Evolution Letters, 2023, **7**(5), 331–338
https://doi.org/10.1093/evlett/qrad030
Advance access publication 19 July 2023
Letter

Pleistocene glaciations caused the latitudinal gradient of within-species genetic diversity

Emanuel M. Fonseca[1], Tara A. Pelletier[2], Sydney K. Decker[1], Danielle J. Parsons[1], Bryan C. Carstens[1]

[1]Department of Evolution, Ecology and Organismal Biology, The Ohio State University, Columbus, OH, United States
[2]Department of Biology, Radford University, Radford, VA, United States

Corresponding author: Department of Evolution, Ecology and Organismal Biology, The Ohio State University, 318 W. 12th Ave., Columbus, OH 43210, United States. Email: emanuelmfonseca@gmail.com

Abstract

Intraspecific genetic diversity is a key aspect of biodiversity. Quaternary climatic change and glaciation influenced intraspecific genetic diversity by promoting range shifts and population size change. However, the extent to which glaciation affected genetic diversity on a global scale is not well established. Here we quantify nucleotide diversity, a common metric of intraspecific genetic diversity, in more than 38,000 plant and animal species using georeferenced DNA sequences from millions of samples. Results demonstrate that tropical species contain significantly more intraspecific genetic diversity than nontropical species. To explore potential evolutionary processes that may have contributed to this pattern, we calculated summary statistics that measure population demographic change and detected significant correlations between these statistics and latitude. We find that nontropical species are more likely to deviate from neutral expectations, indicating that they have historically experienced dramatic fluctuations in population size likely associated with Pleistocene glacial cycles. By analyzing the most comprehensive data set to date, our results imply that Quaternary climate perturbations may be more important as a process driving the latitudinal gradient in species richness than previously appreciated.

Key words: intraspecific genetic diversity, nucleotide diversity, latitudinal gradient, Pleistocene glaciation

Lay Summary

It is well appreciated among biologists that species richness is higher near the equator than at temperate latitudes. Using data from 38,000 species, we demonstrate here that a similar gradient is present for intraspecific genetic diversity. Results indicate that nontropical species were more likely to have undergone significant changes in their historical population demography than were tropical species. Decades of phylogeographic work have suggested that Quaternary glaciation forced temperate species into isolated glacial refugia. Under this scenario, genetic diversity would be lost due to reduced population size, population bottlenecks, and genetic drift. Our analyses represent the most comprehensive test of the Quaternary glaciation hypothesis to date and suggest that the latitudinal gradient in intraspecific genetic diversity is at least partially a result of Pleistocene glaciation.

Introduction

Intraspecific genetic diversity (IGD) is an important component of biodiversity because it provides adaptive potential (Frankham, 2005), facilitates species persistence (Frankham, 2005; Pereira et al., 2012), and provides clues about the historical demography of species (Avise, 2000). Although IGD is of pivotal importance to the long-term maintenance of biodiversity (Hedrick & Kalinowski, 2000), spatial patterns of intraspecific genetic diversity and their causes are not well documented globally. Existing surveys of IGD are often limited in their geographic and/or taxonomic scope (e.g., Barrow et al., 2021) due to challenges associated with assembling and analyzing large data sets. The most comprehensive surveys to date, which sampled ~4,500 species of terrestrial vertebrates (Miraldo et al., 2016) and ~5,400 fish (Manel et al., 2020), found evidence that species in tropical areas contain higher levels of IGD than nontropical species. While these investigations did not include a wide range of taxonomic groups or explore the potential historical causes of the observed pattern, they suggest that a latitudinal gradient in IGD may be present, mirroring the well-known species richness latitudinal gradient (Gaston, 2000; Willig et al., 2003). Establishing that there is a latitudinal gradient in intraspecific genetic diversity across disparate taxonomic groups and discovering causal explanations for this gradient would represent an important step toward understanding the global distribution of biodiversity.

Quaternary glaciation, which has long been recognized as a prominent factor that influences IGD (Hewitt, 2004), is a potential cause of spatial variation in genetic diversity. During the Quaternary, cyclical global cooling coupled with ice sheet advance and retreat promoted considerable demographic change and extinction (Hewitt, 2004; Nogués-Bravo et al., 2010). A recent survey of >2,000 mammal species found that climatic stability, both seasonally and over evolutionary time scales, covaried with intraspecific genetic diversity (Theodoridis et al., 2020), a result consistent with the findings of many early phylogeographic investigations (e.g., summarized in Avise 2000). Since species at

Received April 13, 2022; revisions received June 26, 2023; accepted July 10, 2023

high latitudes are more likely to inhabit regions affected by glaciation, apparent patterns of latitudinal variation in IGD could result from a loss of intraspecific genetic diversity via genetic drift in response to these population bottlenecks at higher latitudes (Avise, 2000; Hedrick & Kalinowski, 2000).

Historically, sequence data from large numbers of samples have been collected by researchers in disciplines ranging from landscape genetics (Holderegger & Wagner, 2008) and phylogeography (Avise, 2000) to taxonomic investigations using DNA barcoding (Hebert & Gregory, 2005) on a species by species basis. Collectively, these data represent a substantial resource that could be used to document spatial patterns of IGD and identify the processes responsible for these patterns. Here, we explore both questions by collecting genetic data from more than 38,000 species across the tree of life using the *phylogatR* database (Pelletier et al., 2021), which uses metadata records to associate DNA sequence accessions from databases such as NCBI GenBank and Barcode of Life Database (BOLD) to natural history and locality data stored in Global Biodiversity Information Facility (GBIF). An expansive quality control protocol was implemented because repurposed data can contain a variety of errors including incorrect taxonomic assignment, incorrect alignment, or artifactual insertions/deletions in the gene sequence (Miraldo et al., 2016; Sidlauskas et al., 2010). After using a stringent set of filters (see Supplementary Materials and Supplementary Figure S1), our data set contained a total of 1,048,698 unique DNA sequences from 38,134 species (28,656 insects, 6,959 vertebrates, 1,559 arachnids, and 960 plants: Supplementary Table S1). These data allowed us to assess the impact that Pleistocene climate change and glaciation had on intraspecific genetic diversity on a global scale.

Methods

Aggregating DNA sequence from open-source repositories

We aggregated georeferenced mitochondrial DNA for animals and chloroplast DNA for plants using *phylogatR* (Pelletier et al., 2021; https://phylogatr.osc.edu), an open-source platform available through the Ohio Supercomputer Center (OSC). *phylogatR* retrieves genetic and geographic data by integrating information from the GBIF (https://www.gbif.org/), GenBank (https://www.ncbi.nlm.nih.gov/genbank/), and the BOLD (http://www.boldsystems.org/index.php). To retrieve georeferenced DNA sequences, *phylogatR* first accesses GBIF records and parses out records with GenBank accession numbers. Only records flagged as *PreservedSpecimen*, *MaterialSample*, *HumanObservation*, and *MachineObservation* and that held *AssociatedSequences* are retrieved. After that, the retrieved accession numbers are used to download DNA sequences from GenBank. To minimize data loss and duplication when downloading data from GenBank, duplicate entries with the same accession number, but different GBIF occurrence numbers are checked for similarities in geographic coordinates, basis of record, species names, and event dates. In addition to the sequences from GenBank, sequence and geographic coordinate data from BOLD are incorporated for records without GenBank or GBIF accessions already in the database.

As part of the pipeline, *phylogatR* standardizes gene names. For example, although the mitochondrial gene cytochrome-b has multiple entries on GenBank such as cytochrome-b, cytb, cyt-b, cytbl, and cyb, all these entries are coded under the unique name of CYTB in *phylogatR*. Next, sequences from individual gene names for each species are concatenated and multiple sequence alignment (MSA) is performed using *adjustdirection* and *inputorder* features in MAFFT v7 (Katoh & Standley, 2013). These features enable the direction of sequences to be changed and preserve the original input order of the sequences. In cases where data from multiple genes were available in a single species, we used the gene with the greatest number of sequences. Because the genetic information available for arachnids were largely concentrated in the Northern Hemisphere (~93.5%), we did not include these data in all analyses. To clean the alignments, trimAl v1.2 (Capella-Gutiérrez et al., 2009) is used by setting residue overlap (*resoverlap*), sequence overlap (*seqoverlap*), and gap threshold (*gt*) parameters to 0.85, 50, and 0.15, respectively. Multiple MSA parameters were tested before reaching the final parameter settings.

Data description

We obtained data for 65,725 species, including amphibians (583), birds (2,034), fish (6,817), mammals (903), reptiles (640), insects (47,164), arachnids (2,093), and plants (5,491), totaling 1,158,261 unique DNA sequences. Mitochondrial DNA was used for vertebrates, arachnids, and insects, while chloroplast DNA was used for plants. Our analyses were limited to these data to avoid the complications that would result from comparing genes with different ploidy levels and because they are by far the most available across species for these groups. We further processed the data and removed species with less than five sequences for downstream analyses. A recent study demonstrated that the variance in π declines sharply with a sample size higher than five sequences (Barrow et al., 2021). After processing and filtering the data, our working data set consisted of a total of 38,134 species. Specifically, we gathered genetic information for 368 amphibians, 1,091 birds, 4,511 fish, 642 mammals, 347 reptiles, 28,656 insects, 1,559 arachnids, and 960 plants, totaling 1,048,698 unique DNA sequences. Next, we split our data into four sets: (a) vertebrates (i.e., amphibians, birds, reptiles, mammals, and fish), (b) insects, (c) arachnids, and (d) plants. Detailed information about the genes used in each data set and their proportion is presented in Supplementary Table S1. For vertebrates, the data are homogeneously distributed across the globe (Supplementary Figure S2). In contrast, for insects and plants, the data are geographically concentrated in temperate regions (Supplementary Figures S3–S5), reflecting well-known biases inherent to molecular ecology investigations (Beheregaray, 2008; Turchetto-Zolet et al., 2013).

Calculating genetic diversity

Nucleotide diversity (π) was calculated to describe patterns of intraspecific genetic diversity. Although other statistics such as the number of haplotypes and haplotype diversity are also used to describe patterns of genetic diversity within species, they are not suitable for many of the sequences retrieved from open-source repositories because many sequences are uploaded to GenBank as unique haplotypes per locality and/or have different lengths, precluding a reliable estimation of those genetic diversity measures (Miraldo et al., 2016). As a result, we choose to describe the patterns of genetic diversity using π, defined as the average number of polymorphic sites in pairwise sequence comparison following Nei and Li (1979) and presented in Equation 1:

$$\Pi = \frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} k_{ij} \quad (1)$$

where $\binom{n}{2}$ is number of pairwise sequence comparisons and k_{ij} is the number of polymorphic sites between sequences i and j. We implemented the slightly modified version of equation 1 that was proposed by Miraldo et al. (2016) to allow for a comparison of sequences of different lengths. In the modified version, k_{ij} is

divided by the length of shared base pairs between sequences (m_{ij}), which contrasts with the overall length used in equation 1. The equation of the modified version of nucletotide diversity is presented below.

$$\Pi = \frac{1}{\binom{n}{2}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{k_{ij}}{m_{ij}} \quad (2)$$

We calculated nucleotide diversity based on equation 2 for sequences with at least 50% of overlap in the comparison. In preliminary analyses, we noticed some high values of sequence divergence (>40%) between some of the sequences that were likely related to mislabeling. We approached this caveat by assuming that any comparisons where observed nucleotide diversity was higher than 10% resulted from inclusion of a sequence with an incorrect species label. These alignments were removed from the analysis, a conservative approach that prevents inflation on the calculation of genetic diversity within species. We used the function *nuc.div* implemented in *pegas* R package (Paradis, 2010) to calculate nucleotide diversity between pairs of sequences and then averaged them to obtain a global estimation of nucleotide diversity for each species.

Assessing demographic change

We used Tajima's *D* (Tajima, 1989) and R_2 (Ramos-Onsins & Rozas, 2002), two statistics that have been widely used to detect departure from neutrality that potentially result from population size change over time. Tajima's *D* is composed of two other metrics: the mean pairwise difference (π) and the number (*S*) of segregating sites, both representing an estimator of θ (theta). While the first metric measures the observed genetic variation in the sequences, the second metric quantifies the expected amount of genetic variation under the null model. Thus, under a scenario of constant population size, the observed amount of genetic variation (π) should be approximately equal to the expected genetic variation (*S*). For example, population expansions increase the number of rare variants (i.e., a polymorphic site present in only one—singletons—or in very few individuals) in the population, which leads to an increase in the expected value of theta. Conversely, rare polymorphisms lead to a lower value of π. The distortion between π and S, which represents a departure from mutation-drift equilibrium, is captured by Tajima's *D*. Negative values are interpreted as an evidence of population expansion and positive values as population bottleneck. Note that while some authors have argued that significant results could result from positive or balancing selection, respectively, the demographic causes are the most common interpretation in phylogeographic studies and likely to be the most relevant for a global comparison. To calculate Tajima's *D* and its associate *p*-value for each species, we used the function *tajima.test* in *pegas* R package (Paradis, 2010). Tajima's *D* is mathematically defined by:

$$D = \frac{\Pi - {}^{S}/_{a_1}}{\sqrt{e_1 S + e_2 S\ (S-1)}} \quad (3)$$

where $a_1 = \sum_{i=1}^{n-1} 1/i$, $a_2 = \sum_{i=1}^{n-1} 1/i^2$, $b_1 = \frac{n+1}{3(n-1)}$, $b_2 = \frac{2(n^2+n+3)}{9n(n-1)}$, $c_1 = b_1 - \frac{1}{a_1}$, $c_2 = b_2 - \frac{n+2}{a_1 n} + \frac{a_2}{a_1^2}$, $e_1 = \frac{c_1}{a_1}$, and $e_2 = \frac{c_2}{(a_1^2+a_2)}$.

The R_2 is a statistic formulated on the patterns of segregating site frequency like Tajima's *D*. However, in the case of R_2, it tests specifically for population expansion by computing the difference between the number of singletons and the average number of nucleotide differences. An excess of singletons is expected in the sample under a population growth scenario. R_2 ranges from 0 to 1 with values close to zero representing departures from neutrality. This is a robust statistic for studies dealing with small sample sizes (Ramos-Onsins & Rozas, 2002). We calculated R_2 and its significance value (i.e., *p*-value) using the function *R2.test* implemented *pegas* R package (Paradis, 2010). R_2 is calculated based on the following equation:

$$R_2 = \frac{\left(\sum_{i=1}^{n} \frac{\left(U_i - \frac{k}{2}\right)^2}{n}\right)^{1/2}}{S} \quad (4)$$

where *n* is the number of samples, U_i is the number of singletons in sequence i, *k* is the average number of nucleotide difference between two sequences, and *S* the total number of segregating sites.

Statistical analysis

We used linear regression to investigate whether genetic diversity increases toward the Tropics. To summarize geographic location, we extracted the centroid of the distribution of sampling localities in each species based on coordinates associated with the genetic data. We used absolute latitude as the predictor variable. Linear regressions were performed separately for vertebrates, insects, and plants in R using the function *lm*. We did not include arachnids because most of the data (>93.5%) were concentrated in the Northern Hemisphere. We also performed logistic regression to evaluate whether genetic diversity is inversely proportional to latitude. First, we split our data set into tropical regions (regions between the Tropic of Cancer [latitude –23.5°] and the Tropic of Capricorn [latitude 23.5°]) and nontropical regions (regions with latitude higher than 23.5° and lower than –23.5°). Then, we fit a logistic regression in R using the function *glm* (family = binomial). Finally, we create a null distribution of mean nucleotide diversity for tropical and nontropical by using a randomization procedure. We randomized the locations of species (i.e., tropical and nontropical regions) 1,000 times while keeping the size of the data set constant. In each data set randomized by location, we calculated the mean nucleotide diversity to create a null distribution. Next, the null distribution was compared to the observed value (i.e., calculated from the empirical data set) and a pseudo *p*-value was computed by counting the number of randomized values of genetic diversity that fell above 95% or below 5% from the observed value and then selecting the lowest value and dividing it by the total number of comparisons (i.e., 1,000 comparisons).

We also asked if the magnitude of population size change was higher in higher latitude. First, we calculated the proportion of species with significant ($p < .05$) and nonsignificant (NS) values of Tajima's *D* and R_2 statistics between tropical and nontropical regions in each group and compared them using Fisher's exact test using the R function *fisher.test*. Next, we performed simple linear regressions between absolute Tajima's *D* (response variable) and absolute latitude (predictor variable) for each group separately. We used the absolute value of Tajima's *D* because we were interested in quantifying the magnitude of the population size change and not the response itself (i.e., bottleneck or expansion), since in many systems glacial advance and subsequent retreat promoted a population bottleneck that was followed by population expansion. However, most of the significant values of Tajima's *D* (>99.5%) suggested population expansion. Finally, we also created a null distribution for Tajima's *D* and R_2 through a randomization procedure that mirrored that used for nucleotide diversity and described above. Both linear regressions and randomization were performed as described for nucleotide diversity. For all analyses, we used a Bonferroni correction to account for multiple comparison issues by using the R function *p.adjust*.

Results and discussion

Results indicate that intraspecific genetic diversity increases as the bands move toward the equator (Figure 1A–C; for arachnids see Supplementary Figure S6). Although we included different mitochondrial genes in our analyses, the vast majority (>0.98) were COI and we did not observe major differences in average genetic diversity among these markers (Supplementary Figure S7). Next, to demonstrate that latitude is predictive of intraspecific genetic diversity, we conducted a simple linear regression of latitude versus π (after a log transformation). Results are significant in vertebrates ($p < .01$; Figure 2A) and plants ($p < .01$; Figure 2C), but not in insects ($p > .01$; Figure 2B). While we did not conduct this regression in arachnids due to the underrepresentation of tropical species, the qualitative pattern is similar (Supplementary Figure S6). Results indicate that a latitudinal gradient in intraspecific genetic diversity is present across vertebrate and plant species, but not in insects. Note that French et al. (2022) have conducted an exploration of the determinants of genetic diversity in insects.

To quantify the differences between tropical and nontropical regions, we divided our data set into tropical regions and nontropical regions. A logistic regression of tropical versus nontropical species showed a higher concentration of IGD in tropical regions in all data sets ($p < .01$ for all after Bonferroni correction (Supplementary Figure S8a–c). We found similar results when regressing absolute latitude against π (Figure 2A–C; all $p < .01$; after Bonferroni correction), with increased nucleotide diversity towards the equator. To assess the statistical significance of the apparently different average levels of π between tropical and nontropical regions, we built a null distribution of π using a randomizing procedure and showed that the IGD in tropical species is higher than expected by chance ($p < .01$, Supplementary Figure S9a–c) and IGD in nontropical species is lower than expected by chance ($p < .01$, Supplementary Figure S9a–c). Taken together, these tests demonstrate that apparent differences in IGD between tropical and nontropical species are not artifacts of uneven sampling but represent a real pattern. Given the importance of IGD to biodiversity conservation (Frankham, 2005; Hedrick & Kalinowski, 2000; Pereira et al., 2012), identifying the causes of the latitudinal gradient in IGD is of paramount concern.

Decades of phylogeographic research have argued that Quaternary glaciation had a profound effect on IGD (Avise, 2000; Hewitt, 2004; Hickerson et al., 2010; Theodoridis et al., 2020) because glaciation reduced the available habitat and shifted species ranges toward the equator such that species were forced to persist in small glacial refugia, for example those that occurred in the southern peninsulas in Europe (Hewitt, 2004) or on the Gulf of Mexico coast in North America (Avise, 2000). This combination of glacial advance and range compression would decrease the population size of affected species and promote a loss of allelic diversity due to genetic drift, leading to an overall reduction in intraspecific genetic diversity for species in temperate regions (Hewitt, 2004). In contrast, although some tropical species have shown clear indications of demographic change in response to climatic oscillations during the Quaternary (Gehara et al., 2017; Smith et al., 2014), the lack of glaciers in tropical regions is expected to result in a less severe demographic change in those species compared to temperate species. Moreover, while species may have responded to the retraction of glaciers by gradually increasing their population sizes, in most cases too few generations have passed since

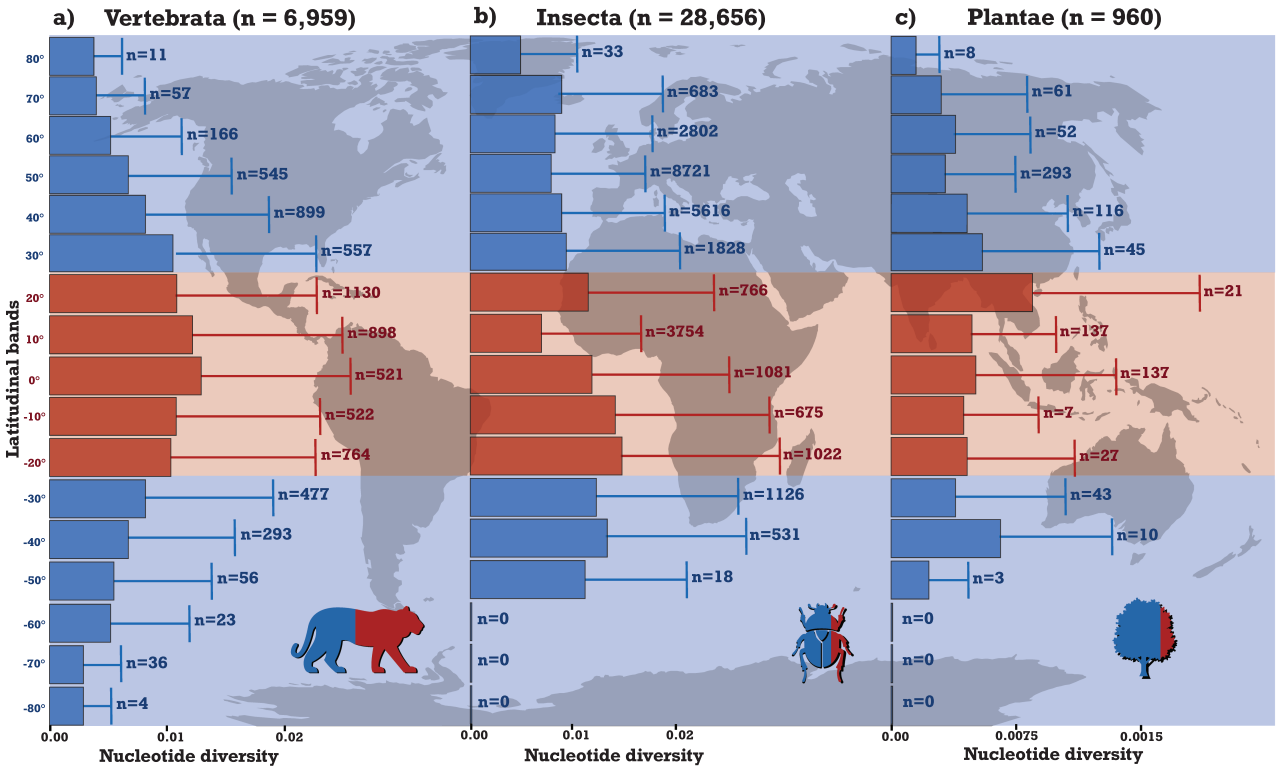

Figure 1. Mean nucleotide diversity across latitudinal bands for (A) vertebrates, (B) insects, and (C) plants. Nucleotide diversity was averaged across species within latitudinal bands of 10°. Error bars represent the confidence intervals of nucleotide diversity across species. Red and blue bars represent tropical and out of the Tropics species, respectively. Numbers close to the bars represent the number of species used to calculate nucleotide diversity on that band. The amount of red and blue on species' silhouettes represent the proportion of tropical and out of the Tropics species on the data set.

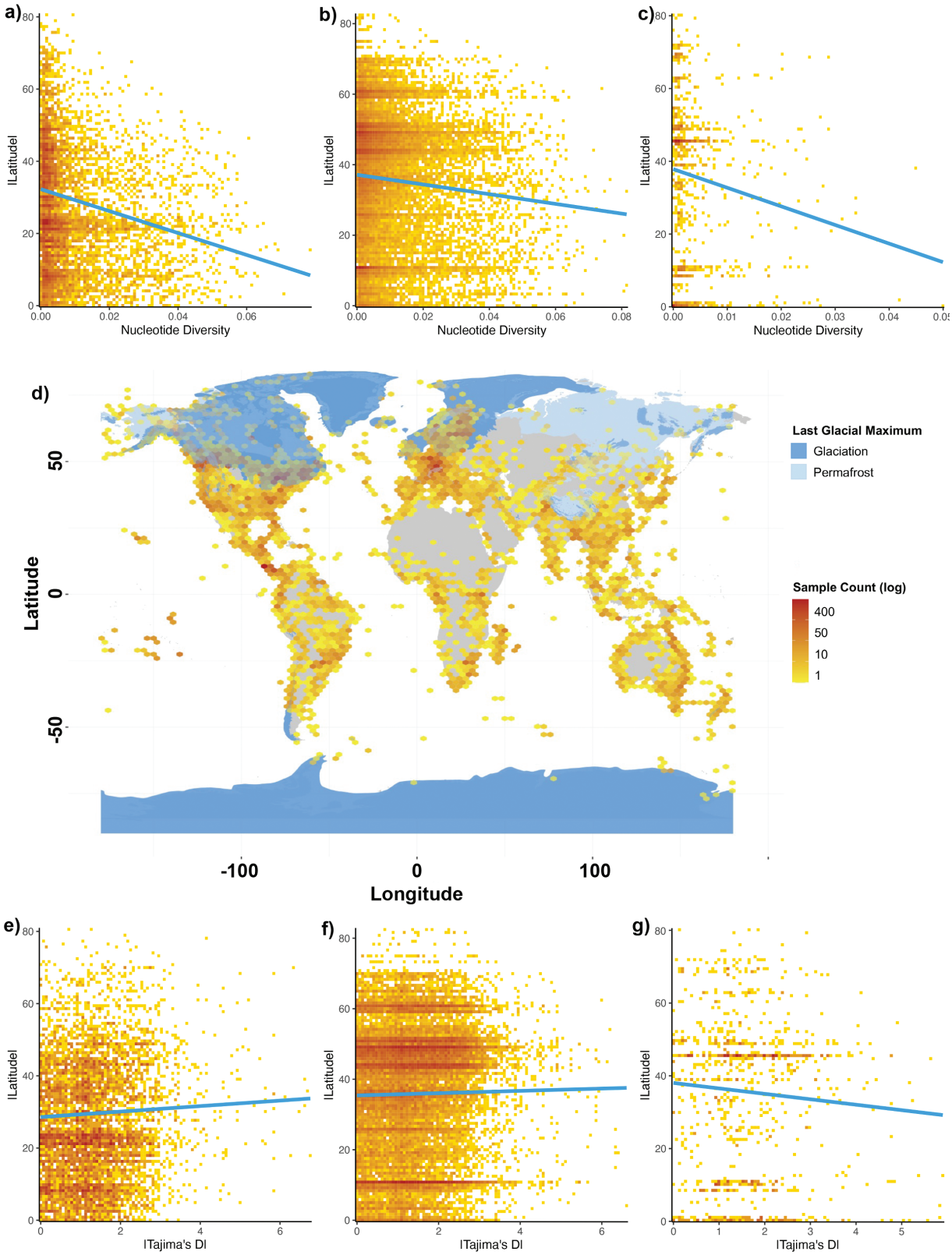

Figure 2. The role of latitude on population demography and intraspecific genetic diversity. The top row shows latitude (absolute) regressed against nontransformed nucleotide diversity (π) for vertebrates (A), insects (B), and plants (C). In the center row, a map showing sampling across the globe and the extension of glaciers during the Last Glacial Maximum (~21 kyr) and current permafrost (D) In the lower row, latitude (absolute) is regressed against Tajima's D (absolute) for vertebrates (E), insects (F), and plants (G). Top and bottom rows depict the density of the raw data as a heatmap, ranging from blue (lower density) to red (higher density).

the last glacial maximum for genetic diversity to return to equilibrium levels (Davis et al., 2018).

Our results support the hypothesis that Quaternary glaciations are an important driver of global patterns of intraspecific genetic diversity. While phylogeographic investigations have used many methodological approaches to evaluate this model, most include the calculation of summary statistics that assess whether the proportion of rare alleles present in the population matches theoretical expectations under neutral models of constant population size and no adaptive change. Here we use two common statistics, Tajima's *D* (Tajima, 1989) and R_2 (Ramos-Onsins & Rozas, 2002) to test this hypothesis. The former is widely used in phylogeographic investigations, with significant results usually attributed to population demographic change. While Tajima's *D* has also been used to detect natural selection on coding sequences, in our data most of the mutations (~66%) are associated with third codon positions which do not change the amino acid sequences of the resultant protein (i.e., silent substitutions), and as such are not exposed to natural selection. Given this high proportion of silent substitutions, population demographic change is a more likely explanation than natural selection as the predominant process responsible for patterns identified in Tajima's *D*. The R_2 statistic is typically used to test the hypothesis that species population sizes have recently increased, with significant values indicative of an excess of singleton alleles consistent with demographic expansion of the type that would occur as species expanded from glacial refugia. Since the Quaternary glaciation hypothesis predicts both a bottleneck during the glacial period and a subsequent expansion after the glacial period, significant values of either statistic are likely associated with demographic response predicted by this hypothesis.

Proportionally more nontropical species had significant Tajima's *D* and R_2 scores than tropical species (Table 1). Using a Fisher's exact test ($p < .05$ after Bonferroni correction), we could reject the null hypothesis that the proportion of significant tests is the same in the tropical and nontropical regions using Tajima's *D* for vertebrates, insects, plants, and arachnids. Similarly, for R_2 in insects, plants, and arachnids, our results showed a higher proportion of significant tests in nontropical regions, but not in vertebrates. Additional evidence for the effect of latitude on Tajima's *D* is provided by linear regression, where latitude is predictive of the magnitude of this statistic in vertebrates ($p < .01$; Figure 2E), insects ($p < .01$; Figure 2F), but not plants ($p = .106$; Figure 2G). The randomization procedure developed here also demonstrates that the results for both statistics are different in tropical and nontropical regions for all groups ($p < .05$), except for Tajima's *D* (both geographic regions) and R_2 (tropical regions) in plants ($p > .05$; Supplementary Figure S9d–i). Although we did not find a higher magnitude of demographic changes in nontropical regions for vertebrates using R_2, the proportion of significant versus nonsignificant values are not statistically different as observed using Fisher's exact test (Table 1). Taken together, these results indicate that nontropical species are more likely to deviate from the patterns of allelic diversity expected under a neutral model, a clear indication of historical population demographic change that likely represents a response to Quaternary glaciation. While this pattern in IGD represents a common finding in single-species phylogeographic studies, as reviewed by Avise (2000), our results demonstrate for the first time that this is a global phenomenon that is partially explained by latitude.

Single-locus data from mitochondrial have been widely used in phylogeographic investigations (Avise, 2000) because they are easy to sequence and contain many variable sites. While levels of variation are lower, similar data from chloroplast genes have been used in plants. While these data have recognized shortcomings (Edwards & Beerli, 2000; Felsenstein, 2006), they are suited to global surveys such as this work because the number of species with single-locus data available in public databases is orders of magnitude greater than species with large genomic data sets. Our inferences rely on establishing that there are real differences in the IGD between tropical and nontropical species and exploring causal factors and are not derived from estimates of population parameters such as divergence time that are expected to be inaccurate using single-locus data (Edwards & Beerli, 2000). Furthermore, it seems unlikely that the correlations identified here between latitude and either π or Tajima's *D* are either a function of the data being from a single locus or from an organellar genome. While some fraction of our significant results in Tajima's *D* may be attributable in part to natural selection, and while in some cases natural selection may be correlated with latitude (e.g., Zhang et al., 2019), our results demonstrate that most of the variable sites in the data analyzed here are associated with silent mutations that likely do not change protein structure. As a result, it seems unlikely that minor latitudinal differences in natural selection could explain the pattern observed here. Population genetic structure can also bias estimates of Tajima's *D*, but since a recent survey demonstrated that latitude is an important predictor of population genetic structure with species at mid-latitudes more likely to exhibit isolation by distance (Pelletier & Carstens, 2018), the patterns identified here would likely be more pronounced if we could remove any effects of population genetic structure on estimates of Tajima's *D*.

Our analyses demonstrate that tropical species contain more IGD than nontropical species in plants and animals. Nucleotide diversity is correlated with latitude (Figure 2A–C)

Table 1. Proportion of population demography. Species with significant ($p < .05$) and nonsignificant (NS) values of Tajima's *D* and R_2 statistics between tropical and nontropical regions for Vertebrata, Insecta, Arachnida, and Plantae.

Clade	Geographic region	Tajima's *D*		R_2	
		$p < .05$	NS	$p < .05$	NS
Vertebrata	Tropical	706 (10.1%)	2,792 (40.1%)	487 (7.0%)	3,011 (43.3%)
	Nontropical	851 (12.2%)	2,610 (37.5%)	497 (7.1%)	2,964 (42.6%)
Insecta	Tropical	1,687 (5.9%)	5,385 (18.8%)	1,044 (3.6%)	6,028 (21.0%)
	Nontropical	6,152 (21.5%)	15,432 (53.9%)	4,525 (15.8%)	17,059 (59.5%)
Arachnida	Tropical	95 (9.9%)	219 (22.8%)	10 (1.0%)	304 (31.7%)
	Nontropical	216 (22.5%)	430 (44.8%)	28 (2.9%)	618 (64.4%)
Plantae	Tropical	22 (1.4%)	79 (5.1%)	16 (1.0%)	85 (5.5%)
	Nontropical	377 (24.2%)	1,081 (69.3%)	370 (23.7%)	1,088 (69.8%)

and results from population demographic changes associated with Quaternary glaciation were less impactful in tropical species. This finding has implications to the ongoing discussion regarding the factors that produce the latitudinal biodiversity gradient (Gaston, 2000; Willig et al., 2003). The latitudinal gradient in intraspecific genetic diversity is consistent with hypotheses that attribute the latitudinal gradient in species richness to historical climatic effects. For example, the climate stability hypothesis predicts that tropical areas contain more species because the stable climate in these regions (over evolutionary time) enables species to adapt into highly partitioned ecological niches (Klopfer, 1959). By forcing species in temperate regions to persist in glacial refugia (Avise, 2000; Hewitt, 2004; Provan & Bennett, 2008), the Quaternary climatic instability had a clear impact on genetic diversity. In extant species, a reduction in range size and genetic diversity put species at risk for extinction (Evans & Sheldon, 2008) and the deficit of genetic diversity limits adaptive potential (Hoffmann & Sgrò, 2011). Given that extinction risk is inversely correlated with IGD (Evans & Sheldon, 2008; Frankham, 2005), the loss of genetic diversity in temperate regions during the Quaternary likely limited the potential for adaptation in these species, potentially leading to higher rates of extinction and/or lower rates of speciation in nontropical species.

The recent emergence of macrogenetics (Leigh et al., 2021) has provided insight into the factors that influence genetic diversity on a global scale (Carstens et al., 2018; Manel et al., 2020; Miraldo et al., 2016). Here, by repurposing data (Sidlauskas et al., 2010) that were originally collected for other uses, we document that a latitudinal gradient in IGD is present and provide evidence that this gradient is caused by Quaternary climate oscillations. Our work demonstrates the potential for macrogenetic analysis to address fundamental questions in ecology and evolutionary biology and supports recent calls to develop databases and protocols that enable additional integration of biodiversity data (Anderson et al., 2020; Heberling et al., 2021; Marden et al., 2021).

Supplementary material

Supplementary material is available online at *Evolution Letters*.

Data availability

All analyses were conducted in R v. 4.0.2 (R Core Team, 2010), and the scripts for data processing and analysis are available on Github https://github.com/emanuelmfonseca/Global_genetic_diversity.

Author contributions

B.C.C., E.M.F., and T.A.P. conceptualize the study. B.C.C., E.M.F., T.A.P., D.J.P., and S.K.D. designed the methodology. E.M.F. conducted the analyses. All authors interpreted the results and participated in the writing of the manuscript and gave final approval for submission.

Conflict of interest: The authors declare no conflict of interest.

Acknowledgments

We thank Eric Frantz, Jeffrey Ohrstrom, and the Ohio Supercomputer Center for their contributions to *phylogatR*. We thank members of the Carstens lab and *phylogatR* beta testers for their helpful comments on this work. Support for this work was provided by the National Science Foundation (NSF) (DBI-1661029 and DBI-1910623 to B.C.C. and DBI-1911293 to T.A.P.) and the Ohio Supercomputing Center (PAA1174). E.M.F. is thankful to the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for his doctoral fellowship (process #88881.170016/2018).

References

Anderson, R. P., Araújo, M. B., Guisan, A., Lobo, J. M., Martínez-Meyer, E., Peterson, A. T., & Soberón, J. M. (2020). Optimizing biodiversity informatics to improve information flow, data quality, and utility for science and society. *Frontiers of Biogeography*, **12**(3), e47839.

Avise, J. C. (2000). *Phylogeography: The history and formation of species*. Harvard University Press.

Barrow, L. N., Fonseca, E. M., Thompson, C. E. P., & Carstens, B. C. (2021). Predicting amphibian intraspecific diversity with machine learning: Challenges and prospects for integrating traits, geography, and genetic data. *Molecular Ecology Resources*, **21**, 2818–2831.

Beheregaray, L. B. (2008). Twenty years of phylogeography: The state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, **17**(17), 3754–3774. https://doi.org/10.1111/j.1365-294X.2008.03857.x

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Carstens, B. C., Morales, A. E., Field, K., & Pelletier, T. A. (2018). A global analysis of bats using automated comparative phylogeography uncovers a surprising impact of Pleistocene glaciation. *Journal of Biogeography*, **45**(8), 1795–1805. https://doi.org/10.1111/jbi.13382

Davis, M., Faurby, S., & Svenning, J. -C. (2018). Mammal diversity will take millions of years to recover from the current biodiversity crisis. *Proceedings of the National Academy of Sciences of the United States of America*, **115**(44), 11262–11267. https://doi.org/10.1073/pnas.1804906115

Edwards, S., & Beerli, P. (2000). Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, **54**(6), 1839–1854. https://doi.org/10.1111/j.0014-3820.2000.tb01231.x

Evans, S. R., & Sheldon, B. C. (2008). Interspecific patterns of genetic diversity in birds: Correlations with extinction risk. *Conservation Biology*, **22**(4), 1016–1025. https://doi.org/10.1111/j.1523-1739.2008.00972.x

Felsenstein, J. (2006). Accuracy of coalescent likelihood estimates: Do we need more sites, more sequences, or more loci? *Molecular Biology and Evolution*, **23**(3), 691–700. https://doi.org/10.1093/molbev/msj079

Frankham, R. (2005). Genetics and extinction. *Biological Conservation*, **126**(2), 131–140. https://doi.org/10.1016/j.biocon.2005.05.002

French, C. M., Bertola, L. D., Carnaval, A. C., Economo, E. P., Kass, J. M., Lohman, D. J., Marske, K. A., Meier, R., Overcast, I., Romiger, A. J., Staniczenko, P., & Hickerson, M. J. (2022). Global determinants of insect mitochondrial genetic diversity. *bioRxiv*, 2022-02. https://doi.org/10.1101/2022.02.09.479762

Gaston, K. J. (2000). Global patterns in biodiversity. *Nature*, **405**(6783), 220–227. https://doi.org/10.1038/35012228

Gehara, M., Garda, A. A., Werneck, F. P., Oliveira, E. F., Fonseca, E. M., Camurugi, F., Magalhães, F. D. M., Lanna, F. M., Sites, J. W., Marques, R., Silveira-Filho, R., São Pedro, V. A., Colli, G. R., Costa, G. C., & Burbrink, F. T. (2017). Estimating synchronous

demographic changes across populations using hABC and its application for a herpetological community from northeastern Brazil. *Molecular Ecology*, **26**(18), 4756–4771. https://doi.org/10.1111/mec.14239

Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B., & Schigel, D. (2021). Data integration enables global biodiversity synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, **118**(6), e2018093118. https://doi.org/10.1073/pnas.2018093118

Hebert, P. D. N., & Gregory, T. R. (2005). The promise of DNA barcoding for taxonomy. *Systematic Biology*, **54**(5), 852–859. https://doi.org/10.1080/10635150500354886

Hedrick, P. W., & Kalinowski, S. T. (2000). Inbreeding depression in conservation biology. *Annual Review of Ecology and Systematics*, **31**(1), 139–162. https://doi.org/10.1146/annurev.ecolsys.31.1.139

Hewitt, G. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 183–195.

Hickerson, M. J., Carstens, B. C., Cavender-Bares, J., Crandall, K. A., Graham, C. H., Johnson, J. B., Rissler, L., Victoriano, P. F., & Yoder, A. D. (2010). Phylogeography's past, present, and future: 10 Years after Avise, 2000. *Molecular Phylogenetics and Evolution*, **54**(1), 291–301. https://doi.org/10.1016/j.ympev.2009.09.016

Hoffmann, A. A., & Sgrò, C. M. (2011). Climate change and evolutionary adaptation. *Nature*, **470**(7335), 479–485. https://doi.org/10.1038/nature09670

Holderegger, R., & Wagner, H. H. (2008). Landscape genetics. *Bioscience*, **58**(3), 199–207. https://doi.org/10.1641/b580306

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, **30**(4), 772–780. https://doi.org/10.1093/molbev/mst010

Klopfer, P. H. (1959). Environmental determinants of faunal diversity. *American Naturalist*, **93**(873), 337–342. https://doi.org/10.1086/282092

Leigh, D. M., van Rees, C. B., Millette, K. L., Breed, M. F., Schmidt, C., Bertola, L. D., Hand, B. K., Hunter, M. E., Jensen, E. L., Kershaw, F., Liggins, L., Luikart, G., Manel, S., Mergeay, J., Miller, J. M., Segelbacher, G., Hoban, S., & Paz-Vinas, I. (2021). Opportunities and challenges of macrogenetic studies. *Nature Reviews Genetics*, **22**(12), 791–807. https://doi.org/10.1038/s41576-021-00394-0

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.

Marden, E., Abbott, R. J., Austerlitz, F., Ortiz-Barrientos, D., Baucom, R. S., Bongaerts, P., Bonin, A., Bonneaud, C., Browne, L., Alex Buerkle, C., Caicedo, A. L., Coltman, D. W., Cruzan, M. B., Davison, A., DeWoody, J. A., Dumbrell, A. J., Emerson, B. C., Fountain-Jones, N. M., Gillespie, R., … Rieseberg, L. H. (2021). Sharing and reporting benefits from biodiversity research. *Molecular Ecology*, **30**(5), 1103–1107. https://doi.org/10.1111/mec.15702

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. https://doi.org/10.1126/science.aaf4381

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, **76**(10), 5269–5273. https://doi.org/10.1073/pnas.76.10.5269

Nogués-Bravo, D., Ohlemüller, R., Batra, P., & Araújo, M. B. (2010). Climate predictors of Late Quartenary extinctions. *Evolution*, **64**(8), 2442–2449. https://doi.org/10.1111/j.1558-5646.2010.01009.x

Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, **26**(3), 419–420. https://doi.org/10.1093/bioinformatics/btp696

Pelletier, T. A., & Carstens, B. C. (2018). Geographical range size and latitude predict population genetic structure in a global survey. *Biology Letters*, **14**(1), 20170566. https://doi.org/10.1098/rsbl.2017.0566

Pelletier, T. A., Parsons, D. J., Decker, S. K., Crouch, S., Franz, E., Ohrstrom, J., & Carstens, B. C. (2021). Phylogeographic data aggregation and repurposing. *Molecular Ecology Resources*, **22**, 2830–2842.

Pereira, H. M., Navarro, L. M., & Martins, I. S. (2012). Global biodiversity change: The bad, the good, and the unknown. *Annual Review of Environment and Resources*, **37**(1), 25–50. https://doi.org/10.1146/annurev-environ-042911-093511

Provan, J., & Bennett, K. (2008). Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology and Evolution*, **23**(10), 564–571. https://doi.org/10.1016/j.tree.2008.06.010

Ramos-Onsins, S. E., & Rozas, J. (2002). Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**(12), 2092–2100. https://doi.org/10.1093/oxfordjournals.molbev.a004034

R Core Team. (2010). *R: A language and environment for statistical computing*.

Sidlauskas, B., Ganapathy, G., Hazkani-Covo, E., Jenkins, K. P., Lapp, H., McCall, L. W., Price, S., Scherle, R., Spaeth, P. A., & Kidd, D. M. (2010). linking big: The continuing promise of evolutionary synthesis. *Evolution*, **64**(4), 871–880. https://doi.org/10.1111/j.1558-5646.2009.00892.x

Smith, B. T., McCormack, J. E., Cuervo, A. M., Hickerson, M. J., Aleixo, A., Cadena, C. D., Pérez-Emán, J., Burney, C. W., Xie, X., Harvey, M. G., Faircloth, B. C., Glenn, T. C., Derryberry, E. P., Prejean, J., Fields, S., & Brumfield, R. T. (2014). The drivers of tropical speciation. *Nature*, **515**(7527), 406–409. https://doi.org/10.1038/nature13687

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**(3), 585–595. https://doi.org/10.1093/genetics/123.3.585

Theodoridis, S., Fordham, D. A., Brown, S. C., Li, S., Rahbek, C., & Nogues-Bravo, D. (2020). Evolutionary history and past climate change shape the distribution of genetic diversity in terrestrial mammals. *Nature Communications*, **11**(1), 2557. https://doi.org/10.1038/s41467-020-16449-5

Turchetto-Zolet, A. C., Pinheiro, F., Salgueiro, F., & Palma-Silva, C. (2013). Phylogeographical patterns shed light on evolutionary process in South America. *Molecular Ecology*, **22**(5), 1193–1213. https://doi.org/10.1111/mec.12164

Willig, M. R., Kaufman, D. M., & Stevens, R. D. (2003). Latitudinal gradients of biodiversity: Pattern, process, scale, and synthesis. *Annual Review of Ecology, Evolution, and Systematics*, **34**(1), 273–309. https://doi.org/10.1146/annurev.ecolsys.34.012103.144032

Zhang, M., Suren, H., & Holliday, J. A. (2019). Phenotypic and genomic local adaptation across latitude and altitude in *Populus trichocarpa*. *Genome Biology and Evolution*, **11**(8), 2256–2272. https://doi.org/10.1093/gbe/evz151