



Bats of a feather: Range characteristics and wing morphology predict phylogeographic breaks in volant vertebrates

Sydney K. Decker¹, Kaiya L. Provost², Bryan C. Carstens¹

¹ Museum of Biological Diversity and Department of Evolution, Ecology and Organismal Biology, The Ohio State University, Columbus, OH 43210, USA

² Department of Biology, Adelphi University, Garden City, NY 11530, USA

Corresponding author: Sydney K. Decker (decker.391@osu.edu)

Editors EditorName

Received 23 October 2024 ♦ Accepted 25 March 2025 ♦ Published 15 April 2025

Abstract

Intraspecific genetic variation and phylogeographic structure can be influenced by factors such as landscape features, environmental gradients, historical biogeography and organismal traits such as dispersal ability. Since deep genetic structure is often considered a precursor to speciation, identifying the factors that are associated with genetic structure can contribute to a greater understanding about diversification. Here, we use repurposed data to perform a global analysis of volant vertebrates (i.e. bats and birds) to estimate where intraspecific phylogeographic breaks occur and to identify the factors that are important predictors of these breaks. We estimate phylogeographic breaks using Monmonier's maximum difference barrier algorithm and conduct a random forests analysis using the presence of a phylogeographic break as a response variable. In bats, phylogeographic breaks are concentrated in biodiversity hotspots while breaks estimated in bird species are more widespread across temperate and tropical zones. However, for both clades, geographical features such as maximum latitude, measures of wing morphology and organismal traits associated with feeding ecology were found to be important predictors of phylogeographic breaks. Our analysis identifies geographical areas and suites of organismal traits that could serve as a starting point for more detailed studies of biodiversity processes.

Highlights

- Phylogeographic breaks, particularly in species with high dispersal ability, are of interest to evolutionary biologists as a potential precursor to speciation.
- Repurposed data can be used to identify large-scale trends in biodiversity.

- Monmonier's maximum differences algorithm can be employed to estimate phylogeographic breaks using georeferenced sequence data.
- For volant vertebrates, phylogeographic breaks are predicted by aspects of geographic range, wing morphology and feeding ecology.
- In bats, breaks are concentrated in tropical biodiversity hotspots while phylogeographic breaks in birds are more evenly distributed across temperate zones.

Keywords

bats, birds, dispersal, macrogenetics, Monmonier's algorithm, phylogeography, Random Forests

Introduction

Intraspecific genetic variation is a featured aspect of various models of speciation (Lande 1980; Orr 1996; Nosil 2012). One central tenet of evolutionary theory is that natural selection acts on variation within species to drive diversification (e.g. Grant (1981)). For this reason, entire disciplines have been devoted to characterising intraspecific genetic variation and identifying the landscape features that may influence this variation (e.g. Avise (2000); Manel and Holderegger (2013)). While synthesis has been impeded by confounding factors such as life history traits, environmental variation and biogeographic history, such work has resulted in the accumulation of millions of sequences in databases such as NCBI GenBank (e.g. Garrick et al. (2015)). Given the availability of biodiversity data in natural history collections (e.g. iDigBio, idigbio.org), the aggregation of specimen collection records in data portals (e.g. GBIF, The Global Biodiversity Information Facility (2024)) and collections of

phenotypic data (e.g. MorphoBank, O’Leary and Kaufman (2011)), it now appears to be an attainable goal to understand the determinants of intraspecific genetic diversity across broad taxonomic groups (Barrow et al. 2021).

One key aspect of life history that is likely to influence the development of intraspecific genetic variation is dispersal ability, particularly the intrinsic and extrinsic factors that limit the ability of individuals of a species to move throughout the landscape. A lack of dispersal can lead to the accumulation of genetic differentiation due to genetic drift and other forces (Slatkin 1987) and, as such, dispersal limitation has been identified as a key factor in biological processes ranging from speciation rates (e.g. Polato et al. (2018)) to community assembly (e.g. Pigot and Tobias (2015)). Research into empirical systems has indicated that extrinsic landscape features such as mountains (e.g. Kane et al. (2019)), rivers (e.g. Myers et al. (2020)) and environmental gradients (e.g. Sebastián et al. (2021)) can limit dispersal and lead to the formation of intraspecific genetic variation.

Long-term limitation of gene flow between populations, with or without a physical barrier to gene flow (e.g. Irwin (2002)), can result in phylogeographic structure. Phylogeographic structure is often characterised using summary statistics that describe genetic variation in populations (e.g. F_{ST}), via the identification of population genetic clusters (e.g. PCA) or by testing for spatial autocorrelation (e.g. isolation by distance). Considerably less effort has been devoted to developing theory and methodology for estimating where barriers to gene flow occur. One notable exception is Monmonier’s algorithm (Monmonier 1973), which can be applied to infer spatial genetic structure by converting the genetic data into a network, identifying the edges of the network that are associated with the highest degree of genetic differentiation and projecting the maximum difference boundary on to a spatial connection network (e.g. Manni et al. (2004)). The linear boundary produced by Monmonier’s algorithm is placed on a map without any input from external information such as geospatial or environmental layers, making the algorithm well-suited for macrogenetic investigations which seek to identify the factors that influence genetic structure, but may rely on data that are unevenly sampled across species. Since geographic distance is typically correlated with both environmental distance and landscape features (e.g. Wang (2013); Pelletier and Carstens (2018)), any inference method that incorporates geographic distance amongst samples will likely be influenced by these same factors. By circumventing this problem, Monmonier’s algorithm enables environmental distance and landscape features to be evaluated as potential factors which may contribute to the formation of phylogeographic breaks.

Due to the perception that volant species have the capacity for long distance dispersal, phylogeographic breaks in these species may be more likely to correspond to environmental conditions, differences in food resources and overall habitat suitability than to geophysical barriers. In bats, patterns of phylogeographic structure vary across species on a global scale, but are thought to be

influenced by factors including social structure, mating behaviour (e.g. autumnal swarming), migration patterns, habitat connectivity, ecological gradients and geographic barriers (Moussy et al. 2013). In a recent review, Hernández-Canchola et al. (2021) noted that Quaternary climate fluctuations, geographic features such as oceans and rivers and ecological processes such as niche differentiation were the most referenced contributors to phylogeographic structure in bat species. In birds, feeding ecology and annual dispersal were the strongest predictors of high diversification rates (Phillimore et al. 2006), suggesting that species ecology plays a large role in diversification and speciation. Morphological proxies of dispersal show a nuanced response with respect to diversification (Claramunt et al. 2011; Tobias et al. 2020). Social behaviour, particularly song, is also thought to be important at maintaining diversity across the landscape (Uy et al. 2018).

To assess global phylogeographic patterns in volant vertebrates and to identify factors that are important in predicting phylogeographic breaks, we repurposed georeferenced mitochondrial sequence data available from *phylogatR* (Pelletier et al. 2022). This database that aggregates geographic information from specimen repositories such as GBIF (The Global Biodiversity Information Facility 2024) and sequence data from GenBank (Benson et al. 1993) and BOLD (Ratnasingham and Hebert 2007) stores these data in a convenient framework. While challenges related to data acquisition, quality control and unevenness can be substantial and should be accounted for (Leigh et al. 2021; Pelletier et al. 2022), we apply an automated analysis pipeline modified from previous investigations (Pelletier and Carstens 2018; Barrow et al. 2021; Parsons et al. 2022) to estimate phylogeographic breaks in hundreds of bird and bat species. Bats and birds were chosen for the study as they are both volant vertebrates with high ecological diversity and are generally considered to have high dispersal ability. Additionally, birds tend to be well studied in comparison to bats and, as such, we hoped that the additional avian data would corroborate models of intraspecific genetic structure in bats. We then use machine learning to identify traits that are predictive of phylogeographic breaks in these taxa.

Methods

Data processing

Data for global Chiroptera and Aves species were downloaded via *phylogatR* (<https://phylogatr.org>; Pelletier et al. 2022). Custom R scripts (available on <https://github.com/skdecker/PhylogeographicBreaks>) were developed to further filter occurrence and sequence data with the goal of removing data that were likely to be inaccurate. To accomplish this, we removed geographic data that fell outside species range maps. Coordinates were cleaned, based on IUCN range maps for bats (IUCN 2022) and BirdLife range maps for birds (BirdLife International 2021) with a

1° (approximately 110 km) buffer to account for animal movement and range map inaccuracies. To detect poor quality genetic data, sequence data were processed to remove sequences with greater than 20% missing sites from alignments, as such data were likely to lead to failures in downstream analyses. We also removed sequences that exhibited high genetic distance from other sequences in the alignment (> 5% in birds and > 10% in bats) because data such as these might be indicative of cryptic species (e.g. Parsons et al. (2022)) or misidentified individuals. These thresholds are based on the distribution of average intraspecific genetic distances in our data (Suppl. material 1: fig. S1) and historical quantifications of mitochondrial genetic distances typical of biological species (Baker and Bradley 2006). Due to the historical use of mtDNA in phylogeographic studies and for ease of comparison across species, we relied exclusively on mtDNA for these analyses. Since over 86% of the sequence alignments downloaded from *phylogatR* for birds and bats were from mitochondrial genes, this also allowed us to include the largest possible sample size. Species with 15 or more sampled individuals in the alignment and at least three unique localities after processing were included in downstream analyses. In species with data for multiple genes, the gene with the highest number of sequences was used.

Estimating phylogeographic breaks

To identify species traits that are associated with the formation of genetic structure, we first need to identify species that exhibit spatial population genetic structure. A quantitative approach was preferred since this could be applied across hundreds of species and would not be biased by the interpretations of individual researchers, as would be the case if we conducted a synthetic literature review. We found it expedient to apply a method that does not require *a priori* division of samples, such as calculating F_{ST} values. We did not use a test of genetic isolation by distance (Wright 1943) due to the anticipated correlation between geographic and environmental distance (e.g. Dillon (1984); Lee and Mitchell-Olds (2011)) and the association of genetic isolation by distance with other factors such as latitude (Pelletier and Carstens 2018). Rather, Monmonier's algorithm (Monmonier 1973) was applied to estimate phylogeographic breaks.

Monmonier's algorithm was implemented through the `optimize.monmonier` function in the R package *adegenet* v.2.1.4 (Jombart 2008; Jombart and Ahmed 2011). Since Monmonier's algorithm is incompatible with duplicate localities, a small amount of noise was added to duplicate coordinates with the R function `jitter` (base R, v.4.1.1; R Core Team 2021). The input data for the function is a connection network created from jittered sample coordinates and a matrix of genetic distances (Box 1). For our analyses, we used Euclidean distances for the genetic data and geographic connection networks built using Delaunay triangulation via the `chooseCN` function in *adegenet*. Unlike

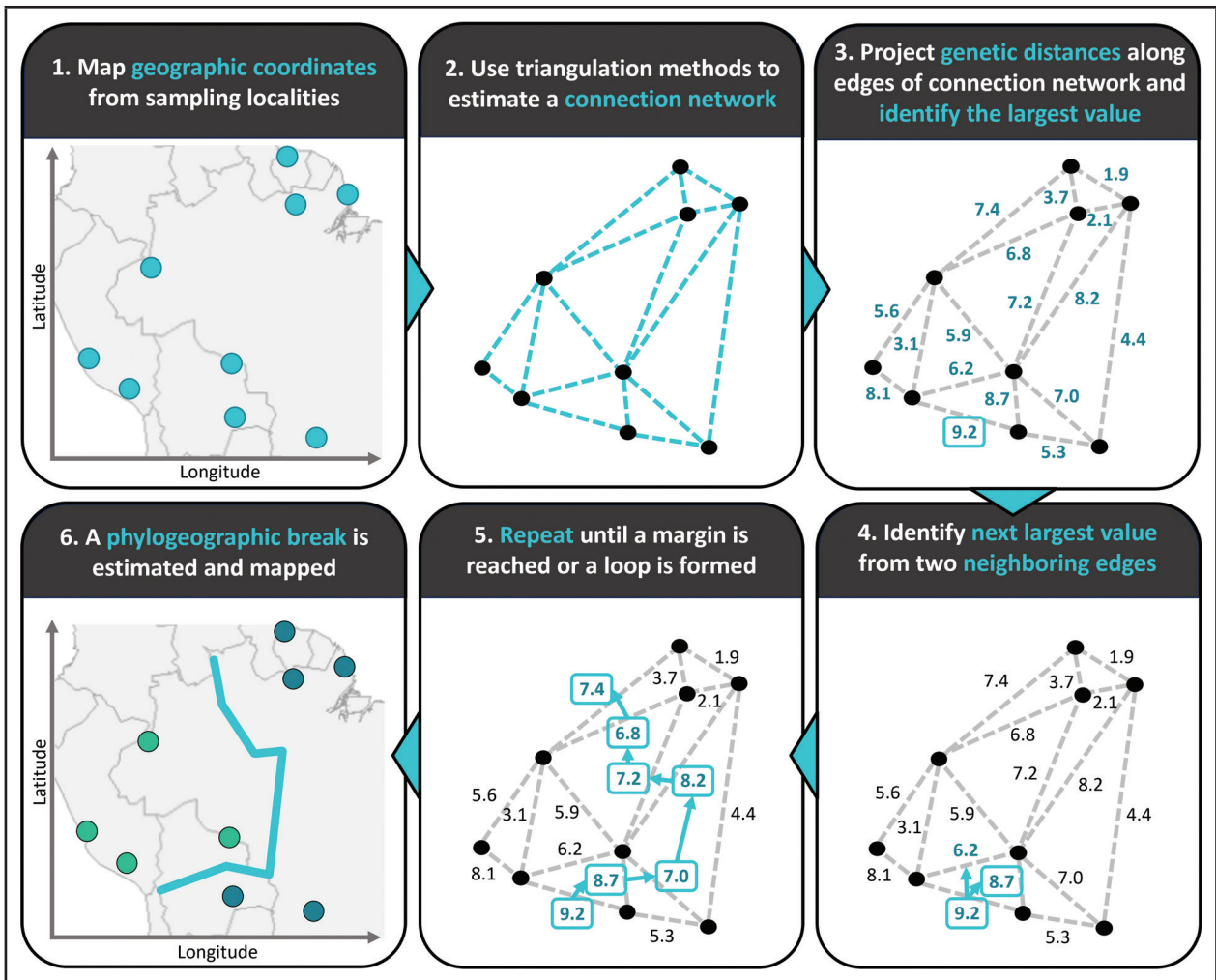
the standard monmonier function, `optimize.monmonier` tries several starting points within the connection network and returns the best boundary, avoiding single strong local differences (e.g. Dupanloup et al. (2002)). The number of tries used in the algorithm was proportional to the number of individuals for each species. A maximum of 10 tries was used for species with greater than 100 sequences for computational efficiency. Two possible directions of breaks were assessed, allowing for the estimation of multiple breaks per species.

The results of the function are output as a list of coordinates comprising discovered boundaries, where boundaries are drawn perpendicular to edges in the connection network (see Box 1). These returned coordinates were plotted as lines on a map with the sampling localities and the jittered sampling coordinates for each species individually to check that estimated phylogeographic breaks clearly divided the sampling range, isolating individuals/populations from one another. We considered a species to not exhibit a phylogeographic break if only a single set of coordinates were returned (i.e. no dividing line is drawn between multiple specimens). In cases where the presence of a phylogeographic break was ambiguous, the local distances plot from the Monmonier's algorithm run was evaluated: a sharp decrease in the plot suggests a break is present while a steady cline suggests that no break is present. This assignment was used as a binary response variable in the following random forests classifier.

Traits

Once genetic data were acquired and phylogeographic breaks were estimated, an organismal trait dataset (Suppl. material 2: Data Sheets 1, 2) was compiled from published, open-source datasets. For bat traits, we used PanTHERIA (Jones et al. 2009), COMBINE (Soria et al. 2021) and Crane et al. (2022). Some missing values for wing morphology metrics were supplemented by a literature search (Suppl. material 2: Data Sheet 1). AVONET (Tobias et al. 2022) was used for bird traits, following the BirdLife taxonomic system. The datasets consisted of geographic (e.g. maximum latitude), taxonomic (family/order), morphological (e.g. mass), ecological (e.g. trophic level) and environmental (e.g. habitat) traits for each species. As random forests cannot be trained with missing data, traits with more than 40% missing data across species or species with more than 40% missing trait values were removed from the analysis and missing values for remaining traits were imputed using the R package *mice* v.3.15.0 (van Buuren and Groothuis-Oudshoorn 2011). For bats, the taxonomic family was used as a phylogenetic predictor in the imputation. Penone et al. (2014) found little bias in methods such as *mice* for imputing missing trait variables, especially when taxonomic information is used as a predictor. Imputation was conducted over 30 iterations. For each variable with imputed data, the imputed values were checked for bias against observed data with

Box 1. Phylogeographic break estimation using Monmonier's maximum difference algorithm.



the densityplot function (Sarkar 2008; Suppl. material 1: fig. S2). Missing data were only imputed for the bat trait dataset as AVONET reports inferred values from the closest relative in cases of missing values (Tobias et al. 2022). An additional trait describing if a species is predicted to harbour cryptic diversity was used for bats based on Parsons et al. (2022) analysis of cryptic diversity in mammals. Linear morphological traits that are highly correlated with overall body size (e.g. wingspan, beak length) were body-size corrected by regressing the trait with body length (bats) and tarsus length (birds; Rising and Somers (1989); Senar and Pascual (1997)) and retaining the residual. Ratio measurements such as wing loading or hand-wing index were not corrected as these measures are effectively already correcting for wing size (Sheard et al. 2020).

To test for statistical difference in quantitative traits between species with a phylogeographic break and those without, we conducted Mann-Whitney U tests using the `wilcox.test` function in the `stats` R package v.4.1.1 (R Core Team 2021; Suppl. material 1: tables S1, S2). This non-parametric test was used as it is more robust to data with skewed distributions, as is observed in most of the traits in our dataset (Suppl. material 1: figs S7, S8).

Random forests classifier

To identify traits important in predicting whether a species will exhibit a phylogeographic break, we built a random forests (RF) classifier with the R package `caret` v. 6.0-88 (Kuhn 2021) to identify which of the trait data were responsible for a particular species being classified in either category. To improve fit of the RF model, we used recursive feature elimination (RFE) to test subsets of between 1 and 25 predictor variables with 25-fold cross validation and five repetitions with the `rfe` function from `caret`. Additionally, the Pearson correlation of predictor variables was assessed with the `cor` function from the R package `stats` v.4.1.1 (R Core Team 2021) and the package `corrplot` v.0.92 (Wei and Simko 2021; Suppl. material 1: figs S3, S4). Predictor variables identified by RFE were retained for the classifier. We split data into training (85%) and testing (15%) datasets and 2000 decision trees were used in training the classifier with 5-fold cross validation. Out of bag (OOB) error rate, within-class error and variable importance metrics such as mean decrease in accuracy (MDA) were averaged over 50 random seeds of RF and accuracy of the final models was assessed using the testing datasets.

Results

Results of data processing

Genetic and locality data for 383 species of bats and 1971 species of birds were downloaded from *phylogatR*. After filtering for missing data, high sequence divergence, occurrences outside of published geographic ranges and species with fewer than 15 sequences per alignment, we retained 126 bat species and 214 bird species. Though there were substantially more bird species than bat species in the original dataset, a larger proportion of bird species (~ 82%) were excluded as a result of filtering steps due to low number of sequences per species. The average length of the alignments was 688 base pairs (bp) for bats and 787 bp for birds.

Results of Monmonier's algorithm

A conspicuous phylogeographic break was estimated in 68 bat species (54%) and 95 bird species (44.4%) using Monmonier's algorithm. In bats, these phylogeographic

breaks were primarily clustered in areas of high diversity and sampling, specifically southern Central America, northern South America and southeast Asia (Fig. 1). Breaks estimated in birds were more widespread across temperate zones, South America and southeast Asia (Fig. 1). Longer transcontinental and intercontinental phylogeographic breaks estimated for birds are likely explained by the larger geographic ranges of birds in our dataset (mean: 11,803,801 km²) compared to range size of bats included (mean: 6,930,099 km²).

Results of random forests Classifier

Missing data were imputed for the bat trait dataset and density plots indicated that bias in imputed values was minimal (Suppl. material 1: fig. S2). In the bat dataset, RFE indicated that 18 predictor variables should be included in the classifier after removal of highly correlated non-morphological variables (Suppl. material 2: Data Sheets 3, 5). For birds, the best model included 10 variables. We assessed preliminary versions of the trained models and found variation in accuracy and in which variables were

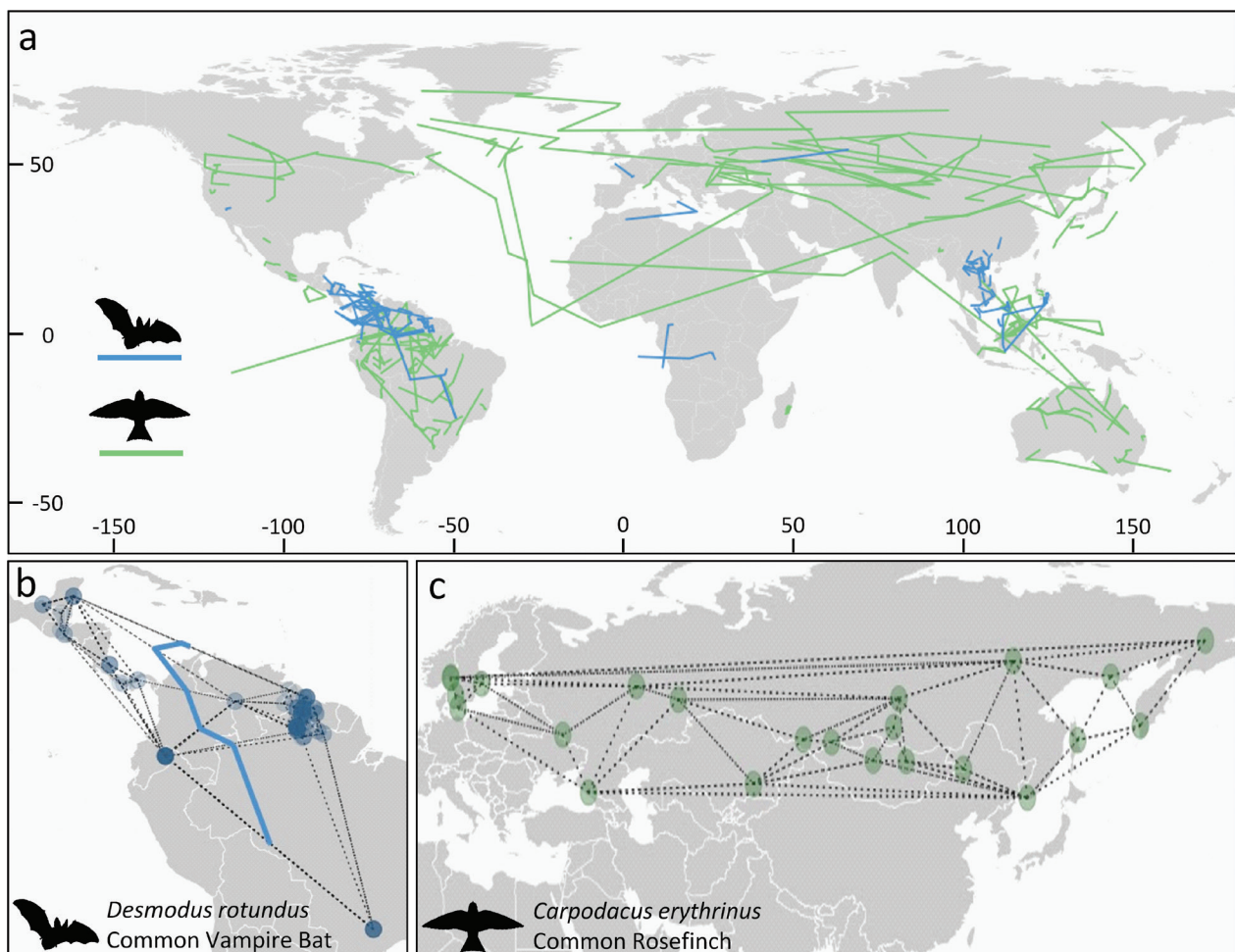


Figure 1. A) Map of phylogeographic breaks estimated by Monmonier's algorithm for global bats (blue) and birds (green); B) Example of species in which a break (thick blue line) was estimated, *Desmodus rotundus*, with connection network between occurrences; C) Example of species determined to not contain a phylogeographic break, *Carpodacus erythrinus*.

most important in predicting if species exhibited a phylogeographic break; thus, 50 independent classifiers were trained for each dataset.

Across 50 random seeds, the OOB error rate for the bat dataset classifier ranged between 0.2461 and 0.3909, with an average balanced accuracy of 69.28% (Table 1, Suppl. material 2: Data Sheet 12). The classifier for the bird dataset had an average balanced accuracy of 65.05%, with OOB error rate ranging from 0.2945–0.3930 (Table 1, Suppl. material 2: Data Sheet 13). The bat classifier was more accurate in predicting bat species that contained a phylogeographic break (class error 0.2043) compared to bat species that did not exhibit a break. In contrast, the bird classifier more accurately predicted species that did not exhibit a phylogeographic break (class error 0.2587) than those that did (Table 1). The bat classifier had higher sensitivity and precision, indicating that the model slightly over-predicted that species would contain breaks. The bird model had higher specificity, indicating a tendency to predict that species would not exhibit phylogeographic breaks (Table 1). This result is expected considering the slightly uneven distributions of the response variable in the bat and bird datasets (Adam et al. 2014).

Table 1. Error rates for the trained bat and bird random forests phylogeographic break classification models.

	OOB	Class Error (No Break)	Class Error (Break)	Sensitivity	Specificity	Precision
Bats	0.3271	0.4696	0.2043	0.8180	0.5675	0.7078
Birds	0.3489	0.2587	0.4626	0.5657	0.7353	0.6440

Important predictors

In bats, the ten top predictor variables as measured by MDA included occurrence area, cryptic diversity predicted, latitudinal range, maximum latitude, mountains within range, brain mass, wing loading, body-size corrected wingspan, carnivory and change in human population density across their range (Fig. 2A). Occurrence area ($p = 1.7 \times 10^{-5}$), latitudinal range ($p = 2.4 \times 10^{-5}$) and maximum latitude ($p = 5.9 \times 10^{-4}$) were significantly different in species estimated to contain a phylogeographic break compared to those without a phylogeographic break with species that exhibit a phylogeographic break having larger occurrence areas at higher latitudes. Species with a phylogeographic break also exhibited significantly smaller brain mass and shorter wingspans compared to species without a phylogeographic break ($p = 0.0092$, $p = 0.028$, respectively; Suppl. material 1: table S1). In birds, invertivory, status as migratory, granivory, use of semi-open habitat, adult mass, hand-wing index, maximum latitude, tarsus length and belonging to the taxonomic orders Procellariiformes and Caprimulgiformes were the ten most important predictors (Fig. 2C). Bird species with a phylogeographic break exhibited maximum latitudes closer to the Equator ($p = 0.031$). Species with breaks were also smaller with less mass ($p = 0.040$) and a smaller hand-wing index ($p = 0.036$).

Discussion

Consistent with previous investigations into population genetic structure (e.g. Martin and McKay (2004); Pelletier and Carstens (2018)), many top predictors of phylogeographic breaks in bats and birds include range characteristics such as occurrence area and measures of latitude. In our study, bird species that contained phylogeographic breaks exhibit maximum latitudes closer to the Equator, consistent with the view that tropical bird species are expected to be more sedentary, have decreased dispersal ability (Sheard et al. 2020) and have more genetic population structure (e.g. Brawn et al. (1996)). Conversely, bat species with a higher maximum latitude more often exhibit phylogeographic breaks, a pattern likely driven by low sampling south of the Equator and the concentration of breaks found north of the Equator in southern Central America and southeast Asia (Suppl. material 1: fig. S5). Additionally, the presence of physical barriers such as a mountain range can still play an important role in structuring genetic diversity even in volant vertebrates as such barriers were a top predictor of phylogeographic breaks in bats. While a recent meta-analysis of phylogeographic breaks in North American mammals found no significant association between the location of phylogeographic breaks in bats and distance to mountains (Jensen et al. 2024), our results indicate that mountains can still be important drivers of phylogeographic structure in highly vagile species. However, many species in both the bird and bat datasets are widely distributed across mountainous regions and lack a phylogeographic break. The organismal traits that facilitate connectivity across mountain ranges are likely related to those required by powered flight, particularly features of the wing that influence flight performance and efficiency.

Most of the organismal traits that were found to be important predictors of phylogeographic breaks in bats and birds are associated with wing morphology (Fig. 3) and feeding ecology. Wing loading (i.e. the ratio of mass to wing surface area) and wingspan are likely to be important because these factors influence the efficiency of flight in flying objects that range in size from *Drosophila* to the Airbus A380. In bats, higher wing loading is positively correlated with flight speed and distances (Norberg and Rayner 1987) and genetic connectivity (Taylor et al. 2012), whereas high aspect ratios are found in species with long, thin wings that exhibit lower drag and increased aerodynamic efficiency (Norberg 1995). Both characteristics are likely correlated with higher dispersal ability, larger geographic ranges (e.g. Luo et al. (2019)) and greater potential to traverse physical barriers in the landscape (Burns and Broders 2014). We found that body-size corrected wingspan was a top predictor of phylogeographic breaks in bats and is significantly smaller in species which exhibit a phylogeographic break (Fig. 2B). Wing loading was amongst the top ten important predictors in bats and was higher in bats that were estimated to not contain a phylogeographic break, though this relationship was not statistically significant ($p = 0.201$; Suppl. material 1: table S1).

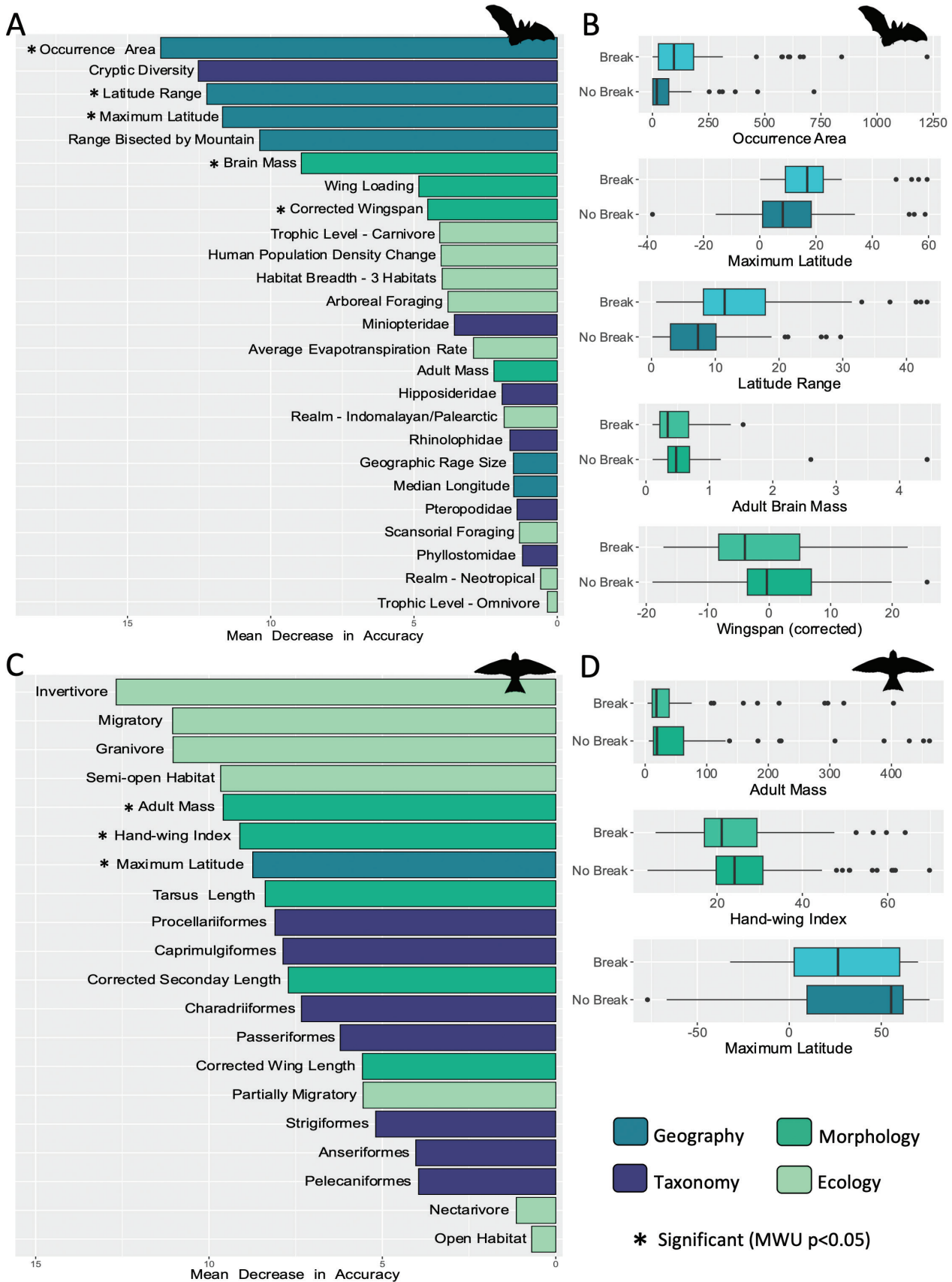


Figure 2. A) Predictor variable importance by mean decrease in accuracy (MDA) for bats, averaged across 50 independently trained classifiers; **B)** boxplots for significant variables in the bat dataset; **C)** predictor variable importance by MDA for 50 independently trained bird model; **D)** boxplots for Mann-Whitney U test significant variables in the bird dataset. Outliers for adult mass in birds above 500 grams were removed to visualise the boxplot.

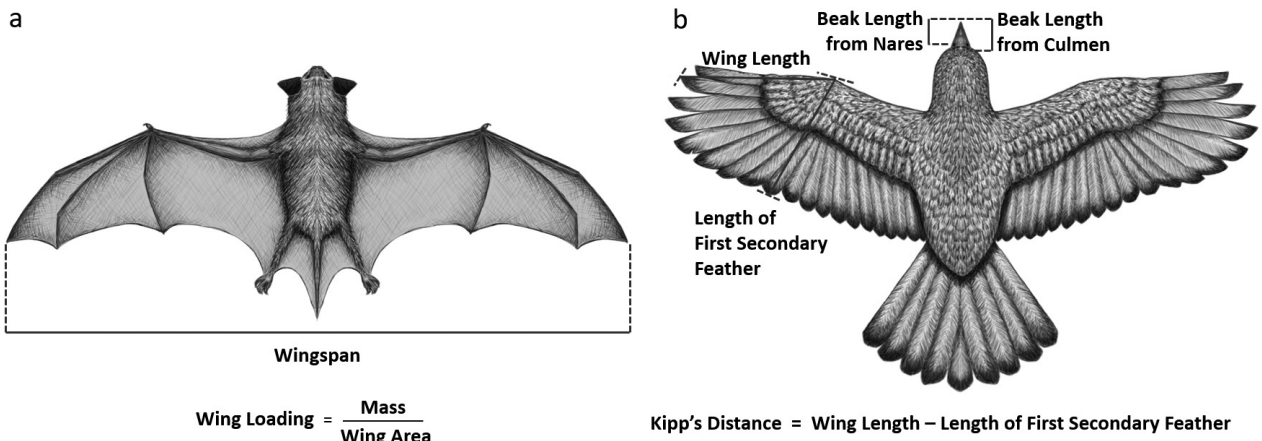


Figure 3. A) Bat illustration with wingspan measurement indicated and equation for wing loading; **B)** Bird with important morphological measurements indicated and the equation for Kipp's distance. Illustrations by Sydney K. Decker.

Similarly, in birds, increased dispersal ability with more efficient long-distance flight is associated with elongated wings with higher hand-wing index, an approximation of aspect ratio (Claramunt et al. 2011). Our results are consistent with findings that bats and birds with wing morphologies adapted for more efficient flight, such as higher hand-wing index, exhibit fewer phylogeographic breaks, possibly due to increased dispersal ability and genetic connectivity across their geographic ranges (Fig. 2D).

Wing shape is also implicated in bat and bird feeding ecology as determinants of how species can navigate their environment to forage (e.g. Norberg and Rayner (1987); Sheard et al. (2020)). In birds, habitat specialisation and diet specialisation are interrelated (Reif et al. 2016). We suspect that this is also true of bats, although we cannot identify a comparable study. However, it is well established that habitat variation is associated with dietary differences within species (e.g. Clare et al. (2011)), that habitat influences the guild structure of bat communities (e.g. Rocha et al. (2018)) and that wing shape influences both the extent of and types of specialisation in feeding ecology (e.g. Fenton (1982)). In birds, invertivory and granivory trophic niche categories were included in top predictors of phylogeographic breaks. Wing morphology in passerines is associated with diet, particularly as it relates to foraging behaviour and food availability (Sheard et al. 2020). Variation in genetic differentiation across biogeographic barriers in South American bird species is also explained by foraging stratum (Burney and Brumfield 2009), suggesting a considerable interaction between feeding ecology and genetic diversity.

Phylogeographic breaks are not necessarily indicative of species-level diversity, as phylogeographic patterns can be caused by social structure or sex biased dispersal (e.g. Dávalos and Russell (2014)), historically allopatric populations (e.g. Fleming et al. (2010)), limited dispersal or small population sizes (Irwin 2002). Some phylogeographic breaks estimated here correspond to boundaries of named subspecies, for example, the eastern and western subspecies of the North American Wilson's Warbler (*Cardellina pusilla*; species specific results available on <https://github.com/skdecker/PhylogeographicBreaks>). Nevertheless, phy-

logeographic structure and intraspecific genetic diversity is positively associated with speciation rates in some taxa, for example, in New World birds (Harvey et al. 2017), while no association is apparent in other groups such as the diverse radiation of Australian lizards (Singhal et al. 2018). Furthermore, a large portion of phylogeographic breaks estimated here occur in the Tropics, indicating that, within species, phylogeographic structure follows similar latitudinal gradients as intraspecific genetic diversity (e.g. Smith et al. (2017); Fonseca et al. (2023)), predictions of cryptic species (Parsons et al. 2022) and species richness (e.g. Gaston (2000); Mittelbach et al. (2007)).

While repurposed, single locus, mitochondrial data may not be representative of overall patterns of population genetic structure (e.g. Bazin et al. (2006); Galtier et al. (2009)), extensive data quality control steps and established automated analysis pipelines allow initial investigations of global biodiversity patterns (e.g. Barrow et al. (2021)). We used rigorous filtering to exclude samples that were highly differentiated from other sequences in species alignments, which could cause an underestimation of deep phylogeographic breaks. However, it was our observation that most sequences excluded at this step were from species with known cryptic lineages currently undergoing taxonomic revision and likely reflect species-level boundaries rather than intraspecific phylogeographic breaks (e.g. *Kerivoula hardwickii*; Tu et al. (2018)).

The use of Monmonier's algorithm to identify phylogeographic breaks has been fairly limited despite its long history as an estimate of population structure. For example, Dupanloup et al. (2002) compared Monmonier's algorithm with spatial analysis of molecular variance (SAMOVA) and found that the algorithm performed better than SAMOVA in finding genetic breaks of simulated single locus data. However, across different methods of phylogeographic break estimation, isolation by distance (IBD) can be misidentified as a break if sampling is sparse across the range of a species (e.g. Templeton et al. (1995); Irwin (2002)). As we used repurposed data with imperfect sampling, it is likely that some estimated breaks represent sampling gaps; however, breaks were often estimated between samples that were not the

most geographically distant. For example, if sampling gaps were the largest contributor to estimation of phylogeographic breaks, we would expect a break to be estimated that separate *Desmodus rotundus* samples in southern Brazil from all other samples, rather than the estimation of a break between eastern and western populations (Fig. 1B).

Once phylogeographic breaks were estimated, 50 independent RF classifiers were trained due to low accuracy and instability of variable importance across preliminary models, likely caused by the use of small datasets. It was our intention that analysis of phylogeographic breaks in birds would yield a more accurate model due to increased availability of data compared to bats; however, over 89% of bird species in the original download from *phylogatR* were removed by our data quality control filters. To verify that low model accuracy was due to small sample size and intrinsic variation, we trained a second RF classifier using IBD, a more widely used metric of how intraspecific genetic diversity is structured across the landscape, as the response variable against the trait datasets (Suppl. material 1). The models trained with IBD also exhibited low accuracy (66.67% for the bat classifier, 60.88% for birds; Suppl. material 1: table S3), this secondary analysis suggesting that low accuracy in preliminary models is not due to the phylogeographic break response being uninformative. Though overall accuracy of both bat and bird models was still relatively low, averaging across multiple independent classifiers allowed us to have more confidence in the measures of predictor variable importance. As RF models predict a complex of traits that interact to affect the response variable, measures such as MDA can help tease apart these relationships to infer biological meaning of influential traits even when there is no statistical difference in trait variables between classes.

Moreover, traits do not evolve independently of one another and it is important to consider the effects of phylogeny in a comparative framework. In the RF analysis, we used taxonomic classification as a proxy for phylogeny and found that only the avian orders Procellariiformes and Caprimulgiformes were amongst the top ten important predictors of phylogeographic breaks in birds (Fig. 2C), while no taxonomic family was a top ten predictor of breaks in bats (Fig. 2A). Notably, the Order Procellariiformes exhibit extreme wing morphologies because they are adapted to foraging over the open ocean and diving for prey (Warham 1996). This specialised ecological niche may facilitate population connectivity, as none of the Procellariiformes species in our analysis was estimated to exhibit a phylogeographic break. Of the Caprimulgiformes species in our analysis, all are predicted to contain a phylogeographic break possibly corresponding to the presence of 6–13 named subspecies in each of these species. This strongly implies that phylogenetic effects are not the primary explanation for the results reported here. Our analysis focuses on traits that influence breaks comprehensively across species, while there likely exist more context-dependent phylogeographic responses to environmental conditions and dispersal ability within individual species. To further explore the effect of phylogeny on the variables in our dataset, we conducted a

phylogenetic factor analysis (PFA; Suppl. material 1). PFA is a multivariate approach to understanding trait evolution under a Bayesian phylogenetic framework. While the PFA did find that some morphological, geographic and ecological trait variables likely did combine to influence diversification, the phylogeographic break variable did not contribute significantly to any of the identified latent evolutionary factors (see Suppl. material 1). This supports the assertion that phylogeographic breaks, as we have estimated here, are not strongly influenced by phylogeny.

Our investigation is a contribution to the growing body of research described as “macrogenetics” (Blanchet et al. 2017) or “automated comparative phylogeography” (Gratton et al. 2017). This research relies on the collection and re-analysis of genetic data that were originally collected for investigation into a single species in order to make inferences about large-scale evolutionary processes. Facilitated by the accession of these data in databases such as NCBI GenBank (Benson et al. 1993) and the development of data aggregators such as *phylogatR* (Pelletier et al. 2022), researchers now have the capacity to ask in-depth questions about genetic diversity in a particular clade (e.g. Manel et al. (2020); Theodoridis et al. (2021); French et al. (2023)) or to investigate patterns across clades on a global scale (Miraldo et al. 2016; Fonseca et al. 2023). Macrogenetic investigations are subject to important questions about scale and over-interpretation (Paz-Vinas et al. 2021, but see Millette et al. (2021)) and it is important to recognise the limitations of single locus data (Edwards and Bensch 2009; Knowles 2009). Given these difficulties and limitations, we recognise that macrogenetic investigations enable novel synthesis of biodiversity data. Our data were drawn from databases containing genetic, environmental, geographic and organismal trait data. These data can be analysed synthetically using machine learning or other AI approaches (e.g. Carstens et al. (2018); Parsons et al. (2022); Yang et al. (2023)). Our study, conducted using data from two clades of volant vertebrate animals, indicates that feeding ecology is likely a more important influence on intraspecific genetic structure than has been previously appreciated. It could be that this result is due to the unique dispersal abilities of the focal clades or it may be that this also holds true for non-volant vertebrates.

Conclusions

Volant vertebrates are not equal in their dispersal capabilities. Our macrogenetic investigation and random forests analyses suggest that species which are capable of flying long distances are less likely to contain intraspecific genetic structure. Phylogeographic breaks tend to be associated with larger ranges at higher latitudes in bats and ranges closer to the Equator in birds. Additionally, ecological traits including habitat breadth in bats and migration, granivory and invertivory in birds were important predictors in our analysis, suggesting an intricate interplay between organismal traits and features of their geographic range contributing to phylogeographic structure. Measures of wing

morphology, such as shorter wingspans in bats and lower hand-wing index in birds, were also important predictors of phylogeographic breaks. Since species-rich clades in bats (i.e. Vespertilionidae) and birds (i.e. Passerines) generally contain species that exhibit these wing morphologies, this suggesting that wing shape may play a role in diversification rate shifts within the Order Chiroptera and Class Aves. While diversification rates appear to be influenced by wing morphology in moths (Aiello et al. 2021) and there has been some exploration of these factors within Class Aves (e.g. Kennedy et al. (2016)), a comprehensive phylogenetic investigation of diversification rates and wing morphology in Chiroptera and Aves is needed.

Acknowledgements

We thank members of the Carstens lab 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 the Ohio Supercomputing Center (PAA1174).

Author contributions

Sydney K. Decker: Conceptualisation, Methodology, Data curation, Analysis, Writing – Original draft, Writing – Review and editing. **Kaiya L. Provost:** Methodology, Data curation, Analysis, Writing – Review and editing. **Bryan C. Carstens:** Conceptualisation, Methodology, Writing – Original draft, Writing – Review and editing, Supervision, Funding acquisition.

Data accessibility statement

Code and source data, including DNA sequence alignments, trait data and analysis files, used in the manuscript are available at <https://github.com/skdecker/PhylogeographicBreaks>. All other data are provided in Suppl. materials 1, 2.

References

- Adam E, Mutanga O, Odindi J, Abdel-Rahman EM (2014) Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. *International Journal of Remote Sensing* 35: 3440–3458. <https://doi.org/10.1080/01431161.2014.903435>
- Aiello BR, Tan M, Bin Sikandar U, Alvey AJ, Bhinderwala B, Kimball KC, Barber JR, Hamilton CA, Kawahara AY, Sponberg S (2021) Adaptive shifts underlie the divergence in wing morphology in bombycoid moths. *Proceedings of the Royal Society B: Biological Sciences* 288: 20210677. <https://doi.org/10.1098/rspb.2021.0677>
- Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, 466 pp. <https://doi.org/10.2307/j.ctv1nznfj7>
- Baker RJ, Bradley RD (2006) Speciation in mammals and the genetic species concept. *Journal of Mammalogy* 87: 643–662. <https://doi.org/10.1644/06-MAMM-F-038R2.1>
- Barrow LN, Masiero da Fonseca E, Thompson CEP, Carstens BC (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>
- Barrowclough GF, Cracraft J, Klicka J, Zink RM (2016) How Many Kinds of Birds Are There and Why Does It Matter? *PLOS ONE* 11: e0166307. <https://doi.org/10.1371/journal.pone.0166307>
- Bazin E, Glémin S, Galtier N (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science* 312: 570–572. <https://doi.org/10.1126/science.1122033>
- Benson D, Lipman DJ, Ostell J (1993) GenBank. *Nucleic acids research* 21: 2963–2965. <https://doi.org/10.1093/nar/21.13.2963>
- BirdLife International and Handbook of the Birds of the World (2021) Bird species distribution maps of the world. Version 2021.1. [Available at] <https://datazone.birdlife.org/species/requestdis>
- Blanchet S, Prunier JG, Kort HD (2017) Time to Go Bigger: Emerging Patterns in Macrogenetics. *Trends in Genetics* 33: 579–580. <https://doi.org/10.1016/j.tig.2017.06.007>
- Brawn JD, Collins TM, Medina M, Bermingham E (1996) Associations between physical isolation and geographical variation within three species of Neotropical birds. *Molecular Ecology* 5: 33–46. <https://doi.org/10.1111/j.1365-294X.1996.tb00289.x>
- Bright JA, Marugán-Lobón J, Cobb SN, Rayfield EJ (2016) The shapes of bird beaks are highly controlled by nondietary factors. *Proceedings of the National Academy of Sciences* 113: 5352–5357. <https://doi.org/10.1073/pnas.1602683113>
- Burney CW, Brumfield RT (2009) Ecology Predicts Levels of Genetic Differentiation in Neotropical Birds. *The American Naturalist* 174: 358–368. <https://doi.org/10.1086/603613>
- Burns LE, Broders HG (2014) Correlates of dispersal extent predict the degree of population genetic structuring in bats. *Conservation Genetics* 15: 1371–1379. <https://doi.org/10.1007/s10592-014-0623-y>
- van Buuren S, Groothuis-Oudshoorn K (2011) mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* 45: 1–67. <https://doi.org/10.18637/jss.v045.i03>
- Carstens BC, Morales AE, Field K, Pelletier TA (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>
- Claramunt S, Derryberry EP, Remsen JV, Brumfield RT (2011) High dispersal ability inhibits speciation in a continental radiation of passerine birds. *Proceedings of the Royal Society B: Biological Sciences* 279: 1567–1574. <https://doi.org/10.1098/rspb.2011.1922>
- Clare EL, Barber BR, Sweeney BW, Hebert PDN, Fenton MB (2011) Eating local: influences of habitat on the diet of little brown bats (*Myotis lucifugus*). *Molecular Ecology* 20: 1772–1780. <https://doi.org/10.1111/j.1365-294X.2011.05040.x>
- Crane M, Silva I, Grainger MJ, Gale GA (2022) Limitations and gaps in global bat wing morphology trait data. *Mammal Review* 52: 165–176. <https://doi.org/10.1111/mam.12270>
- Dávalos LM, Russell AL (2014) Sex-biased dispersal produces high error rates in mitochondrial distance-based and tree-based species delimitation. *Journal of Mammalogy* 95: 781–791. <https://doi.org/10.1644/14-MAMM-A-107>

- Dillon Jr RT (1984) Geographic Distance, Environmental Difference, and Divergence Between Isolated Populations. *Systematic Biology* 33: 69–82. <https://doi.org/10.1093/sysbio/33.1.69>
- Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* 11: 2571–2581. <https://doi.org/10.1046/j.1365-294X.2002.01650.x>
- Edwards S, Bensch S (2009) Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Molecular Ecology* 18: 2930–2933. <https://doi.org/10.1111/j.1365-294X.2009.04270.x>
- Fenton MB (1982) Echolocation, Insect Hearing, and Feeding Ecology of Insectivorous Bats. In: Kunz TH (Ed.) *Ecology of Bats*. Springer US, Boston, MA, 261–285. https://doi.org/10.1007/978-1-4613-3421-7_7
- Fleming TH, Murray KL, Carstens BC (2010) Phylogeography and genetic structure of three evolutionary lineages of West Indian phyllostomid bats. In: Racey PA, Fleming TH (Eds) *Island Bats: Evolution, Ecology, and Conservation*. University of Chicago Press, 116–150.
- Fonseca EM, Pelletier TA, Decker SK, Parsons DJ, Carstens BC (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 CM, Bertola LD, Carnaval AC, Economo EP, Kass JM, Lohman DJ, Marske KA, Meier R, Overcast I, Rominger AJ, Staniczenko PPA, Hickerson MJ (2023) Global determinants of insect mitochondrial genetic diversity. *Nature Communications* 14: 5276. <https://doi.org/10.1038/s41467-023-40936-0>
- Galtier N, Nabholz B, Glémin S, Hurst GDD (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 RC, Bonatelli IAS, Hyseni C, Morales A, Pelletier TA, Perez MF, Rice E, Satler JD, Symula RE, Thomé MTC, Carstens BC (2015) The evolution of phylogeographic data sets. *Molecular Ecology* 24: 1164–1171. <https://doi.org/10.1111/mec.13108>
- Gaston KJ (2000) Global patterns in biodiversity. *Nature* 405: 220–227. <https://doi.org/10.1038/35012228>
- GBIF: The Global Biodiversity Information Facility (2024) What is GBIF? [Available from] <https://www.gbif.org/what-is-gbif>
- Grant PR (1981) Speciation and the Adaptive Radiation of Darwin's Finches: The complex diversity of Darwin's finches may provide a key to the mystery of how intraspecific variation is transformed into interspecific variation. *American Scientist* 69: 653–663.
- Gratton P, Marta S, Bocksberger G, Winter M, Trucchi E, Kühl H (2017) A world of sequences: can we use georeferenced nucleotide databases for a robust automated phylogeography? *Journal of Biogeography* 44: 475–486. <https://doi.org/10.1111/jbi.12786>
- Greenberg R, Cadena V, Danner RM, Tattersall G (2012) Heat loss may explain bill size differences between birds occupying different habitats. *PLOS ONE* 7: e40933. <https://doi.org/10.1371/journal.pone.0040933>
- Harvey MG, Seeholzer GF, Smith BT, Rabosky DL, Cuervo AM, Brumfield RT (2017) Positive association between population genetic differentiation and speciation rates in New World birds. *Proceedings of the National Academy of Sciences* 114: 6328–6333. <https://doi.org/10.1073/pnas.1617397114>
- Hernández-Canchola G, Verde Arregoitia LD, Colunga-Salas P, Gómez-Jiménez YA, León-Paniagua L (2021) A Global Review of Phylogeographic Studies on Bats. In: Lim BK, Fenton MB, Brigham RM, Mistry S, Kurta A, Gillam EH, Russell A, Ortega J (Eds) 50 Years of Bat Research: Foundations and New Frontiers. Fascinating Life Sciences. Springer International Publishing, Cham, 289–309. https://doi.org/10.1007/978-3-030-54727-1_18
- Huber SK, Podos J (2006) Beak morphology and song features covary in a population of Darwin's finches (*Geospiza fortis*). *Biological Journal of the Linnean Society* 88: 489–498. <https://doi.org/10.1111/j.1095-8312.2006.00638.x>
- Irwin DE (2002) Phylogeographic breaks without geographic barriers to gene flow. *Evolution* 56: 2383–2394. <https://doi.org/10.1111/j.0014-3820.2002.tb00164.x>
- IUCN (2022) The IUCN Red List of Threatened Species. 2022-2. <https://www.iucnredlist.org> [Accessed on 13/07/2022]
- Jensen AJ, Cove MV, Goldstein BR, Kays R, McShea W, Pacifici K, Rooney B, Kierepka E (2024) Geographic barriers but not life history traits shape the phylogeography of North American mammals. *Global Ecology and Biogeography* 33: e13875. <https://doi.org/10.1111/geb.13875>
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27: 3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>
- Jones KE, Bielby J, Cardillo M, Fritz SA, O'Dell J, Orme CDL, Safi K, Sechrest W, Boakes EH, Carbone C, Connolly C, Cutts MJ, Foster JK, Grenyer R, Habib M, Plaster CA, Price SA, Rigby EA, Rist J, Teacher A, Bininda-Emonds ORP, Gittleman JL, Mace GM, Purvis A (2009) PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology* 90: 2648–2648. <https://doi.org/10.1890/08-1494.1>
- Kane D, Goodwin S, Verspui GJ, Tump A, Marín GM del M (2019) Reptile diversity of southern Morocco: range extensions and the role of the Djebel Ouarkiz as a biogeographical barrier. *Herpetology Notes* 12: 787–793.
- Kennedy JD, Borregaard MK, Jønsson KA, Marki PZ, Fjeldså J, Rahbek C (2016) The influence of wing morphology upon the dispersal, geographical distributions and diversification of the Corvids (Aves; Passeriformes). *Proceedings of the Royal Society B: Biological Sciences* 283: 20161922. <https://doi.org/10.1098/rspb.2016.1922>
- Knowles LL (2009) Statistical Phylogeography. *Annual Review of Ecology, Evolution, and Systematics* 40: 593–612. <https://doi.org/10.1146/annurev.ecolsys.38.091206.095702>
- Kuhn M (2021) caret: Classification and Regression Training. [Available from:] <https://CRAN.R-project.org/package=caret>
- Lande R (1980) Genetic Variation and Phenotypic Evolution During Allopatric Speciation. *The American Naturalist* 116: 463–479. <https://doi.org/10.1086/283642>
- Lee C-R, Mitchell-Olds T (2011) Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Molecular Ecology* 20: 4631–4642. <https://doi.org/10.1111/j.1365-294X.2011.05310.x>
- Leigh DM, van Rees CB, Millette KL, Breed MF, Schmidt C, Bertola LD, Hand BK, Hunter ME, Jensen EL, Kershaw F, Liggins L, Luikart G, Manel S, Mergeay J, Miller JM, Segelbacher G, Hoban S, Paz-Vinas I (2021) Opportunities and challenges of macrogenetic studies.

- Nature Reviews Genetics 22: 791–807. <https://doi.org/10.1038/s41576-021-00394-0>
- Luo B, Santana SE, Pang Y, Wang M, Xiao Y, Feng J (2019) Wing morphology predicts geographic range size in vespertilionid bats. *Scientific Reports* 9: 4526. <https://doi.org/10.1038/s41598-019-41125-0>
- Manel S, Holderegger R (2013) Ten years of landscape genetics. *Trends in Ecology & Evolution* 28: 614–621. <https://doi.org/10.1016/j.tree.2013.05.012>
- 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: 692. <https://doi.org/10.1038/s41467-020-14409-7>
- Manni F, Guérard E, Heyer E (2004) Geographic Patterns of (Genetic, Morphologic, Linguistic) Variation: How Barriers Can Be Detected by Using Monmonier's Algorithm. *Human Biology* 76: 173–190. <https://doi.org/10.1353/hub.2004.0034>
- Martin PR, McKay JK (2004) Latitudinal variation in genetic divergence of populations and the potential for future speciation. *Evolution* 58: 938–945. <https://doi.org/10.1111/j.0014-3820.2004.tb00428.x>
- Millette KL, Fugère V, Debyser C, Greiner A, Chain FJJ, Gonzalez A (2021) Refining analyses of existing data sets is valuable for macrogenetics: a response to Paz-Vinas, Jensen et al., (2021). *Ecology Letters* 24: 1285–1286. <https://doi.org/10.1111/ele.13733>
- Miraldo A, Li S, Borregaard MK, Flórez-Rodríguez A, Gopalakrishnan S, Rizvanovic M, Wang Z, Rahbek C, Marske KA, Nogués-Bravo D (2016) An Anthropocene map of genetic diversity. *Science* 353: 1532–1535. <https://doi.org/10.1126/science.aaf4381>
- Mittelbach GG, Schemske DW, Cornell HV, Allen AP, Brown JM, Bush MB, Harrison SP, Hurlbert AH, Knowlton N, Lessios HA, McCain CM, McCune AR, McDade LA, McPeck MA, Near TJ, Price TD, Ricklefs RE, Roy K, Sax DF, Schluter D, Sobel JM, Turelli M (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters* 10: 315–331. <https://doi.org/10.1111/j.1461-0248.2007.01020.x>
- Monmonier MS (1973) Maximum-Difference Barriers: An Alternative Numerical Regionalization Method. *Geographical Analysis* 5: 245–261. <https://doi.org/10.1111/j.1538-4632.1973.tb01011.x>
- Moussy C, Hosken DJ, Mathews F, Smith GC, Aegerter JN, Bearhop S (2013) Migration and dispersal patterns of bats and their influence on genetic structure. *Mammal Review* 43: 183–195. <https://doi.org/10.1111/j.1365-2907.2012.00218.x>
- Myers EA, McKelvy AD, Burbrink FT (2020) Biogeographic barriers, Pleistocene refugia, and climatic gradients in the southeastern Nearctic drive diversification in cornsnakes (*Pantherophis guttatus* complex). *Molecular Ecology* 29: 797–811. <https://doi.org/10.1111/mec.15358>
- Norberg UM (1995) How a long tail and changes in mass and wing shape affect the cost for flight in animals. *Functional Ecology* 9: 48–54. <https://doi.org/10.2307/2390089>
- Norberg UM, Rayner JMV (1987) Ecological morphology and flight in bats (Mammalia; Chiroptera): wing adaptations, flight performance, foraging strategy and echolocation. *Philosophical Transactions of the Royal Society B* 316: 337–419. <https://doi.org/10.1098/rstb.1987.0030>
- Nosil P (2012) *Ecological Speciation*. OUP Oxford, 300 pp. <https://doi.org/10.1093/acprof:osobl/9780199587100.001.0001>
- O'Leary MA, Kaufman S (2011) MorphoBank: phylophenomics in the 'cloud'. *Cladistics* 27: 1–9. <https://doi.org/10.1111/j.1096-0031.2011.00355.x>
- Orr HA (1996) Dobzhansky, Bateson, and the Genetics of Speciation. *Genetics* 144: 1331–1335. <https://doi.org/10.1093/genetics/144.4.1331>
- Parsons DJ, Pelletier TA, Wieringa JG, Duckett DJ, Carstens BC (2022) Analysis of biodiversity data suggests that mammal species are hidden in predictable places. *Proceedings of the National Academy of Sciences* 119: e2103400119. <https://doi.org/10.1073/pnas.2103400119>
- Paz-Vinas I, Jensen EL, Bertola LD, Breed MF, Hand BK, Hunter ME, Kershaw F, Leigh DM, Luikart G, Mergeay J, Miller JM, Van Rees CB, Segelbacher G, Hoban S (2021) Macrogenetic studies must not ignore limitations of genetic markers and scale. *Ecology Letters* 24: 1282–1284. <https://doi.org/10.1111/ele.13732>
- Pelletier TA, Carstens BC (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 TA, Parsons DJ, Decker SK, Crouch S, Franz E, Ohrstrom J, Carstens BC (2022) phylogatR: Phylogeographic data aggregation and repurposing. *Molecular Ecology Resources* 22: 2830–2842. <https://doi.org/10.1111/1755-0998.13673>
- Penone C, Davidson AD, Shoemaker KT, Di Marco M, Rondinini C, Brooks TM, Young BE, Graham CH, Costa GC (2014) Imputation of missing data in life-history trait datasets: which approach performs the best? *Methods in Ecology and Evolution* 5: 961–970. <https://doi.org/10.1111/2041-210X.12232>
- Phillimore AB, Freckleton RP, Orme CDL, Owens IPF (2006) Ecology predicts large-scale patterns of phylogenetic diversification in birds. *The American Naturalist* 168: 220–229. <https://doi.org/10.1086/505763>
- Pigot AL, Tobias JA (2015) Dispersal and the transition to sympatry in vertebrates. *Proceedings of the Royal Society B: Biological Sciences* 282: 20141929. <https://doi.org/10.1098/rspb.2014.1929>
- Polato NR, Gill BA, Shah AA, Gray MM, Casner KL, Barthelet A, Messer PW, Simmons MP, Guayasamin JM, Encalada AC, Kondratieff BC, Flecker AS, Thomas SA, Ghalambor CK, Poff NL, Funk WC, Zamudio KR (2018) Narrow thermal tolerance and low dispersal drive higher speciation in tropical mountains. *Proceedings of the National Academy of Sciences* 115: 12471–12476. <https://doi.org/10.1073/pnas.1809326115>
- R Core Team (2021) R: A language and environment for statistical computing. [Available from:] <https://www.R-project.org/>
- Ratnasingham S, Hebert PDN (2007) bold: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Molecular Ecology Notes* 7: 355–364. <https://doi.org/10.1111/j.1471-8286.2007.01678.x>
- Reif J, Hořák D, Křištín A, Kopsová L, Devictor V (2016) Linking habitat specialization with species' traits in European birds. *Oikos* 125: 405–413. <https://doi.org/10.1111/oik.02276>
- Rising JD, Somers KM (1989) The Measurement of Overall Body Size in Birds. *The Auk* 106: 666–674. <https://doi.org/10.2307/4087777>
- Rocha R, Ovaskainen O, López-Baucells A, Farneda FZ, Sampaio EM, Bobrowiec PED, Cabeza M, Palmeirim JM, Meyer CFJ (2018) Secondary forest regeneration benefits old-growth specialist bats in a fragmented tropical landscape. *Scientific Reports* 8: 3819. <https://doi.org/10.1038/s41598-018-21999-2>
- Sarkar D (2008) *Lattice: Multivariate Data Visualization with R*. Springer, New York, NY. <https://doi.org/10.1007/978-0-387-75969-2>

- Sebastián M, Ortega-Retuerta E, Gómez-Consarnau L, Zamanillo M, Álvarez M, Arístegui J, Gasol JM (2021) Environmental gradients and physical barriers drive the basin-wide spatial structuring of Mediterranean Sea and adjacent eastern Atlantic Ocean prokaryotic communities. *Limnology and Oceanography* 66: 4077–4095. <https://doi.org/10.1002/lno.11944>
- Senar JC, Pascual J (1997) Keel and tarsus length may provide a good predictor of avian body size. *ARDEA* 85: 269–274.
- Sheard C, Neate-Clegg MHC, Alioravainen N, Jones SEI, Vincent C, MacGregor HEA, Bregman TP, Claramunt S, Tobias JA (2020) Ecological drivers of global gradients in avian dispersal inferred from wing morphology. *Nature Communications* 11: 2463. <https://doi.org/10.1038/s41467-020-16313-6>
- Sheard C, Street SE, Evans C, Lala KN, Healy SD, Sugawara S (2023) Beak shape and nest material use in birds. *Philosophical Transactions of the Royal Society B: Biological Sciences* 378: 20220147. <https://doi.org/10.1098/rstb.2022.0147>
- Singhal S, Huang H, Grundler MR, Marchán-Rivadeneira MR, Holmes I, Title PO, Donnellan SC, Rabosky DL (2018) Does population structure predict the rate of speciation? A comparative test across Australia's most diverse vertebrate radiation. *The American Naturalist* 192: 432–447. <https://doi.org/10.1086/699515>
- Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* 236: 787. <https://doi.org/10.1126/science.3576198>
- Smith BT, Seeholzer GF, Harvey MG, Cuervo AM, Brumfield RT (2017) A latitudinal phylogeographic diversity gradient in birds. *PLOS Biology* 15: e2001073. <https://doi.org/10.1371/journal.pbio.2001073>
- Soons J, Genbrugge A, Podos J, Adriaens D, Aerts P, Dirckx J, Herrel A (2015) Is beak morphology in Darwin's finches tuned to loading demands? *PLOS ONE* 10: e0129479. <https://doi.org/10.1371/journal.pone.0129479>
- Soria CD, Pacifici M, Di Marco M, Stephen SM, Rondinini C (2021) COMBINE: a coalesced mammal database of intrinsic and extrinsic traits. *Ecology* 102: e03344. <https://doi.org/10.1002/ecy.3344>
- Taylor PJ, Goodman SM, Schoeman MC, Ratrimomanarivo FH, Lamb JM (2012) Wing loading correlates negatively with genetic structuring of eight Afro-Malagasy bat species (Molossidae). *Acta Chiropterologica* 14: 53–62. <https://doi.org/10.3161/150811012X654268>
- Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics* 140: 767–782. <https://doi.org/10.1093/genetics/140.2.767>
- Theodoridis S, Rahbek C, Nogues-Bravo D (2021) Exposure of mammal genetic diversity to mid-21st century global change. *Ecography* 44: 817–831. <https://doi.org/10.1111/ecog.05588>
- Tobias JA, Ottenburghs J, Pigot AL (2020) Avian Diversity: Speciation, Macroevolution, and Ecological Function. *Annual Review of Ecology, Evolution, and Systematics* 51: 533–560. <https://doi.org/10.1146/annurev-ecolsys-110218-025023>
- Tobias JA, Sheard C, Pigot AL, Devenish AJM, Yang J, Sayol F, Neate-Clegg MHC, Alioravainen N, Weeks TL, Barber RA, Walkden PA, MacGregor HEA, Jones SEI, Vincent C, Phillips AG, Marples NM, Montaña-Centellas FA, Leandro-Silva V, Claramunt S, Darski B, Freeman BG, Bregman TP, Cooney CR, Hughes EC, Capp EJ, Varley ZK, Friedman NR, Korntheuer H, Corrales-Vargas A, Trisos CH, Weeks BC, Hanz DM, Töpfer T, Bravo GA, Remeš V, Nowak L, Carneiro LS, Moncada RAJ, Matysioková B, Baldassarre DT, Martínez-Salinas A, Wolfe JD, Chapman PM, Daly BG, Sorensen MC, Neu A, Ford MA, Mayhew RJ, Fabio Silveira L, Kelly DJ, Annorabah NND, Pollock HS, Grabowska-Zhang AM, McEntee JP, Carlos T, Gonzalez J, Meneses CG, Muñoz MC, Powell LL, Jamie GA, Matthews TJ, Johnson O, Brito GRR, Zyskowski K, Crates R, Harvey MG, Jurado Zevallos M, Hosner PA, Bradfer-Lawrence T, Maley JM, Stiles FG, Lima HS, Provost KL, Chibesa M, Mashao M, Howard JT, Mlamba E, Chua MAH, Li B, Gómez MI, García NC, Päckert M, Fuchs J, Ali JR, Derryberry EP, Carlson ML, Urriza RC, Brzeski KE, Prawiradilaga DM, Rayner MJ, Miller ET, Bowie RCK, Lafontaine R-M, Scofield RP, Lou Y, Somaratna L, Lepage D, Illif M, Neuschulz EL, Templin M, Dehling DM, Cooper JC, Pauwels OSG, Analuddin K, Fjeldså J, Seddon N, Sweet PR, DeClerck FAJ, Naka LN, Brawn JD, Aleixo A, Böhning-Gaese K, Rahbek C, Fritz SA, Thomas GH, Schlegel M (2022) AVONET: morphological, ecological and geographical data for all birds. *Ecology Letters* 25: 581–597. <https://doi.org/10.1111/ele.13898>
- Tokita M, Yano W, James HF, Abzhanov A (2017) Cranial shape evolution in adaptive radiations of birds: comparative morphometrics of Darwin's finches and Hawaiian honeycreepers. *Philosophical Transactions of the Royal Society B: Biological Sciences* 372: 20150481. <https://doi.org/10.1098/rstb.2015.0481>
- Tu VT, Hassanin A, Furey NM, Son NT, Csorba G (2018) Four species in one: multigene analyses reveal phylogenetic patterns within Hardwicke's woolly bat, *Kerivoula hardwickii*-complex (Chiroptera, Vespertilionidae) in Asia. *Hystrix, the Italian Journal of Mammalogy* 29: 111. <https://doi.org/10.4404/hystrix-00017-2017>
- Uy JAC, Irwin DE, Webster MS (2018) Behavioral isolation and incipient speciation in birds. *Annual Review of Ecology, Evolution, and Systematics* 49: 1–24. <https://doi.org/10.1146/annurev-ecolsys-110617-062646>
- Wang IJ (2013) Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution* 67: 3403–3411. <https://doi.org/10.1111/evo.12134>
- Warham J (1996) *The Behaviour, Population Biology and Physiology of the Petrels*. Academic Press, 623 pp. <https://doi.org/10.1016/B978-012735415-6/50012-0>
- Wei T, Simko V (2021) corrrplot: Visualization of a correlation matrix. R package version 0.92. <https://github.com/taiyun/corrrplot>
- Wright S (1943) Isolation by Distance. *Genetics* 28: 114–138. <https://doi.org/10.1093/genetics/28.2.114>
- Yang J, Carstens BC, Provost KL (2023) Machine learning reveals relationships between song, climate, and migration in coastal *Zonotrichia leucophrys*. *bioRxiv* 2023.03.08.531720. <https://doi.org/10.1101/2023.03.08.531720>

Supplementary materials

Supplementary material 1

Supplementary methods and results, references, figs S1–S12, tables S1–S6 (.docx)

Link: <https://doi.org/10.21425/fob.18.139911.suppl1>

Supplementary material 2

Supplementary datasheets 1–23 (.xlsx)

Link: <https://doi.org/10.21425/fob.18.139911.suppl2>