can be included.

H2: Overlapping Range Hypothesis

Areas that have extensive species range overlap may have lower PopPerSpp due to higher competition, smaller range sizes, etc. (Kennedy et al., 2018; Pelletier & Carstens, 2018). For instance, lower latitudes are more likely to have high species richness, moderate TotPopR, and lower PopPerSpp. Species richness is to the point of oversaturation at low latitudes (Schluter, 2016), and tropical species are generally restricted to smaller ranges (Currie et al., 2004; Stevens, 1989). The combination of small range sizes and fewer open niches would lower the number of intraspecific populations able to differentiate (or speciate with time), because fewer opportunities for local adaptation or population differentiation are available to occur across a species range (Schluter & Pennell, 2017; Weir & Schluter, 2007). Collectively, one might expect TotPopR to increase as species richness increases, but PopPerSpp to decrease with increasing species richness (**Overlapping Range Hypothesis)**.

To test the Overlapping Range Hypothesis, we calculated the species richness in 500 km^2 equal area grid cells generated in the Behrmann projection across the American continents and correlated it using a linear model with both the absolute population richness (TotPopR) and the number of sampled populations of each species (PopPerSpp) (Figure 2.3c, Table A2.3; see Appendix 2 Supplementary Methods). The relationship between the number of species in an area was positively correlated with TotPopR ($p<0.001$, $R^2=0.75$; Figure 2.3c, Table A2.3), and PopPerSpp ($p<0.001$, $R^2=0.22$, Figure 2.3d). While our analysis does not show the expected

trend for PopPerSpp, this may again be due to incomplete population sampling for each species in the dataset. We note that the slopes of the two relationships (4.74 and 0.08 for TotPopR and PopPerSpp, respectively; Table A2.3) do provide some indication that the trends between the two population richness metrics are different and that different mechanisms may underpin them. Perhaps as species richness increases, PopPerSpp does not increase at a corresponding rate, indicating that species richness has some impact on the capacity for evolution of population richness within a species. If the actual number of populations within a species range was known, we expect this positive relationship between PopPerSpp and species richness to break down further, showing the negative relationship as predicted, or a very weak relationship.

H3: Range-Restricted Gene Hypothesis

If species range size influences population size and gene flow between populations (Currie et al., 2004; Fine, 2015), and range size is also correlated with PopPerSpp (H2), then genetic diversity will be more strongly associated with range size than with latitude (**Range-Restricted Gene Hypothesis) –** although some latitudinal patterns may occur as a result of this association. Previous studies have found latitudinal trends for genetic diversity, where higher alpha and beta genetic diversity (equivalent to GenPerSpp and TotGenDiv, respectively) were observed at low latitudes (Adams & Hadly, 2013; Miraldo et al., 2016; Schluter & Pennell, 2017). Even when spatial autocorrelation (Gratton et al., 2017a), number of DNA sequences, and species identity (GenPerSpp) (Schluter & Pennell, 2017) were accounted for, authors found a latitudinal gradient in genetic diversity – although the slope of the relationship was very small (e.g. -0.002, Appendix 2 Supplementary Methods; Schluter & Pennell, 2017). However, these data were based on mitochondrial genetic diversity (mtDNA) rather than nuclear DNA. mtDNA may not be selectively neutral (Bazin et al., 2006) which is important for standardized comparisons across species and populations. Moreover, mtDNA may not reflect genetic variation in the nuclear genome which is integral for adaptation to environmental change (Ballard & Whitlock, 2004; Bazin et al., 2006; Ghalambor et al., 2007; Hurst & Jiggins, 2005; Sgrò et al., 2011). Conversely, microsatellite nuclear DNA variation can be a reasonable metric of genome-wide variation, and the polymorphic nature of microsatellite loci is able to better resolve population structure at fine scales (Angers & Bernatchez, 1998; Jarne & Lagoda, 1996; Väli et al., 2008). Microsatellite-based estimates of GenPerSpp may show a weaker latitudinal pattern than the TotGenDiv metric adopted in past mtDNA studies (e.g. Miraldo et al., 2016). Although non-neutrality has been

observed in some studies involving nuclear microsatellite loci (Ranathunge et al., 2018; Selkoe & Toonen, 2006; Wiehe, 1998), this does not appear to be widespread in *MacroPopGen* (Lawrence et al., 2019; see also Selkoe & Toonen, 2006).

We tested the Range-Restricted Gene Hypothesis by using generalized linear mixed models and model selection where one model was constructed for each population-level genetic diversity metric as the dependent variable (observed heterozygosity, Ho, and mean number of alleles, MNA); taxonomic identity was accounted for with random effects. Fixed effects for both models included range size, latitudinal extent, Ho or MNA (i.e. Ho for MNA, MNA for Ho), and other study-specific metrics (for details see Appendix 2 Supplementary Methods). All population-specific data can be found in the *MacroPopGen* database (Lawrence et al., 2019). After model selection, the mixed model for heterozygosity included the interaction between MNA and the number of microsatellite loci as well as taxonomic class, and the random effects for study, species, and family (Table A2.2):

$$Ho \sim MNA * Microsatellite\ loci + MNA{:}Class + (1|Study) + (1|Species) + (1|Family)$$

The model for MNA only included interactions between Ho and number of microsatellite loci, as well as taxonomic class, and the random effects for study, species, and family:

$$MNA \sim Ho * Microsatellite + Ho{:}Class + (1|Study) + (1|Species) + (1|Family)$$

The retention of Ho, MNA, and the number of microsatellite loci in the models is not entirely surprising and indicates that these factors are more associated with genetic diversity than range size or latitudinal extent, although this effect varies according to taxonomic grouping (Figure 2.3f). While these measures of genetic variation are sometimes (weakly) correlated (Comps et al., 2001), the two metrics still indicate differences in population processes, as we discussed in the ecological framework, where decreases in MNA do not always correspond with decreases in Ho (Allendorf, 1986). Additionally, we used variance inflation factors to test for collinearity between variables and found no evidence for any statistically significant collinearity (see Appendix 2 Supplementary Methods). Thus, we wanted to include both metrics in model selection to test how the effects of range size would compare to the effect of each metric on each other. Indeed, when we tested a model that only included range size and latitudinal extent, only latitudinal extent (not range size) was significant. Figure 2.3e demonstrates this lack of a significance for range size, while a positive relationship is found in Figure 2.3f (note only MNA

is shown but results were similar for Ho). Varying relationships among taxa were also supported by the different slopes of linear relationships shown in Figure 2.3f. The inclusion of Ho, MNA, and number of microsatellite loci could indicate that genetic diversity metrics are sensitive to the number of alleles present within a population, where more alleles and loci being present increases the likelihood of being heterozygous and vice-versa (Figure 2.3f). Together, the results of these models suggest that genetic diversity is not particularly influenced by range size or the latitudinal breadth of a species' range.

New Analysis Summary

We proposed three hypotheses relating range size with population richness and genetic diversity, taking inspiration from a synthesis of species richness theories. However, we found minimal support for our hypotheses, highlighting the idiosyncrasies in intraspecific diversity patterns previously found between taxonomic groups (DeWoody & Avise, 2000; Martinez et al., 2018; Medina et al., 2018; Millette et al., 2019; Willoughby et al., 2017). While we have not explicitly considered taxa-specific traits (e.g. migratory behaviour, age at maturity, body size), the differences found between taxonomic groups may indicate that such data could further explain trends in intraspecific diversity.

Overall, we found marginal support for two of our three hypotheses. This is likely due to a number of factors, one being that accurate data for population richness is under-developed, as many populations are under-sampled. Additionally, large range sizes may not necessarily correspond with more genetic diversity. For example, animals with larger body sizes may have large range sizes but relatively lower population sizes simply because they need more space per individual or per population. This could mean that a large-ranged animal may still have fewer individuals per population, resulting in fewer populations overall and potentially lower genetic diversity. Future analyses should consider factors such as body size in conjunction with range size to better explain variation in genetic diversity.

Of the taxa examined in our analyses, fishes had the strongest, most significant positive relationships between range size, population richness (Figure 2.3b), and the genetic diversity metrics (Figure 2.3e-f). This latter relationship was particularly steep for anadromous fish, consistent with previous works that have found that anadromous fishes tend to have higher genetic diversity than freshwater fishes (DeWoody & Avise, 2000; Martinez et al., 2018). All other taxonomic groups did not show significant relationships between range size and population

richness. This is likely due to incomplete sampling across species ranges relative to many of the fish species in this database, leading to an underrepresentation of population richness (e.g. average PopPerSpp for anadromous fish = 109, amphibians = 20, Table A2.4). This underrepresentation could also be affecting our results for range size with genetic diversity – perhaps the populations that were sampled from species with large ranges happened to be lower (or higher) in genetic diversity then otherwise expected. This is a sort of sampling bias that could be corrected if we had complete data on populations for a few large- and small-ranged species to investigate further.

Our results contribute to the idea that disentangling intraspecific diversity patterns can be much more complicated than species richness as many factors require simultaneous consideration (see (Blanchet et al., 2017; Marchesini et al., 2018; Martinez et al., 2018; Medina et al., 2018; Millette et al., 2019; Paz-Vinas et al., 2018; Willoughby et al., 2017). The limited scope in the scale of genetic diversity, and perhaps the minimum and maximum degree of genetic diversity required for viable populations (e.g. 0 to 1 for heterozygosity; Ellegren & Galtier, 2016; Leffler et al., 2012) could also have a major impact on the detection of broad scale patterns. The magnitude of differences in genetic diversity across a latitudinal gradient would additionally not be as large as seen in the species richness gradient. For example, there are at least ~143% more species in tropical relative to temperate countries (e.g. Brazil: ~170,000-210,000 known species, Canada: ~70,000 known species; Canadian Endangered Species Conservation Council, 2001; Lewinsohn & Prado, 2005). In contrast, Miraldo et al. (2016) only found 27% more total mitochondrial genetic diversity in the tropics, summed across terrestrial mammals and amphibians (i.e. TotGenR). The influence of these factors could explain why our analyses of intraspecific diversity do not show as clear a pattern as species richness, warranting further exploration in tandem with environmental properties, anthropogenic factors, and species- or population-specific functional/life history traits.

Overall Conclusion

Although there has been some recent support for latitudinal gradients in intraspecific diversity (Adams & Hadly, 2013; Gratton, et al., 2017b; Martin & Mckay, 2004; Millette et al., 2019; Miraldo et al., 2016; Schluter & Pennell, 2017), no study has generated latitudinal expectations for both population richness and genetic diversity by drawing from species-level literature – indeed there is an admitted lack of theoretical foundation (Millette et al., 2019). We demonstrate that the distinct latitudinal patterns found in species richness are much more complicated at the intraspecific level. Our synthesis suggests that, species richness, population richness, and genetic diversity within species will be uncoupled in many instances due to a combination of historical and contemporary factors. Factors such as range size (i.e. Rapoport's rule) and biological differences between and within taxonomic groups appear to play a larger role in moderating population richness and genetic diversity gradients. These inferences have implications for the fundamental understanding of the species richness gradient and for biodiversity conservation, as they shed light on what may drive changes to species distributions and species adaptability at different latitudes in the future.

Our focus on population richness and genetic diversity was complemented by the usage of microsatellite data obtained from the *MacroPopGen* database (Lawrence et al., 2019). This database does not include adaptive, functional, or phylogenetic diversity, as standardized phylogenies below the species level, for example, do not exist for most populations studied with nuclear DNA. We expect future analyses that include these other aspects of intraspecific diversity will only clarify the patterns described here further and perhaps account for some of the noise in the data. As mentioned, the increased sampling of populations within species would also be useful to test latitudinal gradient theories with more certainty. While the relationships presented here may not be very strong, the results are likely to be strongly affected by lack of full sampling within species ranges. As technology advances, results collated from genome-wide assessments will also help refine our hypotheses further and more fully represent genetic diversity and population richness.

While we have largely focused our discussion on the theories for latitudinal patterns in biodiversity, our results also have conservation implications. As larger range sizes are typically associated with greater population richness and genetic diversity, species with small ranges are likely to be at greater risk (Fine, 2015), whereas population rich species are likely to be less at

risk to changing conditions. This is reminiscent of the theory of island biogeography where just as smaller areas are associated with fewer species, so are small areas generally associated with fewer genetically distinct populations. Our conclusion may not seem novel, but our study is the first to fully discuss this with respect to populations as a quantifiable unit. These results may have consequences for conservation management where only assessing an area's species richness may not capture the extent of biodiversity in that area. Assessing population richness for each species and their genetic diversity may give a better indication of ecosystem health and the species' ability to remain intact (Martinez et al., 2018; Paz-Vinas et al., 2018).

We urge for a more holistic approach in biodiversity science and conservation where all aspects of biodiversity are considered together (ecosystem diversity, species diversity, functional diversity, intraspecific diversity), especially as future technology refines and improves our understanding of intraspecific diversity even further.

Glossary

Genetic diversity: Defined in this review as neutral genetic diversity within a population or species. Often assessed with microsatellite data as observed heterozygosity or allelic diversity/mean number of alleles per locus (MNA).

Observed heterozygosity: A measure of genetic diversity representing the percentage of heterozygous loci of individuals within a population. Declines in isolated populations as effective population size decreases (Coltman & Slate, 2003; Frankham, 1996; Frankham et al., 2002).

MNA: Mean number of alleles – a measure of genetic diversity where the number of alleles are counted for each locus and averaged across individuals in a population. Shows a more rapid response than heterozygosity in decline with effective population size decreases (Coltman & Slate, 2003; Frankham, 1996; Frankham et al., 2002).

Population richness: In general, the number of genetically distinct populations – either across all species (TotPopR), within a species (PopPerSpp), or averaged across many species in an area (AvgPopSpp).

TotPopR: Total population richness – the *total* number of populations within a given area across species, e.g. (Hughes et al., 1997).

PopPerSpp: Populations per species – refers to how many distinct populations one species has across its range or within an area. For example, an area with many populations would be considered "population rich" according to TotPopR but might be classified as "population poor" by PopPerSpp if each species is represented by only a small number of populations (Figure 2.1). TotPopR and PopPerSpp have different implications. Analyzing both TotPopR and PopPerSpp outlines more clearly which species or taxonomic groups may have more populations, and gains an understanding of the genetic history, along with the vulnerability or level of endemism characterising a certain species or taxonomic group.

AvgPopSpp: Average number of populations per species within a given area. Calculated by first determining the PopPerSpp for each species in an area, and then averaging these values for all species in the area.

TotGenDiv: Total genetic diversity –reported in previous large-scale syntheses as a sum or mean of genetic diversity across all species and their populations within a given area (Gratton et al., 2017a; Miraldo et al., 2016; Willoughby et al., 2015). Does not reflect the genetic diversity *between* species, and masks idiosyncrasies between lower levels of taxonomic groups, identifiable when assessed in individual species, as in GenPerSpp (Adams & Hadly, 2013; Martin & McKay, 2004). For simplicity in our discussions, we define TotGenDiv as the sum of neutral genetic diversity across all species and their populations. Note that an additional measure to analyze TotGenDiv patterns would be to assess the variance of genetic diversity across species within an area. This would identify regions with abnormal levels of variability in genetic diversity, indicating that the TotGenDiv of the area may be skewed by a certain species. Alternatively, taking the weighted average of genetic diversity across species (e.g. Millette et al., 2019) and populations in an area would account for differences among sample size and/or number of populations in the area (Schluter & Pennell, 2017). Then, assessing sum, variance, and mean genetic diversity together for broad scale analyses yields more refined insights than simply totalling across species.

GenPerSpp: Refers to the sum of neutral genetic diversity within a single species across all its populations in an area – i.e. species-specific genetic diversity. Provides a more realistic representation of genetic diversity, allows for idiosyncrasies between groups to be identified, and avoids over simplification at large scales.

Effective population size: Represents the number of individuals in a population that are contributing to the next generation (Wright, 1931); also gives an indication of how quickly loss of genetic diversity occurs in a finite-sized population through random genetic drift (Belmar-Lucero et al., 2012; Frankham et al., 2002).

Data Accessibility Statement

The data used in analyses are available on figshare: https://doi.org/10.6084/m9.figshare.7207514.v2 (Lawrence et al. 2019).

Tables

Table 2.1. Latitudinal theories, which of the three frameworks they fall under, and their predictions for species richness, population richness, and genetic diversity. Definitions for population richness and genetic diversity refer to their general definitions unless otherwise specified. Hist = Historical; Ecol = Ecological; Evol = Evolutionary; GD = genetic diversity; GenPerSpp = genetic diversity per species; TotGenDiv = total genetic diversity across species; PopPerSpp = populations per species; TotPopR = total population richness for a given area.

Frame-work	**Theory Description & Explanation**	**Predictions**		
		Species	**Population**	**Genetic**
Hist	Time and area hypothesis: Tropics are older, historically larger geographically, and climatically stable, allowing for more diversification to occur over time *Explanation:* Older low latitude communities have had more time and area for mutations to accumulate as well as populations within species to differentiate into new species, causing fewer populations per species, but perhaps retaining a similar TotPopR to high latitudes (barring nuances as discussed in text) *References:* Wallace, 1878; Pianka, 1966; Mittelbach et al., 2007	**Low Latitudes:** higher **High Latitudes:** lower	**Low Latitudes:** lower PopPerSpp; similar TotPopR **High Latitudes:** higher PopPerSpp; similar TotPopR	**Low Latitudes:** higher **High Latitudes:** lower
Hist	Tropical/Phylogenetic niche conservatism*:* Species that originate in a region, whether tropical or temperate, are	**Low Latitudes:** higher	**Low Latitudes:** lower PopPerSpp	**Low Latitudes:** higher

	more likely to stay within that climate, but older clades may diversify outwards through niche evolution *Explanation:* The typically older age of low latitude species indicates they will have fewer populations but more GD (see time-area hypothesis) as they have remained in tropical environments longer, diversifying over time *References:* Wiens & Donoghue, 2004	**High Latitudes:** lower	**High Latitudes:** higher PopPerSpp	**High Latitudes:** lower
Hist Ecol	Heterogeneous area: Increased ecological heterogeneity in large areas leads to fragmentation and speciation across species' range. Related to time and area hypotheses but more focused on the notion of larger areas having more heterogeneous habitat *Explanation:* Increased fragmentation at high latitudes in large-ranged species allows for populations to differentiate, but not fully enough to lead to new species; larger areas maintain GD within a species due to gene flow between populations *References:* Terborgh, 1973; Rosenzweig, 1995; Mittelbach et al., 2007	**Low Latitudes:** higher **High Latitudes:** lower	**Low Latitudes:** lower PopPerSpp **High Latitudes:** higher PopPerSpp	**Low Latitudes:** higher **High Latitudes:** lower
Ecol	Species range size (Rapoport's rule)*:* Low latitude species experience smaller ranges in climatic variation, therefore more specialization and smaller range sizes; the opposite	**Low Latitudes:** higher **High Latitudes:**	**Low Latitudes:** lower PopPerSpp **High Latitudes:**	**Low Latitudes:** higher TotGenDiv lower GenPerSpp

	phenomenon occurs at high latitudes *Explanation:* More specialization leads to more species with smaller range sizes, fewer populations per species, and more GD across many species, although may result in lower GD within a species due to specialization *References:* Jansen, 1967; Stevens, 1989; Mittelbach et al., 2007	lower	higher PopPerSpp	**High Latitudes:** lower TotGenDiv higher GenPerSpp
Ecol	Genetic drift: Low latitude populations are smaller and tend to experience more genetic drift that differentiates populations and species *Explanation:* More genetic drift leads to speciation and more species at low latitudes, but less distinct populations at high latitudes so PopPerSpp is maintained at higher than low latitudes. GD is higher across species at low latitudes due to different populations accumulating different alleles, but perhaps lower GenPerSpp if alleles are lost through drift *References:* Dynesius & Jansson, 2000; Mittelbach et al., 2007	**Low Latitudes:** higher **High Latitudes:** lower	**Low Latitudes:** lower PopPerSpp **High Latitudes:** higher PopPerSpp	**Low Latitudes:** higher TotGenDiv lower GenPerSpp **High Latitudes:** lower TotGenDiv higher GenPerSpp
Ecol	Energy-diversity hypothesis: Regions of high primary productivity should support more individuals, therefore increased likelihood of more species	**Low Latitudes:** higher **High Latitudes:**	**Low Latitudes:** lower PopPerSpp **High Latitudes:**	**Low Latitudes:** higher **High Latitudes:**

	Explanation: Higher productivity at low latitudes leads to more individuals and more species; but more individuals in general leads to smaller population sizes per species and increased risk of inbreeding for areas with high species richness, affecting both metrics of GD *References:* Pianka, 1966; Currie et al., 2004; Storch et al., 2005	lower	higher PopPerSpp	lower
Ecol	Biotic interactions*:* Biotic interactions are stronger and represent a greater fraction of natural selection for low latitude species; abiotic interactions exert stronger evolutionary forces for higher latitude species *Explanation:* More speciation at low latitudes as biotic interactions drive specialization; general adaptations at high latitudes from abiotic factors maintains gene flow among populations within species, elevating PopPerSpp; may lead to similar levels of TotPopR as different factors drive population richness; specialized adaptations at low latitudes decrease within species GD, but increase GD across many species *References:* Pianka, 1966; Currie et al., 2004; Mittelbach et al., 2007	**Low Latitudes:** higher **High Latitudes:** lower	**Low Latitudes:** lower PopPerSpp similar TotPopR **High Latitudes:** higher PopPerSpp similar TotPopR	**Low Latitudes:** higher TotGenDiv lower GenPerPop **High Latitudes:** lower TotGenDiv higher GenPerPop
Evol	Diversification rates: Diversification rates were historically	**Low Latitudes:**	**Low Latitudes:**	**Low Latitudes:**

	faster at low latitudes, now are becoming faster at higher latitudes; but there are still elevated extinction rates at high latitudes relative to lower latitudes *Explanation:* Low latitudes had a head start with higher diversification rates and lower extinction rates so there is higher species richness and genetic diversity at low latitudes; as high latitudes are experiencing increasing diversification rates, populations are still undergoing differentiation so higher PopPerSpp at high latitudes, but more similar TotPopR across latitudes since many populations already established among species at low latitudes, while many are still differentiating at high latitudes *References:* Weir & Schluter, 2007; Schluter, 2016	higher **High Latitudes:** lower	lower PopPerSpp similar TotPopR **High Latitudes:** higher PopPerSpp similar TotPopR	higher **High Latitudes:** lower
Evol	Evolutionary speed*:* Higher temperatures lead to higher mutation rates, therefore increase genetic divergence (may only apply to ectotherms) *Explanation:* Higher temperatures at low latitudes result in more mutations leading to speciation, therefore less populations per species, but more genetic diversity across species due to accelerated mutation rates across different populations	**Low Latitudes:** higher **High Latitudes:** lower	**Low Latitudes:** lower PopPerSpp **High Latitudes:** higher PopPerSpp	**Low Latitudes:** higher TotGenDiv **High Latitudes:** lower TotGenDiv

	References: Rohde, 1992; Mittelbach et al., 2007; Schluter, 2016			
Evol	Climate change: Milankovitch cycles stronger at high latitudes, thus high latitude species have better dispersal and less speciation than low latitude species *Explanation:* Less dispersal and mixing at low latitudes leads to populations differentiating more in tropics leading to fewer populations at low latitudes. GD maintained within species at high latitudes due to more gene flow *References:* Pianka, 1966; Dynesius & Jansson, 2000; Mittelbach et al., 2007	**Low Latitudes:** higher **High Latitudes:** lower	**Low Latitudes:** lower PopPerSpp **High Latitudes:** higher PopPerSpp	**Low Latitudes:** lower GenPerSpp **High Latitudes:** higher GenPerSpp
Evol	Ecological opportunity hypothesis: Higher speciation rates due to more ecological niches stemming from higher solar energy and annual productivity, reduced temperature seasonality, or stronger biotic interactions at low latitudes *Explanation:* Many niches already filled at low latitudes from speciation, whereas in higher latitudes more niches are becoming open, thus populations have begun to differentiate – but not fully; lower latitudes have accumulated more GD across many niches, but less within a given species *References:* Schluter, 2016	**Low Latitudes:** higher **High Latitudes:** lower	**Low Latitudes:** lower **High Latitudes:** higher	**Low Latitudes:** higher TotGenDiv lower GenPerSpp **High Latitudes:** lower TotGenDiv

Figures

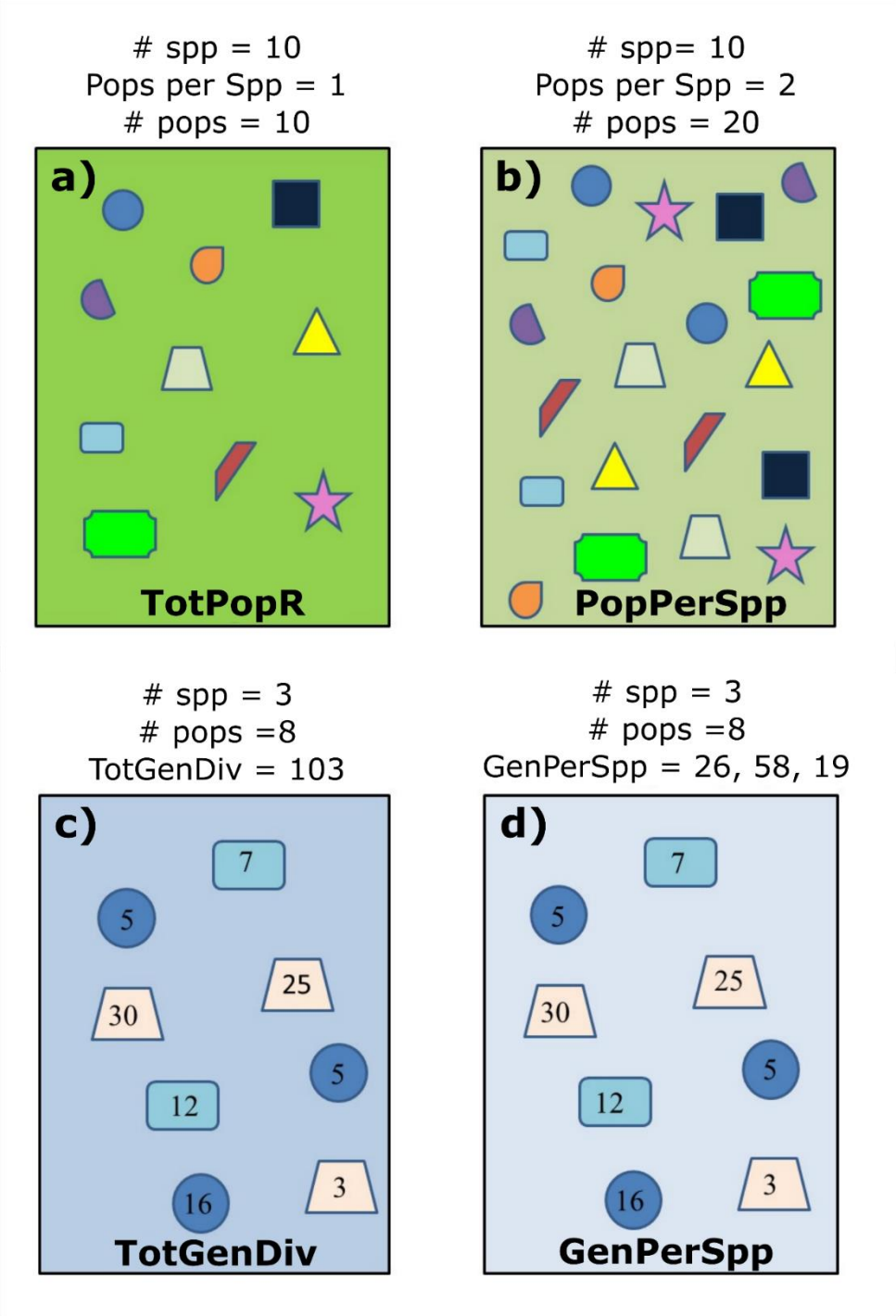

Figure 2.1. Demonstration that intraspecific diversity can provide insight into the biodiversity of an area rather than simply looking at the species ("spp") richness. Two areas that have the same number of species may not have the same number of populations (a and b) or genetic diversity (c

and d). If an area (a) has fewer populations ("pops") per species (PopPerSpp) than another area with the same number of species (b) then that area has less population richness, even though species richness is the same. Likewise, if genetic diversity, given as values of MNA here, is summed across all the species and populations in an area (TotGenDiv, c), this value masks the nuances of genetic diversity of the species present (GenPerSpp). When each individual species' total genetic diversity is considered, nuances of the genetic diversity in an area are more apparent.

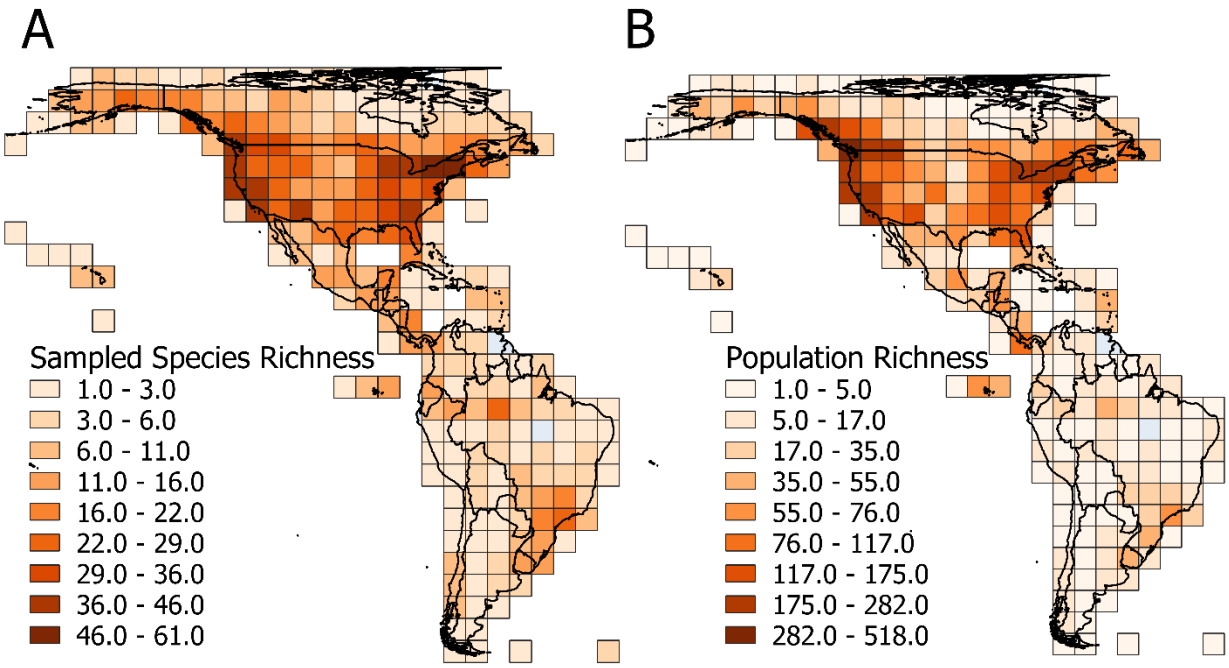

Figure 2.2. A) Number of vertebrate species sampled in each 500 x 500km^2 grid cell. B) Number of genetically distinct populations across vertebrate species in each grid cell. Data obtained from *MacroPopGen* database (Lawrence et al., 2019) and projected with the World Behrmann projection.

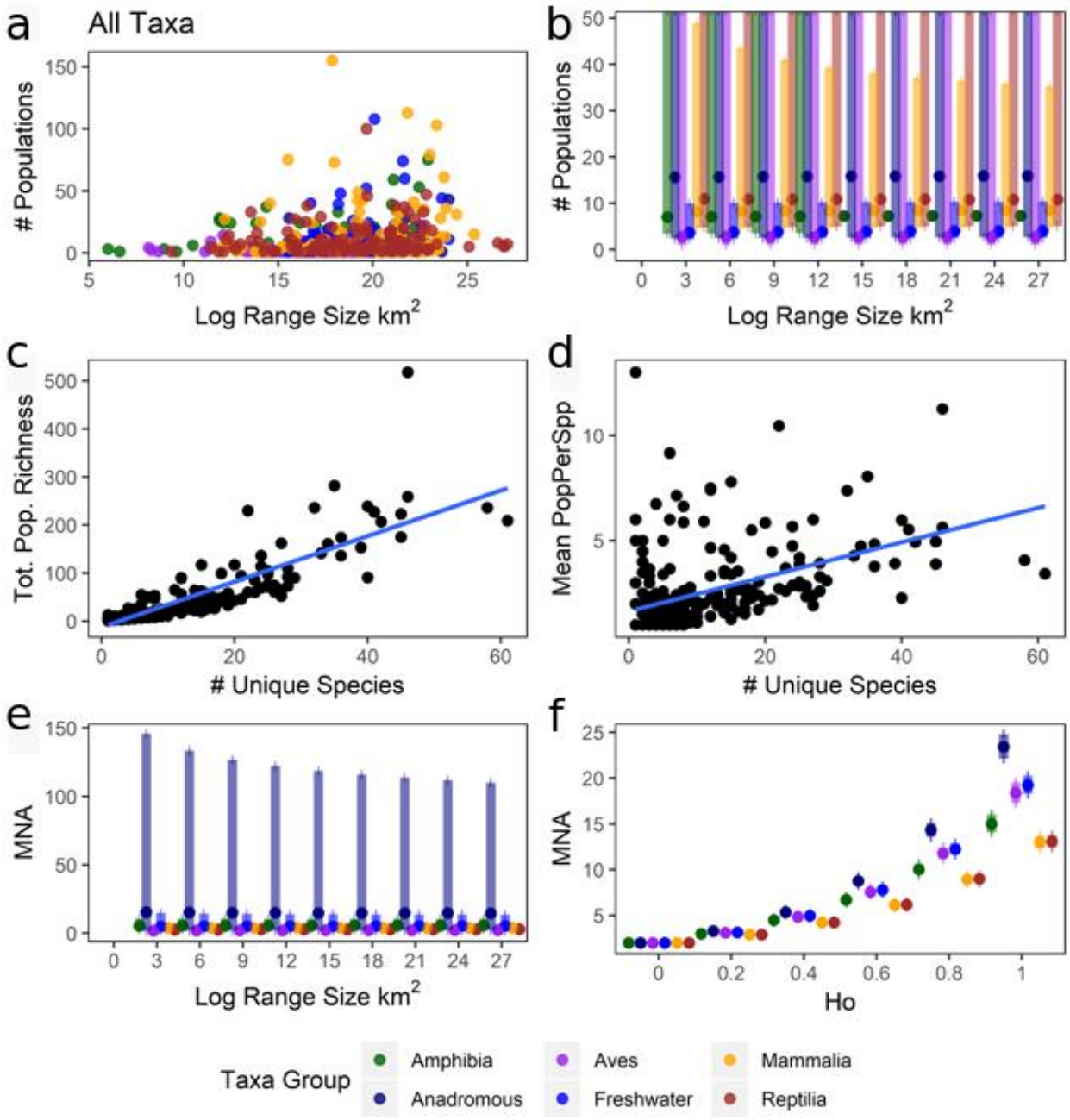

Figure 2.3. Results for testing the Geographic Distribution Hypothesis (a, b), the Overlapping Range Hypothesis (c, d), and the Range Restricted Gene Hypothesis (e, f). A) Log of species range size and the number of genetically distinct populations within a species (PopPerSpp) for all taxonomic groups. B) Linear prediction estimates from a GLMM for the relationship between range size and number of populations for each taxonomic group. Error bars represent upper and lower confidence intervals. C-D) The number of unique species within grid cells (n=250) of an area of 500km^2 (x axis) and C) the total number of populations (Tot .Pop. Richness, R^2=0.75,

p<0.001) or D) the average number of populations within each species (PopPerSpp; $R^2$0.22, p<0.001) for each grid cell. Solid line represents linear regression between the two variables. E-F) Linear prediction estimates from a GLMM of the relationship between E) log of range size and genetic diversity, measured as mean number of alleles (MNA) or F) observed heterozygosity (Ho) and MNA. Error bars represent upper and lower confidence intervals. Results for the other genetic diversity metric, observed heterozygosity (Ho), not shown as relationships were very similar as MNA.

Chapter 3: Weak latitudinal and environmental influences on vertebrate population genetic diversity across the Americas

Submitted June 15, 2020 as:

Lawrence, E.R., and Fraser, D.J. (in review). Weak latitudinal and environmental influences on vertebrate population genetic diversity across the Americas. *Nature Communications.*

Abstract

Species diversity gradients are well established, but the latitudinal distribution of population-specific genetic diversity (PGD) remains unstudied despite vast quantities of genetic data now available. We tested alternative predictions for latitudinal or environmental gradients in PGD (e.g. negative gradients mediated by environmental variables), while accounting for variation among and within taxa. Using nuclear DNA data from ~900 vertebrate species, we found weak latitudinal and variable environmental influences on PGD that were taxa-dependent across the Americas. We suggest the weak gradient is partly due to opposing processes that diminish patterns across latitudes; similarly, large-scale genetic gradients can be flattened when assessing across species versus within species. While species diversity follows a negative latitudinal gradient, PGD does not appear to follow the same pattern, suggesting different processes structure large-scale patterns in vertebrate PGD. Our results indicate that conservation efforts targeting high species diversity regions may not capture high genetic diversity regions.

Keywords: genetic diversity, latitudinal gradient, vertebrate

Introduction

The latitudinal gradient in species diversity is one of the most studied phenomena in ecology and biogeography. Recent work is striving to understand whether this gradient holds for other aspects of biodiversity, such as functional, phylogenetic, or genetic diversity. Although the species diversity gradient is clear, intraspecific diversity patterns and expectations are less apparent and either differ from theoretical predictions or vary across taxa (Buckley et al., 2010; Lamanna et al., 2014; Lawrence & Fraser, 2020; Usinowicz et al., 2017). Functional diversity appears to be lower than otherwise expected in the tropics, whereas phylogenetic diversity may follow species diversity more closely (Huang et al., 2012; Lamanna et al., 2014; Safi et al., 2011; Usinowicz et al., 2017). Conversely, genetic diversity gradients remain poorly understood. While there are hints of latitudinal gradients in species communities (Adams & Hadly, 2013; Manel et al., 2020; Millette et al., 2019; Miraldo et al., 2016), the incorporation of mechanisms to structure these gradients is lacking.

Herein we focus on the mechanistic factors that may structure contemporary patterns of genetic diversity within natural populations of individual species, i.e. population genetic diversity (PGD), across broad geographic scales. When scaling up to view PGD at broad scales this definition allows us to control for differences among and within taxonomic levels (Lawrence & Fraser, 2020). Additionally, patterns in PGD are often sensitive to mechanisms acting at relatively recent timescales, whereas species diversity patterns are likely more affected by deep time processes. Accordingly, we focus our discussions concerning the mechanisms for PGD on more recent processes.

A useful point of departure for considering mechanisms underlying PGD gradients is to consider processes underlying species gradient theories through the lens of PGD. Such processes may include time for tropical species to diversify and accumulate, ecological limits including range area and niche specialization, and increased diversification rates due to abiotic factors (Lawrence & Fraser, 2020; Pontarp et al., 2019). Whether these mechanisms hold for genetic diversity in general is uncertain, as previous empirical works have been largely exploratory, often without considering underlying mechanisms *a priori* (Millette et al., 2019; Miraldo et al., 2016). If species and genetic diversity are distributed the same way, then similar underlying processes may structure these two aspects of biodiversity (e.g. environment-regulating

evolution). To begin, we discuss mechanisms that may result in three distinct patterns of PGD: a 'negative' latitudinal gradient, a 'positive' gradient, or no gradient at all.

Decreasing PGD away from the equator may be the most expected for several reasons, including temporal, geographic, and climatic explanations (Figure 1a) (Brown, 2014; Fine, 2015; Mittelbach et al., 2007; Pianka, 1966). Drawing from species gradient literature, low latitudes have had more time for genetic diversity to accumulate within and across species as they evolve and specialize into available habitats compared to high latitude species (Mittelbach et al., 2007; Pereira, 2016). Additionally, low latitudes are associated with higher mean annual temperature, precipitation, and primary productivity in conjunction with less seasonal fluctuations in temperature (Brown, 2014; Currie et al., 2004; Ghalambor, 2006; Rohde, 1992; Willig et al., 2003; Zhang et al., 2018), which could be related positively with PGD. For example, higher temperatures increase mutation and evolution rates (Adams & Hadly, 2013; Allen et al., 2006; Gillooly et al., 2005; Rohde, 1992), while more productive environments (i.e. higher net primary productivity) support larger population sizes in most taxa allowing productive environments to maintain genetic diversity across species (Santini et al., 2018; but see Botero et al., 2014; Thuiller et al., 2020). Overall, low latitude taxonomic groups have experienced more time in warmer, climatically stable, and productive environments, and this could generate a 'negative', or hump-shaped, latitudinal gradient in PGD (Figure 1a).

Other species diversity theories lead to the opposite prediction of a positive, or U-shaped, latitudinal gradient with greater PGD at high latitudes (Figure 1b). One example is the tendency for species' geographic range sizes to decline at low latitudes due to intolerance to climatic fluctuations and varying dispersal capabilities (i.e. Rapoport's rule; Brown, 2014; Ghalambor, 2006; Stevens, 1989). These typically smaller range sizes in the tropics may be associated with smaller population sizes, more genetic drift, genetic bottlenecks from undergoing niche specialization (i.e. exploiting an ecological opportunity) (Bernos & Fraser, 2016; Currie et al., 2004; Fine, 2015; Siqueira et al., 2020), and therefore perhaps lower PGD. In contrast, high latitude species commonly have larger range sizes and better dispersal capabilities, thus likely more gene flow connecting populations to maintain PGD across their ranges (Ellegren & Galtier, 2016; Fan et al., 2019; Ghalambor, 2006; Martin & Mckay, 2004; Medina et al., 2018; Pelletier & Carstens, 2018). Additionally, climatic variance may be positively associated with PGD but negatively associated with species diversity. Regions of large climatic or temperature

fluctuations typically have lower species diversity (e.g. Northern hemisphere) but climatic variation may favour higher standing PGD to deal with such fluctuating environments (Barrett & Schluter, 2008; Botero et al., 2014; Brennan et al., 2019). These processes in turn may result in a positive latitudinal gradient for PGD.

Finally, much of the discussion above provides opposing processes that, when accounted for, could ultimately cancel out broad-scale PGD patterns, rendering an additional prediction of no overall latitudinal PGD gradient (Figure 1c). Smaller population size and niche specialization may constrain PGD within individual species at low latitudes, even though perhaps more time has passed for it to accumulate (Lawrence & Fraser, 2020; Mittelbach et al., 2007; Siqueira et al., 2020). While high latitude species typically have larger ranges (Fan et al., 2019), many of these species have experienced colonization bottlenecks after deglaciation events, in addition to having less time to accumulate new genetic diversity through mutation (Adams & Hadly, 2013; Green et al., 1996; Hewitt, 2004; Provan & Bennett, 2008; Stewart et al., 2010). Additionally, several genetic diversity gradients have been found within individual species (Adams & Hadly, 2013; Martin & Mckay, 2004) but due to taxonomic differences, such gradients may be flattened when viewed across many species at once (Millette et al., 2019). Together, these opposing processes may lead to no latitudinal PGD gradient at all.

Previous studies attempting to identify latitudinal gradients in genetic diversity used data from mitochondrial DNA (mtDNA) markers and found some support for a decrease in genetic diversity across species away from the equator (Adams & Hadly, 2013; Gratton, et al., 2017a; Martin & Mckay, 2004; Millette et al., 2019; Miraldo et al., 2016). This pattern was often relatively weak, and stronger in certain taxa (e.g. mammals; Adams & Hadly, 2013; Miraldo et al., 2016). While these studies provide an initial exploration on genetic diversity gradients upon which to build, their measure of genetic diversity was summed across species, representing a "community-level" of genetic diversity which does not contain information founded at the population level like PGD. Additionally, their use of mtDNA data may not reflect nuclear or genome-wide genetic diversity, important for adaptation to environmental change (Ballard & Whitlock, 2004; Bazin et al., 2006; Ghalambor et al., 2007; Hurst & Jiggins, 2005; Sgrò et al., 2011). Conversely, variation in nuclear DNA, e.g. microsatellites, can be a reasonable metric of genome-wide variation (Jarne & Lagoda, 1996; Väli et al., 2008) and is, appealingly, quantified commonly at the population level.

To test the mechanisms that may structure patterns in nuclear PGD, we utilized a large vertebrate database containing genetic data anchored to the population level, which make it particularly suitable for such analyses (Lawrence et al., 2019). This database reports metrics of presumably neutral genetic diversity for each geo-referenced, genetically distinct population. Conversely, previous studies either geo-referenced individual sequences (Miraldo et al., 2016), or grouped sampling localities into geographic "populations" (e.g. "high" or "low" latitude groups; Adams & Hadly, 2013; Martin & Mckay, 2004; Millette et al., 2019). Such *a priori* grouping could lead to biases in subsequent analyses and does not account for population-level dynamics.

Here, we investigate the relationship between latitude and PGD in vertebrate species across the American continents, and how environmental factors might mediate or influence this relationship. We tested the three aforementioned potential outcomes: a negative, positive, or no latitudinal gradient. To distinguish between these alternative predictions and to assess the importance of contemporary environmental variables for influencing PGD, we used vertebrate data, derived from microsatellite studies from ~900 species.

Methods

Data acquisition

We used georeferenced vertebrate population genetic data from *MacroPopGen* (Lawrence et al., 2019), collected from 895 species, 1308 studies published between 1993 and 2017, and from 9090 genetically distinct localities across the American continents. Each population has information on observed heterozygosity (Ho), mean number of alleles (MNA), sample size, and taxonomic grouping (i.e. Class, Family, Genus, Species), as well as a unique identifier for each study reference (RefID). For our analysis, we focused on using the metrics MNA and Ho for anadromous and freshwater fish, amphibians, birds, mammals, and reptiles. We mapped populations and PGD using QGIS v3.2.2 by taking the count, mean, and standard deviation of PGD from populations within 500km x 500km grid cells.

To test whether environmental factors mediate patterns of PGD as they appear to for species diversity, we obtained the following climatic variables from CHELSA for the period 1979–2013 by extracting raster values based on the point data of each of the populations: mean annual temperature (MAT, °C), annual precipitation (AP, mm/year), temperature annual range

(TAR, °C), and temperature seasonality (TS, standard deviation of monthly mean temperatures) (Karger et al., 2017a,b). Note that these "modern" climatic variables overlap roughly with the years the population-genetic data were collected, so they are a reasonable estimate for climates experienced by these populations. We also collected climatological data from the Last Glacial Maximum (LGM, 21,000 years ago) for each climatic variable (Karger et al., 2017b, 2017a) to account for historical effects of climate, particularly for high-latitude populations. The elevation (m) of each geo-referenced population was obtained using the R package rgbif v1.2.0 with the srtm3 model; ocean areas with no data were assigned an elevation of zero. The srtm3 model is based on data collected during the Shuttle Radar Topography Mission from the Space Shuttle Endeavour using the onboard radar system data; it provides an estimate based on a sample area of 90m x 90m. To obtain productivity data, we extracted raster data of net primary productivity (NPP, units of elemental carbon $x10e^{-11}$) from Imhoff et al. (Imhoff et al., 2004; Imhoff & Bounoua, 2006) for each population point.

Model Selection

To analyze the relationship between PGD, latitude, and contemporary environmental variables we used generalized additive mixed models (GAMMs), using the gam() function in R package mgcv v1.8-31. Data were first partitioned for each PGD metric and then trimmed such that only genera with ≥10 populations were retained. Additionally, populations with unavailable values (NA) for any variable were removed from each metric's dataset. After partitioning, 3475 and 4636 genetically distinct populations remained for Ho and MNA datasets, respectively. Response variables were Ho and MNA, modeled with beta and gamma distributions, respectively. Ho is a continuous variable bounded between zero and one, thus a beta distribution was deemed most appropriate; MNA values are positive, continuously distributed, and right skewed, so a gamma distribution was most appropriate for these data.

For each of the MNA and Ho datasets, we conducted model selection by first testing which taxonomic level was most important in null models by including a random effect for Class, Family, Genus, or no taxa effect at all. Importance of taxa-specific random effects were tested to account for variance among taxonomic groups, as previous works have found that not all groups may show the same genetic diversity pattern (Adams & Hadly, 2013; Hirao et al., 2017). We chose not to test below the Genus level to avoid loss of data. We used the information

theoretic approach (AIC; Akaike, 1974; Anderson & Burnham, 2002) to compare null models that only included the random effect for RefID and the taxonomic level. The model with the lowest AIC then identified the best random effect structure based on fit and complexity. RefID and the identified taxonomic level were included as a random effect in subsequent models. All models were weighted by population-specific sample size of genotyped individuals to account for sample size differences between populations. Before incorporating variables together to test full models, we tested for multicollinearity using variance inflation factors (Zuur et al., 2009). LGM and modern climatic variables were highly correlated, as were TS and TAR. Thus, we decided to focus only on modern variables, since contemporary PGD is likely to be more related to modern climatic data. Additionally, we chose to include TAR over TS so the units would be more comparable to MAT, and to reflect the range of temperatures a population may experience. In all models, continuous variables were also smoothed using cubic regression splines with shrinkage applied. Shrinkage allows for a smoother to have zero degrees of freedom, thus a smoother can be dropped from the model during model selection (Zuur et al., 2009). Interaction terms were fitted with tensor products and thin-plate regression splines using the te() function in mgcv package, according to Pederson et al. (2019).

Following Zuur et al.'s (2009) approach for model selection, we considered models within 2 ΔAIC points as equivalent. We proceeded with forward model selection by sequentially adding one of the six modern and non-collinear variables (latitude, elevation, NPP, MAT, AP, TAR) to the null models until addition did not improve model fit, as reflected by a decrease in AIC. We chose this approach to minimize risk of overfitting models, but also tested backwards model selection by starting with a full model and testing the sequential removal of variables. Interactions between fixed effects were tested to account for the effect one variable might have on another - for example, high elevations at low latitudes tend to experience similar variations in temperature variation as high latitude regions (Ghalambor, 2006; Janzen, 1967). Such interactions might reveal micro-niche influences on PGD (Ghalambor, 2006; Janzen, 1967). Thus, in our model selection process we included the biologically relevant interactions of elevation with MAT, TAR, and NPP.

After model selection, we performed cross-validation on the selected model using the validation set approach with caret package v6.0-86. We first trained the model on a random 50% subset of the data, and then tested how well it predicted results using the other 50% of the

dataset. Since AIC can sometimes select overfitted models (Pedersen et al., 2019; Zuur et al., 2009), we also assessed variable significance after model selection – if a variable was not significant and/or its degrees of freedom were reduced to zero, it was removed from the model.

Taxa-Specific Patterns

After identifying which taxa level was most appropriate, we tested a taxa-by-latitude interaction using tensor products. This allowed us to identify latitudinal patterns among specific groups, and how accounting for taxa can influence the overall PGD patterns in latitude.

Results

Data acquisition

The distribution of population points, and gridded heatmap of PGD metrics across the Americas are found in Figure A3.1. Taxonomic classes varied in their mean PGD and mean environmental variables experienced (Table 3.1). Anadromous fishes showed the highest mean values for both Ho (0.70) and MNA (14.97) and tended to have populations at higher latitudes (mean latitude 50.55; Table 3.1). Reptiles, birds, and amphibians experienced the highest MAT (17.97, 14.25, and 12.51°C respectively), and AP (1113, 1230, and 1258 mm/year, respectively), whereas birds and reptiles showed the lowest TAR (22.20, 23.19°C). Amphibians experienced the highest mean NPP ($3.78e^{11}$ units of elemental carbon) whereas anadromous fishes experienced the lowest NPP ($2.56e^{11}$ units of elemental carbon).

Model Selection

After model comparison for taxonomic-level random effects, the Genus model had the lowest AIC (Table A3.1). Thus, Genus was selected for subsequent models. Both forward and backward model selection identified the full model with all variables and the three interactions as the model with the lowest AIC (Table A3.1). These full models explained a high degree of deviance for Ho and MNA (83.9% and 85.5% respectively). The random effects of RefID and Genus were both significant, indicating an important aspect of study- and genus-specific responses. While all variables were significant for the Ho model, only MAT, elevation, AP, and the interaction between TAR and elevation were significant for MNA (Table A3.2). Latitude was marginally significant ($p=0.092$) in the MNA model; since it was a main variable of interest, we retained it in the model and then tested the exclusion of NPP, NPP's interaction with elevation, TAR, AP, and the interaction between elevation and MAT. Latitude became significant after the

removal of these five variables (p=0.0021). We compared this model to the saturated model to assess how well each predicted the testing data after cross validation.

Cross validation determined that the model without the non-significant terms had the lowest Root Mean Square Error (RMSE) and lower prediction error rate (7.6795 and 0.97214 compared to 7.6796 and 0. 97215, respectively); although the differences were minimal, this was selected as the final model. The Ho model had a RMSE of 0.5562 and a prediction error rate of 1.05855. Since the units of RSME are in the unit of the dependent variable – i.e. MNA and Ho – this would indicate that the selected models have RMSE values roughly equal to the mean value of each PGD metric. However, even when comparing null or latitude-only models, RMSE only changed by a maximum of 0.005 units.

Effect of Latitude and Environmental Variables

Figure 3.2 and 3.3 show the predicted effect for each significant variable on each PGD metric. The only latitudinal pattern in MNA was that MNA increased at latitudes >30° in the North hemisphere (Figure 3.2). MNA declined at elevations above 1000m, and we found a U-shaped effect of MAT where MNA increased below 0°C and above 15°C. Overall MNA increased with AP, although there were oscillations until ~3800mm/year where the effect steadily increased. The interaction between TAR and Elevation showed an increase in MNA (represented by regions in red Figure 3.2) at low elevations (0 to ~100m) and intermediate TAR (15 to 25°C), as well as at elevations >2000m and TAR >30°C; MNA decreased between 2000-3000m elevation and TAR 20-30°C.

Latitude had a somewhat V-shaped pattern for Ho, where Ho slowly declined towards 0° latitude, and then sharply increased at ~30° latitude (Figure 3.3). Contrary to MNA, Ho increased with both Elevation and NPP, and there a was a hump-shaped relationship with MAT compared to the U-shaped pattern found for MNA. Ho increased between ~-15°C and 10°C and declined below/above these values, respectively. For TAR, Ho only increased >30°C and started to decline again >50°C. As with MNA, Ho oscillated with AP, only increasing between 500 and 1000mm/year, and largely declining >1000mm/year. For the interaction between Elevation and MAT, Ho increased only at low elevations (<1000m), regardless of MAT. Similarly, for the other two interactions with elevation, Ho only increased at elevations <2000m and low values of both TAR (<10°C) and NPP (<$3e^{11}$ units of elemental carbon).

Taxa-Specific Patterns

Because Genus was a significant predictor, we tested models for Ho and MNA that included an additional tensor product interaction term between Latitude and Genus to identify genera-specific patterns. For MNA, this resulted in 15 of 115 genera having significant relationships, 7 of which showed MNA decreasing at latitudes above the equator, 6 showed an increase of MNA above the equator, and 2 showed a slightly hump-shaped pattern where the "hump" occurred at ~30° latitude (Figure A3.2). For Ho, 39 of 104 genera had significant relationships, 15 showed a decline in Ho above the equator, 12 showed an increase in Ho at latitudes above the equator, 1 showed a hump-shape relationship, 5 showed a slightly U- or V-shaped relationship, and 6 showed a relatively flat relationship across latitudes (Figure A3.3). Interestingly, when the effect of Genus was interacted with Latitude, the global smoother of Latitude became non-significant (p=0.076) for MNA, but both MNA and Ho showed the first predicted pattern: a peak of PGD at low latitudes, reflected by a hump-shaped pattern (Figure A3.2).

Discussion

Here we assessed the effect of latitude and the environment on vertebrate nuclear PGD. We predicted potential outcomes by drawing from species gradient theory: a negative gradient, a positive gradient, or no gradient at all. Overall, our results suggested a weak latitudinal pattern (MNA) in nuclear PGD (Figure 3.2 and 3.3). We also found significant effects of environmental variables, particularly for elevation, MAT, and AP (Table A3.2). This suggests that environmental variables together better predict patterns in PGD than any one variable alone. However, we also found that the taxonomic level of Genus accounted for a great deal of the variation in our models, and models including Genus had lower AIC values than those without (Table A3.1). Therefore, genus-specific effects may be one of the strongest describing factors for large-scale patterns of vertebrate genetic diversity, and higher taxonomic levels may not account for this variation.

Past works that have found a latitudinal gradient in genetic diversity generally found it for individual species (Adams & Hadly, 2013; Martin & Mckay, 2004), but the gradient was less pronounced across Classes (Millette et al., 2019; Miraldo et al., 2016; Schluter & Pennell, 2017), particularly when accounting for species identity (Gratton, et al., 2017a). A possible explanation

could be due to the *a priori* grouping these studies used where species were split into low or high latitude groups (Adams & Hadly, 2013; Martin & Mckay, 2004). Thus, a latitudinal gradient would appear when assessing individual species, but when looking across all species, what may be considered low latitude for one species may be considered high latitude for another species. This could explain why, across species, we found little evidence for the "classic" latitudinal gradient, which only appeared in the global smoother when Genus was accounted for, and was only significant for Ho. This supports the notion that genetic diversity across species does not show strong patterns with latitude. Adams & Hadly (2012) presented two main theories for why they found a latitudinal gradient in their mtDNA data: higher temperatures increase mutation rates at the tropics, and glaciation effects cause bottlenecks at higher latitudes. While they provided some of the first mechanistic speculation on what may structure genetic diversity patterns, they did not formally test environmental variables and their data only came from 72 vertebrate species, strongly favouring mammalian species (n=41), a group which has shown the strongest latitudinal gradient relative to others (Millette et al., 2019; Miraldo et al., 2016).

Here we had the temperature data to test such a theory and found a U-shaped pattern with MAT for MNA and a hump-shaped pattern for Ho where, for both metrics, PGD tended to be higher between -10°C and 0°C (Figure 3.2 and 3.3). This could indicate that increasing temperature does not always have a positive effect on PGD, or that there is a potential bias in taxonomic representation in the dataset, since anadromous fishes had the highest PGD, occurred at high latitudes, and had lower MAT (Table 3.1). Nevertheless, of the 535 populations with MAT <0°C, only 56 were anadromous fish (259 were mammals, 180 were freshwater fish, 25 were birds, and 15 were amphibians). We also found a general decrease in MNA at elevations >1000m and oscillating effects of AP for both PGD. The significance of the interactions between elevation with MAT (Ho) and TAR (Ho and MNA) helps to elucidate our findings. Since higher elevations are typically associated with cooler temperatures, more precipitation, greater temperature range, less productivity, smaller range sizes, and overall less genetic variation (Gillman et al., 2009), this could explain the slight negative effect of elevation, and initially negative effect of temperature on its own.

While our study's results differ somewhat from past works (Adams & Hadly, 2013; Martin & Mckay, 2004; Miraldo et al., 2016), the dataset used here is much larger, having a wide taxonomic and latitudinal breadth. Admittedly, most data (~85%) comes from North America,

and there is a disproportionately large number of anadromous fish populations. However, Lawrence et al. (2019) found limited biases across geographic and taxonomic groups as well as no bias associated with microsatellite loci number and type. Furthermore, when inspecting the mean PGD across North America, which is relatively well-sampled, there is no obvious latitudinal gradient across this continent (Figure A3.1). Additionally, even though anadromous fish were well represented, no anadromous genus had significant latitudinal patterns (Figures A3.2. 3.3). When we accounted for the effect of genus the peak of PGD at high latitudes largely disappeared, resulting in a weak increase of PGD at low to mid-latitudes, suggesting that anadromous fish may have been driving the original latitudinal patterns found in Figures 3.2 and 3.3.

Our results show that nuclear PGD is much more nuanced in its distributions than species diversity or in how genetic diversity was measured in past work on latitudinal gradients (Miraldo et al., 2016). Adams & Hadly (2013) suggested genetic diversity might be a precursor to species diversity, and this is congruent with work discussed by Schluter & Pennell (2017). High genetic diversity at low latitudes in genera that do not show the typical species latitudinal gradient may indicate that with enough time, such a gradient will be manifested. This notion could also be applied to high latitude species since speciation rates are increasing at higher latitudes (Schluter & Pennell, 2017). It is possible that there is a transition phase of speciation rates occurring since landscape rearrangements following the last deglaciation, and this transition may be masking any nuclear DNA gradients across latitudes. As speciation rates accelerate at high latitudes, this could lead to an overall increase in PGD across high latitude species. Meanwhile PGD may be maintained or even decrease at low latitudes (due to genetic drift in smaller populations), potentially flattening a PGD gradient. Coupled with this effect may be the role of anthropogenic impacts, such as urban fragmentation (Goossens et al., 2006), the implications of which were beyond the present study but we will be investigating in future works.

The non-significant “typical” latitudinal gradient in PGD has important global biodiversity conservation implications, since conserving regions of high species diversity may not simultaneously conserve regions of high PGD (Paz-Vinas et al., 2018). Therefore, one must determine when the focus should remain on species richness, and when PGD should be incorporated into conservation goals to provide greater resolution. Ideally, a more comprehensive approach to conservation that considers PGD alongside interspecific diversity

(i.e. species richness) is recommended. However, this is not always possible, so we suggest that conservation goals guide the approach taken. If the goal is to conserve the greatest number of species, then species richness should be prioritized. If more narrow targets are to be met, for example the maintenance of specific species or populations, then PGD should be considered to identify which populations are either in need of prioritization or can be used to supplement other populations. In addition to PGD, other aspects of intraspecific diversity that would be valuable to consider include population richness (Lawrence & Fraser, 2020) and adaptive genetic diversity (Lawrence & Fraser, 2020; Stanley et al., 2018). The many different populations across a species range and their genetic diversity may hold the key to future survival of the species. In sum, there is intrinsic value in managing intraspecific diversity to ensure species can adapt to and survive future environmental change (Barrett & Schluter, 2008; Bernatchez, 2016; Rey et al., 2016), and our work identifying the nuanced distribution of genetic diversity only exemplifies this further.

Tables

Table 3.1. Summary of mean environmental and population genetic variables for each of the vertebrate classes assessed across the Americas (before separating into Ho and MNA datasets). Ho = observed heterozgosity, MNA = mean number of alleles, Lat = degrees latitude, MAT = mean annual temperature (°C), AP = annual precipitation (mm/year), Elevation (m), TAR = total annual range (°C), NPP = net primary productivity (units of elemental carbon x10e^{-11}), LGM = Last Glacial Maximum. Values in parentheses represent the standard deviation.

Class	Species	Ho	MNA	Lat	MAT	LGM MAT	AP	LGM AP	Elevation	TAR	LGM TAR	NPP
Amphibia	104	0.57 (0.17)	7.50 (5.56)	32.72 (18.27)	12.51 (7.225)	0.32 (13.54)	1258.83 (893.63)	12700.49 (8044.23)	805.93 (890.74)	26.87 (10.64)	29.45 (12.50)	3.78E^{+11} (2.16E^{+11})
Anadromous	15	0.70 (0.15)	14.97 (7.01)	50.55 (5.95)	7.35 (4.19)	-10.60 (9.10)	1122.44 (743.53)	13404.07 (7194.75)	222.98 (338.72)	26.93 (6.91)	25.10 (8.20)	2.56E^{+11} (1.45E^{+11})
Aves	254	0.59 (0.17)	7.65 (6.16)	25.92 (26.0)	14.25 (8.86)	4.17 (15.59)	1230.72 (866.84)	12819.69 (8281.16)	519.632 (742.59)	22.20 (12.76)	24.63 (14.36)	3.25E^{+11} (2.64E^{+11})
Freshwater	231	0.57 (0.17)	7.12 (4.00)	37.44 (20.42)	10.46 (8.24)	-5.60 (15.93)	988.72 (593.08)	10544.26 (5767.31)	472.06 (605.76)	29.89 (9.78)	32.56 (10.87)	3.42E^{+11} (2.39E^{+11})
Mammalia	158	0.60 (0.14)	6.09 (3.93)	34.19 (27.15)	9.43 (9.10)	-4.28 (15.40)	917.45 (739.85)	9491.57 (6822.33)	736.71 (869.81)	30.08 (10.17)	31.55 (10.42)	2.79E^{+11} (2.16E^{+11})
Reptilia	133	0.59 (0.15)	6.69 (3.63)	27.54 (16.37)	17.97 (6.56)	7.83 (13.81)	1113.82 (651.43)	11560.58 (6122.40)	278.28 (441.53)	23.19 (11.24)	27.81 (13.77)	3.46E^{+11} (2.46E^{+11})

Figures

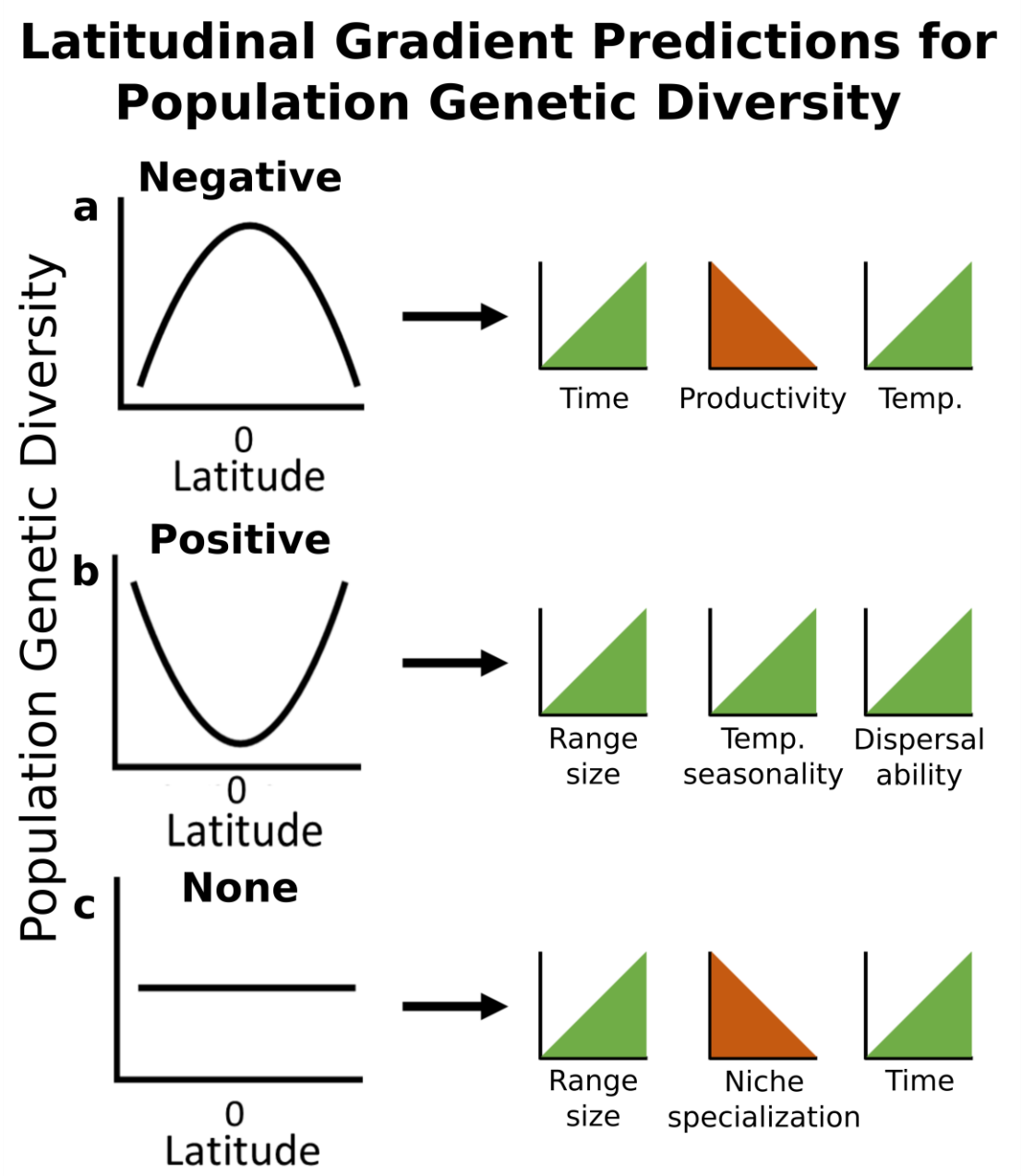

Figure 3.1. Summary of the three predictions for a latitudinal gradient in population genetic diversity, indicating which variables are likely to contribute to expectations for a) positive, b) negative, or c) no latitudinal gradient. All y axes represent population genetic diversity, indicating the generally expected trend for genetic diversity with each variable.

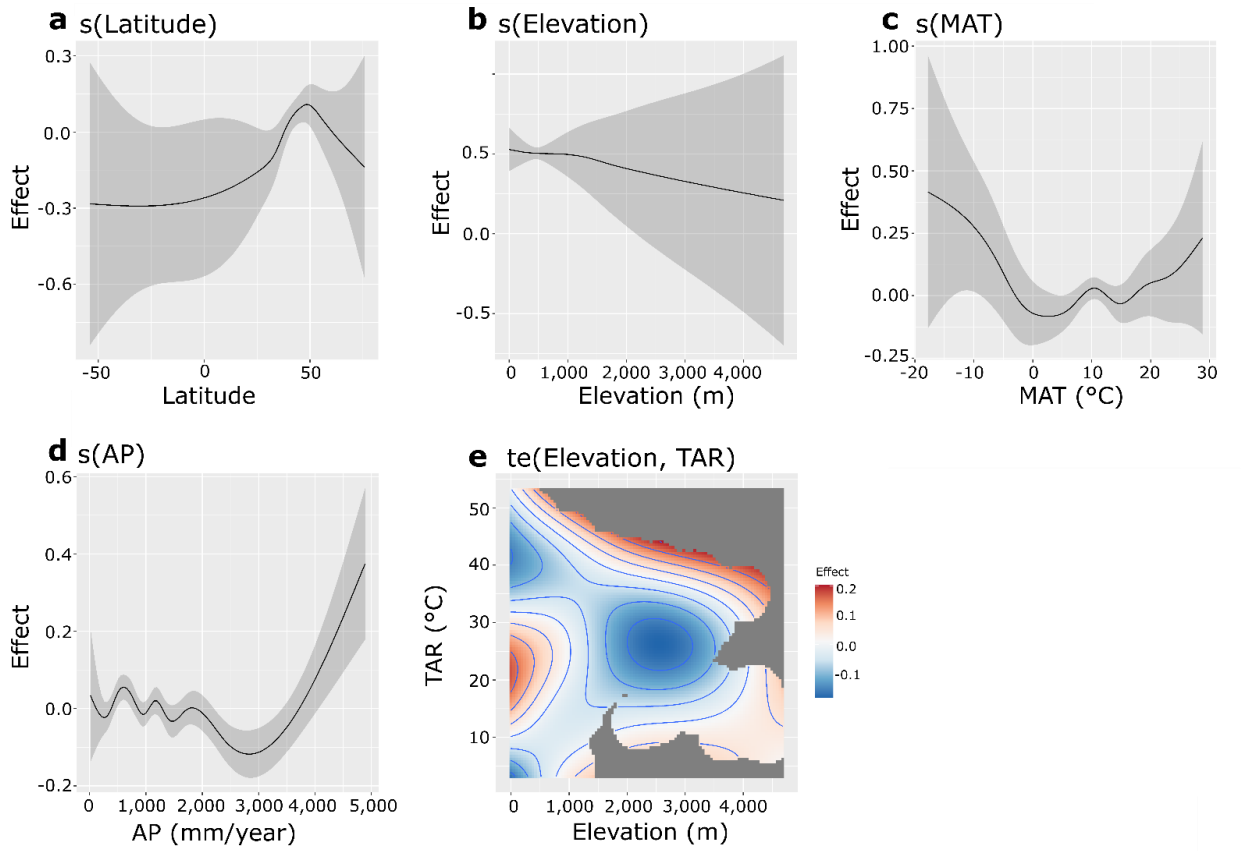

Figure 3.2. The predicted effect of latitude and environmental variables on vertebrate population genetic diversity (mean number of alleles) for vertebrate species across the Americas. Predictors from the selected generalized additive mixed model were fitted by smoothers (s; tensor products (te) for interaction) and include a) Degrees Latitude, b) Elevation (m), c) MAT = mean annual temperature (°C), d) AP = annual precipitation (mm/year), e) the interaction between Elevation and total annual temperature range (°C, TAR). Dark grey zones represent areas that were unable to be estimated.

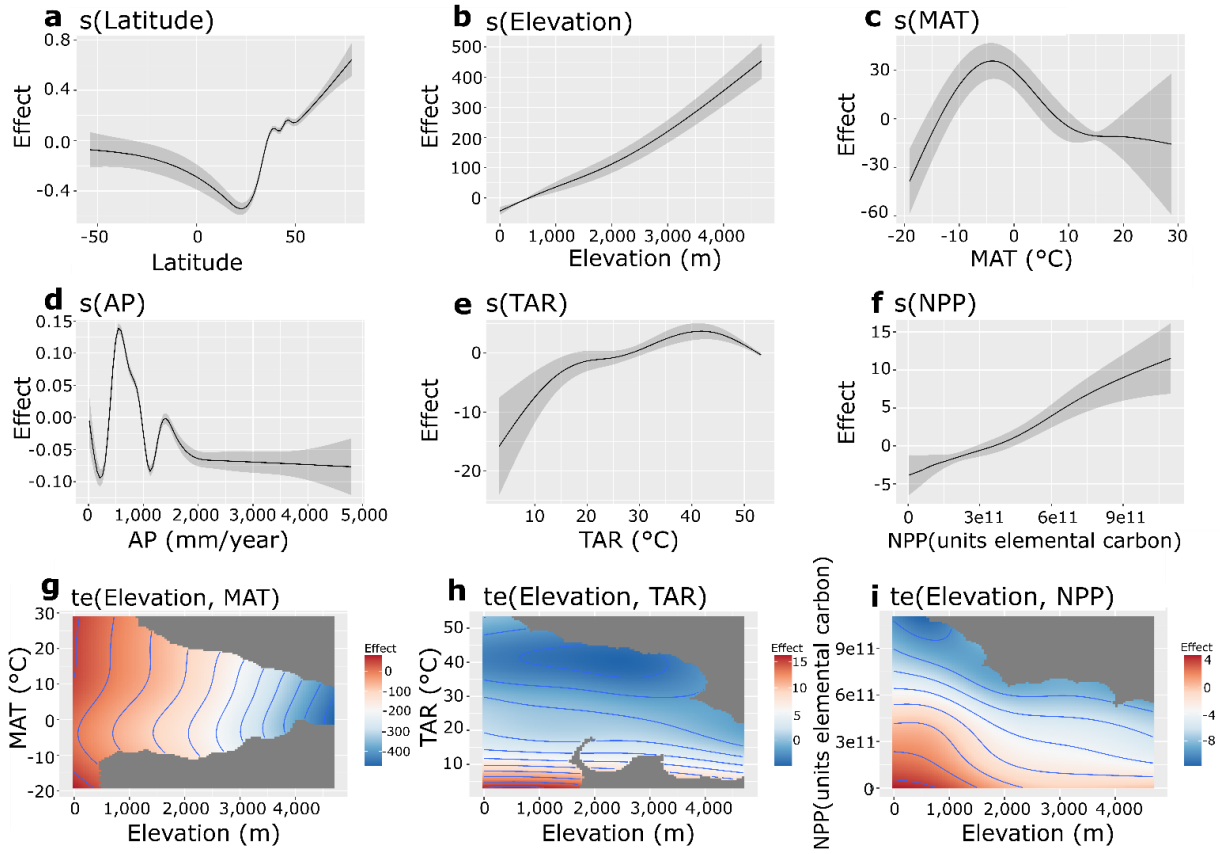

Figure 3.3. The predicted effect of latitude and environmental variables on vertebrate population genetic diversity (observed heterozygosity) for vertebrate species across the Americas. Predictors from the selected generalized additive mixed model were fitted by smoothers (s; tensor products (te) for interactions) and include a) degrees Latitude, b) Elevation (m), c) MAT = mean annual temperature (°C), d) AP = annual precipitation (mm/year), e) TAR= total annual temperature range (°C), f) NPP = net primary productivity (units of elemental carbon x10e$^{-11)}$, g) the interaction between elevation and MAT, h) the interaction between elevation and TAR, and i) the interaction between elevation and NPP. Dark grey zones represent areas that were unable to be estimated.

Chapter 4: Variability in anthropogenic impacts on nuclear vertebrate population genetic diversity across the Americas

Abstract

Humans have varying impacts on biodiversity, although conventionally these impacts are thought to be negative. For genetic diversity within populations, the evidence for negative impacts are, thus far, largely inconsistent across taxonomic groups. However, few studies have accounted for the variable intensity of human impacts, the surrounding heterogeneity in habitats, and the variability among and especially within taxonomic classes. Here we used population-level data to assess how land use (assessed by anthropogenic biomes), human population density (HPD), distance to urban or natural environments, and the surrounding anthropogenic biomes influence vertebrate nuclear population genetic diversity across the American continents. We found limited evidence for significant effects of HPD and anthropogenic biomes on their own. Instead, we found that the composition of surrounding biomes and distance to urban/wild environments had the most significant effects. The more urbanized a region was, the stronger the negative effect on population genetic diversity. Additionally, we found considerable variation among genera of different classes. Our study demonstrates the variable nature of human impacts on genetic diversity and emphasizes that there is no broad brush for developing conservation management regimes to mitigate human impact. We thus highly recommend assessing at least at the genus level, ideally lower (e.g. species) when evaluating human impacts on population genetic diversity.

Introduction

Human influences are various in nature, and as a result have varying effects on biodiversity (DiBattista, 2008; Fahrig et al., 2018; Frankham et al., 2002; Goossens et al., 2006; Nowakowski et al., 2018; Schmidt et al., 2020; Sebastián-González et al., 2019). One important biodiversity component that is increasingly recognized and incorporated into biodiversity planning is genetic diversity (Millette et al., 2019), particularly genetic diversity when measured at the population level (PGD) (Paz-Vinas et al., 2018; Schmidt et al., 2020). Large-scale syntheses of anthropogenic impacts on genetic diversity in general are only just emerging, but highlight a variety of effects that are dependent on taxonomic group and type of impact (DiBattista, 2008; McGill et al., 2015; Millette et al., 2019; Miraldo et al., 2016; Schmidt et al., 2020). Recent syntheses have focused on how human population density (HPD) may influence genetic diversity, but this can have limitations such as not encapsulating the heterogeneous impacts that can occur outside of densely populated regions (Cincotta et al., 2000; Millette et al., 2019; Schmidt et al., 2020). Several other land use activities surrounding a population could account for ecosystem disturbances regardless of HPD (e.g. agricultural crop- or rangelands; Cincotta et al., 2000; Ellis et al., 2010; Ellis & Ramankutty, 2008). It is clear that additional investigations would help to elucidate the specific impacts of human activities on PGD, and lend aid to (i) conservation prioritization (Fonseca et al., 2019; Paz-Vinas et al., 2018), (ii) the development of effective strategies to minimize GD losses (e.g. habitat conversion) (DiBattista, 2008; Millette et al., 2019), and (iii) the improvement of biodiversity monitoring regimes (Leroy et al., 2018; Mimura et al., 2017; Schmidt et al., 2020).

A conventional first hypothesis might be that human impacts have a negative influence on PGD. Human influences can result in habitat reduction, fragmentation, and isolation in many populations (DiBattista, 2008; Johnson & Munshi-South, 2017; Otto, 2018). This in turn would result in population size reductions, which leads to increases in inbreeding and genetic drift, ultimately eroding PGD – the "small population paradigm" (Caughley, 1994; Reed & Frankham, 2003; Willi et al., 2006). However, there are some indications that the impacts from humans are not consistent, and not always negative (Millette et al., 2019; Schmidt et al., 2020).

These inconsistencies lead to a second hypothesis that PGD will vary with the type and intensity of human impact (Habrich et al., *submitted*; Schmidt et al., 2020). When assessing types

of land use and/or human impact on a gradient from urban to "wild", we would expect greater reductions in PGD as the human impact intensity increases. Ellis and Ramankutty's (Ellis & Ramankutty, 2008) "anthropogenic biomes" are a perfect example to use, as these biomes scale with human impact intensity. For example, on average, highly dense urban environments would be expected to have the greatest reduction in PGD of wild populations, while croplands and rangelands may only reduce PGD of certain species. Of course, croplands themselves may have variable impacts since they can range from intensive monocultures to polycultures. Semi-natural and wild regions with minimal anthropogenic development would then be expected to retain the highest levels of PGD in wild species. However, we note that just because a population may be located in a heavily impacted area, this might not correspond to low PGD – due to potential buffering effects of heterogeneous habitat in the surrounding area, a population may be protected from the otherwise negative effects of human impact.

As variation in responses to anthropogenic impacts exist, changes to PGD might also be hypothesized to vary according to taxonomic group (Millette et al., 2019; Schmidt et al., 2020). Mammals are one example that have frequently been shown to be more affected by humans than groups such as birds (Millette et al., 2019; Miraldo et al., 2016; Schmidt et al., 2020). However, recent work has emphasized that tremendous variation not only exists among groups, but also likely within groups (Habrich et al., *submitted*; Lawrence et al., 2019; Lawrence & Fraser, 2020). For example, larger, long-lived species (e.g. Caribou) are likely more negatively affected by anthropogenic impacts than smaller species that might exploit and benefit from human habitats (e.g. Racoons) (Barrueto et al., 2014; Habrich et al., *submitted*). Yet past studies were not only relatively taxonomically limited (two to four groups) (Millette et al., 2019; Miraldo et al., 2016; Schmidt et al., 2020), but also lumped taxonomic groups based on classes, assuming intra-class variation to be less than inter-class. Thus, it is not yet clear which taxonomic level best captures human impacts on genetic diversity.

While previous studies laid the groundwork for investigating human impacts on PGD among animal classes, they did not obtain population-specific estimates of genetic diversity (exception: Habrich et al., *submitted*; Schmidt et al., 2020). Instead, genetic diversity was obtained from individual (sometimes grouped) sequences (Millette et al., 2019; Miraldo et al., 2016) or a single value averaged across populations within a study (DiBattista, 2008). Population identification is an important aspect of within-species variation as it emphasizes genetic

distinctness, biological realism, and accounts for more genetic diversity than a mean across potentially unrelated sequences (Lawrence et al., 2019; Lawrence & Fraser, 2020). Such a measure of genetic diversity allows for population-specific impacts to be incorporated to better identify the human effects on PGD at higher taxonomic levels. Thus, finer-scale resolution, by use of genetic markers such as microsatellites, permits identification of these population-units and their associated PGD.

There is also the question of which genetic tools best measure PGD within populations. Previous studies largely focused on non-nuclear, mitochondrial DNA (Bazin et al., 2006; Millette et al., 2019; Miraldo et al., 2016; exception Schmidt et al., 2020, Habrich et al., *submitted*). Mitochondrial genes, e.g. *co1* and *cytb*, are not selectively neutral (Millette et al., 2019; Pentinsaari et al., 2016) and may not be as influenced by any particular anthropogenic pressures (Hendry et al., 2008; Millette et al., 2019). Conversely, nuclear genetic markers such as microsatellites can better reflect genome-wide genetic diversity, and are also typically selectively neutral (Selkoe & Toonen, 2006; Wiehe, 1998). The polymorphic nature of microsatellites also allows population structure to be more readily resolved at fine scales (Angers & Bernatchez, 1998; Jarne & Lagoda, 1996; Väli et al., 2008). While other nuclear markers such as Single-Nucleotide-Polymorphisms (SNPs) also provide genome-wide information and are increasingly adopted, microsatellites have been the marker of choice for two decades, thus currently providing the greatest abundance of collectable data across taxa (Lawrence et al., 2019). Together, these factors allow for the anthropogenic impacts on nuclear PGD to be more effectively assessed.

To simultaneously test the three aforementioned hypotheses, we used a population-genetics database that provided nuclear microsatellite PGD for genetically distinct populations of vertebrates (Lawrence et al., 2019). The availability of such data allowed us to investigate the effect of anthropogenic influence on PGD within taxa and at the population level, not just across individual sequences. Our objective was to determine if humans have a generally negative effect on vertebrate PGD, and how this effect may vary across impact type and among genera.

Methods

Genetic data acquisition

We obtained vertebrate PGD data for six taxonomic groups (amphibians, anadromous fish, birds, freshwater fish, mammals, and reptiles) from the geo-referenced population-genetics database, *MacroPopGen* (Lawrence et al., 2019). We chose to focus on observed heterozygosity (Ho) and mean number of alleles (MNA) as these are main metrics of nuclear genetic diversity, and because the two metrics reflect different aspects of PGD. For example, population size reductions can cause detectable changes in MNA much more rapidly than in Ho, which only has detectable changes over longer time scales (Allendorf, 1986; Nei et al., 1975). To minimize type I error, our dataset only included genera with a minimum of 10 populations. In total, we used a subset of data from 7951 populations, representing 460 species, 165 genera, and 84 Families, based on 471,817 individual genotypes from 871 studies.

Anthropogenic data acquisition

Anthropogenic influence was measured using two metrics: HPD (persons km^{-2}) and land usage (defined by the anthropogenic biomes described in Ellis & Ramankutty 2008). Both were obtained in raster format at a spatial resolution of 5 arc minutes from Klein Goldewijk et al. (Klein Goldewijk et al., 2017). For both metrics, data were obtained on a per-population basis by overlaying the population point data and extracting the raster value for each population, using the raster() and extract() functions from the R package raster v3.1-5. Thus, each population was identified as being located within a particular anthropogenic biome (hereafter, "Orginating Biome").

As a first way of addressing land usage, anthropogenic biomes (hereafter anthromes) were represented by six groups, with definitions based on global land use (percent area of crops, pasture, urban land, etc.), land cover (percent area of trees and bare earth), and human population statistics (urban versus non-urban; Ellis et al., 2010; Ellis & Ramankutty, 2008). From previous works by Ellis et al., we have adapted these six anthromes, listed from most urban to least urban (numbers of genetically-distinct populations in parentheses): Urban (n=933), Village (n=101), Cropland (n=1611), Rangeland (n=565), Semi-Natural (n=2676), and Wild (n=1779). We additionally added two aquatic biomes: Freshwater (n=54) and Ocean (n=232), to account for populations that were found outside anthromes. These populations were largely represented by

fish (n=121) and reptiles (n=129), while birds (n=64), mammals (n=55), and amphibians (n=12) were represented in lower numbers. HPD and land cover were mapped alongside genetic diversity metrics using the World Behrman projection in ArcGIS v10.7.1.

In general, the impact that humans have had on biodiversity has increased over time, so as a supplementary analysis we wanted account for these changes. To do so, we collected data between 1990 and 2016 for HPD and the anthromes to assess if, and by how much, each of these metrics changed. These years were selected because the years that the genetic data were published ranged between 1993 and 2017, with a mean year of 2010. However, the most recent year for anthropogenic biome data was 2016, so we chose 2016 as our maximum annual range. First, we calculated the difference between years for HPD and anthrome metrics, and then formally tested differences using Wilcoxon signed-rank tests for the years 1990, 2010, and 2016. While we found significant differences between years for both metrics, relatively few populations exhibited a change in HPD >100 persons km^{-2} or shifted to a more "urban" biome (484 and 200 populations of 7951 respectively; see Appendix 4: Supplementary Methods, Table S4.4, Figure S4.1). This may be due to the fact that most populations sampled did not occur in regions of very high HPD (e.g. only 123 populations were found in regions >2000 persons km^{-2}). Thus, we decided to use data from 2010 because it was the mean year the genetic data were published in and is likely to best represent the human impacts experienced by most of the populations at time of sampling.

Accounting for habitat heterogeneity

As mentioned, habitat heterogeneity surrounding a population could have buffering effects on its PGD. For example, a population geo-referenced in an Urban biome may be relatively close to a Semi-Natural biome, and/or in the surrounding area the Urban biome may compose a smaller area relative to non-urban biomes. Accounting for heterogeneity in the surrounding habitats could thus explain why, for instance, a population found in an Urban biome may have higher genetic diversity than expected – the surrounding areas may provide a refuge for such populations. As a first metric to account for this heterogeneity, we calculated the distance from each population to the nearest edge of "Natural" and Urban biomes. This distance gives a first impression on how close a potential refuge or threat may be to a population in question. To do this we first converted the anthrome raster into a shapefile and then subset it into

two shapefiles: one with only Semi-Natural and Wild features (together: Natural), and one with only Urban features (i.e. excluding Villages, Cropland, Rangeland). Then we used the Generate Near Table function in ArcMap to measure the distance in kilometres from each population to the edge of the nearest Natural or Urban biome.

As an additional metric for assessing habitat heterogeneity around populations, we calculated the proportion of different anthromes surrounding a population. To calculate the proportion of surrounding anthromes, we created a 100km buffer around each population point and then used the Tabulate Intersection function to calculate the percent of each biome around a population. We chose a buffer distance of 100km first because impacts from urban centres can put pressure on populations and/or ecosystems at least 100km away (Cincotta et al., 2000; Repetto, 1994); thus 100km represents a minimum distance from human impact for most populations. Second, taxa-specific dispersal distances with which we could base buffer distances were scarce. To transform these percentages into a metric – hereafter "proportion of biomes" (POB) – for use in models, we attributed each biome a rank from 1-7, such that a higher rank was given to more wild biomes. Therefore, 1 represented both aquatic biomes, 2 represented Urban biomes, 3 represented Villages, 4 represented Croplands, 5 represented Rangelands, 6 represented Semi-Natural, and 7 represented Wild biomes. Then, for each population, we summed the product of biome percent by rank. For example, a population with surrounding 80% wild, 10% cropland, and 10% rangeland would be (0.8x7) + (0.1x4) + (0.1x5) = 6.5. Thus, the final POB score generates a relative metric for the degree to which each population is exposed to a refuge (i.e. a "natural" habitat), where the higher the score, the more exposed to a refuge the population is.

Human impacts on population genetic diversity

To test the effect of human impacts on PGD, we created a set of generalized additive mixed models (GAMMs) for each anthropogenic metric considered (HPD, land use, distance to Urban/Natural, POB). Data were partitioned for each PGD metric (MNA and HO) to remove unavailable values, resulting in 4970 and 5470 vertebrate populations for MNA and Ho, respectively. For all GAMMs, Reference ID (RefID) was included as a random effect to account for differences among studies. To test which taxonomic level best accounted for variation within groups, we compared null models that had either taxonomic Class (e.g. Amphibia, Mammalia,

freshwater fish, etc.), Family, Genus, or no taxonomic grouping as an additional random effect using the information theoretic approach (AIC) (Akaike, 1974; Anderson & Burnham, 2002). The model with the lowest AIC was retained. For both Ho and MNA, the Genus model was selected, thus Genus as a random effect was included in subsequent model selection. Genus was included as a random effect instead of a fixed effect to ensure there were genus-level intercepts as per Pederson et al. (2019). All models were also weighted by population-specific sample size of genotyped individuals. Before conducting model selection, we tested for multicollinearity among variables by using variance inflation factors; no collinearity was found as all variables were below 3 (Zuur et al., 2009).

After determining taxonomic random effect structure for the set of models associated with each PGD metric, we conducted forward model selection using AIC. To do this we sequentially added a fixed effect and tested whether its addition resulted in a decrease in AIC; models within 2 units of each other were considered equal (Zuur et al., 2009). To test the impact of various human influences on both PGD metrics, GAMMs included the following five variables: HPD, the anthrome a population fell into (Originating Biome), distance to nearest Natural biome, distance to nearest Urban biome, and the POB metric as fixed effects; a sixth variable included the interaction between HPD and Originating Biome. HPD and Originating Biome accounted for the direct impacts of increasing human presence on vertebrate PGD, while the distance and POB metrics accounted for the buffering effects of a heterogeneous environment. The interaction between HPD and Originating Biome accounted for the fact that a synergistic effect may exist between the two in certain cases wherein the combination of both is worse on PGD than either has in isolation.

Results

Data acquisition & general trends

Among anthromes, mean HPD was highest in Urban biomes (888.97 persons km^{-2}) and lowest in Wild biomes (1.30 persons km^{-2}; Table 4.1), although standard deviation values were quite high (1773 and 23, respectively; Table 4.1). This was reflected in significant differences for HPD between all anthromes and the Urban biome; there were no significant differences in HPD between anthromes and the Wild biome (Figure 4.1A, Table S4.3). Mean PGD was highest in Freshwater and Semi-Natural biomes for MNA (10.09 and 8.58 respectively, Table 4.1), and in

Croplands for Ho (0.62; Figure 4.1b, d, Table 4.1). Rangeland (mean MNA = 6.28) and Ocean (mean Ho = 0.54) biomes had the lowest PGD (Table 4.1). The Semi-Natural biome had significantly higher MNA than the Wild biome ($p=0.027$, Table S3), and there was significantly lower Ho in the Wild biome compared to the Urban biome ($p=0.001$; Table S4.3, Figure 4.2). As with HPD, both MNA and Ho showed large standard deviations across biomes (Table 4.1).

Among taxa, there was a great deal of variability in HPD and PGD between populations originating from different biomes. Populations of Birds, Reptiles, and Amphibians tended to be found in regions having among the highest mean HPD (167, 166, 157 person km^{-2} on average, respectively), while Anadromous and Freshwater fish populations tended to experience lower HPD (60 and 97 person km^{-2} on average, respectively). Across biomes, Anadromous fish populations tended to have significantly higher mean PGD of all taxonomic groups, whereas Amphibians and Mammals tended to have significantly lower mean PGD across biomes (Figure 4.1, Tables S4.2, S4.3). As with the anthromes, taxa showed a great deal of variation around the mean for HPD, MNA, and Ho (max standard deviation of 947, 7.0, and 0.70, respectively; Table 4.1). Most taxa had a slightly negative, linear relationship between HPD and both PGD metrics (Figure 4.2). This relationship was only significant for Reptiles (Ho and MNA), Birds (MNA), and Mammals (MNA) (Table S4.5). Conversely, anadromous fish had a significantly positive relationship with HPD for Ho and MNA ($p<0.01$), potentially due to the aquatic nature of their habitat.

Habitat heterogeneity

For the metrics assessing habitat heterogeneity, we found that populations located in Natural biomes (Wild and Semi-Natural) were farthest on average from Urban biomes compared to populations in the remaining biomes (121km, 60.2km respectively, Figure 4.1c). Conversely, of the terrestrial biomes, Croplands and Rangelands were the farthest on average from Natural biomes (39km, 40km). Ocean biomes had the highest mean distance to either biome, 280km for Urban and 82.7km for Natural, likely reflecting populations from oceanic islands (Table S4.2). Surprisingly, most taxonomic groups showed an increase in PGD with increasing distance from Natural biomes, and a decrease in PGD as Urban distance increased (Figure 4.2, Table S4.5). The opposite trend was expected – an increase in PGD as distance away from urban biomes increases and a decrease in PGD as distance from wild biomes increases. Anadromous fish were

the only group which showed a significant positive relationship between PGD and distance from urban biomes, and negative relationship between PGD and distance from natural biomes, as expected ($p<0.005$, Table S4.5).

Populations showed an overall weak relationship with PGD as the POB metric increased, suggesting that accounting for the proportion of wild areas surrounding a population affected PGD (Figure 4.2G-H). Across both metrics, Anadromous fish had greater PGD as POB increased (Table S4.5) whereas Reptiles (Ho and MNA), Birds (MNA), and Mammals (MNA) decreased with increasing POB. Additionally, populations originating from one biome tended to be surrounded largely by the same biome. This was reflected in the mean percent per anthrome within each Originating Biome (Table S4.6) and can be visually inspected in Figure 4.3, where the colour of each Originating Biome appears to dominate its particular panel. For example, populations located in Croplands had a corresponding mean percent area of 14.7% for surrounding Croplands, while other biomes were <10% (Table S4.6, Figure 4.3). However, there was still a great deal of variance in the surrounding 100km of a population, demonstrating the extreme landscape heterogeneity an individual population likely experiences.

Human impacts on population genetic diversity

After model selection for the MNA model, the following four variables were selected in addition to the two random effects (Genus and RefID): distance to Urban biomes, distance to Natural biomes, POB, and the interaction between HPD and Originating Biome (Table 4.2, Figure 4.4). All variables were significant except for distance to Natural biomes, reflected by its consistently neutral effect ($p=0.32$; Figure 4.4); although the model including distance to Natural biomes had a lower AIC than the one excluding it (AIC 1304485 compared to 1304945, respectively). Distance to Urban biomes had a positive effect on MNA, where MNA increased as distance to Urban biome increased (Figure 4.4). MNA decreased at POB scores less than 4, increased between scores of 4 and 5, and then became decreased again at scores >6. In the interaction term for HPD and Originating Biome, MNA was disproportionally reduced in Urban, Village, and Rangeland biomes with higher HPD. Conversely, Croplands, Semi-Natural, and Wild biomes had positive or neutral (in Wild) effects on MNA with respect to increasing HPD. Finally, while Genus was shown to be significant and explained a great deal of variation in the

data ($p<0.001$, Table 4.3), RefID explained the most variation. This demonstrated that while there was variance among genera, the most variation in PGD originated between studies.

For Ho, the model selected included four variables in addition to the two random effects: distance to Urban biomes, distance to Natural biomes, POB, and Originating Biome (Table 4.2, Figure 4.5). As distance to Urban biomes increased, Ho was disproportionally reduced, whereas Ho increased positively with distance to Natural biomes, similar to what was found in the linear models (Figure 4.2, Figure 4.5). This was opposite to what was expected since increasing distance from Urban biomes was expected to show a positive effect on Ho, while increasing distance from Natural biomes was expected to have a negative effect. As with MNA, low POB scores showed a reduction in Ho , where Ho was reduced typically below a POB score of ~3, although there were fluctuations above a score of 4. Again, opposite to what we expected, the effect of Originating Biome on Ho was positive for Urban, Village, and Rangeland, and negative for Freshwater, Ocean, Semi-Natural, and Wild biomes (Table 4.3, Figure 4.5). This result would indicate that populations originating from urbanized biomes (Urban, Village, Rangeland) have more Ho than those originating from natural biomes (Semi-Natural, Wild). As with the MNA model, the effects of Genus and RefID were significant ($p<0.001$, Table 4.3) and accounted for a great deal of variability in relationships among and within taxa.

Discussion

While the conventional hypothesis is that humans would have a consistent negative impact on vertebrate genetic diversity, our results found inconsistent support for such an effect across the Americas. Considering our second and third hypotheses – that instead the effects on PGD depend on the type/intensity of human impact and should vary across taxa – perhaps this is not surprising (Millette et al., 2019; Schmidt et al., 2020). Overall, we found variable effects of the anthropogenic metrics on PDG and significant effects for Genus- and study-specific responses.

Similar to previous works, we found inconsistent and not always negative effects of human impacts on vertebrate PGD (Habrich et al., *submitted*; Millette et al., 2019; Schmidt et al., 2020). While HPD and Originating Biome did not have a strong effect on either metric of PGD when considered separately, when considered together they had an effect on MNA. The main trend revealed, as expected, that as HPD increased, MNA was reduced in Urban, Village, and

Rangeland biomes, whereas MNA increased with increasing HPD in Cropland and Semi-Natural biomes (the relationship was neutral for the Wild biome). Conversely, we found that Urban, Village, and Rangeland biomes had positive effects on Ho, but the more natural biomes (Semi-Natural, Wild) had a negative effect. We also found that as distance from an Urban biome increased, PGD decreased in MNA but increased in Ho; additionally, increasing distance away from a Natural biome resulted in an increase in Ho (neutral effect for MNA). That vertebrate Ho tended to decrease with increasing "wildness" (i.e. originating from Semi-Natural biomes and increasing distance from Urban) was a surprising result, since the relationship for MNA largely followed expectations. Since changes in MNA are manifested more quickly under population declines than changes in Ho (Allendorf, 1986; Nei et al., 1975), one explanation for the discrepancy between PGD metrics described above is that insufficient time has elapsed to detect a consistently negative genetic signal. For example, a previous study found only a 6% loss of within-population genetic variation (allelic diversity and expected heterozygosity) since the industrial revolution, a timeline that surpasses the timeline of genetic data assessed here (Leigh et al., 2019). While we attempted to assess temporal changes in human impact, we were unable to account for yearly variation by assigning anthropogenic values from the year of sampling to each population. Including such temporal variation in future studies would assist in identifying human impacts as they occur on PGD. Furthermore, species turnover may have already occurred in urban areas (Millette et al., 2019), which we hypothesize may explain why we found high PGD in those biomes for some taxa. Namely, the species now residing in urban environments may already be urban adapted and thus are not as impacted with respect to their PGD. Such inconsistencies were not entirely unexpected, however, as past studies have found increasing genetic diversity in regions of high human impact for some taxa, e.g. birds and certain mammalian genera (Habrich et al., *submitted*; Millette et al., 2019).

As proposed by our second hypothesis, assessing the type and intensity of human impact is important because certain impacts might reduce natural habitat entirely (i.e. construction of large cities), whereas others may only alter or shift the habitat (e.g. crop- or rangelands). To measure such variability in impact intensity we generated a metric that accounted for relative biome heterogeneity surrounding populations, the "Proportion-Of-Biomes (POB). This metric gave a relative "wildness" score to populations, where the higher the score, the more "wild" a population was. Our models showed that lower POB scores, reflective of less "wild" habitats,

had negative effects on PGD. Populations originating in Urban biomes had, on average, only 9.36% Urban area, 9.36% Croplands, 7.19% Semi-Natural, and 26.02% Ocean in the surrounding 100km (Table S4.6). Such extreme heterogeneity surrounding Urban biomes may suggest that the surrounding environment buffers potential negative effects from anthropogenic impacts. Alternatively, the taxa that were found to have higher PGD close to Urban biomes could be taxa that do well in the presence of humans (e.g. Racoons) and thus may not be as affected by human-induced habitat fragmentation (Habrich et al., *submitted*), or their habitat is not as strongly affected by HPD (e.g. aquatic species). One final explanation for the inconsistent decline of PGD within urban environments may be due to sampling bias, where researchers can only feasibly sample populations that are relatively close to urban biomes – it is difficult to obtain enough samples from remote populations to conduct population genetics studies (Lawrence et al. 2019). Thus, by chance, the populations sampled in the database could have higher genetic diversity than we would have otherwise expected from more remote, wild populations. However, our POB metric only used a buffer distance of 100km as the minimum distance from an urban centre from which most organisms may experience anthropogenic pressure (Cincotta et al., 2000; Repetto, 1994). Unfortunately, due to paucity of data on class-specific dispersal distances, we were unable to generate buffer distances to reflect differences among dispersal capabilities. This is a key improvement that future research should incorporate as it could reveal further nuances in taxonomic responses to human impacts.

Our third hypothesis that there would be taxonomic variability in how groups responded to anthropogenic impact was supported by our models. Previous studies only accounted for differences among classes (Millette et al., 2019; Schmidt et al., 2020), but here we found that models did not support Class or Family as important for explaining variation, instead Genus accounted for a great deal of the variability in the data. While we found some broad patterns across classes (e.g. anadromous fishes having highest PGD and inhabiting areas of lowest HPD), there existed more variation at the Genus level that would be unaccounted for if only Class or Family were considered. Within taxonomic classes there are different life history traits which may make a genus more or less sensitive to human impacts (Habrich et al., *submitted*). For example, while amphibians generally are sensitive to pollution sources due to their biology (i.e. their permeable skin), they may show greater overall declines in PGD in the presence of urbanization relative to classes such as mammals. Mammals have a great deal of variation in life

history traits and some rodents, in particular, may thrive in urban environments relative to other mammals. Behavioural differences between genera within a taxonomic group are also expected to influence response to anthropogenic impacts. Previous works have found that mammals in particular (no other studies have assessed herptiles) are negatively impacted by human impacts compared to birds (Habrich et al., *submitted*; Millette et al., 2019; Schmidt et al., 2020). However, there is still variation among mammalian genera and groups such as Racoons are not negatively influenced by impacts like road density, whereas Caribou are (Habrich et al., *submitted*). The identification of specific patterns and directional relationships among genera is a subject of interest that could further elucidate the exact influences of human impacts. Future studies should include interactions between Genus and each anthropogenic metric to identify such relationships. Due to the intense computational requirements to run such models, we were unable to include such interactions here.

Conclusion

Our assessment of human impacts on vertebrate PGD across the Americas found inconsistent results that vary according to impact type, taxonomic level, and even the metric of genetic diversity. However, we caution the reader not to assume that humans have minimal impact on vertebrate PGD simply due to the inconsistent results presented here. Other studies have found such inconsistencies (Millette et al., 2019; Schmidt et al., 2020), and it is a reminder that it is difficult to apply a broad-brush for targeting conservation. Reducing anthropogenic impact is ideal, but we acknowledge some species may have become adapted and even thrive within human environments. Nevertheless, a key result here was that declines in MNA were detected as HPD increased, but no such decline was detected for Ho. Since changes in MNA happen more quickly than Ho, this may signal rapid PGD declines in specific contexts and populations showing reduced MNA may need conservation prioritization. When developing conservation programs to minimize losses in genetic diversity, we recommend assessing responses in genetic diversity at least at the genus-level, rather than generalizing according to class. We also recognize that population-specific genetic monitoring would enhance such conservation programs such that the identification of at-risk or genetically diverse populations may be possible (Hoban et al. 2020; Paz-Vinas et al. 2018). The continued assessment of changes to PGD would allow for the detection of human-induced reductions, particularly if

genetic baselines can be better compared temporally, as we were not able to fully account for changes in PGD over time. Continued assessment and inclusion of PGD in conservation frameworks is especially important as global conservation targets are going to be set for “post-2020” by the Convention on Biological Diversity (Convention on Biological Diversity, 2010; Hoban et al., 2020). Our work here demonstrates that there is a great deal of variability among and within taxa, and that it is important to consider such variable responses to different types of human impact when developing conservation regimes.

Acknowledgements

We would like to thank Dr. Erle Ellis for providing access to the anthropogenic biome maps. Additionally, our funding sources which supported this work: NSERC Discovery Grant, QCBS Seed Grant, and a Concordia University Research Chair.

Tables

Table 4.1. Summary of genetic diversity metrics and human population density (HPD) for vertebrates across the American continents and anthropogenic biome, as defined by Klein Goldewijk et al. (2017). N = number of populations; sd = standard deviation, MNA= mean number of alleles, Ho = observed heterozygosity; HPD = human population density (person km^{-2}).

Biome	**N**	**MNA**	**sd(MNA)**	**Ho**	**sd(Ho)**	**HPD**	**sd(HPD)**
Urban	933	8.18	5.31	0.61	0.15	888.97	1773.39
Village	101	7.10	2.95	0.60	0.17	293.49	277.84
Croplands	1611	7.88	4.62	0.62	0.15	23.45	45.52
Rangeland	565	6.28	3.71	0.58	0.16	6.48	57.18
Semi-natural	2676	8.58	6.29	0.60	0.17	9.40	31.63
Wild	1779	7.92	5.57	0.58	0.15	1.30	23.37
Freshwater	54	10.09	4.01	0.55	0.12	81.40	234.54
Ocean	232	6.75	6.87	0.54	0.19	72.83	335.36
Amphibia	1042	7.43	5.66	0.57	0.17	157.32	573.83
Anadromous	1291	15.01	6.97	0.70	0.15	59.91	266.80
Aves	197	6.49	3.72	0.59	0.16	166.95	574.25
Freshwater	2444	7.11	4.03	0.57	0.17	97.00	722.39
Mammalia	1755	6.10	4.01	0.61	0.14	143.53	947.18
Reptilia	1222	6.68	3.56	0.59	0.15	165.82	532.19

Table 4.2. AIC comparison and model fit of select Generalized Additive Mixed Models GAMMs during model selection. Ho = observed heterozygosity; MNA = mean number of alleles; HPD = human population density (persons km^{-2}), anthrome, Natural = distance to Natural biomes (km), Urban = distance to Urban biomes (km), POB = weighted metric of proportion of biomes within 100km of a population; Anthrome = the anthrome that a population was located in. Variables are fit with a smoother (s) and denoted as a random effect by bs="re".

Model	**DF**	**AIC**	**Adj R^2**	**Deviance Explained (%)**
MNA ~ 1 + s(RefID, bs="re")	608.37	1390455	0.792	85.7
MNA ~ 1 + s(Genus, bs="re") + s(RefID, bs="re")	620.48	1369744	0.796	86.6
MNA~ s(HPD, by = Anthrome) + s(Genus, bs="re") + s(RefID, bs="re")	620.96	1313782	0.815	87.0
MNA~ s(Urban) + s(HPD, by = Anthrome) + s(Genus, bs="re") + s(RefID, bs="re")	628.56	1305617	0.821	87.4
MNA~ s(Urban) + s(HPD, by = Anthrome) + s(POB) + s(Natural) + s(Genus, bs="re") + s(RefID, bs="re")	636.87	1304485	0.823	87.4
Ho ~ 1 + s(RefID, bs="re")	729.39	-692321	0.768	81.1
Ho ~ 1 + s(Genus, bs="re") + s(RefID, bs="re")	755.88	-703014	0.776	81.8
Ho ~ s(Urban) + s(Genus, bs="re") + s(RefID, bs="re")	764.87	-707505	0.780	82.1
Ho ~ s(Urban) + s(POB) + s(Genus, bs="re") + s(RefID, bs="re")	773.84	-710859	0.782	82.4
Ho~ s(Urban) + s(POB) + Anthrome + s(Natural) + s(Genus, bs="re") + s(RefID, bs="re")	789.99	-713631	0.784	82.5

Table 4.3. Summary of final GAMMs selected through model selection for either observed heterozygosity (Ho) or mean number of alleles (MNA). HPD = human population density (persons km^{-2}), anthrome, Natural = distance to Natural biomes (km), Urban = distance to Urban biomes (km), POB = weighted metric of proportion of biomes within 100km of a population; Anthrome = the anthrome that a population was located in (i.e. Originating Biome). Variables are fit with a smoother s() and denoted as a random effect by bs="re". Bold values indicate statistical significance.

Dependent Variable	**Independent Variable**	**Estimate[a] edf[b]**	**St. Error[c] Ref.df[d]**	**Test statistic**	**p-value**
MNA	s(Urban)	4.645 [b]	9 [d]	485.442 [3]	**0.005**
	s(HPD):AnthromeCroplands	1.871 [b]	1.981 [d]	3.42 [3]	**0.025**
	s(HPD):AnthromeRangeland	1 [b]	1.001 [d]	2.746 [3]	0.097
	s(HPD):AnthromeSemi-natural	1 [b]	1.001 [d]	2.407 [3]	0.121
	s(HPD):AnthromeUrban	1.008 [b]	1.015 [d]	1.178 [3]	0.277
	s(HPD):AnthromeVillage	3.228 [b]	3.674 [d]	12.145 [3]	**<0.001**
	s(HPD):AnthromeWild	1 [b]	1 [d]	0.844 [3]	0.358
	s(POB)	2.816 [b]	9 [d]	72.915 [3]	**0.045**
	s(Natural)	2.231 [b]	9 [d]	14.621 [3]	0.322
	s(Genus, bs="re".)	84.198 [b]	154 [d]	1373.773 [3]	**<0.001**
	s(RefID, bs="re".)	523.965	663	135.84 [3]	**<0.001**
Ho	Intercept	0.428 [a]	0.039 [c]	11.10 [1]	**<0.001**
	AnthromeFreshwater	-0.025 [a]	0.010 [c]	-2.41 [1]	**0.016**
	AnthromeOcean	-0.056 [a]	0.009 [c]	-6.14 [1]	**<0.001**
	AnthromeRangeland	0.131 [a]	0.004 [c]	35.56 [1]	**<0.001**
	AnthromeSemi-natural	-0.027 [a]	0.004 [c]	-7.28 [1]	**<0.001**
	AnthromeUrban	0.021 [a]	0.004 [c]	5.02 [1]	**<0.001**
	AnthromeVillage	0.078 [a]	0.007 [c]	10.48 [1]	**<0.001**
	AnthromeWild	-0.052 [a]	0.004 [c]	-13.09 [1]	**<0.001**
	s(Urban)	8.868 [b]	9 [d]	11434875 [2]	**<0.001**
	s(POB)	8.923 [b]	9 [d]	19088482 [2]	**<0.001**

s(Natural)	8.885 [b]	9 [d]	3001017 [2]	**<0.001**
s(Genus, bs="re".)	87.458 [b]	158 [d]	916288019 [2]	**<0.001**
s(RefID, bs="re".)	666.431 [b]	729 [d]	411921271 [2]	**<0.001**

[1]z value; [2]Chi-square; [3]F value

Figures

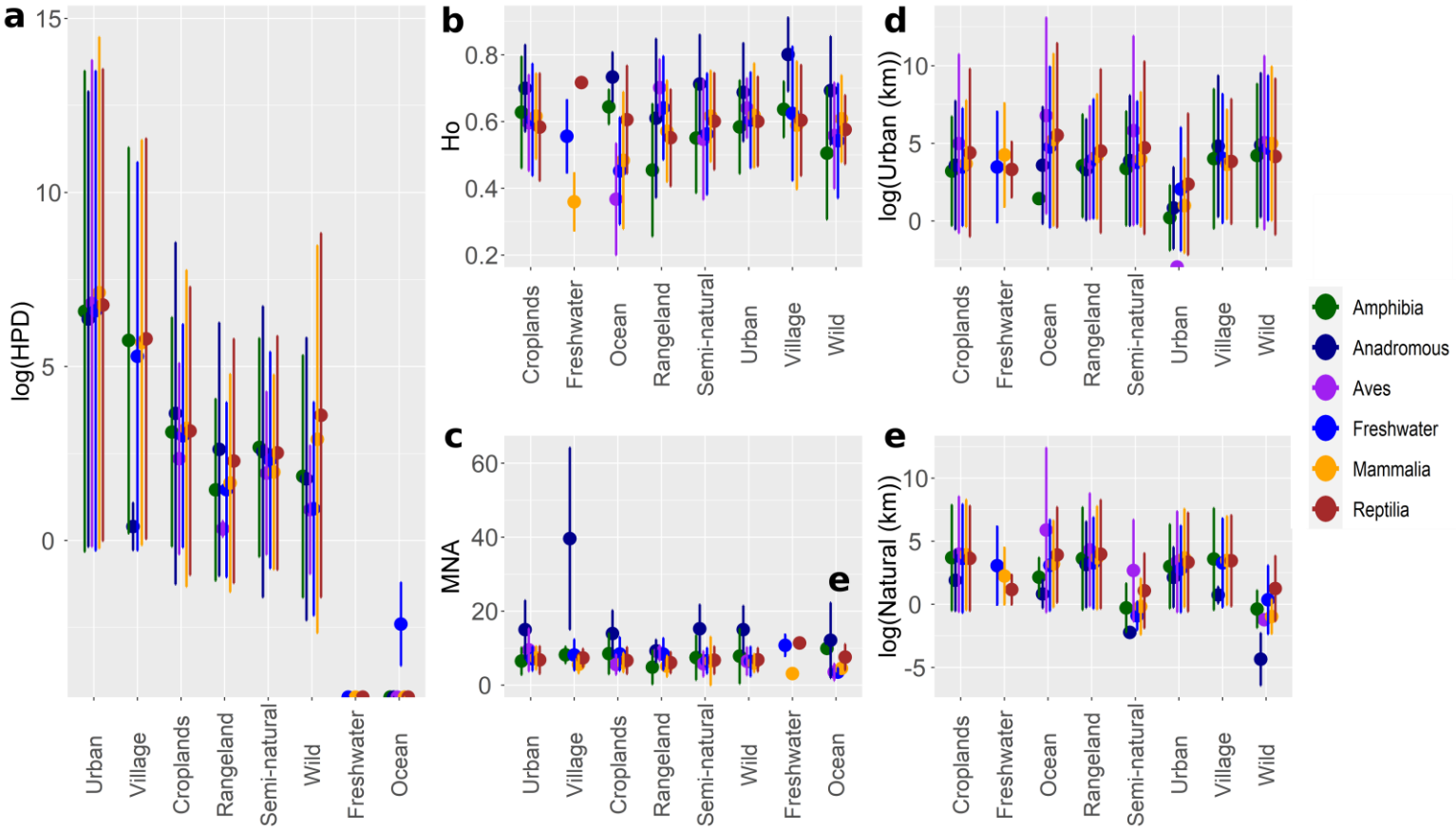

Figure 4.1. Mean (a) human population density (HPD, humans km^{-2}), (b) observed heterozygosity (Ho), (c) mean number of alleles (MNA), (d) distance to Urban biomes (Urban, km), and (e) distance to Natural biomes (Natural, km) for each anthropogenic biome and for each taxonomic group of vertebrates across the American continents (see Table S1 for sample size per group). Error bars represent standard deviation. For full statistical comparisons between groups for each genetic diversity metric see Table S4.3.

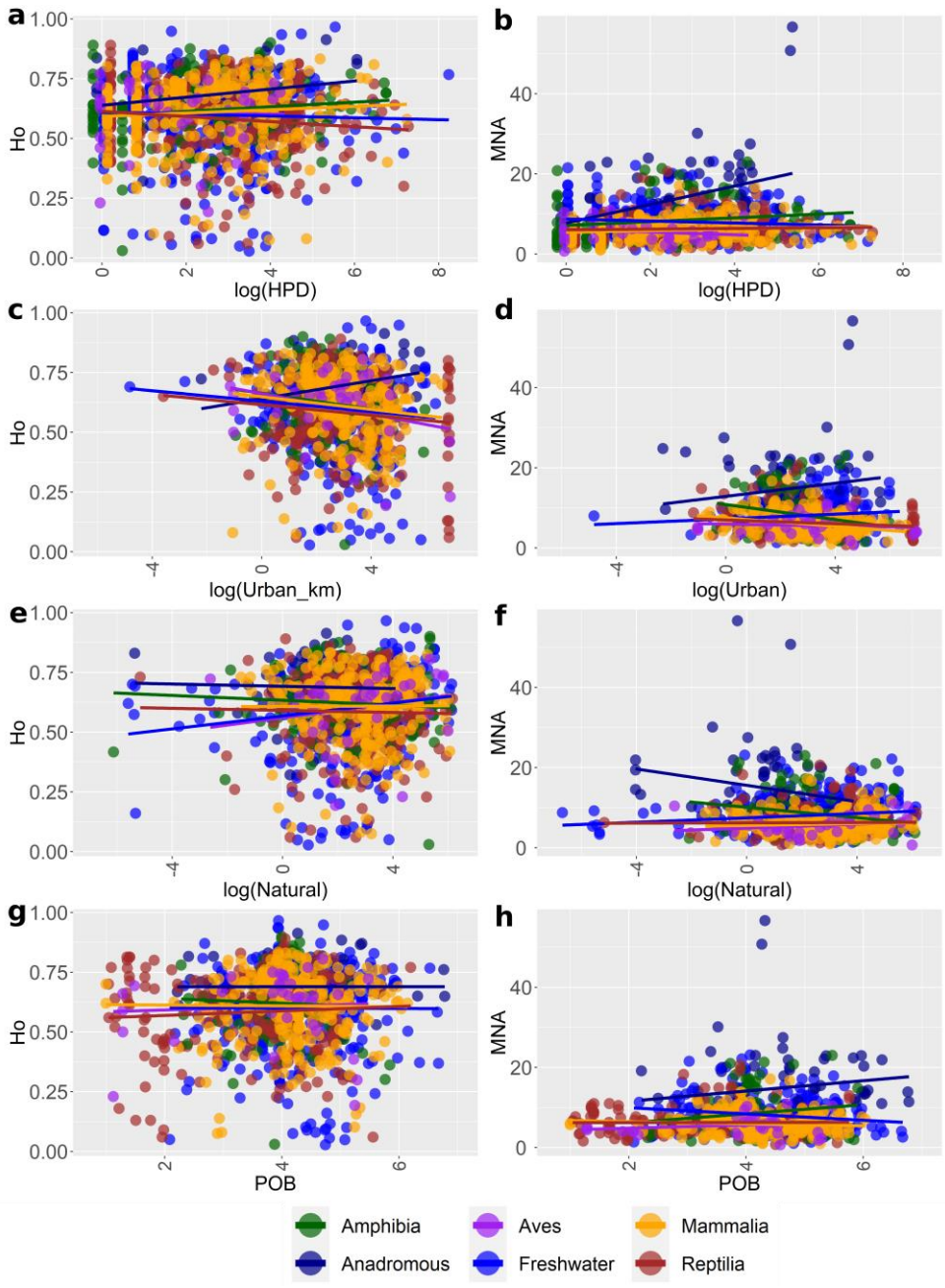

Figure 4.2. Linear relationships between metrics of anthropogenic impacts and genetic diversity metrics for vertebrates across the American continents. Genetic diversity metrics include (a, b) observed heterozygosity (Ho), and (c, d) mean number of alleles (MNA). Anthropogenic metrics include (a, b) log of human population density (HPD), (c, d) log of distance (km) to Urban biomes, (e, f), log of distance to Natural biomes (Semi-Natural and Wild together), and (g, h) the Proportion of Biome metric, providing a measure of "urbanization" within 100km of populations (see text for details).

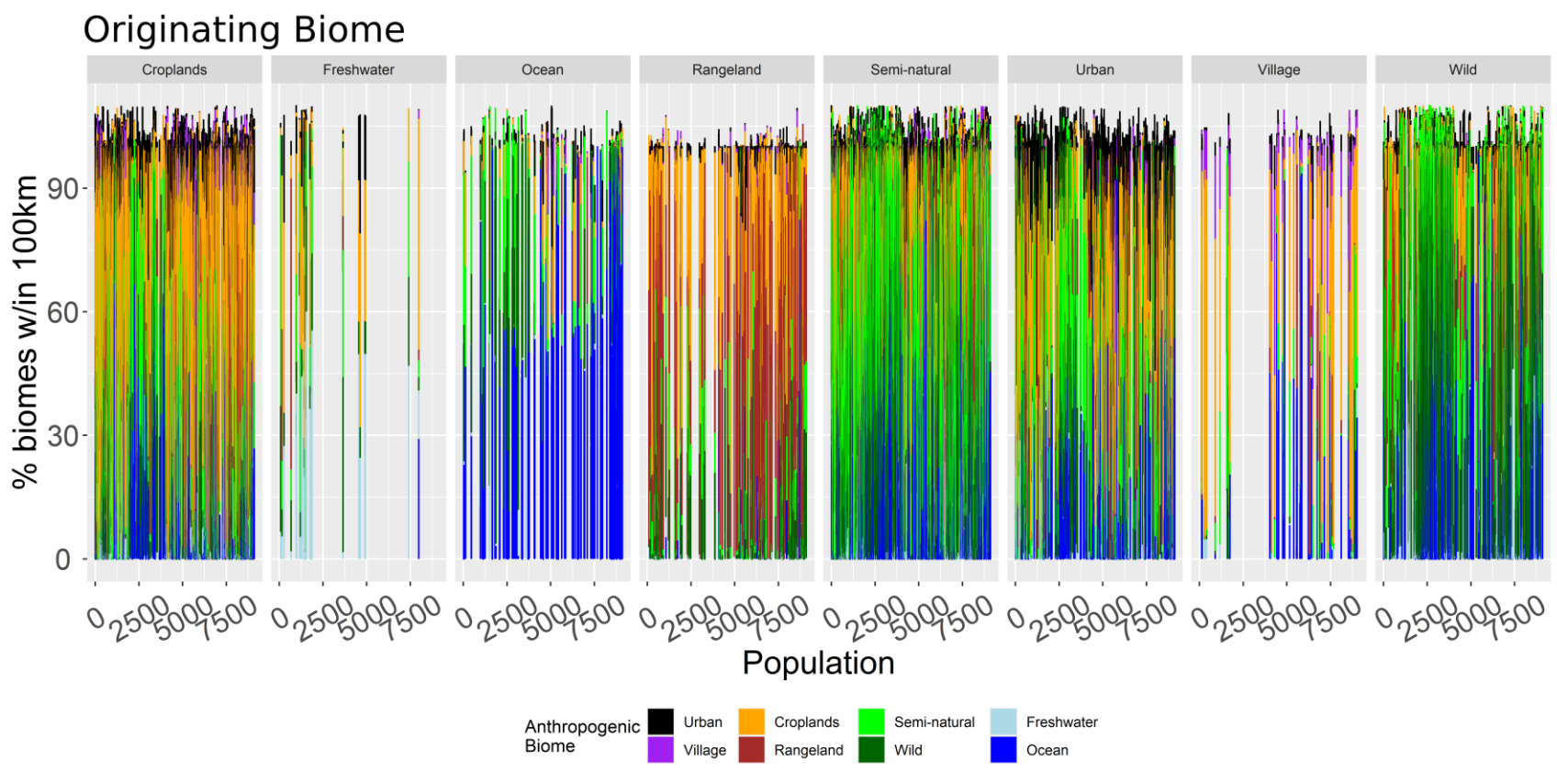

Figure 4.3. Percent of anthropogenic biomes, as defined by Klein Goldewijk et al. (2017), within 100km surrounding a vertebrate population within the American continents. Originating Biome indicates the biome a population was found in in 2010. The x axis indicates ordering of populations. Note some populations exceed 100% due to overlapping layers within the associated shapefiles.

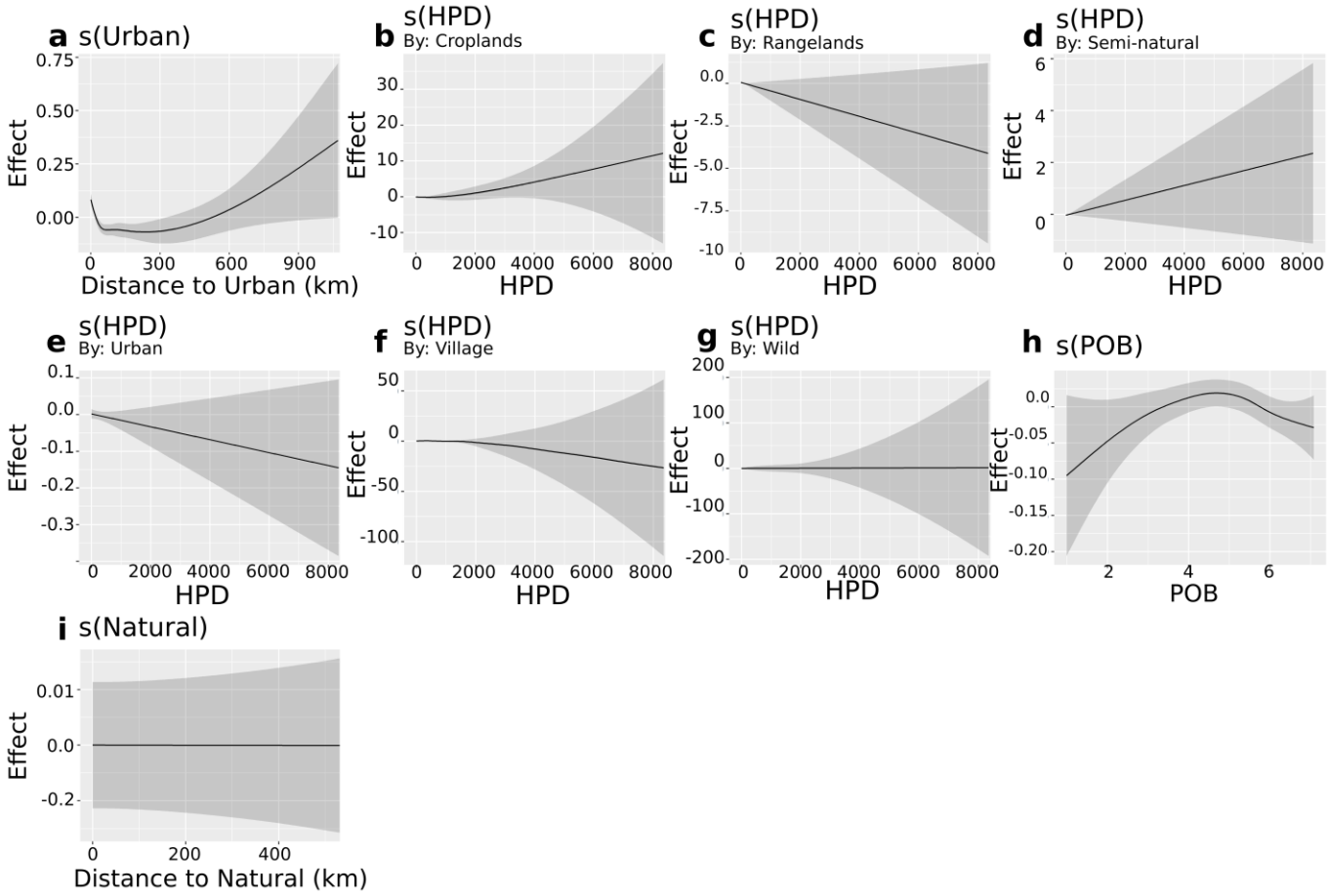

Figure 4.4. The predicted effect of anthropogenic variables selected through model selection for mean number of alleles MNA. Variables were fitted by smoothers (s) and include: (a) distance to nearest Urban biome (Urban, km), (b-g) the interaction between human population density (HPD, persons km^{-2}) and Originating Biome (Croplands, Rangelands, Semi-Natural, Urban, Village, Wild), (h) Proportion of Biome (POB), and (i) distance to nearest Natural biome (Natural, km). Confidence intervals represent standard error.

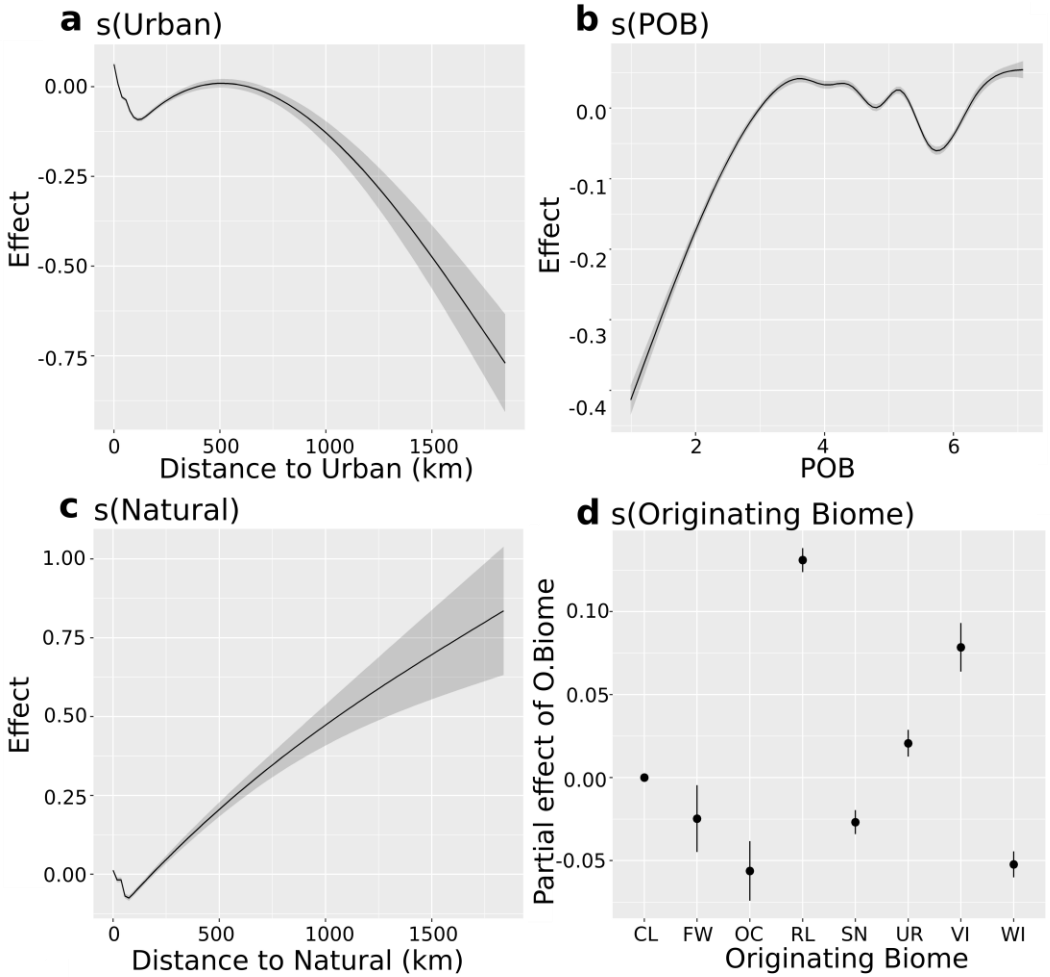

Figure 4.5. The predicted effect of anthropogenic variables selected through model selection for observed heterozygosity. Variables were fitted by smoothers (s) and include: (a) distance to nearest Urban biome (Urban, km), (b) Proportion of Biome (POB), (c) distance to nearest Natural biome (Natural, km), and (d) Originating Biome (CL = Croplands; FW = Freshwater; OC = Ocean; RL = Rangelands; SN = Semi-Natural; UR = Urban; VI =Villages; WI = Wild). Confidence intervals represent the standard error.

General Discussion

Intraspecific diversity is of increasing relevance and interest as technology has improved to properly assess genomic variation. As a result, this field is relatively young compared to the study of species diversity, particularly in the context of conceptual and empirical considerations across broad scales. These technological developments are especially important as Earth is currently undergoing a period of rapid change in the Anthropocene. Information from genomic data may provide insight and guidance for approaching conservation, particularly when derived at the population level. Many works are now demonstrating the usefulness of considering genetic diversity, for example, in conservation management regimes (Miraldo et al., 2016; Paz-Vinas et al., 2018; Willoughby et al., 2015). However, at the time of conducting this thesis, no syntheses of nuclear genetic diversity data had been conducted (chapter 1) to advance understanding of the distribution of genetic diversity across broad scales. Furthermore, no research had sought to explicitly examine two aspects of intraspecific diversity together: genetic diversity and population richness.

Evaluating broad-scale patterns in genetic diversity

There are many inconsistencies in the way previous studies have defined genetic diversity and how it has been measured at broad scales. Previous works that have tried to assess broad-scale genetic diversity not only have inconsistently defined genetic diversity, but often these data were collated from potentially unrelated sequences and/or is summed across spatial scales, as in the assessment of species diversity (Manel et al., 2020; Millette et al., 2019; Miraldo et al., 2016). The assumption that the summation of genetic diversity is of equivalent meaning/value as the summation of species diversity at broad scales poses an issue because genetic diversity is a metric that is derived from individuals forming populations. Genetic diversity is more than just a count that can be summed across species; it is a metric of variation and when it is summed in this way, variation among and within species is overlooked. A sum does not account for the possibility that a particular species or population may be disproportionately contributing to the high levels of genetic diversity in an area.

In Chapter two, I laid the conceptual groundwork for explicitly defining and predicting latitudinal gradients for population richness and genetic diversity. In this chapter it is clear that different ways of defining intraspecific diversity can change predictions, thus it is imperative that

definitions are explicit. For example, I contrasted genetic diversity when measured across species (TotGenDiv), compared to genetic diversity within a single species (GenPerSpp). I showed that these two ways of measuring genetic diversity have inverse latitudinal expectations. Species at high latitudes typically have larger range sizes than tropical species, thus they may have accumulated more genetic diversity across populations from local adaptation than a low-latitude species (Currie et al., 2004; Fine, 2015; Mittelbach et al., 2007; Siqueira et al., 2020). The most glaring point is that sums of genetic diversity across regions will produce results that are skewed by species-rich regions. As ecologists we must ask ourselves if that is a useful way of assessing genetic diversity at broad-scales, and in Chapter two I argue that it is not. Instead, assessing GenPerSpp – or incorporating a way to measure variance – is likely a better approach of assessing genetic diversity at broad scales. This way, patterns are not obfuscated by sampling design and instead more meaningful patterns of genetic diversity can be revealed.

In Chapter three, I tested the latitudinal distribution of genetic diversity – specifically population genetic diversity (PGD). I found limited evidence for the bell-shaped distribution that is often found for species diversity. However, I found that environmental variables have mediating effects on PGD, particularly for mean annual temperature, elevation, annual precipitation, and the interacting effect of elevation with mean annual temperature. These results contribute to the idea that genetic diversity must be considered and assessed differently than species diversity. Additionally, different aspects of genetic diversity – e.g. adaptive genetic diversity – may show differing patterns than what I have found for neutral PGD and merit attention in future investigations. Finally, as my work has focused on genetic diversity rooted at the population level, it thus provided an additional aspect for assessing biodiversity: population richness.

Evaluating population richness

Collectively, my thesis demonstrated both the importance and difficulties in studying the population unit as a richness metric. I also showed that the identification of broad-scale patterns for population richness is currently extremely limited. The lack of thorough sampling data for most taxa and the biases for which taxa are assessed leads to constraints for analysis. Drawing upon work by Hughes et al., which crudely estimated global population richness (Hughes et al., 1997), I initially attempted to construct rarefaction curves (similar to those used for species richness) to more thoroughly approximate population richness (Siegel, 2006). However, this

would have required a handful of well sampled representative species to extrapolate patterns onto under-sampled species with similar biology.

Chapter one was the first to explore broad-scale patterns in population richness and it quickly became obvious that thorough sampling of populations within species is needed to evaluate this level of intraspecific diversity properly. Unfortunately, this level of sampling is not available for most taxa and is largely prevalent in fishes which are often managed at the population level (Morellet et al., 2007; Reiss et al., 2009; Stephenson, 1999). Due to these constraints, I was unable to effectively evaluate a latitudinal gradient in population richness as I did for population genetic diversity.

Chapter two further demonstrated the importance of defining and assessing population richness. As population richness is the level of biodiversity between genetic diversity and species diversity, it can provide valuable information on species maintenance. Species that are more population rich are more likely to be able to persist, while species with fewer populations are most likely to be at risk of extinction. As with genetic diversity, I show that the definition used for population richness can change predictions – whether it is a count across all species (TotPopR), or a count within individual species (PopPerSpp). Using an explicit definition removes ambiguity when making and testing predictions; using shared language increases the ability for researchers to cooperate in a cohesive manner. With less time wasted on disagreeing about definitions, more can be spent on exploring the nuances of variability within an ecological context.

The importance of variability

A common result from my thesis work was that the variation among and within taxonomic groups is key to understanding broad-scale patterns. Each chapter demonstrates the importance of accounting for and investigating variability in data. Chapter one showed that there was significant variation of genetic diversity among taxonomic groups and among geographic regions. This was the first indication that variability can provide additional information that simply taking a mean or count cannot. Chapter two did not explicitly test for variability in data, but rather discussed variability in definitions for facets of biodiversity. This chapter clearly revealed that consistent, unambiguous definitions can remove variability between studies when assessing the same idea. It is important that definitions do not vary between studies aiming to evaluate the same metric, otherwise inconsistent conclusions will be common. Chapter three and

four accounted for variability within taxa, finding that Genus consistently accounted for a large proportion of variability in population genetic diversity. Indeed, when Genus was accounted for in generalized additive models, the latitudinal patterns became clearer (although statistically insignificant). Chapter four not only investigated taxonomic variability, but also variability in the habitat surrounding populations by including metrics that accounted for such heterogeneity (e.g. the Proportion-of-Biome metric). Together, all chapters indicated that further analyses should account for variation in similar ways, including the addition of more complicated analyses.

General Conclusion

Overall, my thesis disputes key assumptions drawn from the species-richness literature. Latitudinal gradients in genetic diversity are complicated, and my models found limited support for a bell-shaped latitudinal gradient in population genetic diversity. Instead, I found support for environmental variables mediating its distribution, and, perhaps more importantly, that genera-specific effects account for the most variability. These results would not be possible without the construction of the population-genetics database outlined in Chapter one.

My thesis highlights the importance of databases such as MacroPopGen. Databases are invaluable tools for analyses, including (but not limited to) the influences of geographic parameters, environmental variables, other biodiversity facets, and anthropogenic impacts. While my work was based on neutral genetic diversity due to the number of available studies, it reinforces the notion that future syntheses should include adaptive genetic diversity as this body of research grows. Adaptive genetic diversity may exhibit different patterns than neutral genetic diversity and could provide additional information for broad-scale biodiversity patterns generally, as well as implications for conservation (Brennan et al., 2019; Kirk & Freeland, 2011; Mittell et al., 2015; Paaby & Rockman, 2014; Paz-Vinas et al., 2018; Selkoe & Toonen, 2006).

Finally, from a conservation standpoint, my work adds to the growing body of research suggesting that in order to capture the vast variability in biodiversity, we must take a systematic approach to conservation (Paz-Vinas et al., 2018). Such an approach would entail collecting data to identify priority areas on a species-by-species basis and is increasingly supported when intraspecific diversity is incorporated into the species diversity perspective. My thesis demonstrated that hotspots of PGD within species do not overlap with species diversity, which may have implications for developing effective conservation strategies. The assessment of PGD allows for identification of populations that can be targeted for conservation – either populations with more or less PGD that could be preserved or rescued, respectively. Assessing broad-scale PGD within species allows for a more realistic representation of regions that are likely in need of conservation by identifying areas deficient in PGD that may be subject to anthropogenic pressures or otherwise. To best understand this world of ecology, we must put to use every tool available to us. This thesis has demonstrated the utility of a population-level assessment in understanding diversity and developing conservation strategies accordingly. While this is but a

first step, I strongly recommend further exploration and incorporation of other aspects of genetic diversity which are sure to elucidate patterns further.

Bibliography

Abell, R., Thieme, M. L., Revenga, C., Bryer, M., Kottelat, M., Bogutskaya, N., Coad, B., Mandrak, N., Balderas, S. C., Bussing, W., Stiassny, M. L. J., Skelton, P., Allen, G. R., Unmack, P., Naseka, A., Ng, R., Sindorf, N., Robertson, J., Armijo, E., Higgins, J. V., Heibel, T. J., Wikramanayake, E., Olson, D., López, H. L., Reis, R. E., Lundberg, J. G., Sabaj Pérez, M. H., Petry, P. (2008). Freshwater Ecoregions of the World: A New Map of Biogeographic Units for Freshwater Biodiversity Conservation. *BioScience*, *58*(5), 403.

Ackerly, D. D. (2003). Community assembly, niche conservatism, and adaptive evolution in changing environments. *International Journal of Plant Sciences*, *164*(S3), 164–184.

Adams, R. I., & Hadly, E. A. (2013). Genetic diversity within vertebrate species is greater at lower latitudes. *Evolutionary Ecology*, *27*(1), 133–143.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Allen, A. P., Gillooly, J. F., Savage, V. M., & Brown, J. H. (2006). Kinetic effects of temperature on rates of genetic divergence and speciation. *Proceedings of the National Academy of Sciences*, *103*(24), 9130–9135.

Allendorf, F.W., & Luikart, G. (2009). *Conservation and the Genetics of Populations*. John Wiley & Sons.

Allendorf, Fred W. (1986). Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology*, *5*(2), 181–190.

Allendorf, Fred W. (2017). Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology*, *26*(2), 420–430.

Anderson, D. R., & Burnham, K. P. (2002). Avoiding pitfalls when using information-theoretic methods. *The Journal of Wildlife Management*, 912–918.

Angers, B., & Bernatchez, L. (1998). Combined use of SMM and non-SMM methods to infer fine structure and evolutionary history of closely-related brook charr (*Salvelinus fontinalis*, Salmonidae) populations from microsatellites. *Molecular Biology and Evolution*, *15*(2), 143–159.

Antonovics, J. (1976). The input from population genetics: “The New Ecological Genetics.” *Systematic Botany*, *1*(3), 233–245.

Antonovics, J. (2003). Toward community genomics? *Ecology*, *84*(3), 598–601.

Araújo, M. B., Pearson, R. G., & Rahbek, C. (2005). Equilibrium of species' distributions with climate. *Ecography*, *28*(5), 693–695.

Ascensão, F., Mata, C., Malo, J. E., Ruiz-Capillas, P., Silva, C., Silva, A. P., Santos-Reis, M., & Fernandes, C. (2016). Disentangle the causes of the road barrier effect in small mammals through genetic patterns. *PLoS ONE*, *11*(3), 1–23.

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*(4), 729–744.

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, *23*(1), 38–44.

Barrueto, M., Ford, A. T., & Clevenger, A. P. (2014). Anthropogenic effects on activity patterns of wildlife at crossing structures. *Ecosphere*, *5*(3).

Bazin, E., Glemin, S., & Galtier, N. (2006). Population size does not influence mitochondrial genetic diversity in animals. *Science*, *312*, 570–572.

Belmar-Lucero, S., Wood, J. L. A., Scott, S., Harbicht, A. B., Hutchings, J. A., & Fraser, D. J. (2012). Concurrent habitat and life history influences on effective/census population size ratios in stream-dwelling trout. *Ecology and Evolution*, *2*(3), 562–573.

Bernatchez, L. (2016). On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. *Journal of Fish Biology*, *89*(6), 2519–2556.

Bernatchez, L., & Wilson, C. C. (1998). Comparative phylogeography of Nearctic and Paleartic fishes. *Molecular Ecology*, *7*, 431–452.

Bernos, T. A., & Fraser, D. J. (2016). Spatiotemporal relationship between adult census size and genetic population size across a wide population size gradient. *Molecular Ecology*, *25*(18), 4472–4487.

BirdLife International. (2017). *Bird species distribution maps of the world. Version 2017.2.* http://datazone.birdlife.org/species/requestdis

Blanchet, S., Prunier, J. G., & De Kort, H. (2017). Time to Go Bigger: Emerging Patterns in Macrogenetics. *Trends in Genetics*, *33*(9), 579–580.

Bohonak, A. J. (1999). Dispersal, Gene Flow, and Population Structure. *The Quarterly Review of Biology*, *74*(1), 21–45.

Botero, C. A., Dor, R., McCain, C. M., & Safran, R. J. (2014). Environmental harshness is positively correlated with intraspecific divergence in mammals and birds. *Molecular Ecology*, *23*(2), 259–268.

Brennan, R. S., Garrett, A. D., Huber, K. E., Hargarten, H., & Pespeni, M. H. (2019). Rare genetic variation and balanced polymorphisms are important for survival in global change conditions. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1904), 20190943.

Brown, J. H. (2014). Why are there so many species in the tropics? *Journal of Biogeography*, *41*(1), 8–22.

Brum, F. T., Graham, C. H., Costa, G. C., Hedges, S. B., Penone, C., Radeloff, V. C., Rondinini, C., Loyola, R., & Davidson, A. D. (2017). Global priorities for conservation across multiple dimensions of mammalian diversity. *Proceedings of the National Academy of Sciences*, *114*(29), 7641–7646.

Buckley, L. B., Jonathan Davies, T., Ackerly, D. D., Kraft, N. J. B., Harrison, S. P., Anacker, B. L., Cornell, H. V., Dänischen, E. I., Grytnes, J. A., Hawkins, B. A., McCain, C. M., Stephens, P. R., & Wiens, J. J. (2010). Phylogeny, niche conservatism and the latitudinal diversity gradient in mammals. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1691), 2131–2138.

Canadian Endangered Species Conservation Council. (2001). *Wild Species 2000: The general status of species in Canada*. Minister of Public Works and Government Services Canada.

Cardillo, M., Purvis, A., Sechrest, W., Gittleman, J. L., Bielby, J., & Mace, G. M. (2004). Human population density and extinction risk in the world's carnivores. *PLoS Biology*, *2*(7), 909–914.

Caughley, G. (1994). Directions in Conservation Biology. *Journal of Animal Ecology*, *63*(2), 215–244.

Ceballos, G. (2002). Mammal Population Losses and the Extinction Crisis. *Science*, *296*(5569), 904–907.

Ceballos, G., Ehrlich, P. R., & Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences*, 201704949.

Cheptou, P.-O., Hargreaves, A. L., Bonte, D., & Jacquemyn, H. (2017). Adaptation to fragmentation: evolutionary dynamics driven by human influences. *Philosophical*

Transactions of the Royal Society B: Biological Sciences, *372*(1712), 20160037.

Cincotta, R. P., Wisnewski, J., & Engelman, R. (2000). Human population in the biodiversity hotspots. *Nature (London)*, *404*(6781), 990–992.

Clark, A. G., Hubisz, M. J., Bustamante, C. D., Williamson, S. H., & Nielsen, R. (2005). Ascertainment bias in studies of human genome-wide polymorphism. *Genome Research*, *15*(11), 1496–1502.

Coltman, D. W., & Slate, J. (2003). Microsatellite measures of inbreeding: A meta-analysis. *Evolution*, *57*(3), 971–983.

Comps, B., Gömöry, D., Letouzey, J., Thiébaut, B., & Petit, R. J. (2001). Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European beech. *Genetics*, *157*(1), 389–397.

Convention on Biological Diversity. (2010). Strategic Plan for Biodiversity 2011–2020, including Aichi Biodiversity Targets. Nagoya, Japan: The Convention on Biological Diversity.

Corander, J., Majander, K. K., Cheng, L., & Merilä, J. (2013). High degree of cryptic population differentiation in the baltic sea herring *Clupea harengus*. *Molecular Ecology*, *22*(11), 2931–2940.

Costello, M. J., May, R. M., & Stork, N. E. (2013). Can we name Earth's species before they go extinct? *Science*, *339*(6118), 413–416.

Currie, D. J., Mitte-lbach, G. G., Cornell, H. V., Field, R., Guégan, J. F., Hawkins, B. A., Kaufman, D. M., Kerr, J. T., Oberdorff, T., O'Brien, E., & Turner, J. R. G. (2004). Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology Letters*, *7*(12), 1121–1134.

Des Roches, S., Post, D. M., Turley, N. E., Bailey, J. K., Hendry, A. P., Kinnison, M. T., Schweitzer, J. A., & Palkovacs, E. P. (2018). The ecological importance of intraspecific variation. *Nature Ecology and Evolution*, *2*(1), 57–64.

DeWoody, J. A., & Avise, J. C. (2000). Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *Journal of Fish Biology*, *56*(3), 461–473.

DiBattista, J. D. (2008). Patterns of genetic variation in anthropogenically impacted populations. *Conservation Genetics*, *9*(1), 141–156.

Dynesius, M., & Jansson, R. (2000). Evolutionary consequences of changes in species'

geographical distributions driven by Milankovitch climate oscillations. *Proceedings of the National Academy of Sciences*, *97*(16), 9115–9120.

Eckert, C. G. G. G., Samis, K. E. E., & Lougheed, S. C. C. C. (2008). Genetic variation across species' geographical ranges: The central-marginal hypothesis and beyond. *Molecular Ecology*, *17*(5), 1170–1188.

Economo, E. P., Narula, N., Friedman, N. R., Weiser, M. D., & Guénard, B. (2018). Macroecology and macroevolution of the latitudinal diversity gradient in ants. *Nature Communications*, *9*(1), 1–8.

Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, *17*(7), 422–433.

Ellegren, H., Moore, S., Robinson, N., Byrne, K., Ward, W., & Sheldon, B. C. (1997). Microsatellite evolution--a reciprocal study of repeat lengths at homologous loci in cattle and sheep. *Molecular Biology and Evolution*, *14*(8), 854–860.

Ellis, E. C., Goldewijk, K. K., Siebert, S., Lightman, D., & Ramankutty, N. (2010). Anthropogenic transformation of the biomes, 1700 to 2000. *Global Ecology and Biogeography*, *19*(5), 589–606.

Ellis, E. C., & Ramankutty, N. (2008). Putting people in the map: Anthropogenic biomes of the world. *Frontiers in Ecology and the Environment*, *6*(8), 439–447.

Fahrig, L., Arroyo-rodríguez, V., Bennett, J. R., Lalonde, V. B., Currie, D. J., Eigenbrod, F., Ford, A. T., Harrison, S. P., Jaeger, J. A. G., Martin, A. E., Martin, J., Metzger, J. P., Morrison, P., Rhodes, J. R., Saunders, D. A., Simberloff, D., Smith, A. C., Tischendorf, L., Vellend, M., & Watling, J. I. (2018). Is habitat fragmentation bad for biodiversity? *Biological Conservation*, *230*, 1–17.

Fan, H., Zhang, Q., Rao, J., Cao, J., & Lu, X. (2019). Genetic diversity-area relationships across bird species. *The American Naturalist*, 1–21.

Fedorov, A. A. (1966). The structure of the tropical rain forest and speciation in the humid tropics. *Journal of Ecology*, *54*(1), 1–11.

Fedorov, V. B., & Stenseth, N. C. (2002). Multiple glacial refugia in the North American Arctic: inference from phylogeography of the collared lemming (Dicrostonyx groenlandicus). *Proceedings of the Royal Society B: Biological Sciences*, *269*(1505), 2071–2077.

Fine, P. V. A. (2015). Ecological and evolutionary drivers of geographic variation in species

diversity. *Annual Review of Ecology, Evolution, and Systematics*, *46*(1), 369–392.

Fonseca, E. M., Werneck, F. P., Gehara, M., Oliveira, E. F., Magalhães, F. de M., Lanna, F. M., Lima, G. S., Marques, R., Mesquita, D. O., Costa, G. C., Colli, G. R., & Garda, A. A. (2019). The role of strict nature reserves in protecting genetic diversity in a semiarid vegetation in Brazil. *Biodiversity and Conservation*, *28*(11), 2877-2890.

Frankham, R. (1996). Relationship of genetic variation to population size in wildlife. *Conservation Biology*, *10*(6), 1500–1508.

Frankham, R. (1997). Do island populations have less genetic variation than mainland populations? *Heredity*, *78*(3), 311–327.

Frankham, R., Ballou, J. D., & Briscoe, D. A. (2002). *Introduction to conservation genetics*. Cambridge Univ. Press.

Fraser, D. J., Walker, L., Yates, M. C., Marin, K., Wood, J. L. A., Bernos, T. A., & Zastavniouk, C. (2019). Population correlates of rapid captive-induced maladaptation in a wild fish. *Evolutionary Applications*, *September 2017*, 1–13.

Galbreath, K. E., & Cook, J. A. (2004). Genetic consequences of Pleistocene glaciations for the tundra vole (*Microtus oeconomus*) in Beringia. *Molecular Ecology*, *13*, 135–148.

Gaston, K. J. (2000). Global patterns in biodiversity. *Nature*, *405*(6783), 220–227.

Gebremedhin, B., Ficetola, G. F., Naderi, S., Rezaei, H. R., Maudet, C., Rioux, D., & Taberlet, P. (2009). Frontiers in identifying conservation units: from neutral markers to adaptive genetic variation. *Animal Conservation*, *12*(2), 107–109.

Ghalambor, C. K. (2006). Are mountain passes higher in the tropics? Janzen's hypothesis revisited. *Integrative and Comparative Biology*, *46*(1), 5–17.

Ghalambor, C. K., McKay, J. K., Carroll, S. P., & Reznick, D. N. (2007). Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, *21*(3), 394–407.

Gibbs, H. K., Ruesch, A. S., Achard, F., Clayton, M. K., Holmgren, P., Ramankutty, N., & Foley, J. A. (2010). Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proceedings of the National Academy of Sciences*, *107*(38), 16732–16737.

Gillman, L. N., Keeling, D. J., Ross, H. A., & Wright, S. D. (2009). Latitude, elevation and the tempo of molecular evolution in mammals. *Proceedings of the Royal Society B: Biological*

Sciences, *276*(1671), 3353–3359.

Gillooly, J., Allen, A., West, G., & Brown, J. (2005). The rate of DNA evolution: Effects of body size and temperature on the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(1), 140–145.

Goossens, B., Chikhi, L., Ancrenaz, M., Lackman-Ancrenaz, I., Andau, P., & Bruford, M. W. (2006). Genetic signature of anthropogenic population collapse in orang-utans. *PLoS Biology*, *4*(2), 285–291.

Species at Risk Act, Pub. L. No. S. C. 2002, c. 29 (2002).

Gratton, P., Marta, S., Bocksberger, G., Winter, M., Keil, P., Trucchi, E., & Kühl, H. (2017a). Which latitudinal gradients for genetic diversity? *Trends in Ecology & Evolution*, *32*(10), 724–726.

Gratton, P., Marta, S., Bocksberger, G., Winter, M., Trucchi, E., & Kühl, H. (2017b). A world of sequences: Can we use georeferenced nucleotide databases for a robust automated phylogeography? *Journal of Biogeography*, *44*(2), 475–486.

Green, D. M., Sharbel, T. F., Kearsley, J., & Kaiser, H. (1996). Postglacial range fluctuation, genetic subdivision and speciation in the Western North American spotted frog complex, Rana pretiosa. *Evolution*, *50*(1), 374–390.

Guo, Q. (2012). Incorporating latitudinal and central– marginal trends in assessing genetic variation across species ranges other factors. *Molecular Ecology*, *21*(22), 5396–5403.

Habrich, A., Lawrence, E. R., & Fraser, D. J. (submitted). Varying genetic imprints of roads and human density in North American mammals. *Evolutionary Applications*.

Hansson, B., & Westerberg, L. (2002). On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology*, *11*, 2467–2474.

Hargreaves, A. L., & Eckert, C. G. (2019). Local adaptation primes cold-edge populations for range expansion but not warming-induced range shifts. *Ecology Letters*, *22*(1), 78–88.

He, F., & Hubbell, S. P. (2011). Species-area relationships always overestimate extinction rates from habitat loss: Supplementary Information. *Nature*, *473*(7347).

Hébert, C., Danzman, R. G., Jones, M. W., & Bernatchez, L. (2000). Hydrography and population genetic structure in brook charr (*Salvelinus fontinalis*, Mitchill) from eastern Canada. *Molecular Ecology*, *9*, 971–982.

Hendry, A. P., Farrugia, T. J., & Kinnison, M. T. (2008). Human influences on rates of

phenotypic change in wild animal populations. *Molecular Ecology*, *17*(1), 20–29.

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*(6789), 907–913.

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.

Hirao, A. S., Watanabe, M., Tsuyuzaki, S., Shimono, A., Li, X., Masuzawa, T., & Wada, N. (2017). Genetic diversity within populations of an arctic–alpine species declines with decreasing latitude across the Northern Hemisphere. *Journal of Biogeography*, *44*(12), 2740–2751.

Hoban, S., Bruford, M., Jackson, J. D. U., Lopes-Fernandes, M., Heuertz, M., Hohenlohe, P. A., Paz-Vinas, I., Sjögren-Gulve, P., Segelbacher, G., Vernesi, C., Aitken, S., Bertola, L. D., Bloomer, P., Breed, M., Rodríguez-Correa, H., Funk, W. C., Grueber, C. E., Hunter, M. E., Jaffe, R., Liggins, L., Mergeay, J., Moharrek, F., O'Brien, D., Ogden, R., Palma-Silva, C., Pierson, J., Ramakrishnanab, U., Simo-Droissart, M., Tani, N., Waits, L., & Laikre, L.(2020). Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. *Biological Conservation*, *248*, 108654.

Huang, S., Stephens, P. R., & Gittleman, J. L. (2012). Traits, trees and taxa: Global dimensions of biodiversity in mammals. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1749), 4997–5003.

Hughes, A. R., Inouye, B. D., Johnson, M. T. J., Underwood, N., & Vellend, M. (2008). Ecological consequences of genetic diversity. *Ecology Letters*, *11*(6), 609–623.

Hughes, J. B., Daily, G. C., & Ehrlich, P. R. (1997). Population diversity: Its extent and extinction. *Science*, *278*(5338), 689–692.

Hurst, G. D. D., & Jiggins, F. M. (2005). Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1572), 1525–1534.

Imhoff, M. L., & Bounoua, L. (2006). Exploring global patterns of net primary production carbon supply and demand using satellite observations and statistical data. *Journal of Geophysical Research Atmospheres*, *111*(22), 1–8.

Imhoff, M. L., Bounoua, L., Ricketts, T., Loucks, C., Harriss, R., & Lawrence, W. T. (2004). *HANPP Collection: Global Patterns in Net Primary Productivity (NPP)*. Palisades, NY:

NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/https://doi.org/10.7927/H40Z715X

IUCN. (2016). *The IUCN Red List of Threatened Species. Version 2016-1*. http://www.iucnredlist.org

Jaenike, J. R. . (1973). A Steady State Model of Genetic Polymorphism on Islands. *The American Naturalist*, *107*(958), 793–795.

Janzen, D. H. (1967). Why mountain passes are higher in the tropics. *The American Naturalist*, *101*(919), 233–249.

Jarne, P., & Lagoda, P. J. L. (1996). Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*, *11*(10), 424–429.

Johnson, M. T. J., & Munshi-South, J. (2017). Evolution of life in urban environments. *Science*, *358*(6363).

Johnson, W. E., Onorato, D. P., Roelke, M. E., Land, E. D., Cunningham, M., Belden, R. C., McBride, R., Jansen, D., Lotz, M., Shindle, D., Howard, J. G., Wildt, D. E., Penfold, L. M., Hostetler, J. A., Oli, M. K., & O'Brien, S. J. (2010). Genetic restoration of the Florida panther. *Science*, *329*(5999), 1641–1645.

Jump, A. S., Marchant, R., & Peñuelas, J. (2009). Environmental change and the option value of genetic diversity. *Trends in Plant Science*, *14*(1), 51–58.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017a). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, *4*, 170122.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017b). *Data from: Climatologies at high resolution for the earth's land surface areas*. Dryad Digital Repository. https://doi.org/10.5061/dryad.kd1d4

Kennedy, J. D., Borregaard, M. K., Marki, P. Z., Machac, A., Fjeldså, J., & Rahbek, C. (2018). Expansion in geographical and morphological space drives continued lineage diversification in a global passerine radiation. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1893), 20182181.

Kirk, H., & Freeland, J. R. (2011). Applications and implications of neutral versus non-neutral markers in molecular ecology. *International Journal of Molecular Sciences*, *12*(6), 3966–

3988.

Klein Goldewijk, K., Beusen, A., Doelman, J., & Stehfest, E. (2017). Anthropogenic land use estimates for the Holocene – HYDE 3.2. *Earth System Science Data*, *9*, 927–953.

Lamanna, C., Blonder, B., Violle, C., Kraft, N. J. B., Sandel, B., Šímová, I., Donoghue, J. C., Svenning, J. C., McGill, B. J., Boyle, B., Buzzard, V., Dolins, S., Jørgensen, P. M., Marcuse-Kubitza, A., Morueta-Holme, N., Peet, R. K., Piel, W. H., Regetz, J., Schildhauer, M., Spencer, N., Thiers, B., Wiser, S K., Enquist, B. J. (2014). Functional trait space and the latitudinal diversity gradient. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(38), 13745–13750.

Lamy, T., Laroche, F., David, P., Massol, F., & Jarne, P. (2017). The contribution of species–genetic diversity correlations to the understanding of community assembly rules. *Oikos*, *126*(6), 759–771.

Laroche, F., Jarne, P., Lamy, T., David, P., & Massol, F. (2014). A neutral theory for interpreting correlations between species and genetic diversity in communities. *The American Naturalist*, *185*(1), 59–69.

Lawrence, E. R., Benavente, J. N., Matte, J.-M., Marin, K., Wells, Z., Bernos, T. A., Krasteva, N., Habrich, A., Nessel, G., Koumrouyan, R. A., & Fraser, D. J. (2019). Geo-referenced population-specific microsatellite data across American continents, the MacroPopGen Database. *Scientific Data*, *6*(1), 14.

Lawrence, E. R., & Fraser, D. J. (2020). Latitudinal biodiversity gradients at three levels: linking species richness, population richness, and genetic diversity. *Global Ecology and Biogeography*, *29*(5), 770–788.

Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Ségurel, L., Venkat, A., Andolfatto, P., & Przeworski, M. (2012). Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biology*, *10*(9), e1001388.

Leigh, D. M., Hendry, A. P., Vázquez-Domínguez, E., & Friesen, V. L. (2019). Estimated six per cent loss of genetic variation in wild populations since the industrial revolution. *Evolutionary Applications*, *12*(8), 1505–1512.

Leroy, G., Carroll, E. L., Bruford, M. W., DeWoody, J. A., Strand, A., Waits, L., & Wang, J. (2018). Next-generation metrics for monitoring genetic erosion within populations of conservation concern. *Evolutionary Applications*, *11*(7), 1066–1083.

Lewinsohn, T. M., & Prado, P. I. (2005). How many species are there in Brazil? *Conservation Biology*, *19*(3), 619–624.

Li, H., & Wiens, J. J. (2019). Time Explains Regional Richness Patterns within Clades More Often than Diversification Rates or Area. *The American Naturalist*, *193*(4), 514–529.

Manel, S., Guerin, P. E., Mouillot, D., Blanchet, S., Velez, L., Albouy, C., & Pellissier, L. (2020). Global determinants of freshwater and marine fish genetic diversity. *Nature Communications*, *11*(1), 1–9.

Marchese, C. (2015). Biodiversity hotspots: A shortcut for a more complicated concept. *Global Ecology and Conservation*, *3*, 297–309.

Marchesini, A., Vernesi, C., Battisti, A., & Ficetola, G. F. (2018). Deciphering the drivers of negative species-genetic diversity correlation in Alpine amphibians. *Molecular Ecology*, *27*(23), 4916-4930.

Marske, K. A., Rahbek, C., & Nogués-Bravo, D. (2013). Phylogeography: Spanning the ecology-evolution continuum. *Ecography*, *36*(11), 1169–1181.

Martin, P. R., & Mckay, J. K. (2004). Latitudinal variation in genetic divergence of populations and the potential for future speciation. *Evolution*, *58*(5), 938–945.

Martinez, A. S., Willoughby, J. R., & Christie, M. R. (2018). Genetic diversity in fishes is influenced by habitat type and life-history variation. *Ecology and Evolution*, *8*(23), 12022–12031.

Matthews, E. (1982). Global vegetation and land use: New high-resolution data bases for climate studies. *Journal of Climate and Applied Meteorology*, *22*, 474–487.

McGill, B. J., Dornelas, M., Gotelli, N. J., & Magurran, A. E. (2015). Fifteen forms of biodiversity trend in the anthropocene. *Trends in Ecology and Evolution*, *30*(2), 104–113.

Medina, I., Cooke, G. M., & Ord, T. J. (2018). Walk, swim or fly? Locomotor mode predicts genetic differentiation in vertebrates. *Ecology Letters*, *21*(5), 638–645.

Meiri, S., Roll, U., Grenyer, R., Feldman, A., Novosolov, M., & Bauer, A. M. (2017). Data from: The global distribution of tetrapodsreveals a need for targeted reptile conservation. *Dryad Digital Repository*. https://doi.org/https://doi.org/10.5061/dryad.83s7k

Meyer, C. F. J., Kalko, E. K. V, & Kerth, G. (2009). Small-scale fragmentation effects on local genetic diversity in two phyllostomid bats with different dispersal abilities in Panama. *Biotropica*, *41*(1), 95–102.

Miller, E. C., & Román-Palacios, C. (2019). Evolutionary time explains the global distribution of freshwater fish diversity. *BioRxiv*, 668079.

Millette, K. L., Fugère, V., Debyser, C., Greiner, A., Chain, F. J. J., & Gonzalez, A. (2019). No consistent effects of humans on animal genetic diversity worldwide. *Ecology Letters*, *23*(1), 55–67.

Mimura, M., Yahara, T., Faith, D. P., Vázquez-Domínguez, E., Colautti, R. I., Araki, H., Javadi, F., Núñez-Farfán, J., Mori, A. S., Zhou, S., Hollingsworth, P. M., Neaves, L. E., Fukano, Y., Smith, G. F., Sato, Y. I., Tachida, H., & Hendry, A. P. (2017). Understanding and monitoring the consequences of human impacts on intraspecific variation. *Evolutionary Applications*, *10*(2), 121–139.

Miraldo, A., Li, S., Borregaard, M. K., Flórez-rodríguez, A., Gopalakrishnan, S., Rizvanovic, M., Wang, Z., Rahbek, C., Marske, K. A., & Nogués-bravo, D. (2016). An Anthropocene map of genetic diversity. *Science*, *353*(6307), 1532–1535.

Mittelbach, G. G., Schemske, D. W., Cornell, H. V., Allen, A. P., Brown, J. M., Bush, M. B., Harrison, S. P., Hurlbert, A. H., Knowlton, N., Lessios, H. A., McCain, C. M., McCune, A. R., McDade, L. A., McPeek, M. A., Near, T. J., Price, T. D., Ricklefs, R. E., Roy, K., Sax, D. F., Schluter, D., Sobel, J. M., Turelli, M. (2007). Evolution and the latitudinal diversity gradient: Speciation, extinction and biogeography. *Ecology Letters*, *10*(4), 315–331.

Mittell, E. A., Nakagawa, S., & Hadfield, J. D. (2015). Are molecular markers useful predictors of adaptive potential? *Ecology Letters*, *18*(8), 772–778.

Morellet, N., GAILLARD, J. M., Hewison, A. M., Ballon, P., Boscardin, Y. V. E. S., Duncan, P., Klein, F., & Maillard, D. (2007). Indicators of ecological change: new tools for managing populations of large herbivores. *Journal of Applied Ecology*, *44*(3), 634-643.

Munguía, M., Townsend Peterson, A., & Sánchez-Cordero, V. (2008). Dispersal limitation and geographical distributions of mammal species. *Journal of Biogeography*, *35*(10), 1879–1887.

Nei, M., Maruyama, T., & Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution*, *29*(1), 1–10.

Nielsen, R. (2004). Population genetic analysis of ascertained SNP data. *Human Genomics*, *1*(3), 218–224.

Nowakowski, A. J., Frishkoff, L. O., Thompson, M. E., Smith, T. M., & Todd, B. D. (2018).

Phylogenetic homogenization of amphibian assemblages in human-altered habitats across the globe. *Proceedings of the National Academy of Sciences*, *115*(15), E3454–E3462.

Orton, M. G., May, J. A., Ly, W., Lee, D. J., & Adamowicz, S. J. (2019). Is molecular evolution faster in the tropics? *Heredity*, *122*(5), 513–524.

Otto, S. P. (2018). Adaptation, speciation and extinction in the Anthropocene. *Proceedings. Biological Sciences*, *285*(1891), 20182047.

Paaby, A. B., & Rockman, M. V. (2014). Cryptic genetic variation: Evolution's hidden substrate. *Nature Reviews Genetics*, *15*(4), 247–258.

Paz-Vinas, I., Loot, G., Hermoso, V., Veyssière, C., Poulet, N., Grenouillet, G., & Blanchet, S. (2018). Systematic conservation planning for intraspecific genetic diversity. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1877), 20172746.

Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: An introduction with mgcv. *PeerJ*, *2019*(5).

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.

Pentinsaari, M., Salmela, H., Mutanen, M., & Roslin, T. (2016). Molecular evolution of a widely-adopted taxonomic marker (COI) across the animal tree of life. *Scientific Reports*, *6*(1), 1–12.

Pereira, H. M. (2016). A latitudinal gradient for genetic diversity. *Science*, *353*(6307), 1494–1495.

Peterson, A. T., Soberon, J., & Sanchez-Cordero, V. (1999). Conservatism of ecological niches in evolutionary time Conservatism of Ecological Niches in Evolutionary Time. *Science*, *285*, 1265–1268.

Pfeiffer, V. W., Ford, B. M., Housset, J., McCombs, A., Blanco-Pastor, J. L., Gouin, N., Manel, S., & Bertin, A. (2018). Partitioning genetic and species diversity refines our understanding of species–genetic diversity relationships. *Ecology and Evolution*, *8*(24), 12351–12364.

Pianka, E. R. (1966). Latitudinal gradients in species diversity: A review of concepts. *The American Naturalist*, *100*(910), 33–46.

Pierce, A. A., Gutierrez, R., Rice, A. M., & Pfennig, K. S. (2017). Genetic variation during range expansion: effects of habitat novelty and hybridization. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1852), 20170007.

Pontarp, M., Bunnefeld, L., Cabral, J. S., Etienne, R. S., Fritz, S. A., Gillespie, R., Graham, C. H., Hagen, O., Hartig, F., Huang, S., Jansson, R., Maliet, O., Münkemüller, T., Pellissier, L., Rangel, T. F., Storch, D., Wiegand, T., & Hurlbert, A. H. (2019). The Latitudinal Diversity Gradient: Novel understanding through mechanistic eco-evolutionary models. *Trends in Ecology & Evolution*, *34*(3), 211–223.

Provan, J., & Bennett, K. D. (2008). Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology and Evolution*, *23*(10), 564–571.

Raffard, A., Santoul, F., Cucherousset, J., & Blanchet, S. (2019). The community and ecosystem consequences of intraspecific diversity: a meta-analysis. *Biological Reviews*, *94*, 648–661.

Ranathunge, C., Wheeler, G. L., Chimahusky, M. E., Kennedy, M. M., Morrison, J. I., Baldwin, B. S., Perkins, A. D., & Welch, M. E. (2018). Transcriptome profiles of sunflower reveal the potential role of microsatellites in gene expression divergence. *Molecular Ecology*, *27*(5), 1188–1199.

Reed, D. H., & Frankham, R. (2003). Correlation between Fitness and Genetic Diversity. *Conservation Biology*, *17*(1), 230–237.

Reiss, H., Hoarau, G., Dickey-Collas, M., & Wolff, W. J. (2009). Genetic population structure of marine fish: mismatch between biological and fisheries management units. *Fish and Fisheries*, *10*(4), 361-395.

Repetto, R. (1994). The" second India" revisited: population poverty and environmental stress over two decades. *WRI Publications Brief*, 163–170.

Rey, O., Danchin, E., Mirouze, M., Loot, C., & Blanchet, S. (2016). Adaptation to global change: A transposable element-epigenetics perspective. *Trends in Ecology and Evolution*, *31*(7), 514–526.

Ricklefs, R. E. ., & Latham, R. E. (1992). Intercontinental Correlation of Geographical Ranges Suggests Stasis in Ecological Traits of Relict Genera of Temperate Perennial Herbs. *The American Naturalist*, *139*(6), 1305–1321.

Riley, S. P. D., Pollinger, J. P., Sauvajot, R. M., York, E. C., Bromley, C., Fuller, T. K., & Wayne, R. K. (2006). A southern California freeway is a physical and social barrier to gene flow in carnivores. *Molecular Ecology*, *15*(7), 1733–1741.

Roberts, J. H., Angermeier, P. L., & Hallerman, E. M. (2013). Distance, dams and drift: What structures populations of an endangered, benthic stream fish? *Freshwater Biology*, *58*(10),

2050–2064.

Rohde, K. (1992). Latitudinal gradients in species diversity: The search for the primary cause. *Oikos*, *65*(3), 514–527.

Roll, U., Feldman, A., Novosolov, M., Allison, A., Bauer, A. M., Bernard, R., Böhm, M., Castro-Herrera, F., Chirio, L., Collen, B., Colli, G. R., Dabool, L., Das, I., Doan, T. M., Grismer, L. L., Hoogmoed, M., Itescu, Y., Kraus, F., Lebreton, M., Lewin, A., Martins, M., Maza, E., Meirte, D., Nagy, Z. T., Nogueira, C. d. C., Pauwels, O. S. G., Pincheira-Donoso, D., Powney, G. D., Sindaco, R., Tallowin, O. J. S., Torres-Carvajal, O., Trape, J-F., Vidan, E., Uetz, P., Wagner, P., Wang, Y., Orme, C. D. L., Grenyer, R., Meiri, S. (2017). The global distribution of tetrapods reveals a need for targeted reptile conservation. *Nature Ecology and Evolution*, *1*(11), 1677–1682.

Rosenzweig, M. L. (1995). *Species diversity in space and time*. Cambridge University Press.

Ruggiero, A., & Werenkraut, V. (2007). One-dimensional analyses of Rapoport's rule reviewed through meta-analysis. *Global Ecology and Biogeography*, *16*(4), 401-414.

Rybicki, J., & Hanski, I. (2013). Species–area relationships and extinctions caused by habitat loss and fragmentation. *Ecology letters*, *16*, 27-38.

Saccheri, I., Kuussaari, M., Kankare, M., Vikman, P., Fortelius, W., & Hanski, I. (1995). Inbreeding and extinction in a butterfly metapopulation. *Nature*, *392*, 491–494.

Safi, K., Cianciaruso, M. V., Loyola, R. D., Brito, D., Armour-Marshall, K., & Diniz-Filho, J. A. F. (2011). Understanding global patterns of mammalian functional and phylogenetic diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1577), 2536–2544.

Sandel, B., Arge, L., Dalsgaard, B., Davies, R. G., Gaston, K. J., Sutherland, W. J., & Svenning, J.-C. (2011). The influence of Late Quaternary climate-change velocity on species endemism. *Science*, *334*(6056), 660–664.

Santini, L., Isaac, N. J. B. B., Maiorano, L., Ficetola, G. F., Huijbregts, M. A. J. J., Carbone, C., & Thuiller, W. (2018). Global drivers of population density in terrestrial vertebrates. *Global Ecology and Biogeography*, *27*(8), 968–979.

Schemske, D. W., & Mittelbach, G. G. (2017). “Latitudinal Gradients in Species Diversity”: Reflections on Pianka’s 1966 Article and a Look Forward. *The American Naturalist*, *189*(6), 599–603.

Schlötterer, C. (2004). The evolution of molecular markers — just a matter of fashion? *Nature Reviews Genetics*, *5*(1), 63–69.

Schluter, D. (2016). Speciation, ecological opportunity, and latitude. *The American Naturalist*, *187*(1), 1–18.

Schluter, D., & Pennell, M. W. (2017). Speciation gradients and the distribution of biodiversity. *Nature*, *546*(7656), 48–55.

Schmidt, C., Domaratzki, M., Kinnunen, R. P., Bowman, J., & Garroway, C. J. (2020). Continent-wide effects of urbanization on bird and mammal genetic diversity. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1920), 20192497.

Sebastián-González, E., Barbosa, J. M., Pérez-garcía, J. M., Morales-Reyes, Z., Botella, F., Olea, P. P., Mateo-tomás, P., Moleón, M., Hiraldo, F., Arrondo, E., Donázar, J. A., Cortés-avizanda, A., Selva, N., Lambertucci, S. A., Bhattacharjee, A., Brewer, A., Anadón, J. D., Abernethy, E., Rhodes Jr, O. E., Turner, K., Beasley, J. C., DeVault, T. L., Ordiz, A., Wikenros, C., Zimmermann, B., Wabakken, P., Wilmers, C. C., Smith, J. A., Kendall, C. J., Ogada, D., Buechley, E. R., Frehner, E., Allen, M. L., Wittmer, H. U., Butler, J. R. A., du Toit, J. T., Read, J., Wilson, D., Jerina, K., Krofel, M., Kostecke, R., Inger, R., Samson, A., Naves-Alegre, L., Sánchez-Zapata, J. A. (2019). Scavenging in the Anthropocene: human impact drives vertebrate scavenger species richness at a global scale. *Global Change Biology*, *25*(9), 3005–3017.

Seidel, H. S., Rockman, M. V, & Kruglyak, L. (2008). Widespread genetic incompatibility in C. elegans maintained by balancing selection. *Science*, *319*(5863), 589–594.

Selkoe, K. A., & Toonen, R. J. (2006). Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. *Ecology Letters*, *9*(5), 615–629.

Servín, J., Sánchez-cordero, V., & Gallina, S. (2003). Distances traveled daily by coyotes, Canis latrans, in a pine-oak forest in Durango, Mexico. *Journal of Mammalogy*, *84*(2), 547–552.

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.

Siegel, A. F. (2006). Rarefaction curves. *Encyclopedia of Statistical Sciences*, *10*, 1–3.

Singhal, S., Huang, H., Grundler, M. R., Marchán-Rivadeneira, M. R., Holmes, I., Title, P. O., Donnellan, S. C., & Rabosky, D. L. (2018). Does population structure predict the rate of speciation? A comparative test across Australia's most diverse vertebrate radiation. *The*

American Naturalist, *192*(4), 432–447.

Siqueira, T., Saito, V. S., Bini, L. M., Melo, A. S., Petsch, D. K., Landeiro, V. L., Tolonen, K. T., Jyrkankallio-Mikkola, J., Soininen, J., & Heino, J. (2020). Community size can affect the signals of ecological drift and niche selection on biodiversity. *Ecology*, e03014.

Sluijs, A., Schouten, S., Pagani, M., Woltering, M., Brinkhuis, H., Damsté, J. S. S., Dickens, G. R., Huber, M., Reichart, G.-J., Stein, R., Matthiessen, J., Lourens, L. J., Pedentchouk, N., Backman, J., Moran, K., & the Expedition 302 Scientists. (2006). Subtropical Arctic Ocean temperatures during the Palaeocene/Eocene thermal maximum. *Nature*, *441*(7093), 610–613.

Smith, B. T., Seeholzer, G. F., Harvey, M. G., Cuervo, A. M., & Brumfield, R. T. (2017). A latitudinal phylogeographic diversity gradient in birds. *PLoS Biology*, *15*(4), 1–24.

Stanley, R. R. E., DiBacco, C., Lowen, B., Beiko, R. G., Jeffery, N. W., Van Wyngaarden, M., Bentzen, P., Brickman, D., Benestan, L., Bernatchez, L., Johnson, C., Snelgrove, P. V. R., Wang, Z., Wringe, B. F., & Bradbury, I. R. (2018). A climate-associated multispecies cryptic cline in the northwest Atlantic. *Science Advances*, *4*(3), eaaq0929.

Stephens, P. R. R., & Wiens, J. J. J. (2003). Explaining species richness from continents to communities: The time-for-speciation effect in emydid turtles. *The American Naturalist*, *161*(1), 112–128.

Stephenson, R. L. (1999). Stock complexity in fisheries management: a perspective of emerging issues related to population sub-units. *Fisheries Research*, *43*(1–3), 247–249.

Stevens, G. C. (1989). The latitudinal gradient in geographical range: How so many species coexist in the tropics. *The American Naturalist*, *133*(2), 240–256.

Stewart, J. R., Lister, A. M., Barnes, I., & Dalen, L. (2010). Refugia revisited: Individualistic responses of species in space and time. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1682), 661–671.

Stork, N. E. (1993). How many species are there? *Biodiversity and Conservation*, *2*, 215–232.

Sutherland, G. D., Harestad, A. S., Price, K., & Lertzman, K. P. (2000). Scaling of natal dispersal distances in terrestrial birds and mammals. *Ecology and Society*, *4*(1).

Tamkee, P., Parkinson, E., & Taylor, E. B. (2010). The influence of Wisconsonian glaciation and contemporary stream hydrology on microsatellite DNA variation in rainbow trout (*Oncorhynchus mykiss*). *Canadian Journal of Fisheries and Aquatic Sciences*, *67*, 919–935.

Tatarenkov, A., Healey, C. I. M., & Avise, J. C. (2010). Microgeographic population structure of green swordail fish: Genetic differentiation despite abundant migration. *Molecular Ecology*, *19*(2), 257–268.

Taylor, E. B. (1999). Species pairs of north temperate freshwater fishes: Evolution, taxonomy, and conservation. *Reviews in Fish Biology and Fisheries*, *9*(4), 299–324.

Terborgh, J. (1973). On the Notion of Favorableness in Plant Ecology. *The American Naturalist*, *107*(956), 481–501.

Thuiller, W., Gravel, D., Francesco, G., Lavergne, S., Münkemüller, T., Pollock, L. J., Zimmermann, N. E., & Mazel, F. (2020). Productivity begets less phylogenetic diversity but higher uniqueness than expected. *Journal of Biogeography*, *47*(1), 44–58.

Tringali, M. D., & Bert, T. M. (1998). Risk to effective population size should be an important consideration in fish stock enhancement programs. *Bulletin of Marine Science*, *62*(2), 641–659.

Underwood, Z. E., Mandeville, E. G., & Walters, A. W. (2016). Population connectivity and genetic structure of burbot (*Lota lota*) populations in the Wind River Basin, Wyoming. *Hydrobiologia*, *765*(1), 329–342.

Endangered Species Act of 1973 As amended through the 108th Congress, Endangered Species Act Of 1973 (2003). papers://a25d78e7-d6e0-4b2b-9b83-8d9ae6ec0592/Paper/p134

Usinowicz, J., Chang-Yang, C. H., Chen, Y. Y., Clark, J. S., Fletcher, C., Garwood, N. C., Hao, Z., Johnstone, J., Lin, Y., Metz, M. R., Masaki, T., Nakashizuka, T., Sun, I. F., Valencia, R., Wang, Y., Zimmerman, J. K., Ives, A. R., & Wright, S. J. (2017). Temporal coexistence mechanisms contribute to the latitudinal gradient in forest diversity. *Nature*, *550*(7674), 105–108.

Väli, Ü., Einarsson, A., Waits, L., & Ellegren, H. (2008). To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology*, *17*(17), 3808–3817.

Vellend, M. (2005). Species diversity and genetic diversity: parallel processes and correlated patterns. *The American Naturalist*, *166*(2), 199–215.

Vellend, M. (2010). Conceptual Synthesis in Community Ecology. *The Quarterly Review of Biology*, *85*(2), 183–206.

Vellend, M., & Geber, M. A. (2005). Connections between species diversity and genetic

diversity. *Ecology Letters*, *8*(7), 767–781.

Vellend, M., Lajoie, G., Bourret, A., Múrria, C., Kembel, S. W., & Garant, D. (2014). Drawing ecological inferences from coincident patterns of population- and community-level biodiversity. *Molecular Ecology*, *23*(12), 2890–2901.

Wallace, A. R. (1878). *Tropical nature, and other essays*. Macmillan and Company.

Waples, R. S. (1998). Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, *89*(5), 438–450.

Waples, R. S., & Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, *15*(6), 1419–1439.

Weider, L. J., Lampert, W., Wessels, M., Colbourne, J. K., & Limburg, P. (1997). Long-term genetic shifts in a microcrustacean egg bank associated with anthropogenic changes in the Lake Constance ecosystem. *Proceedings of the Royal Society B: Biological Sciences*, *264*(1388), 1613–1618.

Weir, B., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370.

Weir, J. T., & Schluter, D. (2007). The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science*, *315*(5818), 1574–1576.

Wiehe, T. (1998). The effect of selective sweeps on the variance of the allele distribution of a linked multiallele locus: hitchhiking of microsatellites. *Theoretical Population Biology*, *53*(3), 272–283.

Wiens, J. J. (2004). Speciation and ecology revisited: Phylogenetic niche conservatism and the origin of species. *Evolution*, *58*(1), 193–197.

Wiens, J. J., & Donoghue, M. J. (2004). Historical biogeography, ecology and species richness. *Trends in Ecology and Evolution*, *19*(12), 639–644.

Willi, Y., Fracassetti, M., Zoller, S., & Van Buskirk, J. (2018). Accumulation of mutational load at the edges of a species range. *Molecular Biology and Evolution*, *35*(4), 781–791.

Willi, Y., Van Buskirk, J., & Hoffmann, A. A. (2006). Limits to the Adaptive Potential of Small Populations. *Annual Review of Ecology, Evolution, and Systematics*, *37*(1), 433–458.

Willig, M. R., Kaufman, D. M., & Stevens, R. D. (2003). Latitudinal gradients of biodiversity: patterns, scale, and synthesis. *Annual Review of Ecology, Evolution, and Systematics*, *34*(1),

273–309.

Willoughby, J. R., Harder, A. M., Tennessen, J. A., Scribner, K. T., & Christie, M. R. (2018). Rapid genetic adaptation to a novel environment despite a genome-wide reduction in genetic diversity. *Molecular Ecology*, *20*(20), 4041–4051.

Willoughby, J. R., Sundaram, M., Wijayawardena, B. K., Kimble, S. J. A., Ji, Y., Fernandez, N. B., Antonides, J. D., Lamb, M. C., Marra, N. J., & DeWoody, J. A. (2015). The reduction of genetic diversity in threatened vertebrates and new recommendations regarding IUCN conservation rankings. *Biological Conservation*, *191*, 495–503.

Willoughby, J. R., Sundaram, M., Wijayawardena, B. K., Lamb, M. C., Kimble, S. J. A., Ji, Y., Fernandez, N. B., Antonides, J. D., Marra, N. J., & Andrew Dewoody, J. (2017). Biome and migratory behaviour significantly influence vertebrate genetic diversity. *Biological Journal of the Linnean Society*, *121*(2), 446–457.

Wofford, J. E. B., Gresswell, R. E., & Banks, M. A. (2005). Influence of barriers to movement on within-watershed genetic variation of coastal cutthroat trout. *Ecological Applications*, *15*(2), 628–637.

World Wildlife Fund. (2017). Living Planet Report Canada: A national look at wildlife loss. *World Wildlife Fund*.

Wright, S. (1931). Evolution in mendelian populations. *Genetics*, *16*(2), 97–159.

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, *15*, 323–354.

Zhang, Q. G., Lu, H. S., & Buckling, A. (2018). Temperature drives diversification in a model adaptive radiation. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1886), 20181515.

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. Springer Science & Business Media.

Appendices

Appendix 1

Full MacroPopGen database can be found at:

Lawrence, Elizabeth R; N. Benavente, Javiera; Matte, Jean-Michel; Marin, Kia; Wells, Zachery; Bernos, Thaïs A.; et al. (2019): MacroPopGen Database: Geo-referenced population-specific microsatellite data across the American continents. figshare. Dataset. https://doi.org/10.6084/m9.figshare.7207514.v1

Appendix 2

Supplementary Methods

Data acquisition

All population data used in the analyses were obtained from the population genetic database, *MacroPopGen* (Lawrence et al., 2019). Briefly, the database collated data from microsatellite studies that investigate population structure of vertebrate species in the American continent, focusing on amphibians, birds, freshwater and anadromous fish, mammals, and reptiles. Note we did not include brackish or catadromous species in any of our analyses. Population-specific genetic data obtained from the database for each population included: observed heterozygosity (Ho), mean number of alleles (MNA), latitude and longitude coordinates, and number of microsatellite loci. This data was supplemented with species range maps obtained from IUCN (IUCN, 2016) as well as from BirdLife International (Bird Life International 2017) and Meiri et al. (2017). We calculated the area of species range size in kilometres squared after projecting in World Behrmann equal area projection.

H1: Geographic Distribution Hypothesis

To test the hypothesis that range size is associated with an increase in the number of populations, we were able to obtain range shapefiles for 625 of the 897 species in the database (Table A2.1; raw file "SppRange-PopsWithin.csv" available on Dryad: https://doi.org/10.5061/dryad.xgxd254ck), and then calculated the area and the latitudinal extent of each range in square kilometres and decimal degrees, respectively. Next, we used a generalized linear model using the R package lme4 where the number of populations per species was our response variable; we chose to model number of populations with the gamma distribution because data were positively skewed. Fixed effects included latitudinal extent of the range, the natural logarithm of range size and Class (Amphibia, Aves, anadromous fish, freshwater fish, Mammalia, and Reptilia) (Table A2.1). The interaction between class and range size was also included to account for differences among taxonomic groups. Latitudinal extent was calculated by determining the decimal degrees latitude between the maximum and minimum latitudinal points of a range.

H2: Overlapping Range Hypothesis

First a 500km by 500km grid size was generated across the American continents in QGIS v3.2.2, using the World Behrman equal area projection; this produced 250 grid cells. Next the

number of unique species within each cell was calculated, along with the total number of populations (TotPopR), and the number of populations for each species within a cell (PopPerSpp) ("PopulationsPerSpp_Grid500km.csv" available on Dryad: https://doi.org/10.5061/dryad.xgxd254ck). Note that this data only represents that of which has been sampled with microsatellites and therefore sampling is not completely representative across species. Using this data, we used two linear models where the number of populations (either PopPerSpp or TotPopR) was the response variable, and the number of species was the fixed effect (Table A2.2). For the PopPerSpp model, TotPopR was an additional fixed effect.

H3: Range-Restricted Gene Hypothesis

Using the species ranges from the aforementioned sets of data (BirdLife International, 2017; IUCN, 2016; Meiri et al., 2017) combined with the genetic data (Lawrence et al. 2019), we completed forwards model selection for two generalized linear mixed models using the glmmTMB packages in R (file "MacroPopGen_Database_final_areas.csv" available on Dryad: https://doi.org/10.5061/dryad.xgxd254ck). MNA and Ho were the response variables and followed a gamma distribution with a log link function and beta distribution, respectively. Fixed effects tested for both models included year study was published, number of microsatellite loci used in study analysis, range size (km^2), class (Amphibia, Aves, anadromous fish, freshwater fish, Mammalia, and Reptilia), and latitudinal extent. Models were weighted by sample size. The Ho model also included MNA as a fixed effect, while the MNA model include Ho as an additional fixed effect. We included MNA and Ho as fixed effects to account for the fact that as MNA increases, the likelihood of being heterozygous also does. We wanted to account for the potential impact that simply having more alleles may be a better predictor of Ho than another factor. Ho was included in the MNA to account for reciprocal effects. Interactions were tested in a stepwise manner when additional variables were added to the model, although the interaction between class and MNA/Ho was included in the final model to account for class-specific differences. Random effects included study and taxonomic grouping (family, genus, species). Additionally, we tested for collinearity by checking variance inflation factor (VIF) scores. No variables with VIF scores above 3 were found and as such we determined there was no collinearity among variables.

Model selection followed guidelines by Zuur et al. (2009) in a forwards stepwise fashion. Briefly, we started with the null model that included only the random effects mentioned above,

then sequentially added a variable and assessed the AIC differences between models using ANOVA comparisons. A variable was only retained in the model if its addition decreased the AIC by >3 units. When two variables were added we tested the interaction between them; interactions were only retained if the interaction model's AIC was lower than the model without.

Code Availability

Code available through Dryad https://doi.org/10.5061/dryad.xgxd254ck.

Tables

Table A2.1*. Number of genetically distinct populations for each species and its corresponding range size (km^2). Population data obtained from *MacroPopGen* (Lawrence et al., 2019), range size information obtained from IUCN (IUCN, 2016), BirdLife International (Bird Life International 2017), and Meiri et al. (2017).

*Lawrence, Elizabeth; Fraser, Dylan (2020), Data from: Latitudinal biodiversity gradients at three levels: linking species richness, population richness, and genetic diversity, v2, Dryad, Dataset, https://doi.org/10.5061/dryad.xgxd254ck

Table A2.2. Model summary for (unscaled) fixed effects used in the generalized (mixed) models used in hypothesis testing; H1: Geographic distribution; H3: Range-restricted gene. Bold values indicate statistical significance. PopPerSpp = the total number of populations for a given species; MNA = mean number of alleles; Ho = observed heterozygosity; TaxaClass = taxonomic grouping.

Hypothesis	Response (Distribution)	Fixed	Coefficient	SE	Test statistic	p-value
H1	PopPerSpp	Intercept	0.1597091	0.0803738	1.987[a]	**0.047359**
	(Gamma)	Ln(Range Size)	-0.0028512	0.0041626	-0.685	0.493619
		TaxaClassAnadromous	-0.0807709	0.2032756	-0.397[a]	0.69125
		TaxaClassAves	0.210914	0.3559205	0.593[a]	0.553676
		TaxaClassFreshwater	0.1336673	0.1201573	1.112[a]	0.266388
		TaxaClassMammalia	-0.0227346	0.0950046	-0.239[a]	0.810953
		TaxaClassReptilia	-0.0543627	0.1005144	-0.541[a]	0.588811
		Lat_extent	-0.0006936	0.0001812	-3.828[a]	**0.000143**
		Ln(Range Size):TaxaClassAnadromous	0.0023087	0.0096802	0.239[a]	0.811572
		Ln(Range Size):TaxaClassAves	0.0046694	0.0203692	0.229[a]	0.818759
		Ln(Range Size):TaxaClassFreshwater	-0.0073394	0.0059697	-1.229[a]	0.219375
		Ln(Range Size):TaxaClassMammalia	0.0012602	0.0048375	0.261[a]	0.794552
		Ln(Range Size):TaxaClassReptilia	0.0032446	0.0050928	0.637 [a]	0.524304
H3	MNA	Intercept	1.062504	0.027877	38.11[b]	< **2e-16**
	(Gamma)	Ho	1.573037	0.026945	58.38 [b]	< **2e-16**

		Msats	-0.03334	0.001334	-25 [b]	< **2e-16**
		Ho:Msats	0.03986	0.001294	30.8 [b]	< **2e-16**
		Ho:TaxaClassAnadromous	0.445443	0.03054	14.59 [b]	< **2e-16**
		Ho:TaxaClassAves	0.202773	0.035106	5.78 [b]	**7.65E-09**
		Ho:TaxaClassFreshwater	0.249416	0.028605	8.72 [b]	< **2e-16**
		Ho:TaxaClassMammalia	-0.142877	0.028673	-4.98 [b]	**6.26E-07**
		Ho:TaxaClassReptilia	-0.13707	0.028083	-4.88 [b]	**1.06E-06**
H3	Ho	Intercept	-3.03E-01	4.03E-02	-7.51 [b]	**5.85E-14**
	(Beta)	MNA	9.90E-02	1.74E-03	56.87 [b]	< **2e-16**
		Msats	-1.71E-02	1.74E-03	-9.82 [b]	< **2e-16**
		MNA:Msats	9.01E-05	1.04E-04	0.86 [b]	0.387
		MNA:TaxaClassAnadromous	3.46E-02	2.01E-03	17.22 [b]	< **2e-16**
		MNA:TaxaClassAves	1.10E-02	2.41E-03	4.57 [b]	**4.92E-06**
		MNA:TaxaClassFreshwater	7.49E-03	1.77E-03	4.23 [b]	**2.33E-05**
		MNA:TaxaClassMammalia	1.93E-02	1.87E-03	10.3 [b]	< **2e-16**
		MNA:TaxaClassReptilia	-1.51E-03	1.79E-03	-0.84 [b]	0.401

[a] Indicates t-value; [b] Indicates z-value

Table A2.3. Summary of the linear models and model statistics for testing Hypothesis 2: Overlapping Range. Bold values indicate significance. See Glossary for acronym definitions. PopPerSpp = average number of populations per species within a grid cell; TotPopR = total number of populations in a grid cell; NumSpp = total number of species in grid cell.

Hypothesis	Response	Fixed	Estimate	SE	t-value	P-value	Adj R^2	F-value	DF	P-value
H2	PopPerSpp	Intercept	1.618	$1.097e^{-1}$	14.750	**< $2e^{-16}$**	0.6702	169.7	3 & 246	**< $2.2e^{-16}$**
		TotPopR	$7.151e^{-2}$	$4.185e^{-3}$	17.087	**< $2e^{-16}$**				
		NumSpp	$-7.614e^{-2}$	$1.247e^{-2}$	-6.105	**$3.97e^{-9}$**				
		TotPopR: NumSpp	$-9.462e^{-4}$	$8.893e^{-5}$	-10.640	**< $2e^{-16}$**				
H2	TotPopR	Intercept	-12.7322	2.6923	-4.729	**$3.79e^{-6}$**	0.7458	731.6	1 & 248	**< $2.2e^{-16}$**
		NumSpp	4.7435	0.1754	27.048	**< $2e^{-16}$**				

Table A2.4. Average number of populations per species for each taxonomic group. Note that means for fish are presented for all as a group, and for freshwater and anadromous species separately.

Taxonomic Group	**Average # Populations per Species**
Amphibians	19.87
Birds	8.43
Fish	32.02
Freshwater	23.36
Anadromous	109.17
Mammals	24.97
Reptiles	16.36

Appendix 3

Table A3.1. AIC comparison and model fit of select GAMMs during model selection. Ho = observed heterozygosity; MNA = mean number of alleles; Lat = degrees latitude; MAT = mean annual temperature (°C); AP = annual precipitation (mm/year); TAR = total annual range (°C); NPP = net primary productivity (units of elemental carbon x10e^{-11}), Elevation (m). Response variables were fitted with smoothing parameters (s or te for interactions); random effects (bs="re") included Reference ID (RefID) and Genus.

Model	**DF**	**AIC**	**Adj R^2**	**Deviance Explained (%)**
Ho ~ 1 + s(RefID, bs="re")	497.83	-415449.5	0.741	80.1
Ho ~ 1 + s(Genus, bs="re") + s(RefID, bs="re")	515.50	-423637.9	0.754	81.1
Ho ~ s(Lat) + s(Genus, bs="re") + s(RefID, bs="re")	258.32	-283959.6	0.763	81.1
Ho ~ s(Lat) + s(MAT_ChelsaC) + s(Genus, bs="re") + s(RefID, bs="re")	267.23	-287399.7	0.767	82.2
Ho~ s(Lat) + s(Elevation) + s(MAT) + s(AP) + s(TAR) + s(NPP) + s(Elevation, MAT) + s(Elevation, TAR) + s(Elevation, NPP) + s(Genus, bs="re")	624.24	-450825.5	0.785	83.9
Ho~ s(Lat) + s(Elevation) + s(MAT) + s(AP) + s(TAR) + s(NPP) + s(Elevation, MAT) + s(Elevation, TAR) + s(Elevation, NPP) + s(Lat, Genus, bs="re")	635.81	-328578.5	0.768	83.5
MNA ~ 1 + s(RefID, bs="re")	530.55	1268395.6	0.758	84.6
MNA ~ 1 + s(Genus, bs="re") + s(RefID, bs="re")	538.94	1250456.9	0.761	85.6
MNA~ s(Lat) + s(Genus, bs="re") + s(RefID, bs="re")	546.78	1243930.5	0.771	85.9
MNA~ s(Lat) + s(MAT_ChelsaC) + s(Genus, bs="re") + s(RefID, bs="re")	554.25	1230967.1	0.785	86.5
MNA~ s(Lat) + s(Elevation) + s(MAT) + s(AP) + s(TAR) + s(NPP) +	318.84	1005970.7	0.797	85.5

te(Elevation, MAT) + te(Elevation, TAR) + te(Elevation, NPP) + s(Genus, bs="re") + s(RefID, bs="re")				
MNA~ s(Lat) + s(Elevation) + s(MAT) + s(AP) + te(Elevation, TAR) + s(Genus, bs="re") + s(RefID, bs="re")	310.84	1007164.8	0.797	85.4
MNA~ s(Lat) + s(Elevation) + s(MAT) + s(AP) + te(Elevation, TAR) + s(Lat, Genus, bs="re") + s(RefID, bs="re")	344.45	994034.7	0.8	86.2

Table A3.2. Summary of selected GAMMs for either Ho (modeled with beta distribution) or MNA (gamma distribution). Predictor variables fitted with a smoother (s) included Lat = latitude, Elevation (m), MAT = mean annual temperature (°C), AP = annual precipitation (mm/year), TAR = total annual range (°C), NPP = net primary productivity (units of elemental carbon $x10e^{-11}$), and relevant interactions fitted with tensor products (te). Response variables were fitted with smoothing parameters (s), and random effects included Genus and Reference ID (RefID).

Dependent Variable	**Independent Variable**	**edf**	**Ref.df**	**Chi sq F***	**p-value**
Ho	s(Lat)	8.897	9	7769170	<0.001
	s(Elevation)	8.782	9	3581353	<0.001
	s(MAT)	8.738	9	2449913	<0.001
	s(Precipitation)	8.967	9	3723226	<0.001
	s(TAR)	8.405	9	2104634	<0.001
	s(NPP)	7.881	9	304730	<0.001
	te(Elevation, MAT)	21.353	21.63	4446	<0.001
	te(Elevation, TAR)	15.742	20	1206249	<0.001
	te(Elevation, NPP)	17.246	20	2940328	<0.001
	s(Genus)	58.042	106	300673798	<0.001
	s(RefID)	455.165	497	41615870	<0.001
MNA	s(Lat)	4.816	9	207.75*	0.021
	s(Elevation)	2.036	9	1.048*	0.028
	s(MAT)	6.934	9	125.742*	0.0068
	s(Precipitation)	7.693	9	60.831*	0.0055
	te(Elevation, TAR)	9.654	11.47	5.787*	<0.001
	s(Genus)	51.719	114	355.899*	<0.001
	s(RefID)	218.755	272	136.385*	<0.001

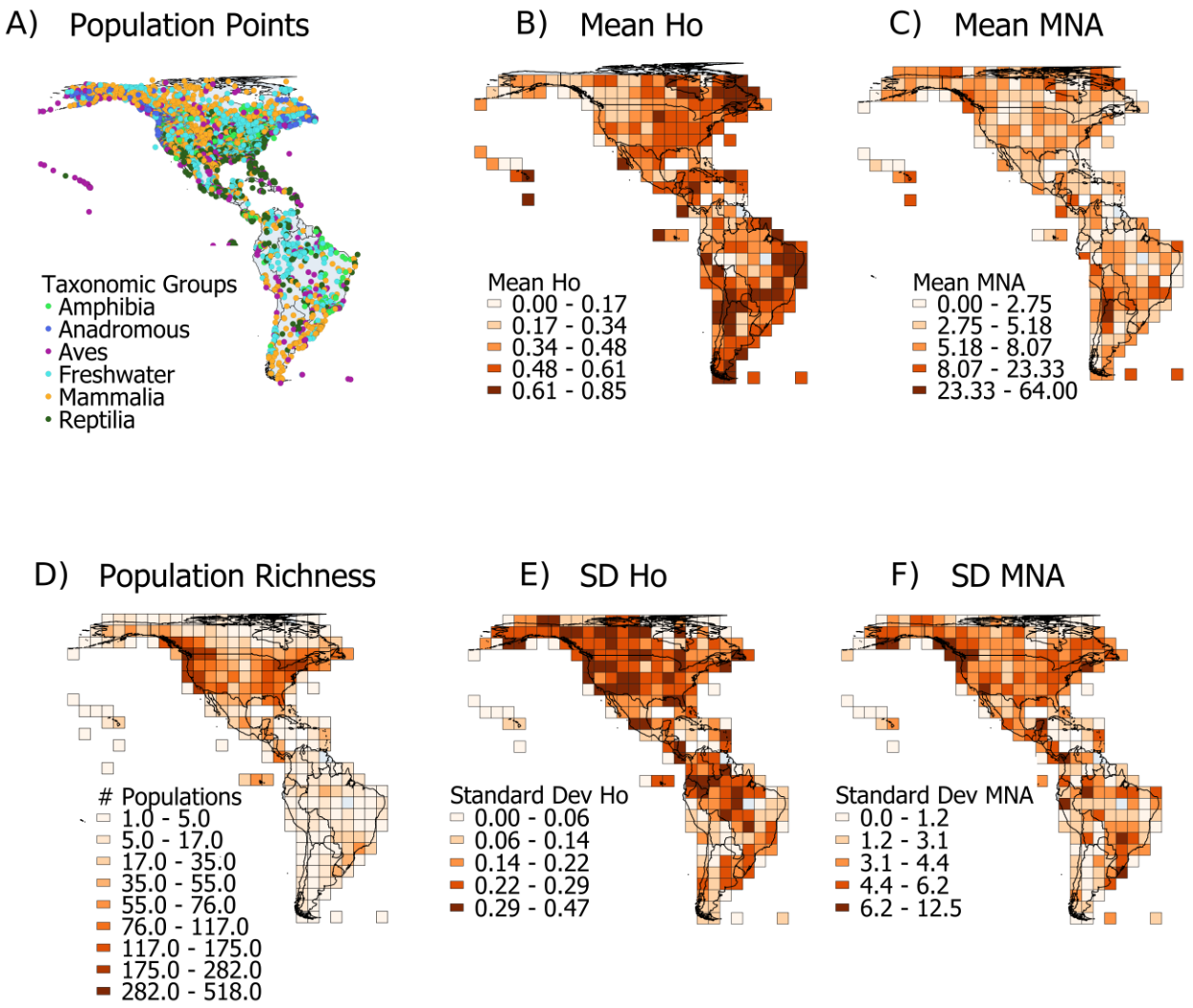

Figure A3.1. Population richness (count of populations) and population genetic diversity – observed heterozygosity (Ho) and mean number of alleles (MNA) – across the American continents. Georeferenced populations are coloured according to taxonomic group (a), and population richness represents the number of populations that are sampled with microsatellite data in each grid cell (d). Population-specific genetic diversity metrics are either averaged (b, c) within each grid cell, or the standard deviation across populations within a grid cell is taken (e, f). Grid size is 500 km by 500 km and maps are projected in World Behrmann.

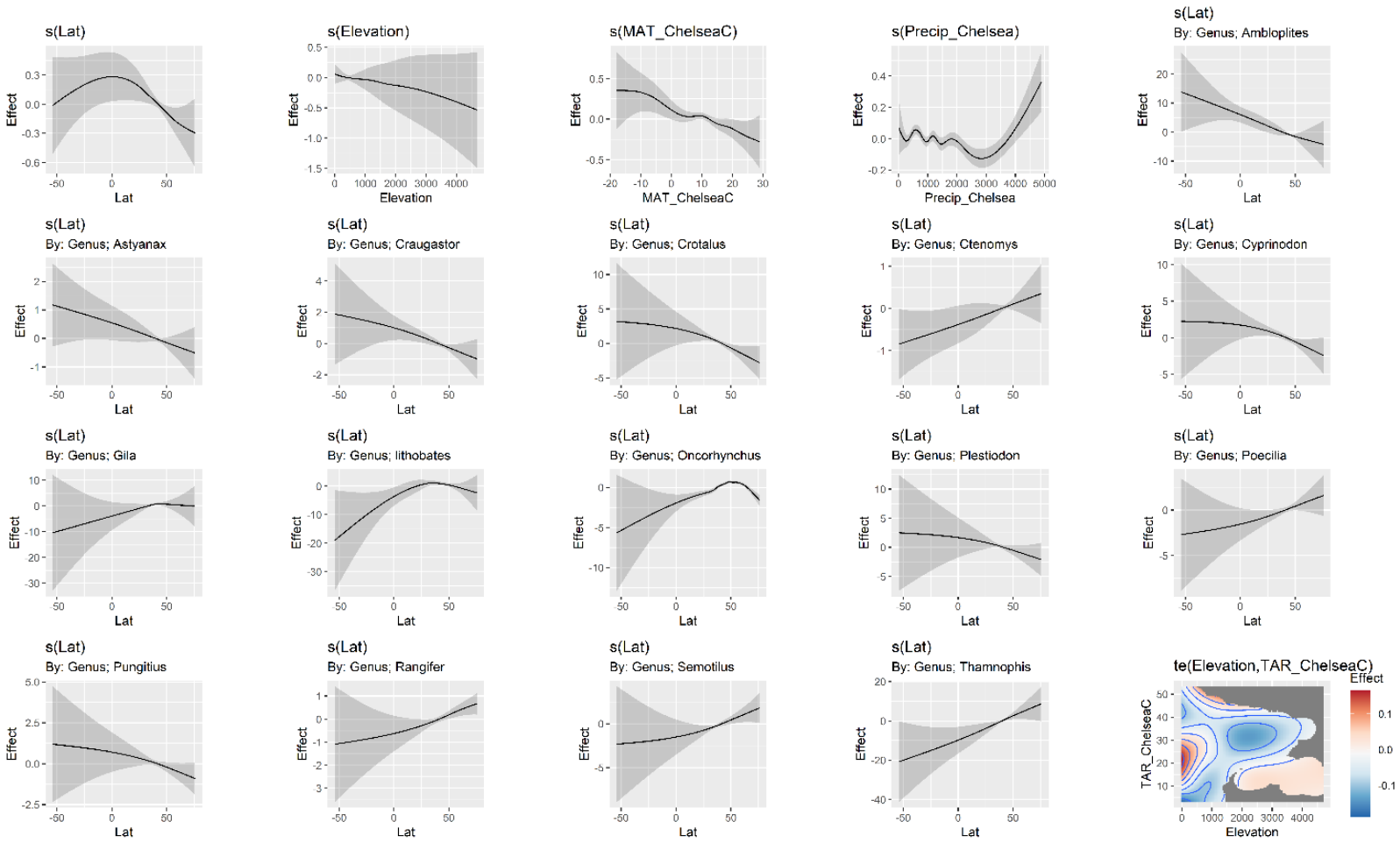

Figure A3.2. The predicted effect of latitude and environmental variables, fitted by smoothers (s; tensor products (te) for interactions), on vertebrate genetic diversity (mean number of alleles) for vertebrate species across the Americas. Genus-specific interactions shown only for the 15 genera with significant relationships. Predictors from the generalized additive mixed model include: Lat = degrees latitude, MAT = mean annual temperature (°C), Precip = annual precipitation (mm/year), npp = net primary productivity (units of elemental carbon x10e^{-11}), TAR= total annual temperature range (°C).

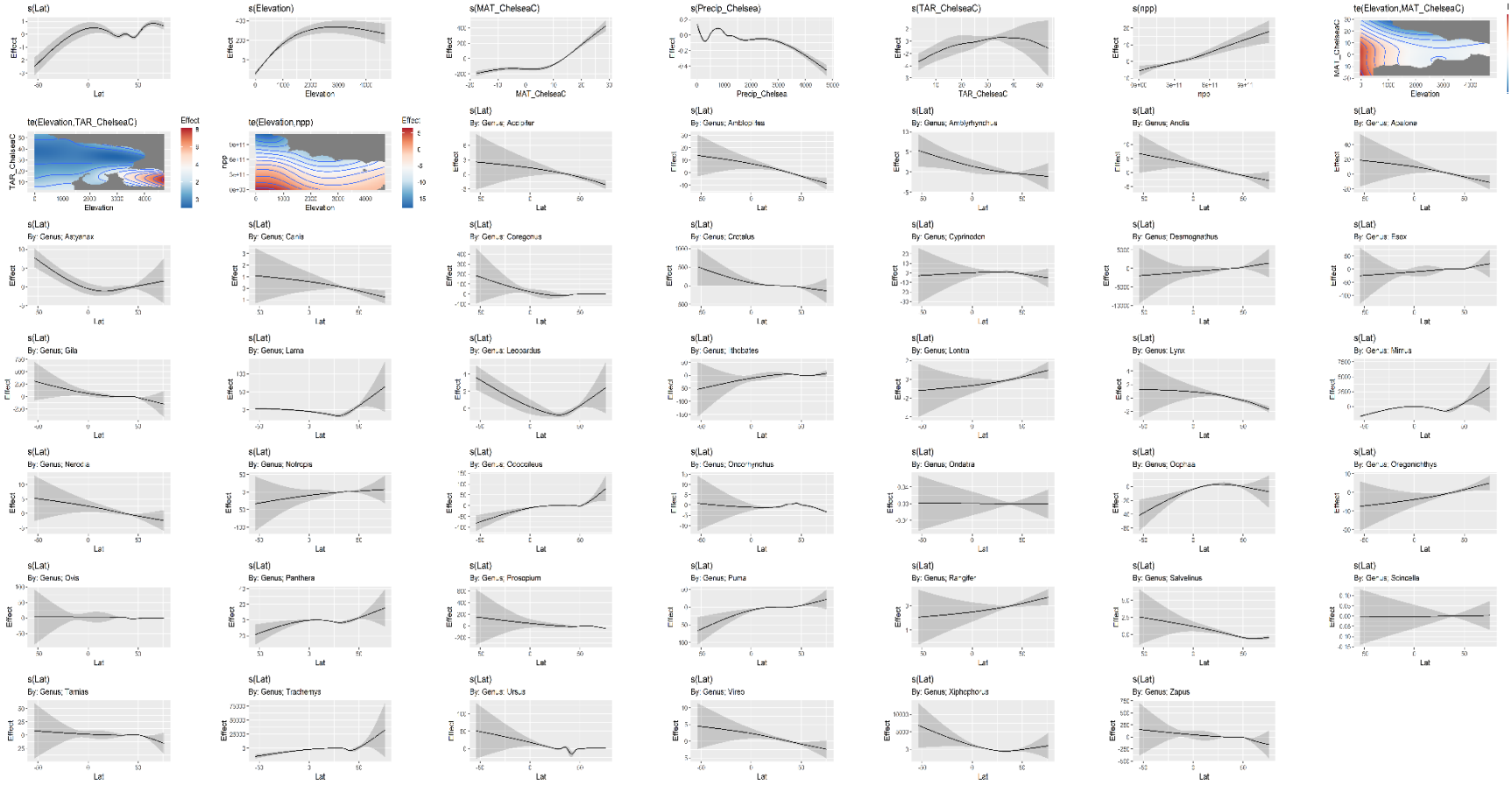

Figure A3.3. The predicted effect of latitude and environmental variables, fitted by smoothers (s; tensor products (te) for interactions), on vertebrate genetic diversity (observed heterozygosity (Ho) for vertebrate species across the Americas. Genus-specific interactions shown only for the 39 genera with significant relationships. Predictors from the generalized additive mixed model include: Lat = degrees latitude, MAT = mean annual temperature (°C), Precip = annual precipitation (mm/year), npp = net primary productivity (units of elemental carbon x10e^{-11}), TAR= total annual temperature range (°C).

Appendix 4

Supplementary Methods

Annual Variation

Generally speaking, the impact that humans have had on biodiversity has increased over time. To account for this, we tested for differences among human impacts for the years that the genetic data was collected (1993-2017) to determine which year of anthropogenic data was best to match with the genetic data. Between 1990 and 2016, 529 vertebrate populations decreased in HPD, although the greatest difference was only a change of -95 persons km^{-2}. Of the 2384 populations which experienced an increase in HPD, 484 populations experienced a change >100 units, up to a maximum difference between years of 11,974 humans/km^2. The remaining 4720 populations did not exhibit a change in HPD during this timeframe. From the Wilcoxon signed-rank test, significant differences were detected in HPD between 1990, 2010, and 2016 ($p<0.0001$, $V=319990$). Given this, we decided that using HPD data from the mean year, 2010, would be the best approach for including this variable in subsequent models.

We found 279 of the 7951 populations showed a change in anthropogenic biome between 1990 and 2016. Of these, 200 were populations that increased in land use intensity – i.e. changed from a more Natural biome to more urban. Only 79 populations shifted towards more Natural biomes. We found significant differences between 1990 and 2016 ($p=0.0015$), 1990 and 2010 ($p=010$), but not between 2010 and 2016 ($p=0.46$). Given this, we decided to use anthropogenic biome data from the mean year, 2010, in subsequent models.

Tables and Figures

Table S4.1. Number of genetically distinct populations per taxonomic group within each anthropogenic biome.

	Amphibia	**Anadromous**	**Aves**	**Freshwater**	**Mammalia**	**Reptilia**
Croplands	290	124	139	552	427	365
Dense Settlements	112	86	46	146	70	85
Rangeland	69	14	40	150	232	118
Semi-natural	395	643	166	952	467	362
Urban	86	41	54	116	97	144
Village	20	0	19	32	29	50
Wild	132	379	77	652	564	91

Table S4.2. Mean genetic diversity (MNA=mean number of alleles, Ho=observed heterozygosity) and distance (km) to Urban or Wild biomes for each anthropogenic biome-taxonomic group combination.

Biome	Taxa	n	MNA	MNAsd	Ho	Hosd	Urban	Urbansd	Natural	Wildsd	HPD	HPDsd
Croplands	Amphibia	264	8.49	5.54	0.63	0.17	24.93	33.92	40.15	66.35	22.65	26.91
Croplands	Anadromous	150	13.91	6.32	0.70	0.13	36.39	63.67	6.71	11.49	38.68	135.70
Croplands	Aves	50	5.54	2.75	0.60	0.14	146.37	318.39	53.29	97.43	10.52	15.64
Croplands	Freshwater	502	8.40	4.56	0.61	0.17	32.35	44.42	37.87	74.64	20.29	24.78
Croplands	Mammalia	407	6.10	2.64	0.62	0.13	40.37	59.63	50.54	80.17	25.10	94.21
Croplands	Reptilia	397	6.64	3.66	0.58	0.16	81.58	224.20	38.22	64.78	23.31	62.73
Freshwater	Freshwater	14	10.75	2.87	0.56	0.11	32.49	34.96	21.47	21.88	0.00	0.00
Freshwater	Mammalia	2	3.09	NA	0.36	0.09	70.40	27.86	9.42	9.41	0.00	0.00
Freshwater	Reptilia	3	11.38	0.18	0.72	NA	27.50	5.92	3.23	3.20	0.00	0.00
Ocean	Amphibia	3	9.90	0.71	0.64	0.05	4.23	0.89	8.55	4.77	0.00	0.00
Ocean	Anadromous	7	12.10	10.19	0.73	0.07	36.14	43.55	2.27	3.06	0.00	0.00
Ocean	Aves	9	3.51	2.29	0.37	0.17	887.40	555.60	359.73	683.41	0.00	0.00
Ocean	Freshwater	11	3.50	1.19	0.45	0.16	116.99	181.65	22.13	37.64	0.09	0.30
Ocean	Mammalia	4	4.52	0.87	0.48	0.21	188.25	256.19	24.75	31.24	0.00	0.00
Ocean	Reptilia	25	7.55	3.53	0.61	0.16	250.22	381.44	50.25	44.34	0.00	0.00
Rangeland	Amphibia	66	4.84	4.61	0.45	0.20	35.09	27.81	37.61	58.80	4.30	13.64
Rangeland	Anadromous	19	9.23	3.04	0.61	0.24	27.29	26.26	23.35	30.59	13.74	38.02
Rangeland	Aves	25	8.65	2.08	0.70	0.09	43.18	38.86	75.60	88.64	1.40	1.26
Rangeland	Freshwater	142	8.39	4.32	0.64	0.16	54.76	47.24	26.47	37.66	4.31	12.33
Rangeland	Mammalia	207	5.08	2.86	0.57	0.15	63.97	56.39	39.84	60.44	5.23	22.84
Rangeland	Reptilia	119	6.06	2.90	0.55	0.15	91.04	196.25	53.04	75.08	9.85	33.34
Semi-natural	Amphibia	361	7.43	6.02	0.55	0.16	29.63	39.68	0.76	7.01	14.52	23.18
Semi-natural	Anadromous	606	15.21	6.56	0.71	0.15	49.05	66.78	0.11	1.24	12.83	65.38
Semi-natural	Aves	60	5.74	3.42	0.55	0.18	338.75	444.21	14.64	55.96	6.95	10.42

Semi-natural	Freshwater	843	6.58	3.49	0.56	0.18	43.31	51.76	0.40	3.10	10.09	22.46
Semi-natural	Mammalia	428	6.49	6.57	0.62	0.14	53.60	76.71	0.84	9.32	7.15	16.42
Semi-natural	Reptilia	351	6.71	3.74	0.60	0.15	112.08	262.43	2.95	19.55	12.41	28.79
Urban	Amphibia	192	6.46	3.73	0.58	0.14	1.24	8.29	20.22	28.25	724.53	997.24
Urban	Anadromous	116	14.99	7.92	0.69	0.15	2.31	14.03	8.54	10.95	579.72	693.49
Urban	Aves	32	9.50	5.84	0.64	0.09	0.00	0.00	29.33	54.80	911.03	1083.32
Urban	Freshwater	236	7.12	3.26	0.60	0.14	7.81	53.48	16.40	31.40	735.67	985.05
Urban	Mammalia	141	7.33	3.16	0.62	0.16	2.69	21.33	39.38	49.61	1239.48	1542.94
Urban	Reptilia	194	6.78	3.75	0.60	0.13	10.72	96.77	28.22	50.87	868.77	879.18
Village	Amphibia	19	8.16	2.27	0.64	0.08	55.03	89.88	36.17	56.43	314.26	255.41
Village	Anadromous	6	39.61	24.58	0.80	0.11	123.94	94.53	2.10	1.95	1.50	1.97
Village	Freshwater	28	8.15	4.20	0.62	0.20	56.19	64.00	26.70	34.04	199.00	263.65
Village	Mammalia	24	5.58	2.39	0.59	0.19	38.43	34.11	32.23	33.69	295.50	335.38
Village	Reptilia	38	7.33	2.48	0.60	0.17	46.29	56.01	31.57	37.09	329.37	315.65
Wild	Amphibia	137	7.82	7.43	0.51	0.20	68.26	101.32	0.69	4.40	6.33	32.56
Wild	Anadromous	387	14.97	6.49	0.69	0.16	133.00	104.54	0.01	0.13	5.87	57.85
Wild	Aves	21	6.46	3.82	0.56	0.16	155.01	268.03	0.30	1.01	2.43	6.34
Wild	Freshwater	668	6.36	4.07	0.54	0.17	110.07	107.17	1.44	15.24	2.50	21.44
Wild	Mammalia	542	5.93	2.51	0.61	0.13	147.22	144.14	0.38	3.93	18.35	261.25
Wild	Reptilia	95	6.82	3.23	0.58	0.10	63.47	153.51	3.51	13.34	36.64	187.18

Table S4.3. Tukey test comparisons for MNA, Ho, and distances to natural or urban biomes among anthropogenic biomes and between taxonomic classes The difference between observed means, lower and upper interval end points, and the adjusted p-value for multiple comparisons are indicated. Bold values indicate statistical significance $p<0.05$.

Metric	Group	Comparison	diff	lwr	upr	p adj
MNA	Biome	Village-Urban	0.602	-1.258	2.462	0.977
MNA	Biome	Croplands-Urban	-0.336	-1.083	0.411	0.873
MNA	Biome	Rangeland-Urban	-1.730	-2.659	-0.800	**0.000**
MNA	Biome	Semi-natural-Urban	0.363	-0.340	1.066	0.771
MNA	Biome	Wild-Urban	-0.213	-0.950	0.523	0.988
MNA	Biome	Freshwater-Urban	1.895	-2.866	6.657	0.930
MNA	Biome	Ocean-Urban	-1.735	-4.112	0.643	0.344
MNA	Biome	Croplands-Village	-0.938	-2.747	0.872	0.768
MNA	Biome	Rangeland-Village	-2.331	-4.223	-0.439	**0.005**
MNA	Biome	Semi-natural-Village	-0.239	-2.030	1.553	1.000
MNA	Biome	Wild-Village	-0.815	-2.620	0.990	0.871
MNA	Biome	Freshwater-Village	1.294	-3.745	6.332	0.994
MNA	Biome	Ocean-Village	-2.336	-5.229	0.557	0.218
MNA	Biome	Rangeland-Croplands	-1.394	-2.217	-0.570	**0.000**
MNA	Biome	Semi-natural-Croplands	0.699	0.145	1.253	**0.003**
MNA	Biome	Wild-Croplands	0.123	-0.474	0.719	0.999
MNA	Biome	Freshwater-Croplands	2.231	-2.510	6.973	0.845
MNA	Biome	Ocean-Croplands	-1.398	-3.737	0.940	0.611
MNA	Biome	Semi-natural-Rangeland	2.093	1.309	2.876	**0.000**
MNA	Biome	Wild-Rangeland	1.516	0.703	2.330	**0.000**
MNA	Biome	Freshwater-Rangeland	3.625	-1.149	8.399	0.292
MNA	Biome	Ocean-Rangeland	-0.005	-2.408	2.398	1.000
MNA	Biome	Wild-Semi-natural	-0.576	-1.116	-0.036	**0.027**
MNA	Biome	Freshwater-Semi-natural	1.532	-3.203	6.267	0.977
MNA	Biome	Ocean-Semi-natural	-2.098	-4.422	0.227	0.112
MNA	Biome	Freshwater-Wild	2.109	-2.632	6.849	0.880
MNA	Biome	Ocean-Wild	-1.521	-3.856	0.813	0.499
MNA	Biome	Ocean-Freshwater	-3.630	-8.881	1.622	0.418
MNA	Taxa	Anadromous-Amphibia	7.449	6.654	8.244	**0.000**
MNA	Taxa	Aves-Amphibia	-0.623	-1.975	0.728	0.777
MNA	Taxa	Freshwater-Amphibia	-0.281	-0.993	0.431	0.871
MNA	Taxa	Mammalia-Amphibia	-1.151	-1.881	-0.420	**0.000**
MNA	Taxa	Reptilia-Amphibia	-0.604	-1.403	0.196	0.261
MNA	Taxa	Aves-Anadromous	-8.073	-9.363	-6.782	**0.000**
MNA	Taxa	Freshwater-Anadromous	-7.730	-8.318	-7.143	**0.000**
MNA	Taxa	Mammalia-Anadromous	-8.600	-9.210	-7.990	**0.000**

MNA	Taxa	Reptilia-Anadromous	-8.053	-8.744	-7.362	**0.000**
MNA	Taxa	Freshwater-Aves	0.342	-0.899	1.583	0.970
MNA	Taxa	Mammalia-Aves	-0.527	-1.779	0.725	0.837
MNA	Taxa	Reptilia-Aves	0.020	-1.274	1.313	1.000
MNA	Taxa	Mammalia-Freshwater	-0.870	-1.366	-0.373	**0.000**
MNA	Taxa	Reptilia-Freshwater	-0.323	-0.916	0.271	0.632
MNA	Taxa	Reptilia-Mammalia	0.547	-0.069	1.163	0.115
Ho	Biome	Village-Urban	0.010	-0.047	0.066	1.000
Ho	Biome	Croplands-Urban	0.000	-0.023	0.023	1.000
Ho	Biome	Rangeland-Urban	-0.031	-0.061	-0.002	**0.027**
Ho	Biome	Semi-natural-Urban	-0.015	-0.037	0.007	0.399
Ho	Biome	Wild-Urban	-0.033	-0.056	-0.010	**0.001**
Ho	Biome	Freshwater-Urban	-0.071	-0.201	0.058	0.707
Ho	Biome	Ocean-Urban	-0.060	-0.134	0.014	0.218
Ho	Biome	Croplands-Village	-0.010	-0.065	0.045	1.000
Ho	Biome	Rangeland-Village	-0.041	-0.099	0.017	0.379
Ho	Biome	Semi-natural-Village	-0.025	-0.080	0.029	0.858
Ho	Biome	Wild-Village	-0.043	-0.098	0.012	0.266
Ho	Biome	Freshwater-Village	-0.081	-0.220	0.058	0.639
Ho	Biome	Ocean-Village	-0.070	-0.159	0.020	0.260
Ho	Biome	Rangeland-Croplands	-0.032	-0.058	-0.005	**0.007**
Ho	Biome	Semi-natural-Croplands	-0.016	-0.033	0.002	0.126
Ho	Biome	Wild-Croplands	-0.033	-0.053	-0.014	**0.000**
Ho	Biome	Freshwater-Croplands	-0.072	-0.201	0.057	0.698
Ho	Biome	Ocean-Croplands	-0.060	-0.133	0.013	0.196
Ho	Biome	Semi-natural-Rangeland	0.016	-0.009	0.041	0.541
Ho	Biome	Wild-Rangeland	-0.002	-0.028	0.025	1.000
Ho	Biome	Freshwater-Rangeland	-0.040	-0.170	0.090	0.983
Ho	Biome	Ocean-Rangeland	-0.028	-0.103	0.047	0.947
Ho	Biome	Wild-Semi-natural	-0.018	-0.036	0.000	0.063
Ho	Biome	Freshwater-Semi-natural	-0.056	-0.185	0.073	0.892
Ho	Biome	Ocean-Semi-natural	-0.044	-0.117	0.028	0.581
Ho	Biome	Freshwater-Wild	-0.038	-0.168	0.091	0.986
Ho	Biome	Ocean-Wild	-0.027	-0.100	0.046	0.954
Ho	Biome	Ocean-Freshwater	0.012	-0.135	0.159	1.000
Ho	Taxa	Anadromous-Amphibia	0.141	0.114	0.168	**0.000**
Ho	Taxa	Aves-Amphibia	0.024	-0.014	0.062	0.459
Ho	Taxa	Freshwater-Amphibia	0.014	-0.007	0.034	0.391
Ho	Taxa	Mammalia-Amphibia	0.046	0.025	0.068	**0.000**
Ho	Taxa	Reptilia-Amphibia	0.022	-0.001	0.045	0.076
Ho	Taxa	Aves-Anadromous	-0.117	-0.156	-0.077	**0.000**
Ho	Taxa	Freshwater-Anadromous	-0.127	-0.150	-0.104	**0.000**
Ho	Taxa	Mammalia-Anadromous	-0.094	-0.118	-0.070	**0.000**

Ho	Taxa	Reptilia-Anadromous	-0.119	-0.145	-0.094	**0.000**
Ho	Taxa	Freshwater-Aves	-0.011	-0.046	0.025	0.957
Ho	Taxa	Mammalia-Aves	0.022	-0.014	0.058	0.490
Ho	Taxa	Reptilia-Aves	-0.002	-0.040	0.035	1.000
Ho	Taxa	Mammalia-Freshwater	0.033	0.017	0.049	**0.000**
Ho	Taxa	Reptilia-Freshwater	0.008	-0.010	0.026	0.801
Ho	Taxa	Reptilia-Mammalia	-0.025	-0.044	-0.005	**0.004**
Natural KM	Biome	Village-Urban	7.012	-7.128	21.152	0.806
Natural KM	Biome	Croplands-Urban	16.264	10.438	22.091	**0.000**
Natural KM	Biome	Rangeland-Urban	17.291	9.693	24.889	**0.000**
Natural KM	Biome	Semi-natural-Urban	-21.618	-27.106	-16.130	**0.000**
Natural KM	Biome	Wild-Urban	-21.864	-27.647	-16.081	**0.000**
Natural KM	Biome	Freshwater-Urban	-5.411	-38.530	27.708	1.000
Natural KM	Biome	Ocean-Urban	59.939	40.745	79.134	**0.000**
Natural KM	Biome	Croplands-Village	9.252	-4.497	23.002	0.455
Natural KM	Biome	Rangeland-Village	10.279	-4.310	24.868	0.392
Natural KM	Biome	Semi-natural-Village	-28.630	-42.240	-15.020	**0.000**
Natural KM	Biome	Wild-Village	-28.876	-42.607	-15.144	**0.000**
Natural KM	Biome	Freshwater-Village	-12.423	-47.807	22.960	0.964
Natural KM	Biome	Ocean-Village	52.927	30.046	75.808	**0.000**
Natural KM	Biome	Rangeland-Croplands	1.026	-5.819	7.871	1.000
Natural KM	Biome	Semi-natural-Croplands	-37.882	-42.269	-33.496	**0.000**
Natural KM	Biome	Wild-Croplands	-38.128	-42.879	-33.378	**0.000**
Natural KM	Biome	Freshwater-Croplands	-21.676	-54.630	11.279	0.486
Natural KM	Biome	Ocean-Croplands	43.675	24.766	62.584	**0.000**
Natural KM	Biome	Semi-natural-Rangeland	-38.909	-45.468	-32.349	**0.000**
Natural KM	Biome	Wild-Rangeland	-39.155	-45.963	-32.346	**0.000**
Natural KM	Biome	Freshwater-Rangeland	-22.702	-56.016	10.612	0.437
Natural KM	Biome	Ocean-Rangeland	42.648	23.120	62.176	**0.000**
Natural KM	Biome	Wild-Semi-natural	-0.246	-4.575	4.083	1.000
Natural KM	Biome	Freshwater-Semi-natural	16.207	-16.690	49.103	0.811
Natural KM	Biome	Ocean-Semi-natural	81.557	62.749	100.364	**0.000**
Natural KM	Biome	Freshwater-Wild	16.453	-16.494	49.400	0.800
Natural KM	Biome	Ocean-Wild	81.803	62.907	100.699	**0.000**
Natural KM	Biome	Ocean-Freshwater	65.350	27.661	103.040	**0.000**
Natural KM	Taxa	Anadromous-Amphibia	-5.998	-11.592	-0.403	**0.027**
Natural KM	Taxa	Aves-Amphibia	26.498	16.061	36.935	**0.000**
Natural KM	Taxa	Freshwater-Amphibia	-1.437	-6.407	3.533	0.963
Natural KM	Taxa	Mammalia-Amphibia	4.236	-1.018	9.489	0.195
Natural KM	Taxa	Reptilia-Amphibia	2.543	-3.122	8.208	0.796
Natural KM	Taxa	Aves-Anadromous	32.496	22.220	42.771	**0.000**
Natural KM	Taxa	Freshwater-Anadromous	4.561	-0.061	9.183	0.056
Natural KM	Taxa	Mammalia-Anadromous	10.233	5.308	15.159	**0.000**

Natural KM	Taxa	Reptilia-Anadromous	8.541	3.179	13.902	**0.000**
Natural KM	Taxa	Freshwater-Aves	-27.935	-37.885	-17.986	**0.000**
Natural KM	Taxa	Mammalia-Aves	-22.262	-32.356	-12.168	**0.000**
Natural KM	Taxa	Reptilia-Aves	-23.955	-34.269	-13.641	**0.000**
Natural KM	Taxa	Mammalia-Freshwater	5.673	1.469	9.876	**0.002**
Natural KM	Taxa	Reptilia-Freshwater	3.980	-0.727	8.687	0.153
Natural KM	Taxa	Reptilia-Mammalia	-1.693	-6.698	3.312	0.929
Urban KM	Biome	Village-Urban	47.280	9.962	84.599	**0.003**
Urban KM	Biome	Croplands-Urban	42.416	27.039	57.793	**0.000**
Urban KM	Biome	Rangeland-Urban	56.600	36.547	76.654	**0.000**
Urban KM	Biome	Semi-natural-Urban	54.949	40.465	69.432	**0.000**
Urban KM	Biome	Wild-Urban	115.494	100.231	130.757	**0.000**
Urban KM	Biome	Freshwater-Urban	30.412	-56.999	117.823	0.966
Urban KM	Biome	Ocean-Urban	275.194	224.534	325.853	**0.000**
Urban KM	Biome	Croplands-Village	-4.864	-41.154	31.425	1.000
Urban KM	Biome	Rangeland-Village	9.320	-29.184	47.825	0.996
Urban KM	Biome	Semi-natural-Village	7.668	-28.252	43.588	0.998
Urban KM	Biome	Wild-Village	68.214	31.972	104.455	**0.000**
Urban KM	Biome	Freshwater-Village	-16.868	-110.255	76.519	0.999
Urban KM	Biome	Ocean-Village	227.913	167.524	288.302	**0.000**
Urban KM	Biome	Rangeland-Croplands	14.184	-3.881	32.250	0.251
Urban KM	Biome	Semi-natural-Croplands	12.533	0.956	24.110	**0.023**
Urban KM	Biome	Wild-Croplands	73.078	60.540	85.617	**0.000**
Urban KM	Biome	Freshwater-Croplands	-12.004	-98.980	74.972	1.000
Urban KM	Biome	Ocean-Croplands	232.778	182.872	282.684	**0.000**
Urban KM	Biome	Semi-natural-Rangeland	-1.652	-18.964	15.660	1.000
Urban KM	Biome	Wild-Rangeland	58.894	40.924	76.863	**0.000**
Urban KM	Biome	Freshwater-Rangeland	-26.188	-114.112	61.735	0.986
Urban KM	Biome	Ocean-Rangeland	218.593	167.054	270.132	**0.000**
Urban KM	Biome	Wild-Semi-natural	60.546	49.120	71.971	**0.000**
Urban KM	Biome	Freshwater-Semi-natural	-24.537	-111.359	62.286	0.990
Urban KM	Biome	Ocean-Semi-natural	220.245	170.607	269.883	**0.000**
Urban KM	Biome	Freshwater-Wild	-85.082	-172.038	1.874	0.060
Urban KM	Biome	Ocean-Wild	159.699	109.828	209.571	**0.000**
Urban KM	Biome	Ocean-Freshwater	244.781	145.309	344.254	**0.000**
Urban KM	Taxa	Anadromous-Amphibia	21.803	7.038	36.568	**0.000**
Urban KM	Taxa	Aves-Amphibia	164.486	136.940	192.031	**0.000**
Urban KM	Taxa	Freshwater-Amphibia	13.672	0.554	26.789	**0.035**
Urban KM	Taxa	Mammalia-Amphibia	30.995	17.128	44.861	**0.000**
Urban KM	Taxa	Reptilia-Amphibia	50.781	35.830	65.731	**0.000**
Urban KM	Taxa	Aves-Anadromous	142.683	115.563	169.802	**0.000**
Urban KM	Taxa	Freshwater-Anadromous	-8.131	-20.330	4.067	0.402
Urban KM	Taxa	Mammalia-Anadromous	9.192	-3.808	22.192	0.333

Urban KM	Taxa	Reptilia-Anadromous	28.978	14.827	43.128	**0.000**
Urban KM	Taxa	Freshwater-Aves	-150.814	-177.073	-124.555	**0.000**
Urban KM	Taxa	Mammalia-Aves	-133.491	-160.132	-106.850	**0.000**
Urban KM	Taxa	Reptilia-Aves	-113.705	-140.926	-86.484	**0.000**
Urban KM	Taxa	Mammalia-Freshwater	17.323	6.229	28.416	**0.000**
Urban KM	Taxa	Reptilia-Freshwater	37.109	24.687	49.531	**0.000**
Urban KM	Taxa	Reptilia-Mammalia	19.786	6.576	32.996	**0.000**
HPD	Biome	Village-Urban	-599.165	-779.592	-418.738	**0.000**
HPD	Biome	Croplands-Urban	-910.650	-983.219	-838.081	**0.000**
HPD	Biome	Rangeland-Urban	-924.504	-1018.893	-830.115	**0.000**
HPD	Biome	Semi-natural-Urban	-920.407	-988.062	-852.751	**0.000**
HPD	Biome	Wild-Urban	-929.241	-1000.566	-857.917	**0.000**
HPD	Biome	Croplands-Village	-311.484	-487.584	-135.385	**0.000**
HPD	Biome	Rangeland-Village	-325.339	-511.495	-139.183	**0.000**
HPD	Biome	Semi-natural-Village	-321.241	-495.373	-147.109	**0.000**
HPD	Biome	Wild-Village	-330.076	-505.666	-154.486	**0.000**
HPD	Biome	Rangeland-Croplands	-13.854	-99.683	71.974	0.997
HPD	Biome	Semi-natural-Croplands	-9.757	-64.842	45.328	0.996
HPD	Biome	Wild-Croplands	-18.592	-78.126	40.943	0.949
HPD	Biome	Semi-natural-Rangeland	4.097	-77.618	85.813	1.000
HPD	Biome	Wild-Rangeland	-4.737	-89.516	80.041	1.000
HPD	Biome	Wild-Semi-natural	-8.835	-62.270	44.600	0.997
HPD	Taxa	Anadromous-Amphibia	1.162	-72.785	75.109	1.000
HPD	Taxa	Aves-Amphibia	22.843	-120.047	165.733	0.998
HPD	Taxa	Freshwater-Amphibia	26.891	-38.990	92.773	0.854
HPD	Taxa	Mammalia-Amphibia	89.244	19.716	158.772	**0.003**
HPD	Taxa	Reptilia-Amphibia	27.611	-47.556	102.778	0.902
HPD	Taxa	Aves-Anadromous	21.681	-119.142	162.503	0.998
HPD	Taxa	Freshwater-Anadromous	25.729	-35.539	86.997	0.838
HPD	Taxa	Mammalia-Anadromous	88.082	22.909	153.255	**0.002**
HPD	Taxa	Reptilia-Anadromous	26.449	-44.709	97.607	0.897
HPD	Taxa	Freshwater-Aves	4.049	-132.711	140.808	1.000
HPD	Taxa	Mammalia-Aves	66.401	-72.152	204.955	0.747
HPD	Taxa	Reptilia-Aves	4.768	-136.699	146.235	1.000
HPD	Taxa	Mammalia-Freshwater	62.353	6.498	118.208	**0.018**
HPD	Taxa	Reptilia-Freshwater	0.720	-62.016	63.455	1.000
HPD	Taxa	Reptilia-Mammalia	-61.633	-128.188	4.921	0.088

Table S4.4. Change in population-specific anthropogenic biomes, as defined by Klein Goldewijk et al. (2017), between 1990 (across) and 2016 (down).

	1990							
2016	Croplands	Freshwater	Ocean	Rangeland	Semi-natural	Urban	Village	Wild
Croplands	1563	0	0	13	22	0	10	3
Freshwater	2	51	0	0	1	0	0	0
Ocean	15	0	211	1	2	1	2	0
Rangeland	19	0	0	542	2	0	0	2
Semi-Natural	30	0	0	4	2600	4	0	38
Urban	32	0	0	4	37	842	6	12
Village	20	0	0	0	0	1	80	0
Wild	7	0	0	0	0	0	0	1772

Table S4.5. Summary of linear models with either observed heterozygosity (Ho) or mean number of alleles (MNA) as the response. All models took the form Y~X+X:Taxa, where we included the interaction between X and taxonomic class. Model estimate, standard error, t-value, and p-value are indicated for each linear model; significant p-values are in bold.

Response	Model	Dependent	Estimate	Std. Error	t value	P value
Ho	HPD	(Intercept)	0.621	0.014	43.150	**0.0000**
Ho	HPD	log(p20101)	0.000	0.003	-0.029	0.9767
Ho	HPD	TaxaClassAnadromous	0.068	0.024	2.804	**0.0051**
Ho	HPD	TaxaClassAves	-0.015	0.028	-0.545	0.5862
Ho	HPD	TaxaClassFreshwater	-0.021	0.014	-1.473	0.1409
Ho	HPD	TaxaClassMammalia	-0.012	0.015	-0.806	0.4205
Ho	HPD	TaxaClassReptilia	-0.036	0.015	-2.377	**0.0176**
MNA	HPD	(Intercept)	7.941	0.490	16.212	**0.0000**
MNA	HPD	log(p20101)	0.181	0.084	2.153	**0.0315**
MNA	HPD	TaxaClassAnadromous	6.255	0.667	9.375	**0.0000**
MNA	HPD	TaxaClassAves	-2.664	0.875	-3.045	**0.0024**
MNA	HPD	TaxaClassFreshwater	-0.266	0.492	-0.540	0.5892
MNA	HPD	TaxaClassMammalia	-2.448	0.504	-4.857	**0.0000**
MNA	HPD	TaxaClassReptilia	-2.141	0.518	-4.136	**0.0000**
Ho	Urban	(Intercept)	0.649	0.014	45.782	**0.0000**
Ho	Urban	log(Urban_km)	-0.011	0.003	-3.773	**0.0002**
Ho	Urban	TaxaClassAnadromous	0.069	0.024	2.829	**0.0047**
Ho	Urban	TaxaClassAves	-0.014	0.028	-0.522	0.6021
Ho	Urban	TaxaClassFreshwater	-0.018	0.014	-1.280	0.2008
Ho	Urban	TaxaClassMammalia	-0.007	0.015	-0.465	0.6417
Ho	Urban	TaxaClassReptilia	-0.033	0.015	-2.202	**0.0278**
MNA	Urban	(Intercept)	8.715	0.492	17.713	**0.0000**
MNA	Urban	log(Urban_km)	-0.111	0.090	-1.230	0.2190
MNA	Urban	TaxaClassAnadromous	6.283	0.668	9.402	**0.0000**
MNA	Urban	TaxaClassAves	-2.786	0.874	-3.188	**0.0015**
MNA	Urban	TaxaClassFreshwater	-0.256	0.494	-0.517	0.6049
MNA	Urban	TaxaClassMammalia	-2.470	0.505	-4.886	**0.0000**
MNA	Urban	TaxaClassReptilia	-2.102	0.520	-4.042	**0.0001**
Ho	Natural	(Intercept)	0.609	0.014	44.275	**0.0000**
Ho	Natural	log(Naturalkm)	0.005	0.003	1.778	0.0757
Ho	Natural	TaxaClassAnadromous	0.075	0.025	3.033	**0.0025**
Ho	Natural	TaxaClassAves	-0.017	0.028	-0.609	0.5429
Ho	Natural	TaxaClassFreshwater	-0.021	0.014	-1.423	0.1551
Ho	Natural	TaxaClassMammalia	-0.014	0.015	-0.909	0.3636
Ho	Natural	TaxaClassReptilia	-0.037	0.015	-2.463	**0.0139**
MNA	Natural	(Intercept)	8.367	0.477	17.557	**0.0000**

MNA	Natural	log(Naturalkm)	0.025	0.076	0.330	0.7418
MNA	Natural	TaxaClassAnadromous	6.358	0.681	9.332	**0.0000**
MNA	Natural	TaxaClassAves	-2.849	0.873	-3.263	**0.0011**
MNA	Natural	TaxaClassFreshwater	-0.300	0.493	-0.608	0.5432
MNA	Natural	TaxaClassMammalia	-2.535	0.505	-5.023	**0.0000**
MNA	Natural	TaxaClassReptilia	-2.158	0.519	-4.161	**0.0000**
Ho	POB	(Intercept)	0.609	0.023	26.014	**0.0000**
Ho	POB	WeightedPercent	0.003	0.005	0.563	0.5738
Ho	POB	TaxaClassAnadromous	0.068	0.024	2.774	**0.0056**
Ho	POB	TaxaClassAves	-0.015	0.028	-0.530	0.5963
Ho	POB	TaxaClassFreshwater	-0.022	0.015	-1.521	0.1286
Ho	POB	TaxaClassMammalia	-0.013	0.015	-0.845	0.3983
Ho	POB	TaxaClassReptilia	-0.035	0.015	-2.311	**0.0210**
MNA	POB	(Intercept)	8.509	0.731	11.639	**0.0000**
MNA	POB	WeightedPercent	-0.020	0.147	-0.133	0.8943
MNA	POB	TaxaClassAnadromous	6.323	0.672	9.409	**0.0000**
MNA	POB	TaxaClassAves	-2.849	0.873	-3.262	**0.0011**
MNA	POB	TaxaClassFreshwater	-0.297	0.496	-0.599	0.5494
MNA	POB	TaxaClassMammalia	-2.522	0.505	-4.997	**0.0000**
MNA	POB	TaxaClassReptilia	-2.164	0.523	-4.140	**0.0000**

Table S4.6. The mean percent anthrome area and standard deviation in the 100km surrounding a population. For example, populations originating in Urban biomes, we show the mean percent of other biomes in the surrounding area. Bolded values indicate the corresponding percent for the Originating biome, values in italic indicate the biome with the highest mean percent.

Originating Biome	**Surrounding Biome**	**Mean Percent (%)**	**Mean SD**
Urban	Urban	**9.36**	1.19
	Village	2.42	2.35
	Croplands	9.36	7.79
	Rangeland	6.39	6.07
	Semi-natural	7.19	2.53
	Wild	4.25	3.02
	Ocean	*26.02*	NA
	Freshwater	10.05	NA
Village	Urban	2.06	0.16
	Village	**3.09**	0.81
	Croplands	12.73	8.53
	Rangeland	5.10	11.27
	Semi-natural	4.78	3.81
	Wild	3.83	6.45
	Ocean	*42.18*	NA
	Freshwater	3.16	NA
Croplands	Urban	4.49	0.70
	Village	2.22	0.87
	Croplands	**14.77**	2.59
	Rangeland	6.47	2.49
	Semi-natural	5.50	1.31
	Wild	4.82	3.62
	Ocean	*24.54*	NA
	Freshwater	6.46	NA
Rangeland	Urban	1.89	0.04
	Village	0.87	0.60
	Croplands	6.98	5.19
	Rangeland	**18.81**	20.03
	Semi-natural	2.61	1.29
	Wild	5.95	1.17
	Ocean	*31.50*	NA
	Freshwater	1.22	NA
Semi-natural	Urban	3.33	0.14
	Village	1.87	1.64
	Croplands	5.43	2.87
	Rangeland	6.50	5.80

	Semi-natural	**10.82**	6.44
	Wild	10.56	9.47
	Ocean	*31.51*	NA
	Freshwater	4.51	NA
Wild	Urban	2.18	0.51
	Village	0.51	2.34
	Croplands	4.56	6.36
	Rangeland	7.21	4.03
	Semi-natural	5.52	3.28
	Wild	**22.38**	3.56
	Ocean	*28.07*	NA
	Freshwater	6.88	NA
Ocean	Urban	3.08	0.55
	Village	1.09	3.16
	Croplands	3.13	7.85
	Rangeland	3.27	2.47
	Semi-natural	5.57	1.56
	Wild	12.11	2.97
	Ocean	***70.67***	NA
	Freshwater	7.18	NA
Freshwater	Urban	6.45	0.38
	Village	1.10	0.35
	Croplands	7.62	2.46
	Rangeland	9.66	8.96
	Semi-natural	5.32	4.47
	Wild	4.72	21.01
	Ocean	29.11	NA
	Freshwater	***41.31***	NA

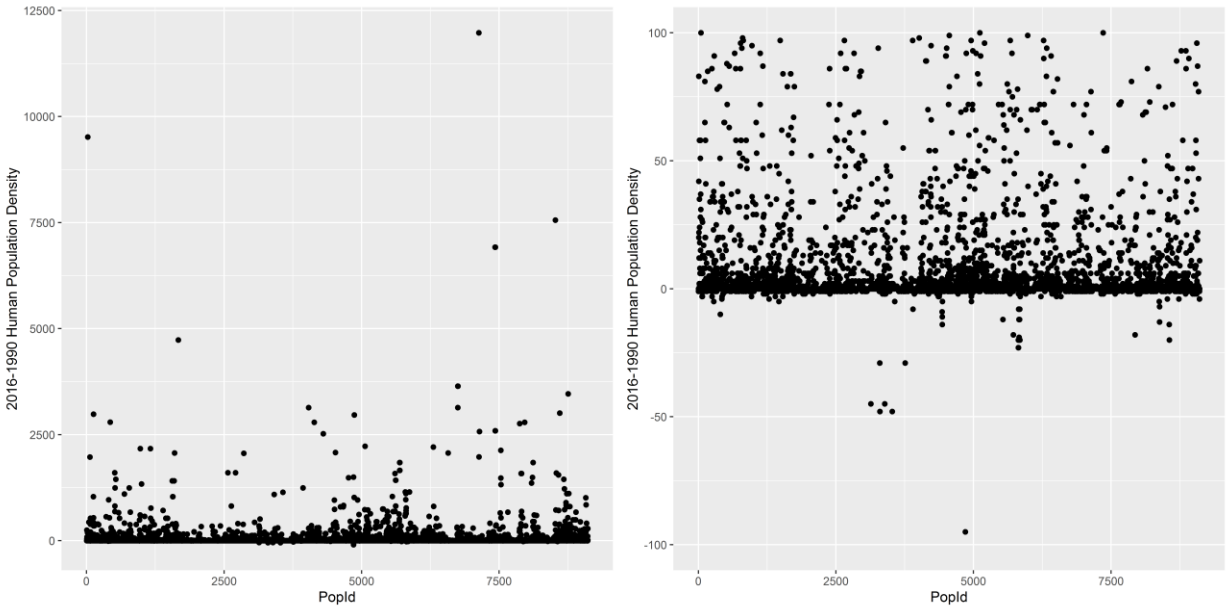

Figure S4.1. A) Change in human population density (HPD) for each vertebrate population represented by the difference between the years 1990 and 2016. B) Zoom in on the y axis (-100, 100).