

## RESEARCH ARTICLE

## Animal Functional Traits

## Is species richness mediated by functional and genetic divergence? A global analysis in birds

Nicholas M. A. Crouch  | David Jablonski 

Department of the Geophysical Sciences,  
The University of Chicago, Chicago,  
Illinois, USA

## Correspondence

Nicholas M. A. Crouch  
Email: [ncrouch@uchicago.edu](mailto:ncrouch@uchicago.edu)

## Funding information

National Aeronautics and Space  
Administration, Grant/Award Number:  
EXOB08-0089; National Science  
Foundation, Grant/Award Number: DEB-  
2049627

Handling Editor: Daniel García

## Abstract

1. Unravelling why species richness shows such dramatic spatial variation is an ongoing challenge. Common to many theories is that increasing species richness (e.g. with latitude) requires a compensatory trade-off on an axis of species' ecology. Spatial variation in species richness may also affect genetic diversity if large numbers of coexisting, related species result in smaller population sizes.
2. Here, we test whether increasing species richness results in differential occupation of morphospace by the constituent species, or decreases species' genetic diversity. We test for two potential mechanisms of morphological accommodation: denser packing in ecomorphological space, and expansion of the space. We then test whether species differ in their nucleotide diversity depending on allopatry or sympatry with relatives, indicative of potential genetic consequences of coexistence that would reduce genetic diversity in sympatry. We ask these questions in a spatially explicit framework, using a global database of avian functional trait measurements in combination with >120,000 sequences downloaded from GenBank.
3. We find that higher species richness within families is not systematically correlated with either packing in morphological space or overdispersion but, at the Class level, we find a general positive relationship between packing and species richness, but that points sampled in the tropics have comparatively greater packing than temperate ones relative to their species richness. We find limited evidence that geographical co-occurrence with closely related species or tropical distributions decreases nucleotide diversity of nuclear genes; however, this requires further analysis.
4. Our results suggest that avian families can accumulate species regionally with minimal tradeoffs or cost, implying that external biotic factors do not limit species richness.

## KEYWORDS

coexistence, latitudinal diversity gradient, morphology, nucleotide diversity, sympatry

## 1 | INTRODUCTION

Many lineages exhibit strong spatial variation in species numbers, commonly manifesting in latitudinal and longitudinal diversity gradients

(Hillebrand, 2004; Schumm et al., 2019). Numerous hypotheses have been proposed for what allows large numbers of species to exist in a single area. High tropical diversity has been attributed to many factors, singly or in concert, including: time, productivity, predation

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Functional Ecology* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

pressures and climatic stability (Dobzhansky, 1950; Fine, 2015; Harvey et al., 2020; MacArthur, 1969; Pianka, 1966). However, common to multiple proposed mechanisms is that, given finite resources, species must partition resources to coexist, a concept often referred to as 'competitive exclusion' or 'limiting similarity' (Gause, 1934; Hutchinson, 1959; Lack, 1973; MacArthur & Levins, 1967). One or many aspects of species' ecology can vary to facilitate this coexistence, often summarized in terms of functional groups (discrete roles in an ecosystem) or ecomorphological traits (as continuous variables), and the intersection of functional trait data with the comprehensive phylogenetic and biogeographic data now available for birds affords new paths for understanding the factors controlling large-scale biodiversity patterns (Tobias et al., 2020). However, local species composition and the formation of species also have a genetic component (Felsenstein, 1981). Therefore, investigations into the mechanisms regulating species richness will likely benefit from considering these different dimensions of species' biology. Here, we test whether the level of co-occurrence of species within a geographic area carries a morphological and/or genetic signature across avian families. If so, this would indicate that diversity is mediated by biotic forces, e.g. competition for limited resources that impose a set of phenotypic and genetic consequences deriving from species numbers. Rejection of this hypothesis would suggest that alternative hypotheses are required to explain patterns of spatial heterogeneity in species richness, such as intrinsic variation in lineage diversification rates, or environmental factors including spatial patterns of net primary productivity and habitat heterogeneity.

For morphology, we target functionally relevant traits, such as wing and beak lengths, which have long been known to correspond to ecological differences among bird species (Ricklefs & Travis, 1980). Analyses encompassing all bird species have shown that such traits can accurately capture even subtle differences in avian functional ecology (Dehling et al., 2016; Pigot et al., 2020), facilitating their use in macroevolutionary studies (Crouch & Ricklefs, 2019; Pigot et al., 2016; Tobias et al., 2022), including quantifying species responses to increasing local or regional diversity. Broadly, species may either pack more densely in morphospace (Pellissier et al., 2018; Pigot et al., 2016), or develop disparate morphologies to allow exploitation of unique resources (Rosamund et al., 2020). The relative position of species in morphospace is most frequently quantified as mean nearest-neighbour

distances (NND, Foote, 1990), which have been used to show that increasing species richness can correlate with denser packing in morphospace (Pigot et al., 2016). Expansion of species' occupation of morphospace can be quantified using convex hulls – a polyhedron bounding a set of points in *n*-dimensional space (Cornwell et al., 2006). Although convex hulls are sensitive to outliers (Guillerme, Cooper, et al., 2020; Guillerme, Puttick, et al., 2020; Podani, 2009), morphospaces constructed using ecomorphological traits provide an intuitive method for quantifying functional richness that can correlate with other metrics of clade disparity (Crouch & Ricklefs, 2019).

Species interactions may also impact genetic diversity. For example, co-occurrence with closely related species may limit population sizes given competition for shared resources (Yamaguchi et al., 2021), referred to as the 'crowding effect' (Gavina et al., 2018, and references therein). This effect scales non-linearly with population density, reducing the possibility of coexistence (Gavina et al., 2018). Moreover, reduced population sizes owing to competition may render species and/or populations more susceptible to extinction for many reasons, including the general expectation that, broadly speaking, nucleotide diversity ( $\pi$ ) tends to correlate with population size (Ellegren & Galtier, 2016). Reduced genetic diversity is expected to increase susceptibility to extinction through intrinsic factors such as mutation load, and reduced adaptive potential (Lanfear et al., 2014; Poon & Otto, 2007), and external factors such as fluctuating climatic conditions (Pauls et al., 2013; Smith et al., 2021), a causal relationship demonstrated most clearly through theoretical models (Johansson, 2018; Zhao et al., 2018). Genetic diversity within populations is therefore a widely-used conservation metric, especially for certain genes (for example CO1, Porter & Hajibabaei, 2018; Petit-Marty et al., 2020, but see also Teixeira & Huber, 2021).

Synthesizing previous research on the mechanisms of coexistence allows a series of predictions for quantitative patterns over broader taxonomic and spatial scales (Table 1). We make predictions at different taxonomic levels as they may yield differing results (Crouch, 2021), which can lead to deeper understanding of evolutionary processes. For example, comparing patterns across many taxa, we might expect the effect of competition to be weaker than for analyses of closely related species (Darwin, 1859; MacArthur & Levins, 1967; Mayfield & Levine, 2010; Webb et al., 2002). For morphology, at the Class level we predict that

| Trait         | Taxonomic level | Prediction   |
|---------------|-----------------|--|
| Morphological | Class           | Nearest-neighbour distances and convex-hull volumes are expected to scale positively with increasing species richness        |
|               | Family          | Increasing family species richness is expected to result in lower nearest neighbour distances or greater convex hull volumes |
| Genetic       | Class           | Mean genetic diversity of species is expected to decrease with increasing species richness                                   |
|               | Family          | Families with greater species richness and/or overlapping ranges are expected to have decreased genetic diversity            |
|               | Species         | Species that are allopatric with confamilial species are expected to have greater genetic diversity                          |

**TABLE 1** Summary of the hypotheses tested in this study with respect to the data type and taxonomic level

higher species richness should lead to an increase in NND and convex hull volumes as shown in previous research restricted to within individual taxonomic orders (Crouch & Ricklefs, 2019). We do, however, predict some deviations; for example, South American rivers have disproportionately higher diversity compared with their immediate vicinity due to an artefact in the spatial data such that rivers are included in the range of species whose limits are defined by its path. Moreover, major rivers separate closely related, often sister species (Crouch et al., 2019) meaning we would expect these points to have comparatively lower NND scores and convex hull volumes due to the greater similarity of the constituent taxa. Within families we expect either (a) a negative relationship between family species richness and NND scores, as species pack in an invariant morphospace to accommodate diversity, or (b) a positive relationship between richness and NND, as species generate novel morphologies to facilitate coexistence and thus increase convex hull volumes. For genetic data we have a general prediction that transcends taxonomic levels – that increasing coexistence should result in lower genetic diversity – but test this hypothesis in different ways at the different taxonomic levels (Table 1). At the class and family level we expect a negative linear relationship between species richness and genetic diversity, while at the species level we expect sympatric species to have significantly lower diversity than when allopatric with confamilials. Some care is required in interpreting results given that allopatric and sympatric species can be significantly different in age (Crouch, 2021), and recent speciation events can be correlated with higher genetic diversity (since both peak at low latitudes, Jetz et al., 2012; Adams & Hadly, 2013; Schluter & Pennell, 2017).

In this study we predict that ecomorphological traits and genetic diversity should show spatial variation consistent with the impact that changes in species richness has on species coexistence. We test this overarching hypothesis at different taxonomic scales (Table 1), leveraging a global data set of avian functional traits, and >120,000 individual genetic sequences downloaded from GenBank, to compare morphological and genetic patterns at a global scale. We find no differential occupation of morphospace and limited depletion in genetic variation with increased species numbers at a location, suggesting negative or displacive interactions among species are secondary at this scale, and discuss our results in the context of previous work investigating the processes governing species richness.

## 2 | MATERIALS AND METHODS

### 2.1 | Morphological analyses

We obtained morphological data from Pigot et al. (2020), who provide principal component (PC) scores for eight external avian measurements (beak length from tip to skull along the culmen; beak length to the nares; beak width at the nares; beak depth at the nares; tarsus length; wing length; first secondary length; and tail length) and body mass (in grams from published sources), along with data on four PC axes describing beak variation. Pigot et al. (2020) generated these data for 9963 species by measuring 52,870 live-caught and museum specimens in combination with data on 2288 species taken

from published sources. The final dataset has a single value for each trait for each species, so that geographical variation in individual species morphology is not incorporated in the analysis; however, the vast majority of variance in trait values is explained between species rather than within (98.25% vs. 1.75%, Pigot et al., 2020), meaning this is highly unlikely to affect our results. We replicated each analysis described below using each of the two data sets, because (a) the beak is a primary axis of ecomorphological differentiation for birds (Chira et al., 2018; Cooney et al., 2017), and (b) analysis of morphological subunits may yield different results, providing additional insight into the process of ecomorphological differentiation.

We quantified morphological trends at two taxonomic and spatial scales: (1) all species analysed concurrently, using a global grid of points at 0.5° intervals (2) family-specific analyses using regions of maximal species richness defined by range polygons of individual species. For the global analyses, we created a matrix for all species present at each point in the grid included in the spatial data (range polygons) of BirdLife International and NatureServe (2016). This sampling strategy is distinct from sampling all species within 0.5° grid cells; sampling within grid cells may artificially record species with narrow or parapatric ranges as coexisting when they do not. Although the point-based approach alleviates potential sampling biases, caution is still required when using these spatial data as resolution varies among species and, in general, distributions may be overestimated (Graham & Hijmans, 2006; Herkt et al., 2017; Hurlbert & Jetz, 2007). Nevertheless, such errors are likely to be small relative to the spatial scale of this analysis and the geographic ranges of most species. There may be instances of artificial overlap if the spatial data define a species distribution as a continuous area, including regions where it is not present (e.g. a mountain top). However, the majority of our analyses focus within families, so that systematic bias would require significant intra-familial variation in the detail of their spatial data, which we believe to be extremely rare. Defining allopatry and sympatry using overlaps in spatial data might be considered simplistic given that species may exploit different parts of the environment; however, given known morphological and ecological similarities between closely related species, it is reasonable to expect them to exploit similar parts of the environment in most instances (see Crouch, 2021 and references therein). For migratory species we analysed the entire range, as it is not clear whether limitation of population sizes on breeding grounds and/or non-breeding grounds is the limiting factor. For example, a considerable number of songbirds have smaller non-breeding than the breeding ranges (e.g. Tennessee Warbler, *Leiothlypis peregrina*).

We calculated two properties of the suite of species inferred at each point: mean NND and mean convex hull size. We calculated NND scores using the R package DISTANCES (Savje, 2019), and hull volumes using the package GEOMETRY (Habel et al., 2019). NND scores are calculated using all PC axes simultaneously, while convex hull volumes are calculated using the first three PC axes (93% of the variation in the data, Pigot et al., 2020). We quantified the effect of species richness on morphological scores using a spatial regression model that accommodates for the non-independence of spatial

data, implemented in the R package *SPAMM* (Rousset & Ferdy, 2014). Species richness was log-transformed prior to its inclusion in the model.

For the family-level analysis we only analysed those families with at least five species ( $n = 140$ , >97% of all avian species, using the taxonomy of Jetz et al., 2012). For each of these families we quantified the maximum number of co-occurring species by converting the spatial polygons into raster layers and taking the sum of the number of species ranges across the range of the family. The area (defined as spatially continuous grid points, i.e. points separated at  $0.5^\circ$  intervals with no intervening sampling points of lesser richness) containing the maximum number of species can vary between families, and there may be multiple, geographically distinct areas with equivalent maximum richness. After extracting the species composition of the maximally diverse area(s) we quantified the mean NND and absolute convex hull size of those species. We evaluated whether these values deviated from approximately random expectation by comparing the empirical values against mean values derived from random draws of species. For each random draw we sampled, without replacement, the same number of confamilial species as in the maximally diverse area, and calculated the mean NND and absolute convex hull size of that sample. We performed 1000 replicates for each family, repeating the procedure if a family had more than one maximally diverse area with all data points included in subsequent analyses (although the results were unchanged if mean values were taken for families with more than maximally diverse area). We quantified the effect of family species richness on NND and convex hull volumes with and without accounting for the shared evolutionary history of families. For the phylogenetically informed analysis we fit models incorporating a  $\lambda$  estimate of phylogenetic signal, implemented in the R package *PHYLOLM* (Ho & Ané, 2014).

To evaluate whether the distribution of the species in morphospace differed between the empirical and null expectations, we repeated the above procedure calculating the skew and kurtosis of the NND scores, rather than the mean. Finally, to evaluate whether dispersion results are biased by the use of convex hulls, we repeated the above sampling procedure calculating multivariate disparity as the sum of variances using the R package *DISPRTY* (Guillerme, 2018). The sum of variances has been shown to accurately capture even subtle changes in trait space occupancy, reducing confounding effects of outliers (Guillerme, Puttick, et al., 2020). Since we expect greater diversity to be accommodated by denser packing in morphospace, we predict a negative relationship between within-family species richness and mean NND scores.

To account for potentially confounding effects of additional variables, we also used path analysis to test apparent correlations of within-family species richness with either packing or overdispersion in morphospace. For example, position in morphospace may not be independent of clade richness, with previous analyses of birds finding that depauperate clades tend to be located more peripherally in the space (Kennedy et al., 2020; Ricklefs, 2005). We evaluated the fit of four models using the R package *PHYLOPATH* (van der Bijl, 2018; von Hardenberg & Gonzalez-Voyer, 2013) that varied the number

of connections between variables (ranging between five and nine, Supplementary Material). As potential explanatory variables we included body mass, range size and position in morphospace (defined as the square root of the sum of the squares of the distances on each PC axis, Ricklefs, 2005). Since this analysis is at the family level, mean values were calculated for each of these variables. *Phylopath* allows for correlations between variables to be tested within a phylogenetic framework to account for the shared evolutionary history of species. Here, we downloaded 1000 phylogenies from the pseudo-posterior distribution of Jetz et al. (2012) using the Hackett et al. (2008) backbone. We created a consensus tree from these trees using *TreeAnnotator* (Drummond & Rambaut, 2007), before simplifying the result to represent inter-familial relationships. Support for each of the four models was evaluated using the C statistic Information Criterion (Cardon et al., 2011), which can accommodate both small sample sizes and path analysis (von Hardenberg & Gonzalez-Voyer, 2013). To generate the final network for each analysis we used model averaging, where the strength of the correlation between variables is weighted by the support of the model. In addition to repeating the analysis for both packing and overdispersion, we performed separate analyses using PC values first from all morphological traits, and then using just the PC values from beak measurements.

## 2.2 | Genetic analyses

We downloaded genetic data from GenBank using the R package *RENTREZ* (Winter, 2017). We analysed mitochondrial and nuclear loci; however, the widespread use of mitochondrial data in resolving species-level phylogenetic relationships means that more of these sequences are available. We chose three mitochondrial genes that maximized both taxonomic coverage and sample size per species (cytb, ND2, and CO1), as well as a separate download of full mitochondrial genomes (mtDNA). We queried GenBank for 10 nuclear genes used in phylogenetic studies, downloading three with more than 500 samples (RAG1, MC1R and EGR1), but ultimately excluded EGR1 owing to insufficient intraspecific sampling. When querying GenBank for gene-specific sequences we restricted results to sequence lengths between 500 base pairs, to filter out low coverage results, and 2000bp to exclude whole-genome sequences, which were analysed independently. We allowed minor grammatical changes in how names are recorded in GenBank; for example, 'RAG1' and 'RAG-1', and, for all downloads, we specified 'Aves' as the organism to ensure we sampled across the entire clade. For all downloads we parsed the raw sequence data into family alignments which were then aligned using *MAFFT* (Katoh et al., 2002) to create balanced alignments such that each species within a family has the same sequence length for each gene. We aligned the data by family to maximize the accuracy of the alignments. Iterating over the family-level alignments we calculated species-specific nucleotide diversity ( $\pi$ ) scores for species with at least two sequences using the R package *PEGAS* (Paradis, 2010). This method specifically accounts for

variation in sequence coverage between species; some species will have missing data to varying extents across each alignment.

As with the analyses of morphology, we analysed the genetic data both across all of Aves and within families. For the analysis of all birds, we created global maps of  $\pi$  using the community presence/absence matrix described above to test for spatial or latitudinal variation in sequence diversity. We quantified the effect of species richness on  $\pi$  by building a spatial regression which also included as predictors the number of sequences used to calculate  $\pi$ , and the gene used to calculate  $\pi$ . Species richness, the number of species  $\pi$  was calculated from, and  $\pi$  itself were log-transformed prior to the analysis. We confirmed that a spatial regression model was necessary by testing for spatial autocorrelation in the residuals of a non-spatial model, repeating this for each gene individually. A significant relationship between species richness and  $\pi$  on a global scale would suggest that the relatedness of species in a community is less important than the total biodiversity in an area.

Next, we compared  $\pi$  between tropical and temperate species. Using the spatial data from Birdlife International and NatureServe (2016), we defined species as tropical if their entire range fell between the Tropic of Cancer and Tropic of Capricorn, with temperate species defined as those whose ranges were entirely outside that tropical band. Out of the 7181 valid species we had data for, this procedure classified 3431 as tropical and 698 as temperate. The remaining species were classified as either bridging northern temperate and tropics ( $n = 1265$ ), bridging southern temperate and tropics ( $n = 1391$ ), or spanning all latitude bands ( $n = 396$ ). Finally, we tested for a genetic effect of sympatry by comparing  $\pi$  scores between sympatric and allopatric species (from all families pooled together) using the two definitions of Crouch (2021): allopatry defined using sister taxa, and all confamilial taxa concurrently. We quantified the differences in binary states (i.e. allopatry/sympatry and temperate/tropical) using Welch's unequal-variances two-sample  $t$ -test and, since  $p$ -values from such tests can be affected by sample size, Cohen's  $D$  estimates of effect sizes, calculated using the R package RSTATIX (Kassambara, 2021).

We performed two tests at the family level. First, we compared within-family species diversity against the difference in  $\pi$  between allopatric and sympatric species of the constituent species. As described above, we believe that these spatial data are sufficient for establishing geographic relationships between species given low intrafamilial variance in species ecology (Crouch, 2021). If there is a genetic cost to coexistence, we would expect a negative relationship; however, if speciose families achieve their richness through an ability to mitigate potential costs, we predict a positive relationship. Second, we tested for a relationship between either the maximum or mean number of overlapping ranges within each family (from Crouch, 2021) and the mean  $\pi$  scores of its constituent species. Again, if there is a cost to coexistence then we predict a negative relationship between these variables. For both tests we performed linear regressions, with per-family species richness, the mean number of overlapping ranges, and  $\pi$  all log-transformed prior to analysis.

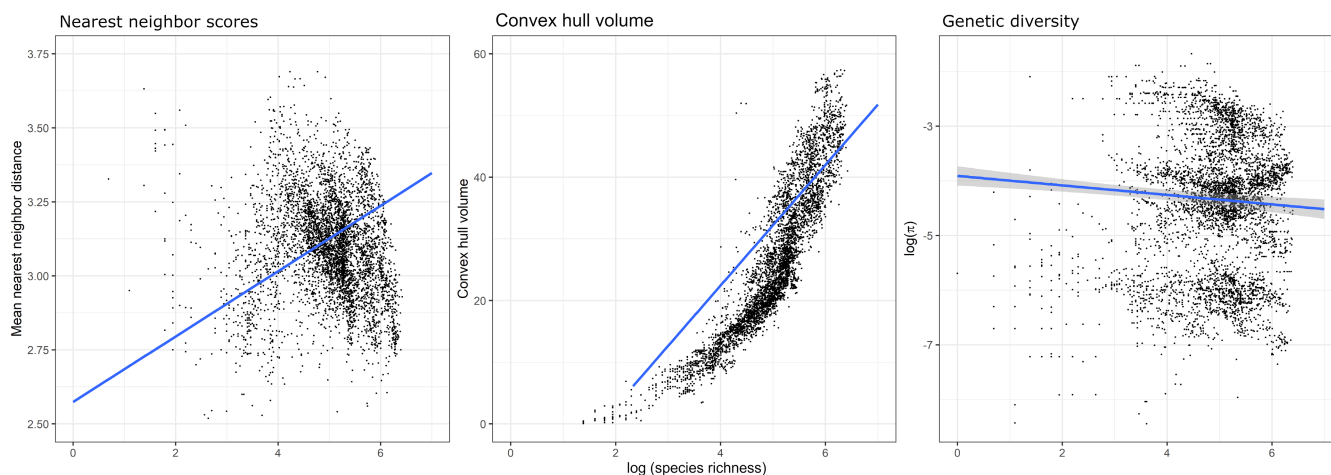
An alternative taxonomic level for the genetic analyses is within a single species, comparing the nucleotide diversity of samples taken from parts of the species range where it occurs in allopatry from either its sister species or con-familial species, and samples where it occurs in sympatry. We examined all CO1 samples with GPS coordinates available on GenBank using the pipeline of Porter and Hajibabaei (2018). The use of CO1 in barcoding studies means that >40% of entries have GPS coordinates, compared with <2% for cytb and ND2. Filtering the data so that only species with at least 4 sequences were retained, as at least 2 sequences are needed in allopatry and sympatry for  $\pi$  calculations, yielded 12,286 sequences representing 1232 species. However, no coordinates were recorded as allopatric from these data (using multiple definitions), likely due to extreme spatial biases in sequences uploaded to GenBank that have recorded GPS coordinates (Figure S1), ruling out intraspecific analyses.

### 3 | RESULTS

#### 3.1 | Morphology

Quantifying the relationship between species richness and NND scores for functional traits on a global scale revealed a strong positive relationship when spatial autocorrelation is taken into account (Figure 1,  $t = 13.09$ ), consistent with previous analyses of avian taxa (Crouch & Ricklefs, 2019). This positive relationship is not seen in the distribution of the raw data (Figure S2), but regions with lower mean NND scores than expected for their species richness (shown in blue in Figure 2a) are concentrated in the tropics. Some of these are artifactual; for example, along the Amazon River (which spuriously contains the distributions of species whose ranges are defined by its course), but are generally concentrated in regions of high diversity. Conversely, regions with high NND scores for their species richness are concentrated in low-diversity, and environmentally harsh regions. We suspect this reflects phylogenetically disparate lineages independently colonizing these regions. While these areas of high residual variance from the regression line are visually striking, they are not abundant in the distribution of data points (Figure S2).

Convex hull volumes and species richness were more strongly correlated than either was with NND scores (Figure 1; Figure S2). Convex hull volumes also varied spatially, but with considerably less clearcut patterns than NND scores (Figure 2). The more ambiguous pattern for convex hulls was due to comparatively few, widely distributed, morphologically extreme species dominating the results. Specifically, across the 61,483 sampling locations only 158 unique species (approximately 1% of all bird species) were recorded as the most peripheral species in morphospace. These taxa are notable for being particularly large (therefore having large ranges) and morphologically quirky, with species including palaeognath taxa (e.g. Ostrich, *Struthio camelus*, and species of kiwi, *Apteryx*), the Marabou stork (*Leptoptilos crumeniferus*) and two species of pelican (*Pelecanus*). As these species are morphologically distinct from



**FIGURE 1** Effect of species richness on nearest neighbour scores (*left*), convex hull volumes (*middle*), and genetic diversity ( $\pi$ , *right*). Black points show the raw data ( $n = 5000$ ) from spatial points analysed. Regression lines shows the results from spatially explicit regression models which also, for the analysis of genetic diversity, considered number of sequences used to calculate  $\pi$  and the gene for which  $\pi$  was calculated. The shaded regions around the lines represent the 95% confidence interval. In the plots of nearest neighbour scores and convex hull volumes this area is contained within the width of the regression line.

most species sampled at each point, they disproportionately influence the calculated convex hull volumes. Again, the analysis of only beak traits produced results indistinguishable from the analysis of all traits (Figure S3).

At a global scale we found fine-scale spatial variation in NND scores (Figure 2). In particular, regions of comparatively low NND scores for their species richness (i.e. blue regions in Figure 2) are generally concentrated in the tropics. We also find some variation within continents, such between the east and western half of North America. At the family level, the number of maximally species-rich regions ranged between 1 and 8 with a mean of 1.8; most families (56%) had a single maximally rich area. There was a weak negative relationship between the species richness of a family and its number of maximally diverse regions (Figure S4). Both mean NND and hull volumes show a general positive relationship with family species-richness (Figure S5), but we found no relationship between a family's total species-richness and either packing or overdispersion (Figure 3, packing estimate =  $-0.004$ ,  $p = 0.75$ ; overdispersion estimate =  $-0.002$ ,  $p = 0.92$ ). In other words, increasing species richness within a single area does not result in differential occupation of morphospace by a family compared with a random sample of species. Analysis of beak traits alone produced the same result as an analysis of all traits combined (Figure S6, packing estimate =  $-0.002$ ,  $p = 0.85$ ; overdispersion estimate =  $0.001$ ,  $p = 0.75$ ). We also find no relationship between family richness and either NND scores or convex hull volumes when accounting for the shared evolutionary history of families (Table S1).

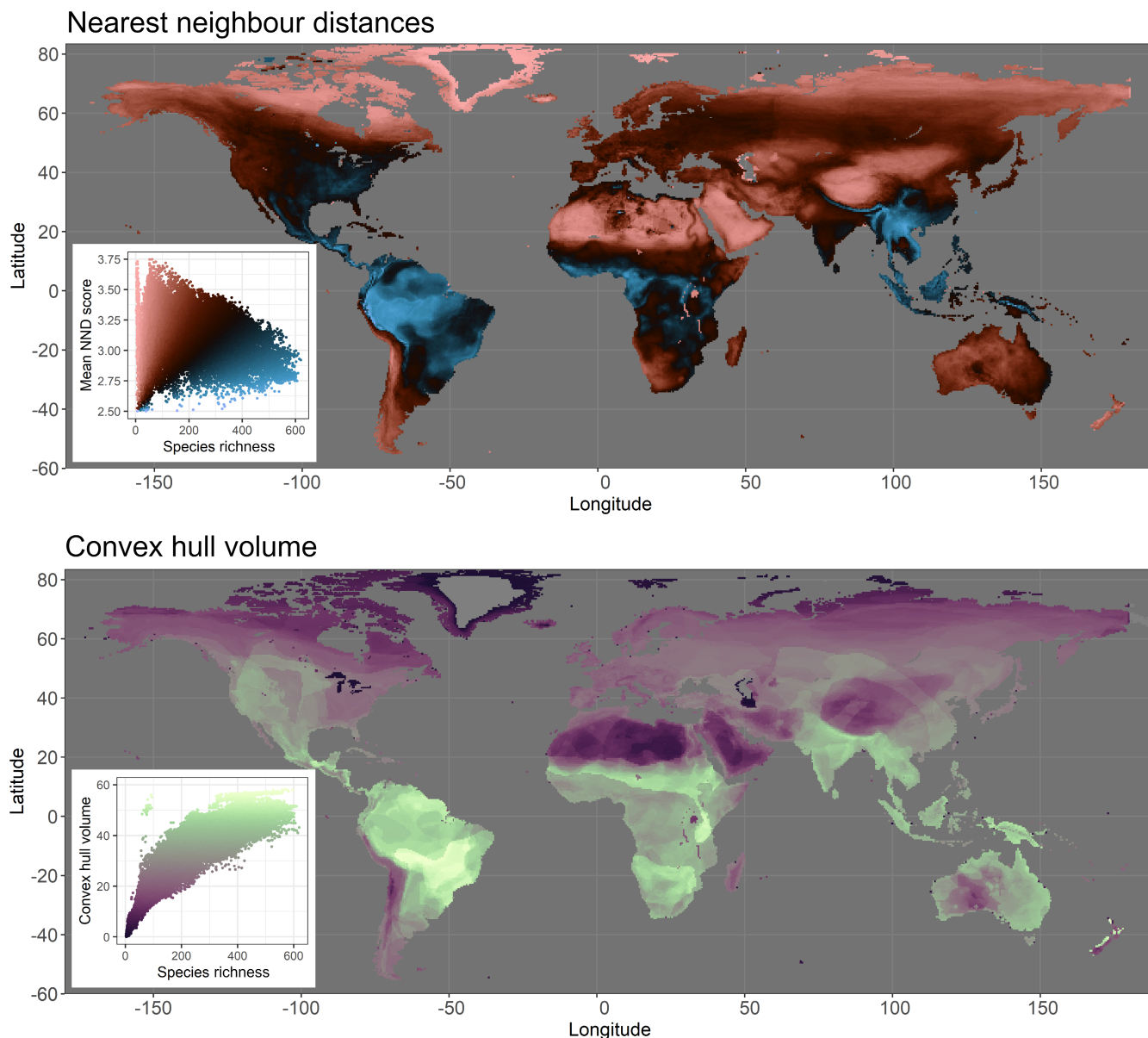
We did not recover any association between the species richness of families and either NND or convex hull volume when analysing the data in a phylogenetic path analysis (Figure S7). The networks did recover established relationships however, such as a positive correlation between mass and range size, and between convex hull volume and number of species per family. One surprising result

was the absence of a relationship between a family's richness and its mean position in morphospace relative to the overall Class centroid (Figure S7), contrary to previous findings (Kennedy et al., 2020; Ricklefs, 2005). As before, analysis of only beak traits produced nearly identical network results (Figure S8). Finally, calculating multivariate disparity rather than convex hull volumes did not change the observed results for the analysis of all traits or just beak traits (Figure S9).

### 3.2 | Genetic results

A total of 124,301 sequences were downloaded from GenBank, with some variation among three target genes (Table 2). As expected, there were considerably fewer nuclear sequences available for analysis than mitochondrial. Using the raw data downloaded from GenBank, genes differed significantly in mean  $\pi$  between genes (Table 2), but with considerable overlap in the distributions. We consolidated or removed 752 species sequences (8% of the total) for species with synonymies or extinct taxa respectively. An example of an extinct species with genetic data available is the Hawai'i 'ō'ō (*Moho nobilis*). Across all genetic datasets, the number of sequences used per species to calculate  $\pi$  ranged from 2, the minimum number of sequences required for calculating  $\pi$ , to 623 (for the ND2 gene for the barn swallow, *Hirundo rustica*), with  $\pi$  uncorrelated with the number of sequences used ( $r = 0.08$ , CI 0.07–0.09).

Global maps of  $\pi$  show pronounced spatial heterogeneity with substantial spatial variation between datasets (Figure S10), with all genes showing significant spatial autocorrelation of scores ( $p < 2.2 \times 10^{-16}$ , Supplementary Material). When accounting for this spatial autocorrelation we found a negative relationship between species diversity and  $\pi$ , but with considerable noise ( $t = -2.46$ , Figure 1). When we compared  $\pi$  between temperate and tropical species we

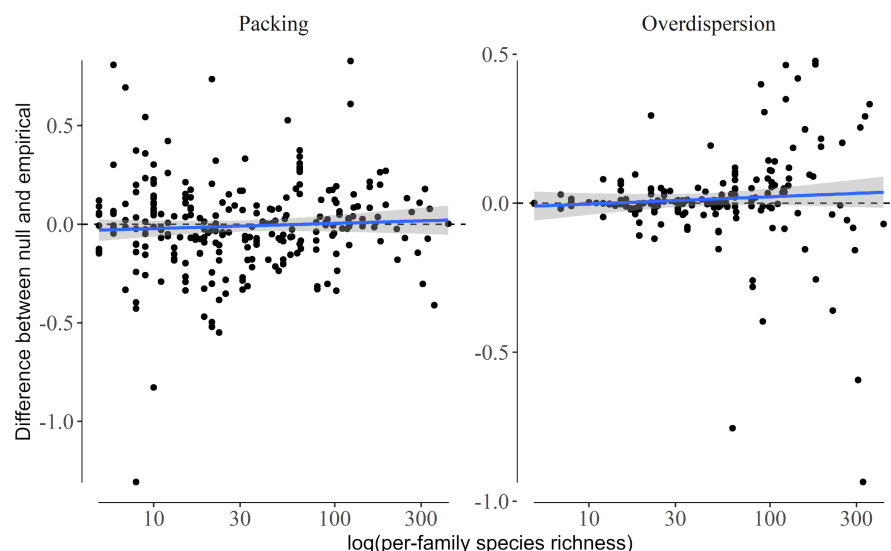


**FIGURE 2** Global variation in nearest-neighbour distances and convex hull volumes. Inset panels in both images show the relationship between the corresponding variable and species richness across the globe. For the NND scores, the black indicates locations showing the expected relationship between the variables (see Figure 1), with brown and blue colours highlighting areas with comparatively higher and lower NND scores for the respective species richness. In bottom panel, dark purple indicates areas of low species richness and small convex hull volumes, and light green areas of high species richness and large convex hull volumes.

identified significant differences in mean values between regions, but with only small or negligible effect sizes (Table S2) as the distributions of values substantially overlap (Figure 4).

For analyses within families, we identified further significant differences among genes, predominantly with largely negligible effect sizes (Table S2; Figure 4). With allopatry defined exclusively using sister taxa, we found allopatric and sympatric species had higher  $\pi$  for different genes, but with largely negligible effect sizes (Table S2). With allopatry defined using all confamilial species, we identified larger effect sizes, but lower sample sizes of allopatric species under this sampling scheme prohibits confident assertion of patterns. We also identified a significant positive relationship between per-family

richness and the difference between allopatric and sympatric species for RAG1 (Figure 4b,  $p = 0.04$ ), but this relationship also seems entirely driven by a single outlier family (Treecreepers, Certhiidae, 10 species). Excluding Certhiidae leaves no significant relationship between the variables, although the families Sittidae and Fringillidae also show some deviation from the general trend (Figure S11). When the same relationship was examined using allopatry defined using confamilial species there was no relationship. However, there were substantially fewer data points for this analysis, as only ~3% of species are allopatric with all of their confamilial species (Crouch, 2021; the proportion remains at ~4% even when 10% overlap of geographic range polygons is treated as allopatric), although 68% of the



**FIGURE 3** Relationship between family richness and packing (left) and overdispersion (right). A linear regression line is shown in blue, with the shaded areas reflecting the 95% confidence intervals.

| Data type     | Gene  | Sequences downloaded | Mean sequences per species | Mean $\pi$ | Std. dev $\pi$ |
|---------------|-------|----------------------|----------------------------|------------|----------------|
| Mitochondrial | CO1   | 30,036               | 6.7                        | 0.023      | 0.069          |
|               | CYTB  | 37,953               | 6.4                        | 0.016      | 0.033          |
|               | ND2   | 47,936               | 7.1                        | 0.016      | 0.030          |
|               | mtDNA | 3054                 | 2.8                        | 0.003      | 0.020          |
| Nuclear       | RAG1  | 2741                 | 1.9                        | 0.024      | 0.079          |
|               | MC1R  | 2106                 | 9.4                        | 0.014      | 0.072          |

**TABLE 2** Summary of sample sizes and nucleotide diversity ( $\pi$ ) across the target genetic data for raw data downloaded from GenBank. mtDNA refers to mitochondrial genomes. All available sequences were also downloaded for EGR1, but there was insufficient data for its inclusion in the final analyses

114 families do have at least one allopatric species. Finally, RAG1 was the only gene identified as having a significant relationship between the mean number of overlapping ranges for a family and  $\log(\pi)$ , with more overlapping ranges corresponding to reduced  $\pi$  (Figure 4c). Using maximum instead of mean overlapping ranges did not change the result, with the two being correlated with each other (Crouch, 2021). These analyses are unlikely to have been affected by differences in the age of species as we found no significant relationships between species age (calculated as the length of the branch leading to the species) and  $\pi$  for any of the genes analysed (Table S3).

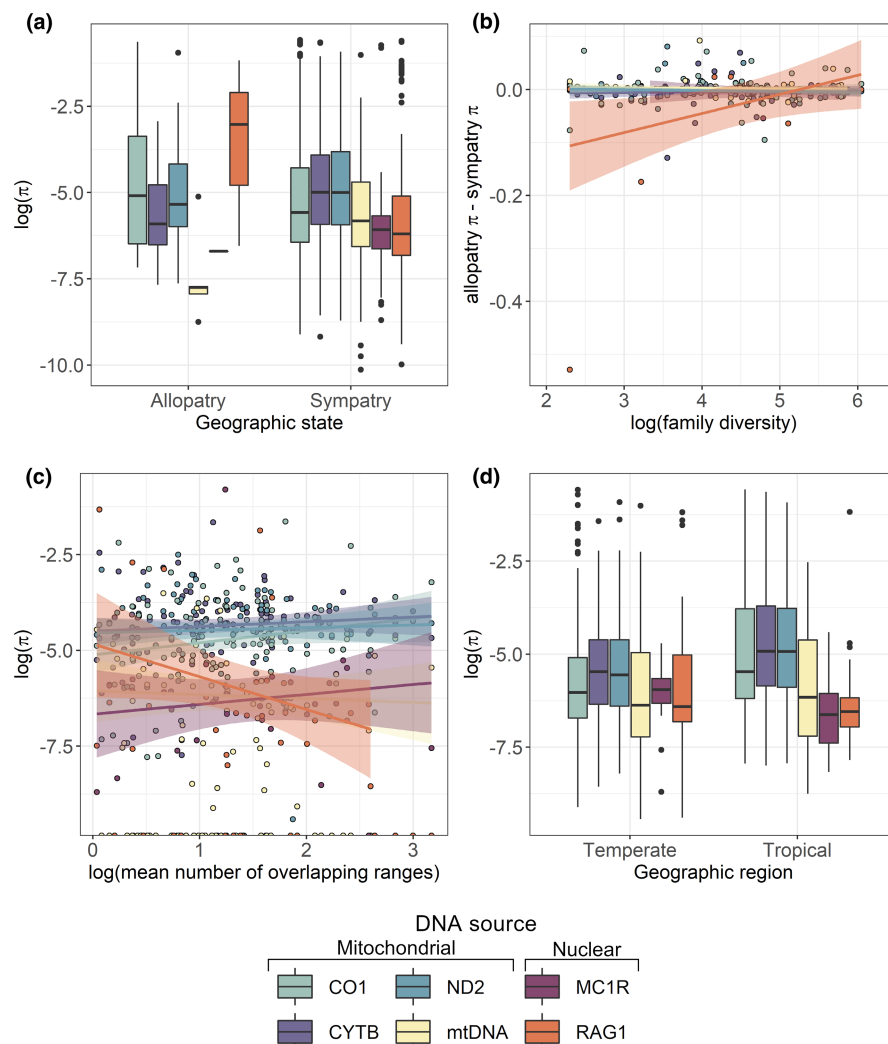
## 4 | DISCUSSION

In this study we predicted that coexistence between species will likely affect functionally important traits and/or genetic diversity, but that the nature of this relationship might vary by taxonomic level (Table 1). At the class level, we expected a positive relationship between species richness and our morphological variables, as including a wider variety of species increases the morphological variation present. We found strong support for this prediction. At the family level, where competition between coexisting species is expected to be greater, we predicted that increasing local diversity of species could be associated either with increased packing in morphological space, indicating finer subdivision of functional attributes, or increasing overdispersion, indicating

greater separation among co-occurring species. Neither expectation was met, as we found no evidence that increasing the number of closely related species in an area produces a morphological signature distinct from a random sample of equivalent species richness. Combined with the observation that the number of overlapping ranges is strongly correlated with family species-richness (Crouch, 2021), these results suggest that clades can accumulate overlapping ranges unimpeded by co-occurring relatives, with functional traits unaltered. Though a surprising conclusion, it is supported by longstanding observation that ecological space is not saturated (i.e. there are 'empty niches', Hutchinson, 1957; Walker & Valentine, 1984; Ashby et al., 2017), a growing body of evidence that ecologically similar species can coexist (Gómez-Llano et al., 2021), and evidence that character displacement (as well as ecological sorting) does not significantly contribute to the evolution of avian beak morphology (a primary axis of ecomorphological differentiation, Tobias et al., 2020; Freeman et al., 2022). These combined works suggest instead that other factors predominantly regulate regional species richness; for example, variation in the intrinsic properties of lineages to speciate.

Lineages differ greatly in their propensity to speciate, often influenced by their intrinsic biotic properties (see Jablonski, 2008 for a general tabulation of such traits at the organismic and higher levels). For example, more elongated, pointed wings can facilitate interregional dispersal, generating additional opportunities for allopatric speciation events (Claramunt et al., 2012; Phillimore

**FIGURE 4** Comparison of genetic diversity ( $\pi$ ) between different geographic comparison. (a) Difference in  $\pi$  between allopatric and sympatric species. (b) the relationship between per-family species richness and  $\pi$  of allopatric and sympatric member species. (c) the relationship between the mean number of overlapping ranges within a family and the mean  $\pi$  of its constituent species. (d) Comparison of  $\pi$  between temperate and tropical species. Lines in panels b and c show linear regressions with 95% confidence intervals.



et al., 2006); in other groups high dispersal may damp gene flow and so curtail allopatric speciation while supporting broad, extinction-resistant geographic ranges (e.g. Jablonski, 2008). Alternatively, behavioural characteristics may regulate species richness. For example, territorial bird species are considerably more likely to have non-overlapping ranges than are non-territorial species along an elevational gradient (Freeman et al., 2019). In this case, separation of species is unrelated to similarity in functional traits, therefore not generating any morphological signal in response to increased species numbers.

The lack of strong effects of bird diversity on morphospace occupation and genetic variation does not mean that ecological opportunity is unimportant for regulating biodiversity patterns. Ecological opportunity can promote dramatic radiations (Grant & Grant, 2008; Losos, 2010; Price et al., 2014; Simpson, 1953; Stroud & Losos, 2016), and its reduction can impede diversification or colonization as resources become less readily available (MacArthur, 1958; Rabosky & Lovette, 2008a; Jablonski, 2017, and on probability of invasion success see Theoharides & Dukes, 2007; Gallien & Carboni, 2016); for example, avian diversity in the Himalayas has been shown to mirror altitudinal changes in resource abundance (Price et al., 2014).

However, present-day regional biodiversity is unlikely to be a stable saturation of ecological opportunity; if it was then it would not explain, for example, Simberloff's (1981) classic observation that invasive species need not affect recipient communities if the invaders exploit unused resources, or the steady production of new species and higher taxa in the tropics where taxonomic richness is highest (e.g. Jablonski et al., 2017). Given that approximately 95% of bird species are sympatric with a confamilial species and that this value increases with greater family species richness (Figure 2 of Crouch, 2021), there does not appear to be an absolute limit to the ability of the existing pool of taxa to accommodate more species. Moreover, biota may be unlikely to reach absolute saturation given that, as diversity nears this point, declining speciation and/or increasing extinction become more pronounced, with background extinction maintaining some fraction of open ecospace (Valentine, 1980; Walker & Valentine, 1984). Therefore, while biotas may find equilibrium points of standing diversity (Valente et al., 2017), it is likely that these are lower than the potential number of species that could be supported. There is considerable discussion as to whether the strong latitudinal and longitudinal variation in species richness regional biotas reflect equilibria or saturated states (Cornell, 2013). It is not a

novel observation that evolution is ongoing, but we must continue to evaluate our hypotheses for what factors regulate biodiversity.

Ecological opportunity is most commonly invoked in regulating biodiversity through quantification of declining diversification rates towards the present in molecular phylogenies (Phillimore & Price, 2008; Price, 2008; Pybus & Harvey, 2000; Rabosky & Lovette, 2008a, 2008b). However, these declines may also be explained by a time-dependent model as well as diversity dependence, and distinguishing between those processes is challenging using extant-only phylogenies (Pannetier et al., 2020). Nevertheless, integrating additional sources of data, such as geographic distributions and trait data can enable greater insight (Tobias et al., 2020). Given that clade size is uncorrelated with slowdowns (Crouch, 2021), and the results here that clades are morphologically and largely genetically unrestricted in their ability to accumulate species, true diversity-dependent processes may be rare in birds. Even so, it is worth reaffirming that diversity dependence and time dependence are likely not mutually exclusive; a clade's diversity is likely a combination of the two with the relative contribution of the two varying temporally (Edie et al., 2018). Given the long-standing history of diversity dependence as a hypothesized regulating factor of species diversity (MacArthur, 1962; Rabosky & Lovette, 2008a), a more confident assessment of the frequency of diversity dependence in regulating extant diversity will likely require integration of the fossil record (see for example Foote et al., 2018; Marshall & Quental, 2016). However, while functional traits have been evaluated for many fossil groups (Bambach et al., 2007; Jablonski, 2017; Novack-Gottshall et al., 2022), the ability to integrate these data will likely vary by clade and by data type. For example, analyses of avian taxa will require use discrete morphological data as the continuous external measurements used here are highly unlikely to fossilize.

Contrary to previous work, we found no relationship between the position of clades in ecomorphological space and their species richness (Kennedy et al., 2020; Ricklefs, 2005). These previous studies use family as the taxonomic unit as we do here, with the analysis of Ricklefs (2005) using two fewer external measurements. However, the more striking difference between our analysis and previously published works is the taxonomic sampling: Ricklefs (2005) sampled 1016 species of passerine birds, and Kennedy et al. (2020) sampled 782 species of corvid passerines. By contrast, we sampled approximately 99% of all birds (those species in families with five or more species,  $n = 9879$ ). We conclude that the relationship between position in morphospace and family species-richness is a property of certain clades (as shown in corvid families), where their unique morphologies restrict their ability to exploit widely distributed resources, ultimately limiting opportunities for allopatric speciation (Kennedy et al., 2017, 2020; Price, 2008; Ricklefs, 2005).

We predicted that increasing species richness in an area would negatively impact the genetic diversity of species, regardless of the taxonomic level. We found only limited evidence that genetic diversity was inversely correlated with the geographic co-occurrence of species when examining all birds concurrently, or that family species richness is inversely correlated with genetic diversity. We suggest

two possible explanations for this result. First is that this result accurately reflects biological reality, which would imply that the negative genetic consequences of coexistence cannot be sustained long term (i.e. over geological periods), so that the species sampled today are survivors of a process that played out in the past. That we find a stronger result when all species are analysed concurrently suggests that the degree of relatedness, and therefore functional trait similarity, is not as important in driving observed patterns. Validation of this hypothesis likely requires sampling of extinct species and/or populations via aDNA (Brüniche-Olsen et al., 2018; Kistler et al., 2015; Smith et al., 2021); however, such sampling efforts can be confounded by anthropogenic effects where the path to extinction may also be extremely rapid, such that species go extinct without detectable genetic decline (but see Pinsky et al., 2021). If the lack of correlation is indeed accurate, then this may also support the idea of that standing genetic variation is overstated in its importance for the long-term persistence of species except at extremely low values, and that simultaneous consideration of other factors such as the demographic history of species is more informative (Smith et al., 2021; Teixeira & Huber, 2021).

As there a strong prior expectation for a negative relationship between species richness and genetic diversity then we must consider whether the weak relationship identified here is a methodological artefact. We predominantly analysed mitochondrial data as these sequences are more abundant on GenBank (Table 2). These mitochondrial genes facilitate differentiation of diversity between species and regions, but may not capture the effect of selection pressures. Instead, the nuclear genes may manifest the effects of coexistence. Of the two nuclear genes examined here, RAG1 (which has a role in immune response) found a pattern consistent with coexistence impacting genetic diversity, but not MC1R (involved in pigment production), although the signal was weak. Therefore, although these results raise the intriguing possibility that mitigating negative effects of coexistence on genetic diversity may play a role in regulating species diversity, far more comprehensive sampling is required. We suggest that future efforts be gene-specific as pooled results may obscure patterns, as our results clearly indicate that spatial patterns also vary by gene (Figure S10); in genomic studies, for example, this might take a sliding-window approach. Another methodological consideration is that sequence data can be mistakenly assigned to the wrong species when uploaded to GenBank (Pentinsaari et al., 2020). Vetting 124,301 sequences is an unwieldy task, but it is worth noting that this may introduce some noise into our analyses. The final methodological consideration is that sampling at the species level may be the wrong taxonomic level, and geographical co-occurrence may drive local populations to extinction, rather than whole species. While we attempted to test for an effect within the range of individual species, there are currently no data available on GenBank that permit testing this hypothesis. We hope that with continued deposition of georeferenced data to GenBank, or even in a dedicated sequencing project, the workflow we developed for testing this hypothesis may be applied in the future.







- von Hardenberg, A., & Gonzalez-Voyer, A. (2013). Disentangling evolutionary cause-effect relationships with phylogenetic confirmatory path analysis. *Evolution*, 67, 378–387.
- Walker, T. D., & Valentine, J. W. (1984). Equilibrium models of evolutionary species diversity and the number of empty niches. *The American Naturalist*, 124, 887–899.
- Webb, C. O., Ackerly, D. D., McPeck, M. A., & Donoghue, M. J. (2002). Phylogenies and community ecology. *Annual Review of Ecology, Evolution, and Systematics*, 33, 475–505.
- Winter, D. J. (2017). rentrez: An R package for the NCBI eUtils API. *The R Journal*, 9, 520–526.
- Yamaguchi, R., Iwasa, Y., & Tachiki, Y. (2021). Recurrent speciation rates on islands decline with species number. *Proceedings of the Royal Society B: Biological Sciences*, 288, 20210255.
- Zhao, X.-F., Buckling, A., Zhang, Q.-G., & Hesse, E. (2018). Specific adaptation to strong competitors can offset the negative effects of population size reductions. *Proceedings of the Royal Society B: Biological Sciences*, 285, 201800007.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Crouch, N. M. A., & Jablonski, D. (2022). Is species richness mediated by functional and genetic divergence? A global analysis in birds. *Functional Ecology*, 00, 1–14. <https://doi.org/10.1111/1365-2435.14153>