Molecular Ecology Resources

WILEY

MOLECULAR ECOLOGY RESOURCES

MINI REVIEW OPEN ACCESS

To Meet Its Potential as an Integrative Discipline, Macrogenetics Must Embrace the Genomic Era

Tara A. Pelletier[1] | Connor N. Ferguson[2] | Jordan D. Satler[2] | Niko J. Zuppas[2] | Seth J. Kane[1] | Ryan M. Gawlik[1] | Bryan C. Carstens[2,3]

[1]Department of Biology, Radford University, Radford, Virginia, USA | [2]Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, Columbus, Ohio, USA | [3]Museum of Biological Diversity, The Ohio State University, Columbus, Ohio, USA

Correspondence: Tara A. Pelletier (tpelletier@radford.edu)

Received: 9 April 2026 | **Revised:** 22 October 2026 | **Accepted:** 29 October 2026

Keywords: biogeography | conservation | genetic diversity | mitochondrial DNA | phylogeography | population genetics | SNP

ABSTRACT

Macrogenetics seeks to uncover the ecological and evolutionary drivers of within- and between-species genetic diversity, population structure, and phylogeographic patterns across broad spatial and taxonomic scales. Advances in automatic pipelines now enable comparative analyses across thousands of species, yet the processes generating these patterns remain poorly resolved. For macrogenetics to mature as a field, it must explicitly integrate the evolutionary mechanisms such as demographic change, range dynamics, and dispersal that are likely to influence intraspecific genetic diversity. Accomplishing this goal likely requires an expansion of macrogenetic data. Organellar DNA sequences have long served as the foundation for phylogeographic inference and have been routinely repurposed for use in macrogenetic investigations. To evaluate the utility of using such organellar sequences as a proxy for genomic diversity, we compiled two datasets from nuclear genes and random SNP data and compared these to mitochondrial DNA sequences of the same species. Results indicate that these estimates of nucleotide diversity are not always correlated, suggesting that single-locus data from the mitochondrial genome are not representative of genome-wide genetic variation within most species. Despite such challenges, discovering the environmental and evolutionary forces shaping intraspecific genetic diversity remains an important goal. To understand the factors that influence genetic diversity within species, macrogenetics must embrace the challenge of including genomic data in its analyses.

1 | Introduction

The relatively new research discipline known as macrogenetics aims to conduct investigations using data from hundreds to thousands of species across large spatial scales, demonstrating geographic patterns in genetic biodiversity (Leigh et al. 2021). One major goal of macrogenetics is to repurpose genetic data (ideally georeferenced) from open-source repositories and analyse these data under a common framework. The repurposed data typically originate from individual landscape genetic or phylogeographic investigations and are analysed using automated bioinformatics pipelines, which has increased the overall efficiency of these studies (e.g., 10,000+ species; Millette et al. 2019; Pelletier et al. 2022). Due to the inclusion of data from potentially hundreds of studies, the potential to increase our understanding of global evolutionary processes via data synthesis is high.

Miraldo et al. (2016) assembled mitochondrial DNA (mtDNA) sequences from 1992 mammal and amphibian species to identify the first evidence of a latitudinal gradient of intraspecific genetic diversity. Smith et al. (2017) found a similar pattern of phylogeographic breaks in birds using mtDNA from 210 species. Their data suggested that older species have more phylogeographic structure than younger species and that

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Molecular Ecology Resources* published by John Wiley & Sons Ltd.

higher extinction rates in temperate regions explain this pattern. Similarly, using mtDNA data from over 38,000 species, Fonseca et al. (2023) documented a latitudinal gradient of genetic diversity within species, where temperate and polar regions contained less genetic diversity than those in tropical regions. This pattern was attributed to glaciation in the temperate regions, which reduced genetic diversity in temperate species. However, some macrogenetic investigations have contradicted these results. French et al. (2023) investigated mean global genetic diversity in over 150,000 insect operational taxonomic units (OTUs). Unlike the investigations in vertebrate species, they report that genetic diversity is lower in the tropics and peaks in temperate areas and suspect this has to do with the range size of insects being smaller in tropical regions, in addition to other taxon-specific traits.

Investigations such as the ones described above illustrate that researchers are attempting to identify the factors that produce global patterns. Such studies provide baseline data on the spatial distribution of intraspecific genetic diversity where explanatory power typically increases with the number of species sampled. However, the broader applicability of these results may be limited by decisions made during the analytical process, including the unit of analysis (e.g., whether to analyse OTUs or nominal species), the taxonomic scale included in the investigation, and spatial considerations (e.g., how data are being partitioned across space).

2 | Taxonomic and Geographic Scale

While spatial patterns are broadly evident across a wide range of taxa, researchers have been less successful in identifying the environmental and evolutionary *processes* that generate these patterns. One strategy has been to include organismal trait data. In fish, Manel et al. (2020) found clear spatial patterns related to factors such as the slope of river drainages and sea surface temperatures for freshwater and marine species, respectively. Decker et al. (2025) identified life history traits associated with wing morphology that were predictive of phylogeographic structure in bats. However, trait data are harder to compile for disparate taxa that do not share the same features, or when considering OTUs as units of analysis. Further, it may be the case that individual taxa are influenced by different processes on regional scales, as examples from amphibians demonstrate. Data from Miraldo et al. (2016) suggested that intraspecific genetic diversity is higher in tropical amphibians and in areas with less human impact, while Barrow et al. (2021) found that intraspecific genetic diversity increased by latitude for Nearctic amphibians. These results can be partially reconciled by acknowledging that range size and elevation also influence intraspecific genetic diversity (Amador et al. 2024). Therefore, investigations on a regional scale may identify factors that are not evident in global surveys. For example, Pelletier and Carstens (2018) analysed sequence data from over 8000 species with broad taxonomic sampling (plants, fungi, and animals) to identify isolation by distance and environment that was more impactful at higher latitudes (i.e., there is more genetic structure further from the equator), which conflicts with results from Amador et al. (2024). The lack of consensus in these findings suggest that the patterns from any given study are dependent on the taxonomic scale and methodology used to assess genetic diversity.

Other researchers have collated detailed collections of species-level data to infer historical demographic history. Data from almost 500 butterflies from the western Palearctic region identified glacial refugia that these species used to survive the Pleistocene glaciations, their post glacial colonization routes, and functional traits associated with endemicity (Dapporto et al. 2024). Carstens et al. (2018) investigated how bats responded to Pleistocene climate change by modelling population demographic expansion and bottlenecks, and then identified organismal traits that were associated with differing responses to shared climatic events. In these cases, by restricting the taxonomic scope, the researchers were able to ask questions pertaining to how organismal traits influenced intraspecific genetic diversity. Rather than documenting global patterns, such investigations engender more targeted research questions into the influence of clade-specific life history traits.

3 | Scale of Genetic Data

For macrogenetics to mature as a discipline, it must fully incorporate the evolutionary processes (e.g., population demographic change, range expansion and contraction, dispersal, and adaptation) that generate biodiversity patterns. Notably, both phylogeographic research and landscape genetics have undergone a similar shift. Both began as descriptive fields documenting how genetic variation is distributed across geographic space (e.g., Avise et al. 1987). Over the last decades, both fields have become increasingly effective at identifying the evolutionary processes shaping biodiversity patterns. Much of this effectiveness results from the dramatic increase in genetic data collected by researchers. These disciplines have benefited from advances in DNA sequencing technologies and generally expanded from using only organellar DNA (in phylogeography) or a limited number of microsatellites (in landscape genetics) to thousands of SNPs and benefited further from statistical methods that enable more powerful tests of hypotheses (Knowles 2009; Hickerson et al. 2010; Garrick et al. 2015; Edwards et al. 2022). Despite this history, most macrogenetic studies that repurpose data have used limited loci to estimate levels of genetic diversity within species, relying largely on mitochondrial and chloroplast data or microsatellite data, not yet taking advantage of genomic-level data (Table 1; see Appendix S1 for literature search details).

Single locus data from organellar genes have recognized shortcomings (Allio et al. 2017). They represent only a single realization of the coalescent process (Hudson and Turelli 2003). Theoretical work has demonstrated that estimates of divergence time from mtDNA are likely to underestimate actual species divergence (Edwards and Beerli 2000) and show clearly that the accuracy of estimates of population genetic parameters comes from increasing the number of independent loci (Felsenstein 2006; Carling and Brumfield 2007). Organellar genealogical patterns can be unrepresentative of organismal history due to adaptive introgression (Llopart et al. 2014) or the pseudogenization of organellar genes in the nuclear genome that can lead to inflated diversity estimates (Bensasson et al. 2001). Finally, because organellar genomes are uniparentally inherited, sex-biassed dispersal can lead to false inference of genetic structure (Cooper et al. 2010) and is more prone to selective sweeps (Galtier et al. 2009), complicating

TABLE 1 | Representative summary of macrogenetic studies.

References	Taxa	Marker	Number of species	Number of sequences	Geographic scale	Pattern
Miraldo et al. (2016)	Amphibians, mammals	mtDNA: Cytb	1992	24,479	Global	IGD is higher in the tropics; IGD is lower for amphibians in areas more affected by humans.
Smith et al. (2017)	Birds	mtDNA: various loci	210	17,573	North and South America	Higher incidence of phylogeographic breaks in the tropics.
Pelletier and Carstens (2018)	Animals, fungi, plants	nDNA; mtDNA; cpDNA	8955	187,652	Global	Higher incidence of isolation by distance and environment in temperate species.
Millette et al. (2019)	Birds, fishes, insects, mammals	mtDNA: COI	17,082	175,247	Global	Human impacts on diversity were taxon specific and scale-dependent, and generally weak or non-significant.
Theodoridis et al. (2020)	Terrestial mammals	mtDNA: Cytb, COI	2128	24,395, 22,570	Global	IGD is higher in the tropics.
Barrow et al. (2021)	Amphibians	mtDNA; nDNA	263	42,067	North America	IGD higher in the tropics and some phylogenetic signal.
Yiming et al. (2021)	Terrestrial mammals	nDNA: MHC II	93	1515	Global	IGD explained by latitude, positive selection, and body mass.
De Kort et al. (2021)	Animals and plants	AFLPs; microsatellites	727	8357	Global	IGD lower in plants than animals; weak latitudinal gradients.
Baranzelli et al. (2022)	Plants	mtDNA: various loci	27	5182	Patagonia	IGD higher in areas with only moderate land use.
Parsons et al. (2022)	Mammals	mtDNA: Cytb, COI	4310	90,759	Global	Hidden species more likely to be found in smaller-bodied taxa and those with large geographic ranges.
French et al. (2023)	Insects	mtDNA: COI	98,417 (OTUs)	2,415,425	Global	IGD did not significantly vary with latitude, but was higher in temperate areas just below LGM ice sheets.
Larkin et al. (2023)	Squamata	mtDNA: Cytb, COI; GBS	22; 22; 30	137, 137; 118	North America	IGD positively correlated with range size.

(Continues)

TABLE 1 | (Continued)

References	Taxa	Marker	Number of species	Number of sequences	Geographic scale	Pattern
Fonseca et al. (2023)	Arachnids, insects, plants, vertebrates	mtDNA, cpDNA: various	38,134	1,048,698	Global	IGD higher in tropics for plants and vertebrates. Statistics summarizing demographic change higher in temperate species.
Lawrence et al. (2023)	Vertebrates	Microsatellites	245	Not reported	North and South America	Hot and cold spots of IGD but no strong latitudinal gradients; taxonomic inconsistencies in spatial and environmental variables.
Amador et al. (2024)	Amphibians	mtDNA	256	4025	Neotropics	IBD explained by elevation; IBE explained by latitude and precipitation.
Figuerola-Ferrando et al. (2023)	Marine invertebrates	Microsatellites	140	9306	Global	Correlation between genetic diversity and scale of ecoregion. Inconsistent pattern across plants and animals.
Clark and Pinsky 2024	Fishes	mtDNA; microsatellites	492	6862	Global	nDNA IGD shows no strong geographic patterns; mDNA IGD peaks at tropics.
Parsons et al. (2024)	Cuadata	mtDNA: Cytb; COI	93	908; 768	Global	Hidden genetic lineages explained by variation in climate variables.
Decker et al. (2025)	Volant vertebrates	mtDNA: various loci	340	19,213	Global	Higher incidence of phylogeographic structure in bats with larger ranges at higher latitudes and more widespread in birds.
Peres and Mantelatto (2025)	Decapods	mtDNA: COI	23	1883	Tropical Atlantic	IGD higher near equator and associated with oxygen levels.

Note: We do not include meta-analyses that do not repurpose existing data. Each study includes information about the focal species (Taxa), genetic data used (Marker), number of species (Number of species), and sequences (Number of sequences), geographic region (Geographic scale), and the major patterns identified by the work (Pattern).

estimates of common summary statistics (Prugnolle and de Meeus 2002). While the prevalence of these phenomena is unknown for most clades, their combined impact could be substantial in investigations that utilize data from thousands of species. Macrogenetics relies on the assumption that estimates of genetic diversity taken from mitochondrial or chloroplast loci are broadly representative of overall genetic diversity, but demographic and stochastic effects question this assumption.

Some published investigations have expressed concerns about the interpretations made from these data (e.g., Paz-Vinas et al. 2021) especially from a conservation perspective (Schmidt and Garroway 2021). However, despite the potential shortcomings of organellar data, there is some evidence to suggest that summaries of genetic diversity derived from these datasets are representative of broader patterns. For example, Exposito-Alonso et al. (2022) gathered full genome data from 20 plant and animal species and found that genetic diversity from these species was positively correlated with geographic area, a pattern similar to those documented for single-locus mtDNA and chloroplast DNA (cpDNA) datasets (Pelletier et al. 2022). However, we identified only one investigation that explicitly examined disparities between the geographical patterns inferred from mtDNA vs. nuclear DNA (nDNA) across a regional scale (i.e., Italian Peninsula; Chiocchio et al. 2024). Failing to find further examples, we explored the extent to which genetic diversity estimates from mtDNA could be predictive of estimates from nDNA. We compiled two separate data sets: one consisting of sequence data from nuclear genes that was targeted, filtered, or transcriptomic (i.e., targeted sequencing dataset), and a second consisting of SNP data collected as part of reduced representation sequencing (i.e., random SNP dataset).

3.1 | Targeted Sequence Data

The targeted sequencing dataset was compiled by conducting a literature search to identify meta-analyses that reported nucleotide diversity (π) estimated from nuclear genes. We were inspired by data from Buffalo (2021) who identified three papers (Corbett-Detig et al. 2015; Leffler et al. 2012; Romiguier et al. 2014) and collected π values from 195 metazoan species (we pulled data from Buffalo 2021; Table S1). We then used Biopython v1.81 to query NCBI Genbank with the search terms 'cytochrome' and 'mitochondrion' to generate mitochondrial data sets for each of these species. A minimum threshold of 2 sequences and a maximum threshold of 50 sequences per species was used, resulting in mtDNA data for 181 of the targeted species. For species with GenBank sequences exceeding the sequence threshold, we randomly selected 50 sequences for download to limit sample size differences across datasets and computational demands (Table S2). Data from each species were aligned with MAFFT v7.471 (Katoh and Standley 2013) using default parameters and ambiguous sites were removed using TRIMAL v1.4.1 (Capella-Gutiérrez et al. 2009). π for each sequence alignment was calculated with Dendropy v4.5.2 (Sukumaran and Holder 2010). Requiring a minimum of 10 sequences per species for analysis, we ended up with 165 species in the final dataset.

Data were checked for normality and Spearman's rank correlation was used to test for a correlation between $\pi_{\text{autosomal}}$ and $\pi_{\text{mitochondrial}}$. Results indicate that there is no correlation between estimates of $\pi_{\text{autosomal}}$ and $\pi_{\text{mitochondrial}}$ ($r = -0.029$; $p = 0.706$; Figure 1a). Subsampling different sample sizes gives similar results, for example, when a minimum of 50 mtDNA sequences is retained ($n = 102$) there is still no correlation between estimates of $\pi_{\text{autosomal}}$ and $\pi_{\text{mitochondrial}}$ ($r = -0.063$; $p = 0.5234$), and there was no correlation between estimates of π and sample size for the mtDNA ($r = -0.0723$; $p = 0.3334$), suggesting sample size is not biassing the results. However, when separated by class, things are a little less clear. Only Insecta ($n = 62$), Mammalia ($n = 36$), Aves ($n = 10$), and Bivalvia ($n = 8$) had sample sizes worth exploring (Table S6). Only Mammalia and Aves had significant correlations between $\pi_{\text{autosomal}}$ and $\pi_{\text{mitochondrial}}$ ($r = 0.36$; $p = 0.029$; $r = -0.48$; $p = 0.019$; respectively), suggesting there is no broad-scale pattern; however, these are not large sample sizes per class and should be explored further.

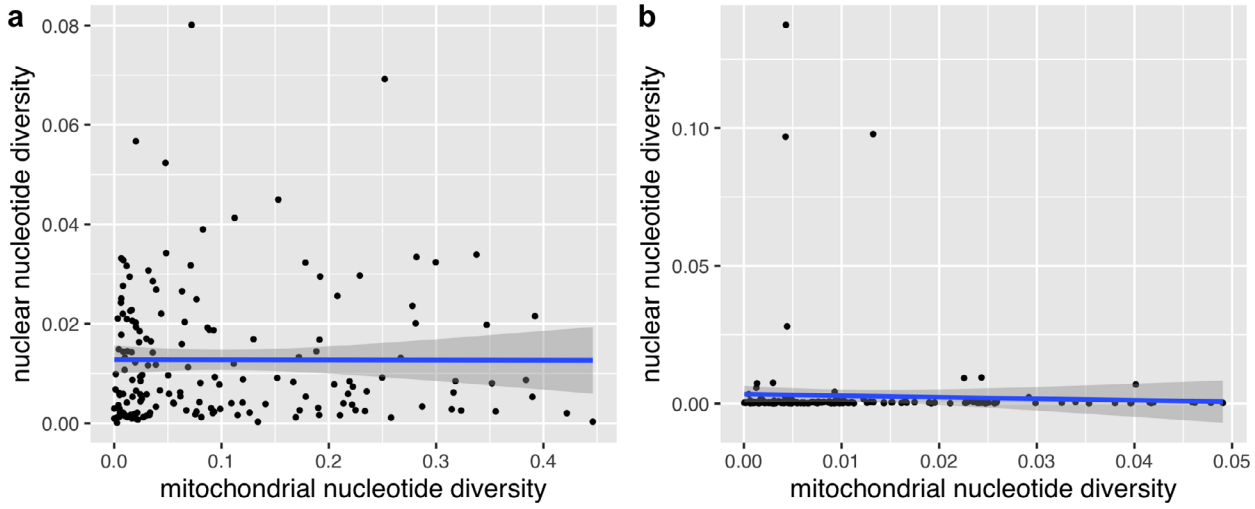

FIGURE 1 | (a) Relationship between measurements of nucleotide diversity for mtDNA sequences and nDNA targeted sequences for 165 species. (b) Relationship between measurements of nucleotide diversity for mtDNA sequences and random SNP data for 172 species.

3.2 | Random SNP Data

The random SNP dataset was compiled by querying three databases (Data dryad, Figbase, and Xenodo) that contain genetic data collected for use in phylogeographic or landscape genomic investigations. We located data in .vcf format that matched search terms ('phylogeograph*', 'landscape gen*', 'population genomic*'). We then verified that these data were associated with published papers and read the abstracts to identify datasets that were collected from natural populations (Table S3). For each species, $\pi_{autosomal}$ was estimated from .vcf files using VCFtools v0.1.16 (Danecek et al. 2011) with non-overlapping 1000 base pair (bp) windows. This window size was chosen to approximate the length of the mitochondrial genes. To obtain mtDNA, we searched the NCBI Nucleotide database via Biopython v.1.85 for sequences annotated as CYTB or COI containing ≥5 and a maximum of 50 sequences per species. For each species, the mitochondrial gene with the greatest number of sequences was downloaded (Table S4). Mitochondrial sequences were aligned using MAFFT v7.526 (Katoh and Standley 2013) using default parameters and checked by eye. Nucleotide diversity was calculated using the R package pegas v1.3 (Paradis 2010). Outliers were removed using IQR and only those with >10 mtDNA sequences were retained resulting in 172 species. Since most species had more SNPs in their mitochondrial data than were present in 1000 bp windows from the .vcf file, we repeated the analysis using 10,000 bp windows (Table S5).

We used Spearman's rank correlation to test for a correlation between $\pi_{autosomal}$ and $\pi_{mitochondrial}$ as above. Results indicate that there is no correlation between estimates of $\pi_{autosomal}$ and $\pi_{mitochondrial}$ using the 1000 bp window ($r=-0.062$; $p=0.4157$; Figure 1b) or the 10,000 bp window ($r=-0.054$; $p=0.4809$). While π should be robust to sample size differences between the autosomal and mitochondrial datasets, we also downsampled corresponding datasets so that both datasets within a species have the same sample size (Table S7). We reran the correlation analyses to ensure that uneven sample sizes did not change our interpretations. The results were similar for both windows (1000 bp window: $r=-0.044$; $p=0.5697$; 10,000 bp window: $r=-0.034$; $p=0.6546$). Again, there was no correlation between estimates of π and sample size for the mtDNA ($r=-0.0379$; $p=0.5925$), suggesting sample size is not biassing our results. Pixy is an additional way to calculate nucleotide diversity that can handle biasses that missing data can generate (Korunes and Samuk 2021). We ran pixy for a limited set of our data, but as our VCFs lack invariant sites, pixy cannot properly adjust the denominator, resulting in estimates of nucleotide diversity 1–2 orders of magnitude above the measure calculated by VCFtools. As such, we have elected to not calculate the values via pixy for more than a subset of this dataset ($n=47$; Table S5). The pixy estimate we calculated results in the same conclusion as our other correlations ($r=0.06747472$; $p=0.6171$). In this dataset, Insecta ($n=23$), Mammalia ($n=24$), Aves ($n=31$) and Teleostei ($n=46$) had sample sizes worth exploring (Table S6). Similarly, only Mammalia had a significant correlation between $\pi_{autosomal}$ and $\pi_{mitochondrial}$ ($r=0.49$; $p=0.014$), again suggesting no broad-scale pattern, and that conclusions based on levels of genetic diversity in mtDNA might be taxon specific.

4 | The Next Generation of Macrogenetic Research

Our results suggest that estimates of genetic diversity from mtDNA are not always correlated with similar estimates from nDNA, directly contradicting the assumption that mitochondrial diversity is broadly representative of the overall genetic diversity of a focal species. This discrepancy could be related to the expected time to coalescence, which is four times greater for autosomal loci when compared to organellar DNA, with increased standard deviation (Hudson and Turelli 2003). These effects can be compounded by the historical demography of a species. A population bottleneck will reduce genetic diversity to a greater extent in organellar than autosomal genomes due to differences in their effective population size (Hudson and Turelli 2003), and post glacial range expansion will involve sequential population reductions at the leading edge, which results in declining genetic diversity away from the origin (Excoffier et al. 2009), and similarly, this pattern will be more pronounced in single locus data. Furthermore, hybrid zones can have elevated levels of genetic diversity in autosomal genes, but this can be difficult to observe in single locus data. Finally, a number of studies deposit only haplotype data to online repositories limiting accuracy for some measures of intraspecific genetic diversity when data are repurposed (Paz-Vinas et al. 2021; Pelletier et al. 2022).

Genomic data are time consuming to collect, expensive to generate, and require bioinformatic processing prior to analysis, but this process is getting easier each year. While large spatial and taxonomic gaps in genomic data are evident (Figure S1), especially in protected areas (Paz-Vinas et al. 2025), these gaps will likely become smaller in the future. However, since collecting any type of genetic data across hundreds to thousands of species is likely to be impossible for individual research labs, macrogenetics will continue to rely on data repurposing, which also increases the value of the original data collection. During the compilation of our SNP dataset, data curation was time consuming yet necessary (Figure 2), where a lack of metadata is a known problem (Crandall et al. 2023; Toczydlowski et al. 2021). Examples of problems that we encountered include species that lacked a valid binomial, taxon names with misspellings, or other metadata concerns that required detailed examination of supplemental files, such as labelling individuals to populations and species. The reconciliation of taxonomic names was particularly time consuming, as many SNP data sets have been collected from a nominal species for the purposes of species delimitation, and in some cases, we had to identify mtDNA that correspond to described taxon boundaries. Our initial search for datasets yielded 502 .vcf files; this was reduced to 439 after manual curation, and to 245 once we identified appropriate matching mtDNA. Additionally, the majority of .vcf files we obtained contained genomic quality control scores (Table S5), and we suggest this should be standard reporting for genomic datasets to allow others to filter based on quality control needs for specific studies. We supplemented this score with the average per site missingness and standard deviation (Table S5) and find that missingness doesn't explain π in our SNP dataset (Adjusted R-squared = 9.024e-05; $p=0.3143$). We also suggest considering including supplemental tables with geographic information to link genetic data from individuals to their geographic locations and/or

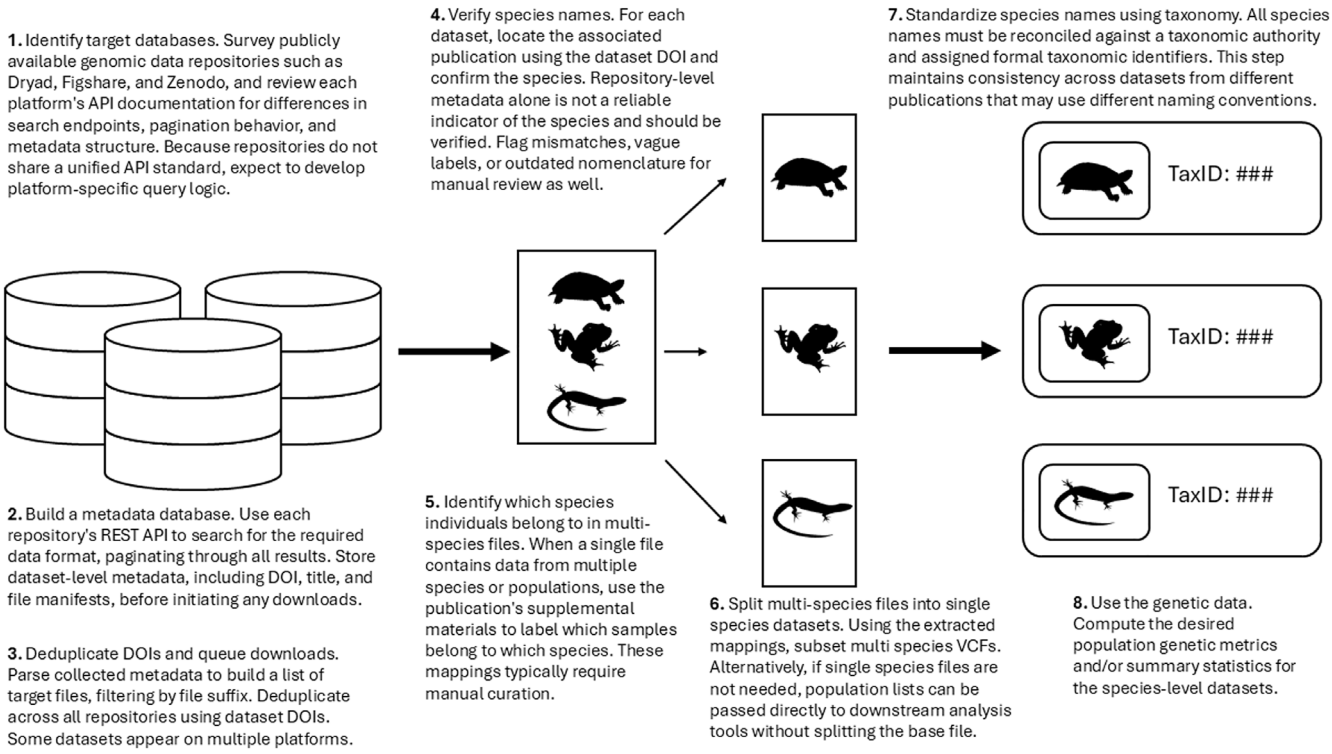

FIGURE 2 | Workflow for gathering SNP data.

including GPS coordinates with online database submissions, greatly increasing the utility of the data.

5 | Conservation Implications

Studies have shown that organellar genes can be a useful marker for conservation purposes (Petit-Marty et al. 2020; Schmidt et al. 2023; Thompson et al. 2021). Baranzelli et al. (2022) used sequence data from non-coding regions of chloroplast genomes to show that genetic diversity would decrease given future climate scenarios and human impact in flowering plants in Patagonia. In animals, mitochondrial data have been widely used for assessing cryptic diversity, and when detected, this undescribed diversity can have substantial implications for conservation status (Bickford et al. 2007). The widespread availability of these data enables large scale investigations to predict where cryptic diversity may be hiding. Parsons et al. (2022) determined that SE Asia is largely under-described for mammals compared to other parts of the globe. Similarly, ~2/3 of salamander species may contain hidden diversity, where hidden species were best explained by variation in climatic factors rather than species traits (Parsons et al. 2024). In mammals, microsatellite genetic diversity was higher in species that were found close to protected areas, demonstrating that genetic diversity can be maintained by protecting geographic space (Schmidt et al. 2025), though currently protected areas do not always house the most intraspecific genetic diversity (Chiocchio et al. 2024; Schultz et al. 2024). Collectively, these investigations suggest that organellar data still play a useful role in biodiversity science and macrogenetics has an important role to play in the conservation of species (e.g., Theodoridis et al. 2020). However, our results imply that researchers should treat inferences derived solely from mtDNA as a first pass at understanding evolutionary patterns and making inferences about levels of genetic diversity.

6 | Conclusions

We did not find estimates of genetic diversity from mtDNA to be overall correlated with similar estimates from nDNA in either our targeted sequencing or random SNP datasets. Taxonomic diversity in these analyses mirrored that observed in existing macrogenetics databases (Figure S1), suggesting that the species sampled for use in these analyses are broadly representative of available data, but still biased. Enthusiasm about the emerging macrogenetics research programme may be tempered by our results, as many inferences from macrogenetics research have resulted from repurposing mitochondrial or chloroplast data. We remain convinced that identifying the environmental and evolutionary forces that shape intraspecific genetic diversity, especially from a global perspective, remains a worthwhile objective. However, researchers should be cognizant of the limitations of the available data and temper their interpretations of results as needed.

Single-locus macrogenetic investigations have contributed important insight into broad-scale global biodiversity patterns. Baseline genetic diversity assessments conducted at a global scale predict loss and implement forward thinking conservation management practices (Hoban et al. 2021; Henry 2025). One estimate suggests that 10% of global genetic diversity may already be lost and that an additional 5% may be lost in the next decade (Exposito-Alonso et al. 2022). As such, it is imperative to develop macrogenetics research even if it is likely to rely on imperfect data. Genomic data should be the goal for macrogenetics

because these data represent most of the additive genetic variance within populations, and this variation is crucial for adaptation to global change (Mable 2019). While investigations that utilize chloroplast or mitochondrial DNA sequences remain useful, researchers should be aware that these data may limit the inferences that result from these analyses.

Acknowledgements

We thank the Carstens lab, MER editor, and anonymous reviewers for comments on this manuscript and feedback on data analysis. We appreciate student support from the Radford University Department of Biology and the Ohio State University Department of Evolution, Ecology, and Organismal Biology. Funding was provided by the National Science Foundation (DBI-1910623) to B.C.C. and the National Science Foundation (DBI-1911293) to T.A.P.

Funding

Funding was provided by the National Science Foundation (DBI-1910623) to B.C.C. and the National Science Foundation (DBI-1911293) to T.A.P.

Ethics Statement

This study relied on publicly available genetic data retrieved from open-source repositories, including NCBI GenBank, Dryad, Figbase, and Xenodo. No live animals or field specimens were sampled or handled. All data aggregation complied with data-use policies of the open-source databases, and all accessions, download DOIs, and data citations are included in Tables S1–S7.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Scripts for the targeted dataset can be found on GitHub (https://github.com/jordansatler/get-seqs-genbank). Scripts for the SNP dataset can be found on GitHub (https://github.com/nzupp/mito-nuclear-diversity). All data from this study are open-source.

References

Allio, R., S. Donega, N. Galtier, and B. Nabholz. 2017. "Large Variation in the Ratio of Mitochondrial to Nuclear Mutation Rate Across Animals: Implications for Genetic Diversity and the Use of Mitochondrial DNA as a Molecular Marker." *Molecular Biology and Evolution* 34: 2762–2772. https://doi.org/10.1093/molbev/msx197.

Amador, L., I. Arroyo-Torres, and L. N. Barrow. 2024. "Machine Learning and Phylogenetic Models Identify Predictors of Genetic Variation in Neotropical Amphibians." *Journal of Biogeography* 51: 909–923. https://doi.org/10.1111/jbi.14795.

Avise, J. C., J. Arnold, R. M. Ball, et al. 1987. "Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics." *Annual Review of Ecology, Evolution, and Systematics* 18: 489–522. https://doi.org/10.1146/annurev.es.18.110187.002421.

Baranzelli, M. C., A. Cosacov, S. M. Sede, M. V. Nicola, and A. N. Sérsic. 2022. "Anthropocene Refugia in Patagonia: A Macrogenetic Approach to Safeguarding the Biodiversity of Flowering Plants." *Biological Conservation* 268: 109492. https://doi.org/10.1016/j.biocon.2022.109492.

Barrow, L. N., E. M. da Fonseca, C. E. P. Thompson, and B. C. Carstens. 2021. "Predicting Amphibian Intraspecific Diversity With Machine Learning: Challenges and Prospects for Integrating Traits, Geography, and Genetic Data." *Molecular Ecology Resources* 21: 2818–2831. https://doi.org/10.1111/1755-0998.13303.

Bensasson, D., D. Zhang, D. L. Hartl, and G. M. Hewitt. 2001. "Mitochondrial Pseudogenes: Evolution's Misplaced Witnesses." *Trends in Ecology & Evolution* 16: 314–321. https://doi.org/10.1016/S0169-5347(01)02151-6.

Bickford, D., D. J. Lohman, N. S. Sodhi, et al. 2007. "Cryptic Species as a Window on Diversity and Conservation." *Trends in Ecology & Evolution* 22: 148–155. https://doi.org/10.1016/j.tree.2006.11.004.

Buffalo, V. 2021. "Quantifying the Relationship Between Genetic Diversity and Population Size Suggests Natural Selection Cannot Explain Lewontin's Paradox." *eLife* 10: e67509. https://doi.org/10.7554/elife.67509.

Capella-Gutiérrez, S., J. M. Silla-Martínez, and T. Gabaldón. 2009. "trimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses." *Bioinformatics* 25: 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

Carling, M. D., and R. T. Brumfield. 2007. "Gene Sampling Strategies for Multi-Locus Population Estimates of Genetic Diversity (Theta)." *PLoS One* 2: e160. https://doi.org/10.1371/journal.pone.0000160.

Carstens, B. C., A. E. Morales, K. Field, and T. A. Pelletier. 2018. "A Global Analysis of Bats Using Automated Comparative Phylogeography Uncovers a Surprising Impact of Pleistocene Glaciation." *Journal of Biogeography* 45: 1795–1805. https://doi.org/10.1111/jbi.13382.

Chiocchio, A., N. L. Santostasi, A. Pezzarossa, R. Bisconti, L. Maiorano, and D. Canestrelli. 2024. "Conserving Genetic Diversity Hotspots Under Climate Change: Are Protected Areas Helpful?" *Biological Conservation* 299: 110828. https://doi.org/10.1016/j.biocon.2024.110828.

Clark, R. D., and M. L. Pinsky. 2024. "Global Patterns of Nuclear and Mitochondrial Genetic Diversity in Marine Fishes." *Ecology and Evolution* 14: e11365. https://doi.org/10.1002/ece3.11365.

Cooper, J. D., P. M. Waser, D. Gopurenko, E. C. Hellgren, T. M. Gabor, and J. A. DeWoody. 2010. "Measuring Sex-Biased Dispersal in Social Mammals: Comparisons of Nuclear and Mitochondrial Genes in Collared Peccaries." *Journal of Mammalogy* 91: 1413–1424. https://doi.org/10.1644/09-MAMM-A-313.1.

Corbett-Detig, R. B., D. L. Hartl, and T. B. Sackton. 2015. "Natural Selection Constrains Neutral Diversity Across A Wide Range of Species." *PLoS Biology* 13: e1002112. https://doi.org/10.1371/journal.pbio.1002112.

Crandall, E. D., R. H. Toczydlowski, L. Liggins, et al. 2023. "Importance of Timely Metadata Curation to the Global Surveillance of Genetic Diversity." *Conservation Biology* 37, no. 4: e14061. https://doi.org/10.1111/cobi.14061.

Danecek, P., A. Auton, G. Abecasis, et al. 2011. "The Variant Call Format and VCFtools." *Bioinformatics* 27: 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Dapporto, L., M. Menchetti, V. Dincă, et al. 2024. "The Genetic Legacy of the Quaternary Ice Ages for West Palearctic Butterflies." *Science* 10: adm8596. https://doi.org/10.1126/sciadv.adm8596.

De Kort, H., J. G. Prunier, S. Ducatez, et al. 2021. "Life History, Climate and Biogeography Interactively Affect Worldwide Genetic Diversity of Plant and Animal Populations." *Nature Communications* 12: 516. https://doi.org/10.1038/s41467-021-20958-2.

Decker, S. K., K. L. Provost, and B. C. Carstens. 2025. "Bats of a Feather: Range Characteristics and Wing Morphology Predict Phylogeographic Breaks in Volant Vertebrates." *Frontiers of Biogeography* 18: e139911. https://doi.org/10.21425/fob.18.139911.

Edwards, S., and P. Beerli. 2000. "Perspective: Gene Divergence, Population Divergence, and the Variance in Coalescence Time in

Phylogeographic Studies." *Evolution* 54: 1839–1854. https://doi.org/10.1111/j.0014-3820.2000.tb01231.x.

Edwards, S. V., V. V. Robin, N. Ferrand, and C. Moritz. 2022. "The Evolution of Comparative Phylogeography: Putting the Geography (And More) Into Comparative Population Genomics." *Genome Biology and Evolution* 14: evab176. https://doi.org/10.1093/gbe/evab176.

Excoffier, L., M. Foll, and R. J. Petit. 2009. "Genetic Consequences of Range Expansions." *Annual Review of Ecology, Evolution, and Systematics* 40: 481–501. https://doi.org/10.1146/annurev.ecolsys.39.110707.173414.

Exposito-Alonso, M., T. R. Booker, L. Czech, et al. 2022. "Genetic Diversity Loss in the Anthropocene." *Science* 377: 1431–1435. https://doi.org/10.1126/science.abn5642.

Felsenstein, J. 2006. "Accuracy of Coalescent Likelihood Estimates: Do We Need More Sites, More Sequences, or More Loci?" *Molecular Biology and Evolution* 23: 691–700. https://doi.org/10.1093/molbev/msj079.

Figuerola-Ferrando, L., A. Barreiro, I. Montero-Serra, et al. 2023. "Global Patterns and Drivers of Genetic Diversity Among Marine Habitat-Forming Species." *Global Ecology and Biogeography* 32: 1218–1229. https://doi.org/10.1111/geb.13685.

Fonseca, E. M., T. A. Pelletier, S. K. Decker, D. J. Parsons, and B. C. Carstens. 2023. "Pleistocene Glaciations Caused the Latitudinal Gradient of Within-Species Genetic Diversity." *Evolution Letters* 7: 331–338. https://doi.org/10.1093/evlett/qrad030.

French, C. M., L. D. Bertola, A. C. Carnaval, et al. 2023. "Global Determinants of Insect Mitochondrial Genetic Diversity." *Nature Communications* 14: 5276. https://doi.org/10.1038/s41467-023-40936-0.

Galtier, N., B. Nabholz, S. Glémin, and G. D. D. Hurst. 2009. "Mitochondrial DNA as a Marker of Molecular Diversity: A Reappraisal." *Molecular Ecology* 18: 4541–4550. https://doi.org/10.1111/j.1365-294X.2009.04380.x.

Garrick, R. C., I. A. S. Bonatelli, C. Hyseni, et al. 2015. "The Evolution of Phylogeographic Data Sets." *Molecular Ecology* 24: 1164–1171. https://doi.org/10.1111/mec.13108.

Henry, R. C. 2025. "Future Projections of Biodiversity Under Global Change Need to Include Genetic Diversity." *Global Change Biology* 31: e70477. https://doi.org/10.1111/gcb.70477.

Hickerson, M. J., B. C. Carstens, J. Cavender-Bares, et al. 2010. "Phylogeography's Past, Present, and Future: 10 Years After." *Molecular Phylogenetics and Evolution* 54: 291–301. https://doi.org/10.1016/j.ympev.2009.09.016.

Hoban, S., M. W. Bruford, W. C. Funk, et al. 2021. "Global Commitments to Conserving and Monitoring Genetic Diversity Are Now Necessary and Feasible." *Bioscience* 71: 964–976. https://doi.org/10.1093/biosci/biab054.

Hudson, R. R., and M. Turelli. 2003. "Stochasticity Overrules the 'Three-Times Rule': Genetic Drift, Genetic Draft, and Coalescence Times for Nuclear Loci Versus Mitochondrial DNA." *Evolution* 57: 182–190. https://doi.org/10.1111/j.0014-3820.2003.tb00229.x.

Katoh, K., and D. M. Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30: 772–780. https://doi.org/10.1093/molbev/mst010.

Knowles, L. L. 2009. "Statistical Phylogeography." *Annual Review of Ecology, Evolution, and Systematics* 40: 593–612. https://doi.org/10.1146/annurev.ecolsys.38.091206.095702.

Korunes, K. L., and K. Samuk. 2021. "Pixy: Unbiased Estimation of Nucleotide Diversity and Divergence in the Presence of Missing Data." *Molecular Ecology Resources* 21, no. 4: 1359–1368. https://doi.org/10.1111/1755-0998.13326.

Larkin, I. E., E. A. Myers, B. C. Carstens, and L. N. Barrow. 2023. "Predictors of Genomic Diversity Within North American Squamates." *Journal of Heredity* 114: 131–142. https://doi.org/10.1093/jhered/esad001.

Lawrence, E. R., E. J. Pedersen, and D. J. Fraser. 2023. "Macrogenetics Reveals Multifaceted Influences of Environmental Variation on Vertebrate Population Genetic Diversity Across the Americas." *Molecular Ecology* 32: 4557–4569. https://doi.org/10.1111/mec.17059.

Leffler, E. M., K. Bullaughey, D. R. Matute, et al. 2012. "Revisiting an Old Riddle: What Determines Genetic Diversity Levels Within Species?" *PLoS Biology* 10: e1001388. https://doi.org/10.1371/journal.pbio.1001388.

Leigh, D. M., C. B. van Rees, K. L. Millette, et al. 2021. "Opportunities and Challenges of Macrogenetic Studies." *Nature Reviews Genetics* 22: 791–807. https://doi.org/10.1038/s41576-021-00394-0.

Llopart, A., D. Herrig, E. Brud, and Z. Stecklein. 2014. "Sequential Adaptive Introgression of the Mitochondrial Genome in Drosophila Yakuba and Drosophila Santomea." *Molecular Ecology* 23: 1124–1136. https://doi.org/10.1111/mec.12678.

Mable, B. K. 2019. "Conservation of Adaptive Potential and Functional Diversity: Integrating Old and New Approaches." *Conservation Genetics* 20: 89–100. https://doi.org/10.1007/s10592-018-1129-9.

Manel, S., P.-E. Guerin, D. Mouillot, et al. 2020. "Global Determinants of Freshwater and Marine Fish Genetic Diversity." *Nature Communications* 11: 692. https://doi.org/10.1038/s41467-020-14409-7.

Millette, K. L. V., C. Fugère, A. Debyser, F. J. J. Chain, and A. Gonzalez. 2019. "No Consistent Effects of Humans on Animal Genetic Diversity Worldwide." *Ecology Letters* 23: 55–67. https://doi.org/10.1111/ele.13394.

Miraldo, A. S., M. K. Li, A. Borregaard, et al. 2016. "An Anthropocene Map of Genetic Diversity." *Science* 353: 1532–1535. https://doi.org/10.1126/science.aaf4381.

Paradis, E. 2010. "Pegas: An R Package for Population Genetics With an Integrated–Modular Approach." *Bioinformatics* 26: 419–420. https://doi.org/10.1093/bioinformatics/btp696.

Parsons, D. J., A. E. Green, B. C. Carstens, and T. A. Pelletier. 2024. "Predicting Genetic Biodiversity in Salamanders Using Geographic, Climatic, and Life History Traits." *PLoS One* 19: e0310932. https://doi.org/10.1371/journal.pone.0310932.

Parsons, D. J., T. A. Pelletier, J. G. Wieringa, D. J. Duckett, and B. C. Carstens. 2022. "Analysis of Biodiversity Data Suggests That Mammal Species Are Hidden in Predictable Places." *Proceedings of the National Academy of Sciences of the United States of America* 119: e2103400119. https://doi.org/10.1073/pnas.2103400119.

Paz-Vinas, I., E. L. Jensen, L. D. Bertola, et al. 2021. "Macrogenetic Studies Must Not Ignore Limitations of Genetic Markers and Scale." *Ecology Letters* 24: 1282–1284. https://doi.org/10.1111/ele.13732.

Paz-Vinas, I., A. G. Vandergast, C. Schmidt, et al. 2025. "Sparse Genetic Data Limit Biodiversity Assessments in Protected Areas Globally." *Frontiers in Ecology and the Environment* 23: e2867. https://doi.org/10.1002/fee.2867.

Pelletier, T. A., and B. C. Carstens. 2018. "Geographical Range Size and Latitude Predict Population Genetic Structure in a Global Survey." *Biology Letters* 14: 20170566. https://doi.org/10.1098/rsbl.2017.0566.

Pelletier, T. A., D. J. Parsons, S. K. Decker, et al. 2022. "phylogatR: Phylogeographic Data Aggregation and Repurposing." *Molecular Ecology Resources* 22: 2830–2842. https://doi.org/10.1111/1755-0998.13673.

Peres, P. A., and F. L. Mantelatto 2025. "Macrogenetics Approach Reveals Spatial Trends and Drivers of Mitochondrial Genetic Diversity at Different Biological Organization Levels in Tropical Western Atlantic

Decapods." *Ecology and Evolution* 15: e71372. https://doi.org/10.1002/ece3.71372.

Petit-Marty, N., M. Vázquez-Luis, and I. E. Hendriks. 2020. "Use of the Nucleotide Diversity in COI Mitochondrial Gene as an Early Diagnostic of Conservation Status of Animal Species." *Conservation Letters* 14: e12756. https://doi.org/10.1111/conl.12756.

Prugnolle, F., and T. de Meeus. 2002. "Inferring Sex-Biased Dispersal From Population Genetic Tools: A Review." *Heredity* 88: 161–165. https://doi.org/10.1038/sj.hdy.6800060.

Romiguier, J., P. Gayral, M. Ballenghien, et al. 2014. "Comparative Population Genomics in Animals Uncovers the Determinants of Genetic Diversity." *Nature* 515: 261–263. https://doi.org/10.1038/nature13685.

Schmidt, C., and C. J. Garroway. 2021. "The Conservation Utility of Mitochondrial Genetic Diversity in Macrogenetic Research." *Conservation Genetics* 22: 323–327. https://doi.org/10.1007/s10592-021-01333-6.

Schmidt, C., S. Hoban, and W. Jetz. 2023. "Conservation Macrogenetics: Harnessing Genetic Data to Meet Conservation Commitments." *Trends in Genetics* 39: 816–829. https://doi.org/10.1016/j.tig.2023.08.002.

Schmidt, C., E. Karachaliou, A. G. Vandergast, et al. 2025. "A Survey of Mammal and Fish Genetic Diversity Across the Global Protected Area Network." *Conservation Letters* 18: e13092. https://doi.org/10.1111/conl.13092.

Schultz, J. T., J. Geldmann, S. Theodoridis, and D. Nogues-Bravo. 2024. "Gaps in Global Protection of Terrestrial Genetic Diversity." bioRxiv: The Preprint Server for Biology. https://doi.org/10.1101/2024.08.23.609342.

Smith, B. T., G. F. Seeholzer, M. G. Harvey, A. M. Cuervo, and R. T. Brumfield. 2017. "A Latitudinal Phylogeographic Diversity Gradient in Birds." *PLoS Biology* 15: e2001073. https://doi.org/10.1371/journal.pbio.2001073.

Sukumaran, J., and M. T. Holder. 2010. "DendroPy: A Python Library for Phylogenetic Computing." *Bioinformatics* 26: 1569–1571. https://doi.org/10.1093/bioinformatics/btq228.

Theodoridis, S., D. A. Fordham, S. C. Brown, S. Li, C. Rahbek, and D. Nogues-Bravo. 2020. "Evolutionary History and Past Climate Change Shape the Distribution of Genetic Diversity in Terrestrial Mammals." *Nature Communications* 11: 2557. https://doi.org/10.1038/s41467-020-16449-5.

Thompson, C. E. P., T. A. Pelletier, and B. C. Carstens. 2021. "Genetic Diversity of North American Vertebrates in Protected Areas." *Biological Journal of the Linnean Society* 132: 388–399. https://doi.org/10.1093/biolinnean/blaa195.

Toczydlowski, R. H., L. Liggins, M. R. Gaither, et al. 2021. "Poor Data Stewardship Will Hinder Global Genetic Diversity Surveillance." *Proceedings of the National Academy of Sciences* 118, no. 34: e2107934118. https://doi.org/10.1073/pnas.2107934118.

Yiming, L., W. Siqi, C. Chaoyuan, et al. 2021. "Latitudinal Gradients in Genetic Diversity and Natural Selection at a Highly Adaptive Gene in Terrestrial Mammals." *Ecography* 2: 206–218. https://doi.org/10.1111/ecog.05082.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Figure S1:** Taxonomic distribution of data availability. Note that these are proportions, not total numbers, for each type of data (PhylogatR/organellar DNA, targeted genes, and random SNPs). **Table S1:** Target sequence data. Contains nucleotide diversity estimates, information about the species, information about the gene, and the citation. **Table S2:** List of accessions for mtDNA associated with the DNA sequence dataset. **Table S3:** Reference information for the random SNP dataset. **Table S4:** List of accessions for mtDNA associated with the random SNP dataset. **Table S5:** Random SNP data. Contains nucleotide diversity estimates and quality control estimates. **Table S6:** Correlations for datasets by Class. **Table S7:** Downsampled random SNP dataset.