Article (Discoveries)

The genomic landscapes of desert birds are structured by contemporary features

Authors: Kaiya Provost[1,2,3,*], Stephanie Yun Shue[4,5], Meghan Forcellati[4,6], Brian Tilston Smith[1]

[1] Department of Ornithology, American Museum of Natural History, New York, New York, USA
[2] Richard Gilder Graduate School, American Museum of Natural History, New York, New York, USA
[3] Evolution, Ecology and Organismal Biology, The Ohio State University, Columbus, Ohio, USA
[4] Bergen County Academies, Hackensack, New Jersey, USA
[5] Biological Sciences, University of California Berkeley, Berkeley, California, USA
[6] Ecology, Evolution, and Environmental Biology, Columbia University, New York, New York, USA

[*]Corresponding author: Kaiya L. Provost, provost.27@osu.edu, Department of Evolution, Ecology and Organismal Biology, The Ohio State University. 318 W. 12th Ave. 300 Aronoff Laboratory Columbus, OH 43210, USA

Abstract

Spatial models show that genetic differentiation can be explained by factors ranging from geographic distance to environmental resistance across the landscape. However, genomes exhibit a landscape of differentiation, which could indicate that multiple factors better explain divergence in different portions of the genome. We test whether the best-predictors of intraspecific differentiation vary across the genome in ten bird species that co-occur in Sonoran and Chihuahuan deserts. Using population-level genomic data, we characterized the genomic landscapes across species and modeled five predictors that represented historical and contemporary mechanisms. The extent of genomic landscapes differed across the ten species, influenced by varying levels of population structuring and admixture between deserts. General dissimilarity matrix modeling indicated that the best-fit models differed from the whole genome and partitions along the genome. The most important predictors of genetic distance were environment and contemporary demography, which each explained 25–38% of observed variation, with paleoclimate and the position of the biogeographic barrier explaining 14–16%, and distance only explaining 9%. In particular, the genome was best explained by the biogeographic barrier in regions where the genome showed high fixation between populations. Similar levels of heterogeneity were observed among species and phenotypic divergence within species. These results illustrate that the genomic landscape of differentiation was influenced by alternative spatial factors operating on different portions of the genome.

Introduction

Levels of nucleotide diversity and the degree of differentiation both vary across genomes (e.g., Ellegren et al., 2012; Li and Ralph 2019). These so-called genomic landscapes are produced by a range of variable processes including ones intrinsic to the genome (meiotic recombination, mutation) and those extrinsic (introgression, selection, and drift). Fluctuating levels of genetic diversity across the genome have been shown to be associated with recombination rate indicating that linked selection reduces variation (e.g., Thom G, Moreira LR, Batista R, Gehara M, Aleixo A, Smith BT, unpublished data, https://www.biorxiv.org/content/10.1101/2021.12.01.470789v1). Likewise, speciation genes, mutation rates, and coalescent times are all known to cause variation in differentiation across the genome (Nosil and Schluter 2011,Benzer 1961; Hodgkinson and Eyre-Walker 2011). In contrast to intrinsic processes, extrinsic processes are mediated through interactions with the adaptive and demographic factors operating across space. Despite evidence of the patterns and processes driving a heterogeneous genomic landscape (e.g., Li and Ralph 2019, Wang et al., 2020), studies examining the spatial predictors of genetic differentiation often treat genomic data as homogeneous. Clarifying the relationship between the heterogeneity of the genomic landscape and spatial predictors of differentiation will elucidate how intraspecific variation arises in the complex physical landscape.

The spatial processes attributed to population differentiation operate over historical through contemporary time scales. For example, population history is often linked to Pleistocene glacial cycles that shifted and fragmented distributions over the last 2.6 million years. An association of genome-wide structuring linked to population fragmentation can be tested under isolation-by-history (IBH), where genetic distances are modeled against paleo-climatic suitability (Vasconcellos et al., 2019; Moreira et al., 2020). There are also atemporal manifestations of historical isolation, such as isolation-by-barrier (IBB; *sensu* Mayr 1942), which posits that population differentiation is best predicted by a landscape feature, for example a mountain range or river. Over shallower evolutionary scales, non-random mating with individuals in closer

geographic proximity can cause genetic differentiation. Geographic distances alone may not be the best predictors of differentiation because adaptation to local climatic conditions causes selection to generate intraspecific differentiation across environmental gradients, which is known as isolation-by-environment (IBE; Wang and Bradburd 2014, Myers et al., 2019, Berg et al., 2015; Zamudio et al., 2016). Because local environmental conditions change rapidly, for example due to species turnover or succession (Phillips 1996, Nuvoloni et al., 2016), associations between differentiation and environment are likely more recent phenomena than historical associations. The increased availability of ecological data for many organisms, such as census data, allows for testing even shallower associations with genetic structuring across the landscape. Contemporary demographic data can be used to test isolation-by-abundance (IBA), where genetic differences are associated with abundance troughs that restrict gene flow (Barton and Hewitt 1981, Hewitt 1989, Barrowclough et al., 2005). Local population size is also known to be a strong driver of genetic structure, especially when compounded with environmental change determining local suitability (Weckworth et al., 2013). While the focus of these models is often on genetic variation, they can also be applied to phenotypic variation (e.g., Moreira et al., 2020). Phenotypic variation is often the product of many loci with little effect that are not always distinguishable from the genome itself. As such, looking directly at phenotype can help reveal whether a particular process is associated with trait variance. Examining the genomic landscape in the context of these alternative spatial models will provide evidence for how factors of varying temporal resolutions influence the peaks and valleys of differentiation. To investigate how landscape features impact genotype and phenotype, we use an exemplar community of co-distributed taxa across the Sonoran and Chihuahuan deserts of the southwestern USA and northern Mexico.

Here we characterize the genomic landscapes of birds occurring across the Sonoran and Chihuahuan deserts and test the relative effect of alternative spatial models in predicting patterns of intraspecific differentiation. To do this, we integrate population-level whole-genome resequencing, specimen-based morphometrics, and comparative sampling across ten co-distributed species that occur across the deserts. We hypothesize that the best-predictors of genetic diversity will vary across species and different partitions of the data, reflecting the multiple extrinsic factors that structure variation across the genomic landscape (Supplementary Figure 1). Alternatively, species could show homogeneous patterns either by the same spatial modeling predicting differentiation in windows across the whole genome or by species exhibiting congruent genomic landscapes shaped by the same geographic barrier. We further evaluate whether summary statistics, reflective of alternative evolutionary processes, could explain alternative spatial predictors of genomic landscapes. This comparative framework will provide resolution to the extent at which peaks and valleys of the genomic landscape correspond to historical through contemporary factors.

Results

Genomic results

We sequenced the genomes of 221 individuals across 10 focal species of passerine distributed in the Sonoran and Chihuahuan deserts. Based on the amounts of missing data, we created three datasets: a complete dataset, a dataset where up to 75% missing data was allowed, and a dataset where up to 50% missing data was allowed. We found that the three missing data partitions did not vary substantially with respect to coverage or number of SNPs. As such, here we describe the results for the complete dataset (for the 75% and 50% missing data partitions, see

Supplementary Information). We recovered sequences with a mean coverage of 2.9 per individual (range 0.4–8.8), 6–25 million reads per individual, and 5–28 million SNPs per species. Mean coverage within species ranged from 2.1–4.2 *Phainopepla nitens* the lowest coverage and *Melozone fusca* having the highest. The average missing data per species ranged from 48–64%. Across individuals, missing data ranged from 13–93% with a mean of 53% (Table 1).

We estimated recombination rates using ReLERNN (Adrion et al., 2020). Mean recombination rates for the entire genome ranged from 8.8–10.6 x 10^{-10} c/bp (where c is the probability of a crossover) across species. Correlations between species in mean recombination across chromosomes range from -0.57 to 0.61 (mean±SD 0.01±0.29). Correlations in mean recombination at the same genomic positions ranged from -0.31 to 0.36 (mean±SD -0.03±0.17).

Table 1: Chromosome-wise values for the recombination rate, F_{ST}, D_{XY}, and proportion of missing data per each species. Values given as mean±standard deviation (number of chromosomes). These are calculated by weighting all chromosome means equally; for size-weighted values see Supplementary Table 1. Note that the number of chromosomes is based off of the pseudo-chromosomes we generated, with a maximum of 36. "Rec"=population recombination rate, or rho. Values are given for the complete dataset; for the 50% and 75% values, see Supplementary Table 2.

Species	**Rec (x 10^{-10})**	**F_{ST}**	**D_{XY}**	**% Missing Sites**
Vireo bellii	9.7±1.2 (33)	0.06±0.09 (35)	0.011 ±0.005 (31)	0.64±0.79 (36)
Amphispiza bilineata	10.0±1.2 (2)	0.02±0.001 (35)	0.018 ±0.005 (20)	0.55±0.43 (36)
Campylorhynchus brunneicapillus	10.4±0.3 (31)	0.03±0.001 (34)	0.011 ±0.008 (31)	0.55±0.02 (36)
Toxostoma crissale	10.5±0.4 (31)	0.04±0.004 (34)	0.01 ±0.006 (31)	0.52±0.41 (36)
Toxostoma curvirostre	10.0±0.5 (34)	0.10±0.023 (34)	0.013 ±0.009 (32)	0.52±0.41 (36)
Auriparus flaviceps	10.2±0.7 (34)	0.05±0.006 (36)	0.015 ±0.007 (35)	0.56±0.47 (36)
Melozone fusca	10.1±0.5 (35)	0.04±0.004 (35)	0.015 ±0.01 (24)	0.51±0.47 (36)
Polioptila melanura	9.7±0.7 (29)	0.03±0.001 (34)	0.014 ±0.01 (23)	0.52±0.43 (36)
Phainopepla nitens	10.0±0.6 (30)	0.02±0.001 (34)	0.012 ±0.007 (28)	0.65±0.01 (36)
Cardinalis sinuatus	9.8±0.6 (36)	0.03±0.005 (36)	0.015 ±0.01 (26)	0.52±0.35 (36)

Lostruct outliers and F_{ST} outliers

We divided the genome into three kinds of partitions. First, we analyzed chromosomes independently. Second, we identified high F_{ST} outliers and analyzed those. Finally, we performed a multidimensional scaling (MSDS) analysis the using R package lostruct version 0.0.0.9000 (Li and Ralph 2019), which subdivided genomes into four partitions, three outliers (LS1, LS2, LS3) and one non-outlier partition (Figure 1; Supplementary Figure 2). Note that outlier groupings of the same color are not analogous across taxa. On average across all species 85.3% of labeled values were non-outliers, and ~4.88% each were LS1, LS2, and LS3.

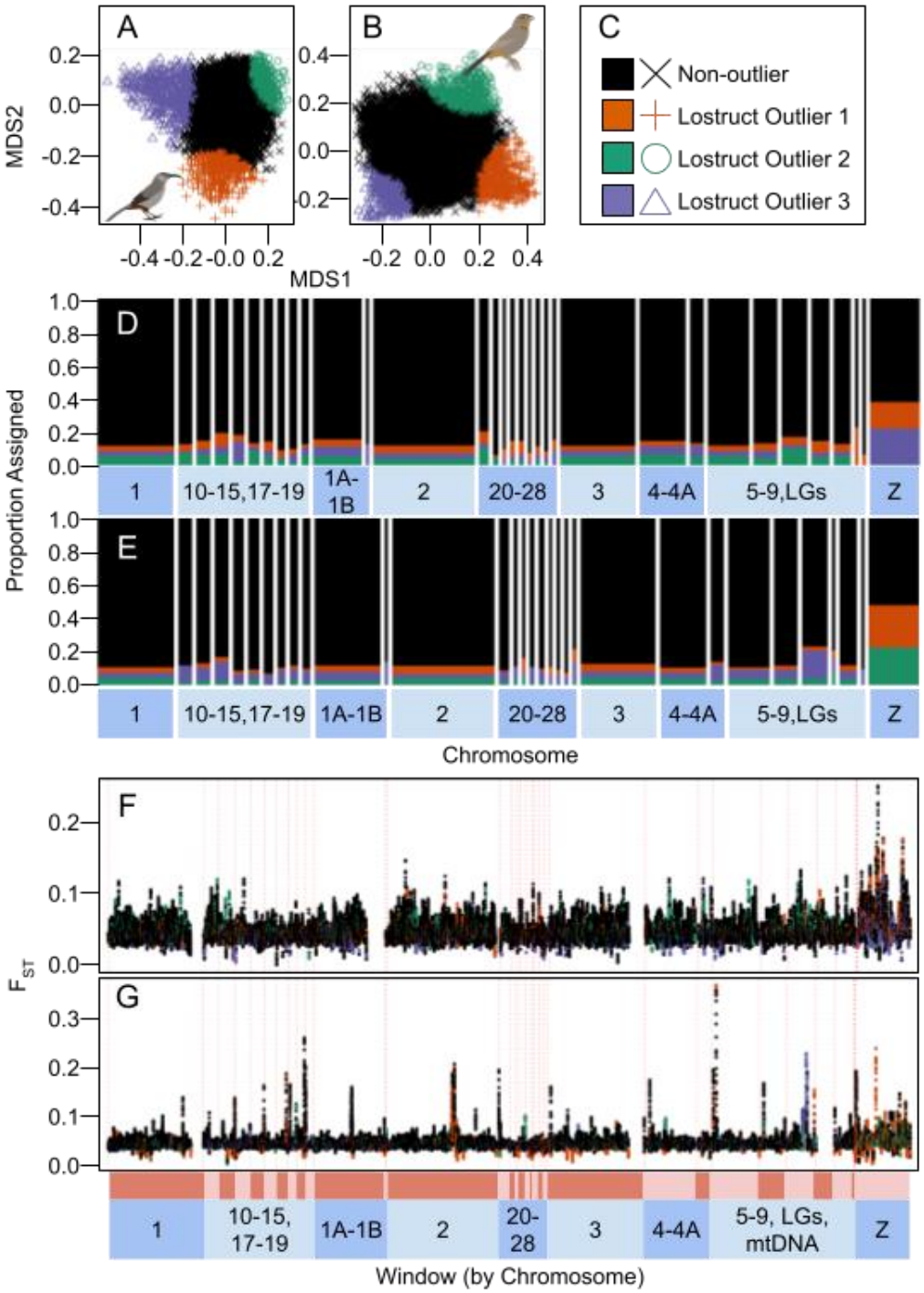

Figure 1: Lostruct partitions vary across species and across chromosomes. Two exemplar species pictured, *Toxostoma crissale* (panels A, D, F) and *Melozone fusca* (panels B, E, G). For all 10 species see Supplementary Figure 2. Panels A and B: Multidimensional scaling coordinates 1 (x-axis) vs 2 (y-axis) for each species, with outlier points highlighted in orange, green, and purple as different partitions, and non-outlier points in black. Panel C: Legend describing colors and shapes in panel C, with black X's showing non-outlier partitions, orange crosses showing lostruct outlier 1, green circles showing district outlier 2, and purple triangles showing lostruct outlier 3. Panels D and E: Proportion of chromosomes assigned to lostruct outliers and non-outliers in panels D and E. Width of bars approximately proportional to length of windows assessed in each chromosome. Panels F and G: F_{ST} values for windows across the genome, colored by lostruct partition, with windows without lostruct data in gray. Note that F_{ST} values are not on the same scale. Chromosomes separated by red lines, with legend at the top.

We calculated F_{ST} values across the genome using ANGSD's realSFS function (Meisner and Albrechtsen 2018). F_{ST} outlier analysis for our species across the datasets with complete, 75%, and 50% missing data found largely congruent results (see Supplementary Information for 75% and 50% datasets). The number of high F_{ST} outliers for the complete dataset ranged from 28–758 across species (with the total number of windows analyzed per species ranging from 100,733–113,555). The outlier lostruct partitions identified above (LS1, LS2, LS3) vary in the proportion of the F_{ST} outliers examined (for the complete dataset), ranging from 0.0%–3.4% (mean 0.2%) for peaks. Though not significant, there appears to be a trend where species with generally higher F_{ST} have more high F_{ST} outliers identified.

Population differentiation

Population differentiation across the Sonoran and Chihuahuan deserts was estimated using PCAngsd in ANGSD (Meisner and Albrechtsen 2018). Species ranged from being highly structured among deserts in four species (*T. curvirostre, V. bellii, A. flaviceps,* and *P. melanura*), showing a gradient of structuring with admixture in three (*T. crissale, M. fusca,* and *Cardinalis sinuatus*), or unstructured in the remaining taxa (*A. bilineata, C. brunneicapillus, P. nitens*; Supplementary Figure 3). F_{ST} values for the species within these three groups varied accordingly: highly structured=0.03–0.10; gradient=0.03–0.04; and unstructured=0.02–0.03. Population differentiation estimated from the chromosomal partitions were generally concordant with genome-level patterns, but smaller chromosomes and/or those with fewer SNPs showed different patterns (Figure 2, Figure 3, Supplementary Figure 4).

After estimating population differentiation, we calculated clines of population assignment across the range of each species, examining cline width and cline center. For cline-based analyses, mean cline width ranges from 6.94–15.89° longitude, where the total area encompassed by each species was ~18° longitude (with zero on the cline defined as 116.10°W longitude; Supplementary Table 3; Figure 2; Figure 3; Supplementary Figure 1). Cline width increases as chromosome size decreases ($p=1.4x10^{-6}$, adjusted $R^2=0.06$), though this varies across species (range p $7.7x10^{-7}$–0.43, range adjusted R^2 -0.01–0.51). Mean cline center location ranges from 3.58° along the cline (~112.52°W) to 12.70° along the cline (~103.4°W). We found that there were negative correlations between the degree of population structure (measured by F_{ST}; see Supplementary Information) and both mean cline width and the standard deviation of cline center locations, which is expected based on how clines are calculated. Species with higher F_{ST} between populations had narrower clines and less variation among partitions in the locations of their clines (Supplementary Figure 5).

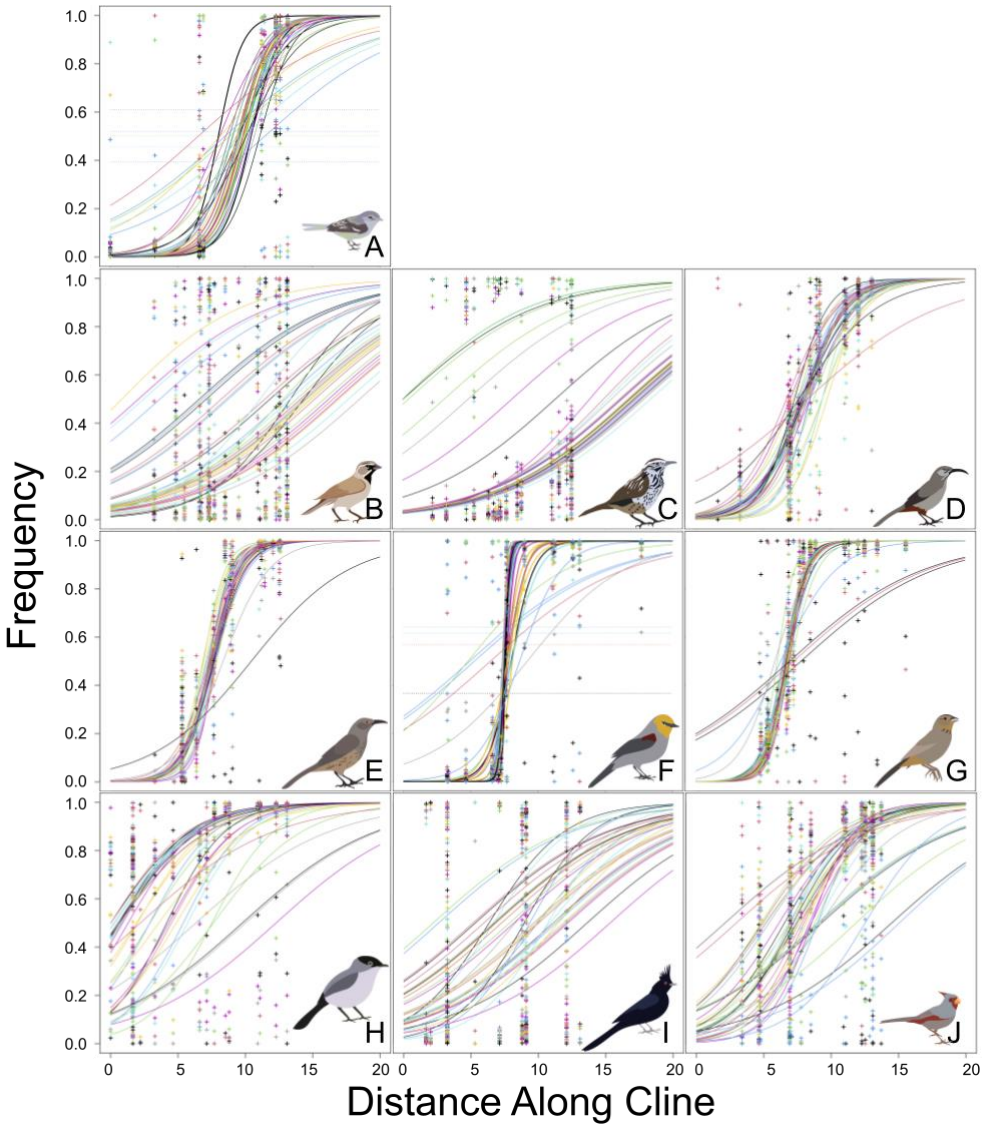

Figure 2: Cline width and center location vary across species and across chromosomes. X-axis shows distance (in degrees longitude) along the sampled area. Y-axis shows the projected cline from population assignments of 0 to 1 in each taxon (panel) and each chromosome (colored lines). Hash marks show population assignments for each individual. Species are as follows: A) *Vireo bellii*, B) *Amphispiza bilineata*, C) *Campylorhynchus brunneicapillus*, D) *Toxostoma crissale*, E) *Toxostoma curvirostre*, F) *Auriparus flaviceps*, G) *Melozone fusca*, H) *Polioptila melanura*, I) *Phainopepla nitens*, J) *Cardinalis sinuatus*.

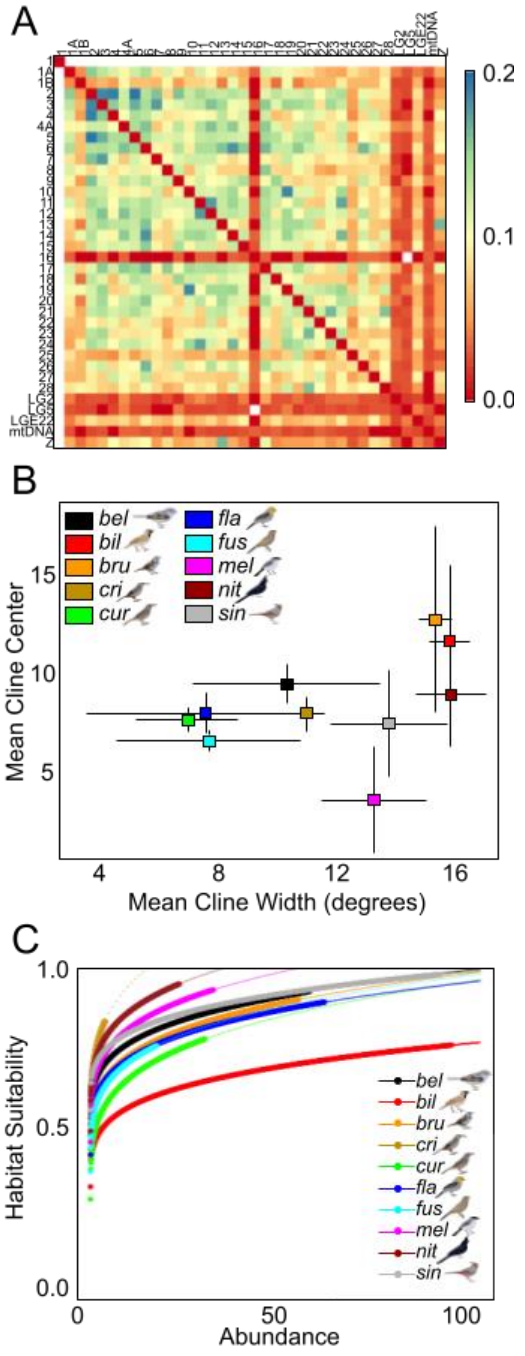

Figure 3. Species vary in their chromosomal structure, population structure, ecology, and abundance across the Sonoran and Chihuahuan deserts. A) Standard deviations of Robinson-Foulds distances across species; see Supplementary Figure 4. Warmer colors indicate lower standard deviations. Chromosomes are arranged in alphanumeric order. B) Mean cline width in degrees vs. mean cline center across chromosomes for each species; see Figure 2. Lines from each point show standard deviations. Species names are shortened for legibility ("bel"=*Vireo bellii*, "bil"=*Amphispiza bilineata*, "bru"=*Campylorhynchus brunneicapillus*, "cri"=*Toxostoma crissale*, "cur"=*Toxostoma curvirostre*, "fla"=*Auriparus flaviceps*, "fus"=*Melozone fusca*, "mel"=*Polioptila melanura*, "nit"=*Phainopepla nitens*, "sin"=*Cardinalis sinuatus*). C) Predicted abundance-habitat suitability relationships for each species; see Supplementary Figure 6, Supplementary Figure 7, Supplementary Figure 8. Colors indicate individual species. Points are large for actual abundance metrics for each species; small points show predicted suitability at non-observed higher abundances. Species names as in Part B.

Morphological variation across the Cochise Filter Barrier

Across the 10 focal species, we measured 294 individuals, including bill, wing, tail, and leg morphology. We collapsed these metrics into a principal components analysis. There were no clear, desert-specific patterns in variation across the Cochise Filter Barrier (N=234), with morphological changes ranging from subtle to significantly different. In our principal components analysis, the first three principal components (PC1, PC2, PC3) explained 74%, 12%, and 6% of the variation in morphology and corresponded approximately to overall body size, bill size/shape, and wing size/shape, respectively (Supplementary Table 4, Supplementary Table 5; Supplementary Figure 9). We found significant differences across the Cochise Filter Barrier in six species in at least one analysis (see Supplementary Information for more details). Between deserts *T. crissale* and *C. sinuatus* differed in body size and bill shape. *Vireo bellii* and *M. fusca* differed in bill shape. *Polioptila melanura* and *A. flaviceps* differed in body size. No species showed significant differences in wing shape.

Climatic suitability and abundance across the Cochise Filter Barrier

Using MaxEnt (Phillips et al., 2006), WorldClim (Hijmans et al., 2005), and other environmental variables (see Methods), we calculated ecological niche models for the present, the mid-Holocene, and the Last Glacial Maximum. During the Last Glacial Maximum, the most suitable areas for all taxa were projected to be further south than the most suitable areas during the present and mid-Holocene. Regions that are predicted to be suitable through all three periods are often reduced compared to current distributions (Supplementary Figure 8; Supplementary Figure 10). We calculated abundance for each species using the Breeding Bird Survey (Pardieck et al., 2019). Abundance was correlated with predicted climatic suitability across all taxa, with adjusted R^2 values of fit lines (log-scaled) ranging from 0.42–0.62 (Figure 3, Supplementary Figure 6, Supplementary Figure 7).

Phenotypic and genotypic datasets are idiosyncratic with respect to landscape features

We chose five metrics of landscape variation (IBA, IBB, IBD, IBE, and IBH) to evaluate against genetic and phenotypic variation within taxa. Differences in variation were attributed to each of these landscape metrics using generalized dissimilarity matrix (GDM) modeling. We evaluated models that were univariate (variation ~ landscape metric), bivariate (variation ~ IBB + landscape metric), and trivariate (variation ~ IBB + IBD + landscape metric); we focus on univariate models. Performance of the GDM models was consistent whether looking at univariate, bivariate, or trivariate data partitions (see Supplementary Information). 2,945/3,030 univariate models converged successfully with an overall 98% convergence. Of the 505 datasets tested, 30.0% selected IBE as the best factor explaining variation, 21.3% selected IBB, 18.2% selected IBA, 14.2% selected IBD, 11.5% selected IBH, and the remainder were ambiguous, with multiple models equally explaining variation. Within the ambiguous models, of which there were 23, 82.6% had IBH as one of the best models, 73.9% had IBE as one of the best models, 56.5% had IBA as one of the best models, 23.1% had IBD as one of the best models, and notably, none of them had IBB as one of the best models.

Across all of the GDMs performed, percent deviance explained by the best model was variable, ranging from 0.1% to 81.9%. The mean±SD percent deviance explained across all datasets was 16.8%±18.2%. Percent deviance explained for the whole genome was lower on average, ranging from 0.5%–6.9% (mean±SD 3.9%±2.2%). F_{ST} outliers, both high and low, tended to have higher percent deviances explained, ranging from 0.14%–69.9% (mean±SD

25.9%±22.4%). Lostruct outliers ranged from 1.0%–54.25% (mean±SD 11.0%±12.6%). Percent deviance explained had the most extreme range in morphology, from 0.3% to 81.9% (mean±SD 17.5%±20.8%). The percent deviance explained varied across taxa, with means ranging from 7.5% (*M. fusca*) to 27.9% (*P. nitens*) and standard deviations ranging from 12.1%–24.0%.

For the models examining signals across the whole genomes, three species had IBB as the most important predictor, three had IBE, two had IBH, one had IBA, and one had IBD. (Figure 4; Supplementary Figure 11). It is notable that all of the genomes identified as having IBE as the best predictor are taxa that are structured across the Cochise Filter Barrier. Chromosome length does not significantly predict any differences between models (p>0.47, n=347).

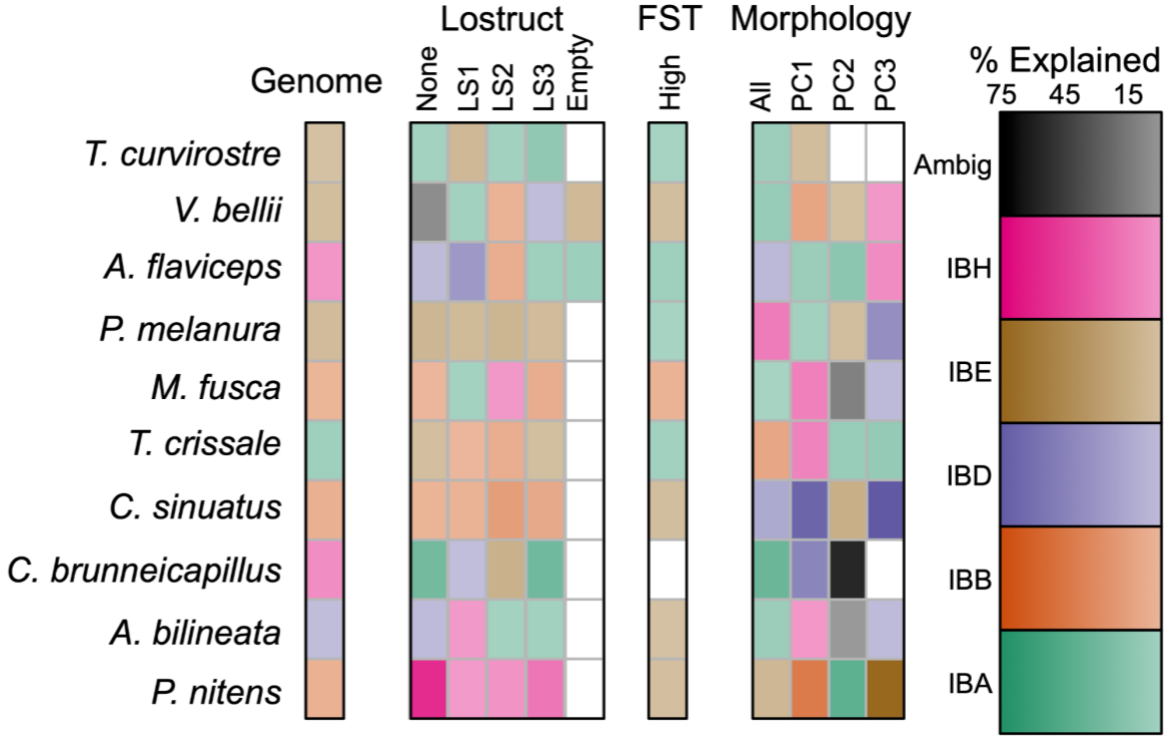

Figure 4: Generalized Dissimilarity Modeling revealed heterogeneous associations between genomic and phenotypic differentiation and alternative spatial hypothesis. Shown is the GDM model summary for each species and partition. Species are along the y-axis and arranged from most to least differentiated across the Cochise Filter Barrier. Individual partitions (genome, F_{ST} high and low outliers, morphology) are along the x-axis. Color indicates the best model. Shade of color indicates how much support the model has (with darkest shade indicating up to 75% support and lightest shade indicating 0%). White boxes have no associated data due to failure of models to converge. The alternative models were as follows: isolation by abundance (IBA), isolation by barrier (IBB), isolation by distance (IBD), isolation by environment (IBE), and isolation by history (IBH). "Ambig" is shorthand for ambiguous partitions that show a mixture of models that best explain the data. White boxes represent models that failed to converge or did not have corresponding datasets. For more partitions of data see Supplementary Figure 11.

For the lostruct partitions, the three outlier partitions (LS1, LS2, LS3) had 5/30 with IBA as the best model, 10/30 IBB, 3/30 IBD, 9/30 IBE, and 3/30 IBH. Most species showed at least some overlap in which model best explained partitions: for example, *A. bilineata*, *T. crissale*, and *C. sinuatus* all have at least two lostruct partitions best explained by IBB. For the non-outlier partitions (LS0, and the "empty" partition for *V. bellii* and *A. flaviceps*), these best model chosen is the same as the best model explaining whole-genome variation in four species (*V. bellii*, *P. melanura*, *C. sinuatus*, *M. fusca*) and that of one of the outlier partitions in all but two species (*V.*

bellii, *A. bilineata*). Notably, for *P. melanura* IBE explains all three outlier partitions, the genome, and the non-outlier lostruct partitions. Likewise, for *C. sinuatus*, all of these are explained by IBB.

All ten species had high-F_{ST} partitions identified (see Supplementary Information for 75% and 50%) across the complete, 75%, and 50% datasets. The genome matched at least one of the high or low partitions in four taxa: *Vireo bellii*, *A. bilineata*, *Toxostoma crissale*, and *Melozone fusca*. With respect to significance, none of the F_{ST} outlier partitions were significantly different (but see Supplementary Information).

There was little congruence across the best landscape predictor of morphological data within species. Overall morphological differentiation had the same explanatory variables as PC3 for *P. nitens* (IBE), and as PC1 for *C. sinuatus* (IBD). Additionally, some individual PCs did match each other: IBD best explained PC1 and PC3 in *C. brunneicapillus*, IBA best explained PC2 and PC3 in *T. crissale* and PC1 and PC2 in *A. flaviceps*, and for *T. curvirostre*, PC2 and PC3 both showed ambiguous results. Neither overall morphology nor the PCs were significantly different than expected in the univariate dataset (though some were in the bivariate and trivariate datasets; see Supplementary Information).

Like overall variation, PC1 (body size) showed an even distribution between all models across the 10 species (i.e., 20% each IBA, IBB, IBD, IBE, and IBH). PC2 (bill shape) was best explained in 30% of species by IBA, 20% by IBE, 10% by IBD, and 40% of the species showed ambiguous results. Lastly, PC3 (wing shape) was best explained in 40% of species by IBD, 20% each by IBA and IBE, 10% IBD, and 10% of species had ambiguous results.

Data characteristics of best-fit models

We looked at whether differences in summary statistics could explain our univariate models (IBA, IBB, IBD, IBE, IBH) across taxa (Supplementary Figure 12; Supplementary Figure 13; Supplementary Figure 14). The summary statistics we examined were recombination rate, missing data, F_{ST}, D_{XY} calculated using ngsTools (Fumagalli et al., 2014), and the length of the chromosome. The clearest pattern was that datasets with ambiguous results among models had more missing data than all others except IBH models ($p<0.0001$). IBH results also tended to have more missing data than most other models ($p<0.02$), but we found that this relationship was not significant when we excluded *P. nitens*, which had both the largest proportion of models explained by IBH and a high proportion of missing data ($p>0.70$). F_{ST} was significant overall ($p<0.04$), with IBB models having significantly higher F_{ST} than IBH models. This relationship was no longer significant in bivariate or trivariate models because IBB was not present (see Supplementary Information). D_{XY} was also significant overall ($p<0.05$), but Tukey's honestly significant difference tests showed that none of the individual comparisons were significant ($p>0.06$). Recombination rate and chromosome length were not significant for univariate models ($p>0.07$) though recombination rate was significant for bivariate and trivariate models (see Supplementary Information).

Landscape predictors are not influenced by habitat suitability

From the ENMs, we calculated habitat suitability for each species across the deserts. Species with more variable suitability across the contact zone have a higher proportion of IBH as the best model (adjusted-$R^2=0.54$, $n=10$, $p<0.01$). As *P. nitens* has both the highest proportion of IBH and the highest variance in suitability, we removed this species in case it was acting as an outlier. After removing this species, the relationship was only nearly significant, but strong

(adjusted-R^2=0.28, n=9, p<0.09). Evaluating this relationship with ANOVA tests finds the same results, where no comparisons are significant without *P. nitens*.

Significance evaluation of hypotheses of evolution across the Cochise Filter Barrier

Species differ more than expected with respect to what spatial models best explain their genotypes and phenotypes. Best-predictors vary across individual species (χ^2=284.0, p~0.0, df=54, simulated p<0.0005), individual partitions of genotype and phenotype differed (χ^2=685.6, p~0.0, df=324, simulated p<0.0005), and with respect to phylogeographic structure across the Cochise Filter Barrier (χ^2=62.9, p<6.5x10^{-9}, df=12, simulated p<0.0005).

Discussion

We found that the best-fit spatial model differed across partitions at multiple scales. Our taxa, which varied in levels of genomic diversity, showed evidence that different spatial processes (reflecting historical through contemporary phenomena) had distinct impacts on the genome compared to targeted subsets of the genome. Similar patterns of heterogeneity were observed among species and with their phenotypic datasets. The disparity in predictors of intraspecific differentiation among the whole genome versus windows and between windows extends the view that evolutionary inferences are dependent on which portions of the genome are examined in a spatial framework. The heterogeneity in model fit across partitions was consistent with the expectation that various evolutionary processes contribute to the peaks and valleys of the genomic landscape. By applying this framework across an assemblage of birds that evolved across a common, dynamic region we showed that at the community-scale, predictors of genomic structure remain idiosyncratic, which may reflect taxa at different stages of the evolutionary histories and responses to the biogeographic barrier.

Extrinsic drivers of the genomic landscape

Our modeling showed that environmental distance was often a strong predictor of levels of intraspecific differentiation, but this pattern was species- and partition-dependent. Genome-wide patterns of differentiation across the Cochise Filter Barrier are partially shaped by environmental adaptation as observed in non-avian taxa distributed across the barrier (Myers et al., 2019). Environmental adaptation is often recovered in taxa who respond to environmental gradients via altered phenotypes (Branch et al., 2017, Dubec-Messier et al., 2018), genotypes (Berg et al., 2015, Manthey and Moyle 2015), or both (Ribeiro et al., 2019). However, our analyses show there was considerable variation among individual regions in the genome, indicating a more nuanced pattern. The species-specific results we found suggests that individual taxa had unique responses to shared aspects of the landscape. Although the focal taxa are co-distributed, we showed how environmental suitability, their general morphologies, and abundances across space varied among species, which may help explain why best-fit models differed. As such, these species-specific factors may explain isolation-by-environment was the best explanatory variable for many, but not all, of the species we investigated.

Individual partitions of the genome also varied with respect to how much environmental variation played a role. At one extreme, environmental variation appears to have little impact on the sex chromosomes. The Z chromosome often showed the barrier (i.e., IBB) as being the most important factor, even in unstructured species such as *Amphispiza bilineata* and *Campylorhynchus brunneicapillus*, perhaps because the locus evolves faster than sites under selection for adaptation to local environmental conditions. Sex chromosomes are known to diverge faster than autosomes

due to their differences in effective population size (Mank et al., 2010), importance in sexual selection (Kirkpatrick 2017), and the presence of speciation genes (Sæther et al., 2007). Given the lack of evidence for environmental variation predicting spatial genetic differentiation on the Z chromosome, this would suggest that any speciation genes present in these taxa may not be involved in adaptation to the environment.

Environment was the most important driver for species with genetic structure. The most intuitive explanation for this was that population structuring in these taxa was facilitated by natural selection to different environments. There was some evidence that this could have happened across other taxa that occur across the Cochise Filter Barrier, as IBE was the best predictor of genome-wide divergence in a community of snakes distributed across the barrier (Myers et al., 2019). However, we must stress that while this explanation was the most intuitive and aligns with predictions, there are numerous processes that can produce IBE (Wang and Bradburd 2014), and it is possible that divergence led to adaptation to these environments secondarily, rather than the reverse, or the patterns are being influenced by some unknown factors that we did not quantify. Nevertheless, at present our results are consistent with the importance of ecologically mediated population differentiation, or isolation-by-environment, in structuring communities across the deserts of North America.

Contemporary versus historical predictors of genomic differentiation

Our finding that the best-fit models varied across species was consistent with the expectations that species idiosyncratically respond, over a range of time scales, to the Cochise Filter Barrier. The spatial patterns we examined vary temporally, with Pleistocene environmental changes being a historical process, while geographic distances, abundances, and environmental variation reflecting more contemporary processes. Historical signatures of Pleistocene isolation are commonly recovered patterns for the Cochise Filter Barrier (Provost et al., 2021) and other communities (Shafer et al., 2010; Ralson et al., 2021), but our data showed that isolation in glacial refugia often did not best explain genome-wide differentiation. This could be due to erosion of historical signals as the Cochise Filter Barrier filters taxa and changes contemporary patterns of gene flow. Alternatively, our proxy for IBH (resistance over projected Pleistocene habitat suitability) may be a poor model for actual historical isolation. For example, paleoenvironmental gradients may no longer be as readily detectable. The presence of the barrier alone was a better predictor despite being atemporal.

In contrast, current environments best explain three genomes and the majority of partitions for five species (with abundances and geographic distances playing a lesser role), suggesting that phenomena operating on more recent timescales influenced genetic and morphological variation across the landscape. If some of the taxa herein are going through incipient speciation, then these contemporary factors should be most potent. Our identification of species abundances as a relatively important predictor of genetic divergence aligns well with landscape genetic studies that use proxies for the effects of contemporary phenomenon and ecological factors on genetic variation (Burney and Brumfield 2009, Paz et al 2015). For example, urbanization, which fragments and reduces population sizes, is well known to impact rates of gene flow and drift, acting as a strong barrier of gene flow since the 20th century (Miles et al., 2019). Our use of available abundance data across large spatial scales shows a more direct relationship between varying abundances across the landscape with levels of differentiation. Further, while both historical and contemporary processes are influencing taxa across this biogeographic barrier, contemporary patterns are seemingly more influential.

Relationship between best-models and window summary-stats

In contrast to the extrinsic drivers of the genomic landscape that we have focused on here, there were no clear associations between partition characteristics and support for a particular model. For example, we found no significant differences in any species between recombination rate across chromosomes and which spatial models were most important on that chromosome. At the phylogeographic-scale, low recombination regions of the genome have been shown to be more likely to reflect population structure (Manthey et al., 2021) and the species tree topology (Thom G, Moreira LR, Batista R, Gehara M, Aleixo A, Smith BT, unpublished data, https://www.biorxiv.org/content/10.1101/2021.12.01.470789v1). The avian recombination rate landscape is thought to be conserved across taxa, even though exact genomic locations of divergence across taxa are not (Singhal et al., 2015, Turbek et al., 2021), with our ten focal species ranging in divergence time from ~75 thousand to ~12 million years between taxa (Harris et al., 2018; Kumar et al., 2017; Barker et al., 2015; Mason and Burns 2013; Price et al., 2014; Pasquet et al., 2014; Hooper and Price 2017; Mitchell et al., 2016; Gibb et al., 2015). Correlations in recombination rates at the same genomic position in these species are greater than 0.37 across chromosomes and always positive (Turbek et al., 2021). The ten desert birds we investigated, in contrast, have estimated divergence times ranging from ~10 to ~60 million years between taxa (Kumar et al., 2017; Barker et al., 2015; Mason and Burns 2013), with correlations in recombination rates at the same genomic position that were often smaller in magnitude and negative. This could reflect a real pattern, where the recombination landscapes are only conserved within more closely related species; our closest taxa, the two non-sister *Toxostoma*, do have the highest correlation in recombination rates across windows and are in the top 25% of the distribution in correlations. However, the differences found could have been caused by coverage depth, differences in the recombination rate estimators used, or missing data allowance. In addition, genetic partitions with higher F_{ST} were more likely to show isolation-by-barrier as the best model. These two metrics should be correlated; the former quantifies the degree of differentiation across the Cochise Filter Barrier, and the latter assigns individuals to their respective sides of the Cochise Filter Barrier. In species where there was differentiation, these two measures should describe the same phenomenon. This likely reflects the gradient in differentiation across species in the community. Given the wide variation across taxa, future work must be done to clarify the relationship between genomic architecture and evolutionary signal at multiple phylogenetic scales.

We explored the signal in our data by examining multiple ways of partitioning genomic windows, using different thresholds of missing data, and evaluating how data attributes influenced model support. We found that genetic partitions with more missing data were more likely to have ambiguous results. Genetic summary methods like PCA are impacted by missing data, particularly when they are imputed, which can cause individuals with disproportionately high levels of missing data to appear like they are admixed between populations (Yi and Latch 2021). It is likely that the reverse is true, where individuals with disproportionately low levels of missing data should fall out as their own populations more readily. For example, in some of our species (namely *Vireo bellii*, *Auriparus flaviceps*, *Polioptila melanura*) the individuals with highest missing data clustered as their own population before detecting any other spatial patterning. We ameliorated this by dropping individuals with too much missing data in some of our datasets. Overall, we did not find qualitative differences in population assignments, but it did generally inflate our fixation values and deflate our genetic diversity values. This is sensible, as reducing the number of

individuals should both increase the likelihood of fixation due to sampling error as well as decrease the overall amount of nucleotide diversity.

Morphological versus genetic associations

We found that in most taxa, genotypic and phenotypic variation within species, and even different aspects of morphological phenotype within species, were not associated with the same landscape factors. Phenotypes were better explained by abundance, whereas genotypes were better explained by the contemporary environment. Discordance between genetic and phenotypic predictors of spatial variation have been observed in other systems, where phenotypic variation was better explained by the environment (Moreira et al., 2020). This discordance could be due to polygenic traits, where genotype-phenotype associations may be mediated by multiple loci of small effect working in concert, either by changing protein structure or regulation (Yusuf et al., 2020, Knief et al., 2017, Duntsch et al., 2020, Aguillon et al., 2021). However, for some phenotypes like plumage color, single genes of large effect have been implicated which should strengthen correlations between genotype and phenotype, at least for those loci (Sin et al., 2020; Toews et al., 2016). For desert birds in particular, phenotypic variation in metabolism (as well as in microbiomes) has been linked to genes that vary with the environment (Ribero et al., 2019). In our study, as with genetic differentiation, the extent of phenotypic structuring varied across species, with bill and body size being significantly different between deserts in a few taxa, but somewhat surprisingly, environmental variation did not usually explain morphological differences. For example, adaptations in bill morphology are frequently observed, such as in Song Sparrows on the Channel Islands that have higher bill surface area in hotter climates (Gamboa et al., 2021). The lack of a tight correlation between environment and phenotype were likely reflective of the shallowness of the evolutionary divergences and the subtlety of the environmental gradient across deserts. The two *Toxostoma* species in our study have previously shown contrasting patterns with respect to climate on beak morphology: *T. crissale* has larger bills in drier habitats, which may aid in cooling while conserving water, while *T. curvirostre* showed a pattern contrary to thermoregulatory predictions with larger bills in cooler climates (Probst et al., 2021), suggesting even in closely related species climate may not have the same role on morphological variation. Even though phenotypic data partitions often did not have the same explanatory factor with respect to the general dissimilarity modeling, there was a correlation between population structure in the genome (and chromosomes to a lesser extent) and phenotypic variation across these ten birds, in that taxa lacking morphological change also lack genetic variation overall.

Conclusion

By quantifying patterns in genotypic and phenotypic variation in communities distributed across a biogeographic barrier, we found that multiple co-occurring processes occur that impact variation within taxa. Although we found that isolation across an environmental gradient was among the most important associations in predicting genetic and phenotypic variation, the best-fit model varied across species and data partitions to reflect these multiple processes. These findings underscore the importance of accounting for heterogeneity in the genome, phenome, and diversification mechanisms acting across time and space to have the most comprehensive picture of spatial structuring in species. This will allow for an assessment of whether best-fit models that are proxies for neutral and adaptive processes are consistent with partitions that are evolving under the same conditions. Without a holistic understanding at each of these levels of organization, as well as the addition of future work that concurrently estimates selection at the organismal and the

nucleotide levels, the actual mechanisms that shape communities will remain obscured. Further, while we did not find consistent predictors of phenotypic divergence, it is still an open question whether other measures of phenotypic variation (e.g., behavioral) may better track divergence, or phenotypic divergence does not follow a deterministic pattern along weak environmental gradients. Overall, this work displays the necessity of integrating spatial predictors of population divergence, differentiation across the genomic landscape, and phenotypic variation in understanding the multiple different mechanisms that have produced the population histories we see across contemporary communities of birds in North America.

Methods and Materials

Study system

The Sonoran and Chihuahuan deserts contain environmental and landscape variation that make them suitable for testing if any of the five discussed spatial models (IBA, IBB, IBD, IBE, and IBH) structure intraspecific variation in taxa. Across the two deserts and the transition zone between them, there is variation in precipitation, elevation, temperature, and vegetation that could result in local adaptation and isolation-by-environment. (Shreve, 1942; Reynolds et al., 2004). Pleistocene glacial cycles repeatedly separated and connected, such that some taxa experienced dramatic range shifts (Zink 2014, Smith et al., 2011), which could have isolated taxa in each desert. Further, there is a well-studied biogeographic barrier separating the deserts, the Cochise Filter Barrier, which is an environmental disjunction that demarcates the transition between the Sonoran and Chihuahuan deserts of southwestern USA and northern Mexico. The barrier is thought to have begun forming during the Oligo-Miocene and completed during the Plio-Pleistocene (Morafka, 1977, Van Devender, 1990; Van Devender et al., 1984, Holmgren et al., 2007, Spencer, 1996) and has formed a community ranging from highly differentiated taxa to unstructured populations (Provost et al., 2021). Demographic troughs caused by spatially varying population abundances could impact the frequency of gene flow across the landscape and the degree of genetic connectivity across the deserts.

Genetic sequencing and genome processing

We performed whole-genome-resequencing across 10 species of birds from the Sonoran and Chihuahuan deserts, obtaining genetic samples from new expeditions and loans from natural history museums (*Cardinalis sinuatus*; *Toxostoma crissale, Toxostoma curvirostre*; *Amphispiza bilineata, Melozone fusca*; *Polioptila melanura; Phainopepla nitens*; *Auriparus flaviceps*; *Campylorhynchus brunneicapillus*; *Vireo bellii*; Supplementary Table 6; Supplementary Figure 15). These species reflect different songbird morphotypes and ecologies in the deserts (e.g., large- to small-bodied, insectivorous to granivorous, migratory to resident). Three of these species (*V. bellii, T. curvirostre, M. fusca*) have shown evidence of structure across the Cochise Filter Barrier, while an additional three (*P. melanura, A. flaviceps, C. brunneicapillus*) have shown evidence of no structure (Zink et al., 2001; Rojas-Soto et al., 2007; Teutimez, 2012; Klicka et al., 2016, Smith et al., 2018).

Using 221 individuals across our 10 focal species, we sequenced 8–14 individuals in both the Sonoran and Chihuahuan deserts per species for a total of 18–25 samples per species. Library preparation and sequencing was performed by RAPiD Genomics (Gainesville, FL). We mapped raw reads of each species to their phylogenetic closest available reference genomes (Supplementary Table 7): notably, *A. bilineata* and *M. fusca* were mapped to the same genome, as

were *C. brunneicapillus, T. crissale, T. curvirostre, P. melanura,* and *P. nitens* (see Supplementary Information). Before mapping, we created pseudo-chromosomal assemblies of these genomes using Satsuma version 3.1.0 (Grabherr et al., 2010) by aligning to the *Taeniopygia guttata* genome (GCF_000151805.1), retaining pseudo-chromosomes with the prefix "PseudoNC". Hereafter, pseudo-chromosomes will be referred to as chromosomes.

We filtered our sequences with FastQ Screen version 0.14.0 (Wingett et al., 2018) to remove contamination by filtering out reads that mapped to PhiX and the following genomes: *Homo sapiens, Escherichia coli, Enterobacteriophage lambda,* and *Rhodobacter sphaeroides*. From our raw reads, we used a pipeline that produced genotype likelihoods using ANGSD version 0.929 (Korneliussen et al., 2014). We converted cleaned FastQ files to BAM using bwa version 0.7.15 (Li and Durbin 2009, Li and Durbin 2010) and picard version 2.18.7-SNAPSHOT from the GATK pipeline (McKenna et al., 2010, DePristo et al., 2011, Van der Auwera et al., 2013). Next, we prepared the BAM files to be used in the ANGSD pipeline using samtools version 1.9-37 (Li et al., 2009; Li 2011), bamUtil version 1.0.14 (Jun et al., 2015), and GATK version 3.8-1-0 (McKenna et al., 2010). This pipeline creates genotype likelihoods to account for uncertainty for low-coverage sequences.

We investigated the impact of missing data on our analyses using three thresholds for retaining sites: a complete dataset, in which all individuals were retained irrespective of missing data; a 75% dataset, in which individuals were only retained if they had less than 75% missing sites; and a 50% dataset, in which individuals were only retained if they had less than 50% missing sites. These different datasets were used for a suite of downstream analyses to assess the sensitivity of the results to individuals with missing data.

Evaluating population structure across the Cochise Filter Barrier

We characterized the degree of population structure across the whole genome and in individual chromosomes across the Cochise Filter Barrier in our focal species. First, using PCAngsd in ANGSD (Meisner and Albrechtsen 2018), which assigns individuals to K clusters and estimates admixture proportions for each individual. To evaluate whether there was structure across the Cochise Filter Barrier, we selected K=2 (though we visualized K values from two to three). We performed this for the complete, 75%, and 50% missing data datasets, but found that these values were largely congruent across the datasets, and so we only use the complete dataset for describing population structure (Supplementary Figure 16, Supplementary Figure 17, Supplementary Figure 18). Second, we plotted population assignment changes over space using a cline analysis via the hzar version 0.2-5 R package (Derryberry et al., 2014) and custom scripts (modified from Burbrink et al., 2021). Analyses were conducted in R version 3.6.1 (R Core Team 2019). We did this to quantitatively evaluate the differences in population structure across chromosomes and in the genome more broadly. We thus were able to calculate the location and width of clines for the entire genome and each chromosome.

Complementing our genome-wide analyses, we ran a local principal components analysis along the genome on the complete dataset using the R package lostruct version 0.0.0.9000 (Li and Ralph 2019). Different chromosomes showed different relationships between individuals (see Supplementary Information). Because of this, we wanted to cluster regions of the genome together that showed similar relationships between individuals in case specific evolutionary processes were causing this pattern. The lostruct method performs principal component analysis on individual windows of the genome, then uses multidimensional scaling (MSDS) to summarize how similar the windows' principal component analyses are when dividing the genome. We extracted three

subsets of outliers for each species, which we designated LS1, LS2, and LS3, and compared it to the remainder of the genome, representing non-outliers.

Genomic summary statistics

We characterized genetic variation across each species' genome and partitions of the genome by calculating a suite of summary statistics and metrics. To quantify genetic differentiation within each species, we calculated pairwise genetic distances between individuals from VCF files using the bitwise.dist function in poppr R package version 2.9.2 (Kamvar et al., 2014; Kamvar et al., 2015), which served as the genetic distance matrices for our generalized dissimilarity matrix models (see below). The function bitwise.dist calculates the Hamming distance of the DNA (i.e., number of differences between two strings). We scaled this distance such that missing data was assumed to match sites without missing data, but final distances were scaled such that comparisons with more missing data would have inflated distances. Neighbor-joining trees were calculated from these matrices to contrast genealogies across the genome. Genealogies across the genome were visualized by calculating pairwise and normalized Robinson-Foulds (RF) distances between all pairs of trees per species (Robinson and Foulds 1981). Recombination rates (in crossovers per base pair, c/bp) across the genome were estimated using the program ReLERNN (Adrion et al., 2020). This program combines simulation with a recurrent neural network to estimate the recombination rate on each chromosome in 100,000 bp windows. We also performed a sliding window D_{XY} analysis using the calcDxy R script included with ngsTools version 1.0.2 (Fumagalli et al., 2014), which gives site-wise DXY values, and then averaged across windows. Windows were overlapping with a size of 100,000 base pairs and offset by 10,000 base pairs. Missing data were calculated using vcftools (Danecek et al., 2011). This was calculated per window, per chromosome, per genome, per site, and per individual.

Using ANGSD's realSFS function, we performed a sliding window F_{ST} analysis by converting SAF output from ANGSD to a site frequency spectrum for both desert populations in each species. Detailed settings can be found in the supplementary information. We performed F_{ST} outlier analysis for our species using the calculated F_{ST} values. Z-scores for F_{ST} for each species were calculated using the formula $ZF_{ST}=(\text{observed}F_{ST}-\text{mean}F_{ST})/SDF_{ST}$. We split the genome into two different partitions based on these z-scores: F_{ST} peaks, for values of F_{ST} greater than five standard deviations above the mean (z-score>5) and F_{ST} troughs for values of F_{ST} greater than five standard deviations below the mean (z-score<-5). We only report the F_{ST} peaks in the main manuscript: for F_{ST} troughs, see the supplementary information. We performed this outlier detection for the complete, 75%, and 50% missing datasets.

Morphological data

We quantified morphological variation in our 10 focal species to assess which of the spatial models best explain morphological variation across the landscape (see *Generalized Dissimilarity Matrix Models*). We measured 366 specimens (19–59 per species), excluding known females and known juveniles to account for any variation attributed to sex and age. Of those, 29 were also present in the genomic dataset, with 0–8 individuals per species.

We generated seven raw plus seven compound morphological measurements, which we designated as proxies for thermoregulation and dispersal, respectively (see Supplementary Information). We reduced the dimensionality of the 14 morphological measurements using a principal components analysis (PCA). We then calculated four distance matrices between individuals: one Euclidean distance matrix for all morphological variables, where we calculated

the euclidean distance between individuals among all raw and calculated measurements; and three euclidean distance matrices for the first three principal components, PC1, PC2, and PC3. We assessed whether there were differences in morphological PCA space between the Sonoran and Chihuahuan Desert populations in each species using DABEST tests in the dabestr package version 0.3.0 (Figure 5; Supplementary Figure 19; Supplementary Figure 20; Ho et al., 2019). Note that this method does not give explicit significance values, instead it shows whether expected confidence intervals overlap zero (i.e., no difference between deserts) or not.

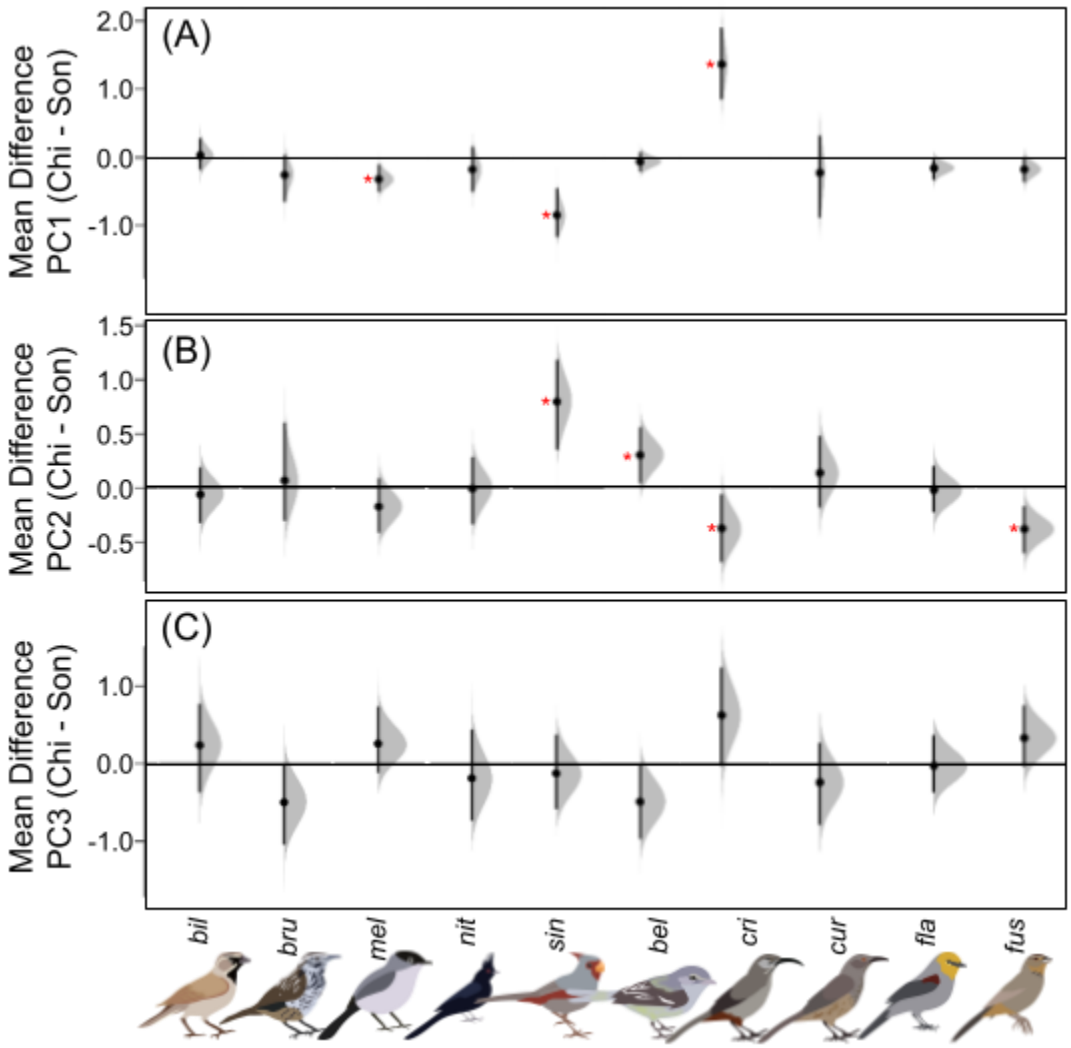

Figure 5: Distribution of unpaired mean differences between Sonoran and Chihuahuan desert individuals for each species from DABEST analysis for morphological PC1 (A), PC2 (B), and PC3 (C). Black horizontal line is at zero, black points and vertical lines show mean and confidence intervals for each distribution in gray. Comparisons that do not cross the zero line are considered significant in DABEST tests, indicated with red asterisk. On the X axis are each species with images (scale does not reflect size differences) with species names are shortened for legibility ("bel"=*Vireo bellii*, "bil"=*Amphispiza bilineata*, "bru"=*Campylorhynchus brunneicapillus*, "cri"=*Toxostoma crissale*, "cur"=*Toxostoma curvirostre*, "fla"=*Auriparus flaviceps*, "fus"=*Melozone fusca*, "mel"=*Polioptila melanura*, "nit"=*Phainopepla nitens*, "sin"=*Cardinalis sinuatus*).

Isolation across the landscape at different temporal resolutions

We calculated IBD matrices by calculating the euclidean geographic distance between the latitude/longitude pair of each specimen in R. We used the WGS84 projection for all data. These variables were somewhat correlated with one another, though less so after accounting for geographic distance (Supplementary Figure 21).

To produce data for the IBH model, we calculated environmental resistances in the Last Glacial Maximum (LGM; ~21,000 years ago) for each species. To do this, we created ecological niche models (ENMs) using 19 layers representing contemporary climate (WorldClim; Hijmans et al., 2005) at a resolution of 2.5 arcminutes. We used MaxEnt (Phillips et al., 2006), with ENMeval version 0.3.1 as a wrapper function for model selection (Muscarella et al., 2014). ENMeval optimizes MaxEnt models based on different sets of feature classes and regularization values (see Supplementary Information). The contemporary ENMs (see IBE section below) were then backprojected to the LGM using WorldClim paleoclimate data (Hijmans et al., 2005). We also backprojected to the Mid-Holocene, but contemporary and Mid-Holocene ENMs were highly correlated, so we excluded the Mid-Holocene values from downstream analyses. We then scaled the LGM suitability values to range between 0–1 and calculated resistances across the environment using the least cost path distance method in ResistanceGA version 4.0–14 (Peterman et al., 2014, Peterman 2018). Regions of high resistance are predicted to reflect poor habitat and be costly to traverse through. The ENMs were thresholded to equal sensitivity-specificity values for visualization (Supplementary Figure 22).

We approximated IBB by assigning individuals based on their location relative to the Cochise Filter Barrier (see Supplementary Information). For proximity to the Cochise Filter Barrier, we assigned individuals to either Sonoran or Chihuahuan populations either based on the results of the K=2 clustering analysis, if there was structure across longitudes, or according to a cutoff of longitude if there was no structure. We chose 108 °W longitude as our cut off—individuals west of this point were deemed Sonoran, and individuals east of this point were deemed Chihuahuan (but see Provost et al., 2021). In some cases, species with genetic breaks had some uncertainty due to unsampled areas or admixed individuals—we labeled these individuals as being unclear with respect to their desert assignment. Georeferencing on some morphological specimens was poor, but all except two specimens (see Results) were identified at least to county level if not to a specific locality. When localities were given, we georeferenced the specimens to the nearest latitude/longitude. Otherwise, we assigned individuals to the centroid of their state or county.

We independently tested IBE by using two datasets: contemporary environmental distance and resistance. For the environmental distances, we used the 19 WorldClim bioclimatic layers (see IBH section). For the latitude/longitude location of each specimen used in both the morphological and genomic analysis, we extracted the values on those WorldClim layers and then calculated the euclidean distances in environmental space between specimens. This gave us an estimate of how different the environments were at each specimen's locality. For the environmental resistances, we created ENMs using the WorldClim layers, then added layers for soil properties, distance to water, terrain features, and vegetation, and occurrence data for the focal species (see Supplementary Information). We then calculated resistances and thresholded as described above.

To assess IBA, which had a temporal scale of the last 50 years, we obtained abundance information from the Breeding Bird Survey (Pardieck et al., 2019). This dataset consists of replicated transects where individual birds are counted across the whole of the United States. The methodology for counting is standardized and covers multiple decades of observations, with our dataset comprising data from 1966–2018. We downloaded raw data for all points, then subsetted

our data to our ten focal species. We averaged the number of individuals across years (though some points only had a single year). We then interpolated across points using inverse distance weighted interpolation in the spatstat version 2.1-0 package in R (idp=5). The interpolations were converted to rasters with extents and resolutions matching those of the ENMs. We then calculated resistances such that regions of high abundance had low resistance, to generate an abundance distance matrix between individuals.

Generalized dissimilarity matrix models

We assessed the relative effect of alternative spatial models on intraspecific variation in our focal species by building generalized dissimilarity matrix models (GDMs). As spatial layers representing our five models, we calculated geographic distances, abundance resistances, environmental distance and resistance, separation by barrier, and paleoenvironmental resistance between all individuals in each species. The models represent different temporal resolutions, with IBH spanning millions to tens of thousands of years ago, IBD spanning thousands to tens of years, IBE spanning hundreds to tens of years, IBA spanning tens to single years, and IBB describing the present-day configuration of the barrier. These predictors served as the input parameters for our GDMs and will be discussed in detail below. With our numerous response matrices (four morphological matrices, three genome matrices for each missing data cutoff, 35 matrices for chromosomes, five matrices for the lostruct partitions, and six matrices for the F_{ST} outliers with missing data cutoffs) and our six predictor matrices (with two for IBE: environmental distance, environmental resistance), we generated generalized dissimilarity matrix models using the gdm package version 1.3.11 in R (Manion et al., 2018). We tested which of IBA, IBB, IBD, IBE, IBH, or a combination best explained the variation in the response matrix (see below). Not all species had all chromosomes sequenced, and not all models converged: we have omitted those data. For each of the 45 response matrices per species, we built a univariate model where the genomic/chromosomal variable was predicted solely by one of the six predictor matrices. We also built models with combinations of two (bivariate) or three variables (trivariate), which we present in the Supplementary Information. Further, we present the GDM results for the chromosomes in the supplementary information. We compared the models based on the highest percent deviance explained.

To identify any overarching patterns with respect to which model of landscape evolution best explained genetic diversity (Supplementary Figure 23), we calculated four summary statistics for each chromosome, each lostruct and F_{ST} outlier partition, and the genome as a whole. We tested whether genomic summary statistics on each chromosome (F_{ST}, D_{XY}, missing data, recombination rate) were correlated with explained percent deviance with an analysis of variance (ANOVA) test and a Tukey's honest significant difference test (Chambers et al., 1992, Miller 1981, Yandell 1997) using the stats v. 3.6.1 package in R. We did this for the complete dataset; for 75% and 50% missing data datasets, see Supplementary Information. We also calculated linear models comparing the proportion of each model to species-wide estimates of habitat suitability across the barrier. For all significance tests, we used an alpha value of 0.05 as our significance cutoff.

We evaluated whether the best-predictors of genomic landscapes varied across species and across partitions of the data using Chi-squared tests of significance, via the chisq.test function in the stats package in R. For each, the expected distributions assuming no differences between species, partitions, or structure were calculated and compared to the observed distributions. Chi-squared tests were performed both with and without Monte Carlo simulations (N=2000 simulations each repeated 1000 times).

Acknowledgements

This work would not have been possible without generous specimen loans from DMNH (J. Woods), UWBM (R. Faucett, J. Klicka, S. Birks), UMMZ (J. Hinshaw, B. Benz), TCWC (G. Voelker), MSB (M. L. Campbell, M. Andersen, C. Witt, A. Johnson, J. McCullough), LSUMZ (D. Dittmann, F. Sheldon), CUMV (V. Rohwer, C. Dardia), AMNH (P. Sweet, P. Capainolo, B. Bird, T. Trombone). We are grateful to numerous State and Federal Collection Permit officers, and many BLM managers (T. Schnell, S. Cooke, M. McCabe, J. Atkinson, M. Daehler, S. Torrez, D. Tersey). Thanks to staff at Dalquest Desert Research Station (N. Horner) and Indio Mountains Research Station (J. Johnson). We thank M. Ingala for illustrating the birds used in many of our figures and for helpful feedback. Helpful input comes from the Smith Lab, S. Simpson, L. Musher, D. Fletcher, F. Burbrink, L. Alter, D. Kelly, I. Overcast, A. Xue, M. Hickerson, M. Blair, P. Galante, R. Harbert, E. Sterling, A. Xue, and E. Myers., the Underrepresented Genders in Museum Ornithology group, and the B. Carstens lab. This work was funded by the AMNH Frank M. Chapman Fund, American Ornithological Society, Society of Systematic Biologists, RGGS Sydney Anderson Travel Award, AMNH Linda H. Gormezano Fund, and AMNH RGGS Graduate Fellowship. BTS was supported by US NSF award DEB-1655736.

Data Availability

These custom functions were deposited into a custom R package, subsppLabelR, which is available at github.com/kaiyaprovost/subsppLabelR, and scripts used to perform these analyses are found at github.com/kaiyaprovost/whole_genome_pipeline. All data used to perform analyses will be available on Dryad upon acceptance.

References

Adrion JR, Galloway JG, Kern AD. 2020. Predicting the landscape of recombination using deep learning. *Molecular Biology and Evolution*. 37:1790-1808.

Aguillon SM, Walsh J, Lovette, IJ. 2021. Extensive hybridization reveals multiple coloration genes underlying a complex plumage phenotype. *Proc Royal Soc B*. 288:20201805.

Backstrom N, Karaiskou N, Leder EH, Gustafsson L, Primmer CR, Qvarnström A, Ellegren H. 2008. A gene-based genetic linkage map of the collared flycatcher (*Ficedula albicollis*) reveals extensive synteny and gene-order conservation during 100 million years of avian evolution. *Genetics*. 179:1479-1495.

Barker FK, Burns KJ, Klicka, J., Lanyon SM, Lovette IJ. 2015. New insights into New World biogeography: An integrated view from the phylogeny of blackbirds, cardinals, sparrows, tanagers, warblers, and allies. *The Auk: Ornithological Advances*. 132:333-348.

Barrowclough GF, Groth JG, Mertz LA, Gutierrez RJ. 2005. Genetic structure, introgression, and a narrow hybrid zone between northern and California spotted owls (*Strix occidentalis*). *Molecular Ecology*. 14:1109-1120.

Barton NH, Hewitt, GM. 1981. Hybrid zones and speciation. *Evolution and Speciation*. 109-145.

Benzer S. 1961. On the topography of the genetic fine structure. *Proceedings of the National Academy of Sciences of the United States of America*. 47:403.

Berg PR, Jentoft S, Star B, Ring KH, Knutsen H, Lien S, Jakobsen KS, Andre C. 2015. Adaptation to low salinity promotes genomic divergence in Atlantic cod (*Gadus morhua L.*). *Genome Biology and Evolution.* 7:1644-1663.

Betancourt AJ, Presgraves DC. 2002. Linkage limits the power of natural selection in *Drosophila*. *Proceedings of the National Academy of Sciences*. 99:13616-13620.

Branch CL, Jahner JP, Kozlovsky DY, Parchman TL, Pravosudov VV. 2017. Absence of population structure across elevational gradients despite large phenotypic variation in mountain chickadees (*Poecile gambeli*). *Royal Society Open Science*. 4:170057.

Burbrink FT, Gehara M, McKelvy AD, Myers EA. 2021. Resolving spatial complexities of hybridization in the context of the gray zone of speciation in North American ratsnakes (*Pantherophis obsoletus* complex). *Evolution*. 75:260-277.

Burney CW, Brumfield RT. 2009. Ecology predicts levels of genetic differentiation in Neotropical birds. *The American Naturalist*. 174(3):358-368.

Chambers JM, Freeny AE, Heiberger RM. 1992. Analysis of variance. In Chambers JM, Hastie TJ, editors. Statistical Models in S. Pacific Grove (CA): Wadsworth and Brooks/Cole. p. 145–194.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean, G. 2011. The variant call format and VCFtools. *Bioinformatics*. 27:2156-2158.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*. 43:491.

Derryberry EP, Derryberry GE, Maley JM, Brumfield RT. 2014. HZAR: hybrid zone analysis using an R software package. *Molecular Ecology Resources*. 14:652-663.

Dubuc-Messier G, Réale D, Perret P, Charmantier, A. 2017. Environmental heterogeneity and population differences in blue tits personality traits. *Behavioral Ecology*. 28:448-459.

Duntsch L, Tomotani BM, de Villemereuil P, Brekke P, Lee KD, Ewen JG, Santure AW. 2020. Polygenic basis for adaptive morphological variation in a threatened Aotearoa | New Zealand bird, the hihi (*Notiomystis cincta*). *Proc Royal Soc B*. 287:20200948.

Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, Künstner A, Mäkinen H, Nadachowska-Brzyska K, Qvarnström A, et al. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*. 491:756-760.

Fumagalli M, Vieira FG, Linderoth T, Nielsen R. 2014. ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics*. 30:1486-1487.

Gamboa MP, Ghalambor CK, Sillett TS, Morrison SA, Funk WC. 2021. Adaptive divergence in bill morphology and other thermoregulatory traits is facilitated by restricted gene flow in song sparrows on the California Channel Islands. *Molecular Ecology*. 00:1-17.

Gibb GC, England R, Hartig G, McLenachan PA, Taylor-Smith BL, McComish BJ, Cooper A, Penny D. 2015. New Zealand passerines help clarify the diversification of major songbird lineages during the Oligocene. *Genome Biology and Evolution*. 7:2983-2995.

Grabherr MG, Russell P, Meyer M, Mauceli E, Alföldi J, Di Palma F, Lindblad-Toh K. 2010. Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics*. 26:1145-1151.

Han F, Lamichhaney S, Grant BR, Grant PR, Andersson L, Webster MT. 2017. Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Res*. 27:1004-1015.

Harris RB, Alström P, Ödeen A, Leaché AD. 2018. Discordance between genomic divergence and phenotypic variation in a rapidly evolving avian genus (*Motacilla*). *Molecular Phylogenetics and Evolution*. 120:183-195.

Hewitt GM. 1989. The subdivision of species by hybrid zones. In: Otte D, Endler J, editors. Speciation and its Consequences. Sunderland (MA): Sinauer Associates. p. 85-110.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*. 25:1965-1978.

Ho J, Tumkaya T, Aryal S, Choi H, Claridge-Chang A. 2019. Moving beyond P values: data analysis with estimation graphics. *Nature Methods*. 16:565-566.

Hodgkinson A, Eyre-Walker A. 2011. Variation in the mutation rate across mammalian genomes. *Nature Reviews Genetics*. 12:756-766.

Holmgren CA, Norris J, and Betancourt JL. 2007. Inferences about winter temperatures and summer rains from the late Quaternary record of C4 perennial grasses and C3 desert shrubs in the northern Chihuahuan Desert. *Journal of Quaternary Science: Published for the Quaternary Research Association*. 22:141-161.

Hooper DM, Price TD. 2017. Chromosomal inversion differences correlate with range overlap in passerine birds. *Nature Ecology Evolution*, 1:1526-1534.

Jun G, Wing MK, Abecasis GR, Kang HM. 2015. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res*, 25:918-925.

Kamvar ZN, Tabima JF, Grünwald NJ. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*. 2:e281.

Kamvar ZN, Brooks JC, Grünwald NJ. 2015. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*. 6:208

Kirkpatrick M. 2017. The evolution of genome structure by natural and sexual selection. *Journal of Heredity*. 108: 3-11.

Klicka LB, Kus BE, Burns KJ. 2016. Conservation genomics reveals multiple evolutionary units within Bell's Vireo (*Vireo bellii*). *Conservation Genetics*. 17:455-471.

Knief U, Schielzeth H, Backström N, Hemmrich-Stanisak G, Wittig M, Franke A, Griffith SC, Ellegren H, Kempenaers B, Forstmeier W. 2017. Association mapping of morphological traits in wild and captive zebra finches: reliable within, but not between populations. *Molecular Ecology*. 26:1285-1305.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics*. 15:356.

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Molecular Biology and Evolution*. 34:1812-1819.

Laine VN, Gossmann TI, Schachtschneider KM, Garroway CJ, Madsen O, Verhoeven KJ, De Jager V, Megens HJ, Warren WC, Minx P, et al. 2016. Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nature Communications*. 7:1-9.

Langley C H, Stevens K, Cardeno C, Lee Y C, Schrider DR, Pool JE, Langley SA, Suarez C, Corbett-Detig RB, Kolaczkowski B, et al. 2012. Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics*. 192:533–598.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 27:2987-2993.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 25:1754-1760.

887 Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform.
888 *Bioinformatics*. 26:589-595.
889 Li H, Ralph P. 2019. Local PCA shows how the effect of population structure differs along the
890 genome. *Genetics*. 211:289-304.
891 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R.
892 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. 25:2078-2079.
893 Manion G, Lisk M, Ferrier S, Nieto-Lugilde D, Mokany K, Fitzpatrick MC. 2018. gdm:
894 Generalized dissimilarity modeling. R package version, 1(11).
895 Mank JE, Vicoso B, Berlin S, Charlesworth B. 2010. Effective population size and the Faster-X
896 effect: Empirical results and their interpretation. *Evolution*. 64:663-674.
897 Manthey JD, Klicka J, Spellman GM. 2021. The genomic signature of allopatric speciation in a
898 songbird is shaped by genome architecture (Aves: *Certhia americana*). *Genome Biology*
899 *and Evolution.* 13(8):evab120.
900 Manthey JD, Moyle RG. 2015. Isolation by environment in White-breasted Nuthatches (*Sitta*
901 *carolinensis*) of the Madrean Archipelago sky islands: a landscape genomics approach.
902 *Molecular Ecology*. 24:3628-3638.
903 Martin SH, Jiggins CD. 2017. Interpreting the genomic landscape of introgression. *Current*
904 *Opinion in Genetics & Development.* 47:69-74.
905 Mason NA, Burns KJ. 2013. Molecular phylogenetics of the Neotropical seedeaters and seed-
906 finches (*Sporophila*, *Oryzoborus*, *Dolospingus*). *Ornitología Neotropical*. 24:139-155.
907 Mayr E. 1942. Systematics and the origin of species, from the viewpoint of a zoologist. New York:
908 Columbia University Press.
909 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler
910 D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework
911 for analyzing next-generation DNA sequencing data. *Genome Res*. 20(9):1297-1303.
912 Meisner J, Albrechtsen A. 2018. Inferring population structure and admixture proportions in low-
913 depth NGS data. *Genetics*. 210:719-731.
914 Miles LS, Rivkin LR, Johnson MT, Munshi-South J, Verrelli BC. 2019. Gene flow and genetic
915 drift in urban environments. *Molecular Ecology*. 28:4138-4151.
916 Miller RG. 1981. Simultaneous Statistical Inference. New York: Springer.
917 Mitchell KJ, Wood JR, Llamas B, McLenachan PA, Kardailsky O, Scofield RP, Worthy TH,
918 Cooper A. 2016. Ancient mitochondrial genomes clarify the evolutionary history of New
919 Zealand's enigmatic acanthisittid wrens. *Molecular Phylogenetics and Evolution*. 102:295-
920 304.
921 Morafka DJ. 1977. A Biogeographical Analysis of the Chihuahuan Desert through its
922 Herpetofauna. Dordrecht: Springer.
923 Moreira LR, Hernández-Baños, BE, Smith BT. 2020. Spatial predictors of genomic and
924 phenotypic variation differ in a lowland Middle American bird (*Icterus gularis*). *Molecular*
925 *Ecology*. 29(16):3084-3101.
926 Muscarella R, Galante PJ, Soley-Guardia M, Boria RA, Kass JM, Uriarte M, Anderson RP. 2014.
927 ENM eval: An R package for conducting spatially independent evaluations and estimating
928 optimal model complexity for Maxent ecological niche models. *Methods in Ecology and*
929 *Evolution*. 5:1198-1205.
930 Myers EA, Xue AT, Gehara M, Cox CL, Davis Rabosky AR, Lemos-Espinal J, Martínez-Gómez
931 JE, Burbrink FT. 2019. Environmental heterogeneity and not vicariant biogeographic

barriers generate community-wide population structure in desert-adapted snakes. *Molecular Ecology*. 28(20):4535-4548.

Normand P, Lapierre P, Tisa LS, Gogarten JP, Alloisio N, Bagnarol E, Bassi CA, Berry AM, Bickhart DM, Choisne N, et al. 2007. Genome characteristics of facultatively symbiotic *Frankia sp.* strains reflect host range and host plant biogeography. *Genome Res*. 17(1):7-15.

Nosil P, Schluter D. 2011. The genes underlying the process of speciation. *Trends in Ecology & Evolution*. 26:160-167.

Nuvoloni FM, Feres RJ. F., Gilbert B. 2016. Species turnover through time: colonization and extinction dynamics across metacommunities. *The American Naturalist*. 187:786-796.

Pardieck KL, Ziolkowski Jr DJ, Lutmerding M, Aponte V, Hudson MA. R. 2019. North American Breeding Bird Survey Dataset 1966–2018, version 2018.0. US Geological Survey, Patuxent Wildlife Research Center, Laurel, MD.

Pasquet E, Barker FK, Martens J, Tillier A, Cruaud C, Cibois A. 2014. Evolution within the nuthatches (Sittidae: Aves, Passeriformes): molecular phylogeny, biogeography, and ecological perspectives. *Journal of Ornithology*. 155:755-765.

Paz A, Ibáñez R, Lips KR, Crawford AJ. 2015. Testing the role of ecology and life history in structuring genetic variation across a landscape: A trait-based phylogeographic approach. *Molecular Ecology*. 24(14):3723-3737.

Peterman WE. 2018. ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. *Methods in Ecology and Evolution*. 9:1638-1647.

Peterman WE, Connette GM, Semlitsch RD, Eggert LS. 2014. Ecological resistance surfaces predict fine-scale genetic differentiation in a terrestrial woodland salamander. *Molecular Ecology*. 23:2402-2413.

Phillips OL. 1996. Long-term environmental change in tropical forests: increasing tree turnover. *Environmental Conservation*. 23(3):235-248.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190:231-259.

Price TD, Hooper DM, Buchanan CD, Johansson US, Tietze DT, Alström P, Olsson U, Ghosh-Harihar M, Ishtiaq F, Gupta SK, et al. 2014. Niche filling slows the diversification of Himalayan songbirds. *Nature*. 509(7499):222-225.

Probst CM, Ralston J, Bentley I. 2021. Effects of climate on bill morphology within and across *Toxostoma* thrashers. *Journal of Avian Biology*.

Provost KL, Myers EA, Smith BT. 2021. Community phylogeographic patterns reveal how a barrier filters and structures taxa in North American warm deserts. *Journal of Biogeography*. 48:1267-1283.

R Core Team. 2019. R: a language and environment for statistical computing, version 3.0.2. Vienna, Austria: R Foundation for Statistical Computing; 2013.

Ralston J, FitzGerald AM, Burg TM, Starkloff NC, Warkentin IG, Kirchman JI. 2021. Comparative phylogeographic analysis suggests a shared history among eastern North American boreal forest birds. *The Auk* 138(3): ukab018.

Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlović M, Noor MAF, Mehlig B, Westram AM. 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary Biology*. 30(8):1450-1477.

Reynolds J.F., Kemp P.R., Ogle K, Fernández RJ. 2004. Modifying the 'pulse-reserve' Paradigm for deserts of North America: precipitation pulses, soil water, and plant responses. *Oecologia*. 141:194-210.

Ribeiro ÂM, Puetz L, Pattinson NB, Dalén L, Deng Y, Zhang G, da Fonseca RR, Smit B, Gilbert MTP. 2019. 31° South: the physiology of adaptation to arid conditions in a passerine bird. *Molecular Ecology*. 28(16):3709-3721.

Robinson DF, Foulds LR. 1981. Comparison of phylogenetic trees. *Mathematical Biosciences*. 53:131-147.

Rojas-Soto OR. 2003. Geographic variation of the curve-billed thrasher (*Toxostoma curvirostre*) complex. *The Auk*. 120:311-322.

Rojas-Soto OR, De Los Monteros AE, Zink RM. 2007. Phylogeography and patterns of differentiation in the curve-billed thrasher. *The Condor*. 109:456-463.

Romanov MN, Dodgson JB, Gonser RA, Tuttle, EM. 2011. Comparative BAC-based mapping in the white-throated sparrow, a novel behavioral genomics model, using interspecies overgo hybridization. *BMC Research Notes*. 4(1):1-13.

Sæther SA, Sætre GP, Borge T, Wiley C, Svedin N, Andersson G,... Qvarnström A. 2007. Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. *Science*. 318:95-97.

Shafer AB, Cullingham CI, Cote SD, Coltman DW. Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. 2010. *Molecular Ecology* 19(21):4589-4621.

Shreve F. 1942. The desert vegetation of North America. *The Botanical Review*. 8:195-246.

Sin SYW., Lu L, Edwards SV. 2020. De Novo assembly of the northern cardinal (*Cardinalis cardinalis*) genome reveals candidate regulatory regions for sexually dichromatic red plumage coloration. *G3: Genes|Genomes|Genetics.* 10(10):3541-3548.

Singhal S, Leffler EM, Sannareddy K, Turner I, Venn O, Hooper DM, Strand AI, Li Q, Raney B, Balakrishnan CN, et al. 2015. Stable recombination hotspots in birds. *Science*. 350(6263):928-932.

Smith BT, Smith BT, Bryson Jr RW, Mauck III WM, Chaves J, Robbins MB, Aleixo A, Klicka J. 2018. Species Delimitation and Biogeography of the Gnatcatchers and Gnatwrens (Aves: Polioptilidae). *Molecular Phylogenetics and Evolution*. 126: 45-57.

Smith BT, Escalante P, Baños BEH, Navarro-Sigüenza AG, Rohwer S, Klicka J. 2011. The role of historical and contemporary processes on phylogeographic structure and genetic diversity in the Northern Cardinal, *Cardinalis cardinalis*. *BMC Evolutionary Biology*. 11:136.

Spencer JE. 1996. Uplift of the Colorado Plateau due to lithosphere attenuation during Laramide low-angle subduction. *Journal of Geophysical Research: Solid Earth*. 101:13595-13609.

Teutimez MR. 2012. The cactus wren (*Campylorhynchus brunneicapillus*) in southern California: haplotype comparisons among coastal and inland populations. Long Beach (CA): California State University.

Toews DP, Taylor SA, Vallender R, Brelsford A, Butcher BG, Messer PW, Lovette IJ. 2016. Plumage genes and little else distinguish the genomes of hybridizing warblers. *Current Biology*. 26:2313-2318.

Turbek SP, Browne M, Di Giacomo AS, Kopuchian C, Hochachka WM, Estalles C, Lijtmaer DA, Tubaro PL, Silveira LF, Lovette IJ, et al. 2021. Rapid speciation via the evolution of pre-mating isolation in the Iberá Seedeater. *Science*. 371(6536):1312-1312.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*. 43:11-10.

Van Devender TR. 1990. Late quaternary vegetation and climate of the Sonoran Desert, United States and Mexico. In Betancourt JL, Van Devender TR, Martin PS, editors. Packrat middens: the last 40,000 years of biotic change. Tucson (AZ): University of Arizona Press. p. 134-165.

Van Devender TR, Betancourt JL, Wimberly M. 1984. Biogeographic implications of a packrat midden sequence from the Sacramento Mountains, south-central New Mexico. *Quaternary Research*. 22(3):344-360.

Vasconcellos MM, Colli GR, Weber JN, Ortiz EM, Rodrigues MT, Cannatella DC. 2019. Isolation by instability: Historical climate change shapes population structure and genomic divergence of treefrogs in the Neotropical Cerrado savanna. *Molecular Ecology*. 28(7):1748-1764.

Wang IJ, Bradburd GS. 2014. Isolation by environment. *Molecular Ecology*. 23:5649-5662.

Wang J, Street NR, Park EJ, Liu J, Ingvarsson PK. 2020. Evidence for widespread selection in shaping the genomic landscape during speciation of *Populus*. *Molecular Ecology*. 29(6):1120-1136.

Weckworth BV, Musiani M, DeCesare NJ, McDevitt AD, Hebblewhite M, Mariani S. 2013. Preferred habitat and effective population size drive landscape genetic patterns in an endangered species. *Proc Royal Soc B: Biological Sciences*. 280:20131756.

Williams D, Gogarten JP, Lapierre P. 2010. Filling the gaps in the genomic landscape. *Genome Biology*. 11(2):1-2.

Wingett SW, Andrews S. 2018. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000 Research*. 7:1338.

Wright S. 1943. Isolation by distance. *Genetics*. 28(2):114.

Yandell BS. 1997. Practical data analysis for designed experiments. New York: Chapman & Hall/CRC.

Yi X, Latch EK. 2021. Nonrandom missing data can bias Principal Component Analysis inference of population genetic structure. *Molecular Ecology Resources.* 00:1-10.

Yusuf L, Heatley MC, Palmer JP, Barton HJ, Cooney CR, Gossmann TI. 2020. Noncoding regions underpin avian bill shape diversification at macroevolutionary scales. *Genome Res*. 30:553-565.

Zamudio KR, Bell RC, Mason NA. 2016. Phenotypes in phylogeography: Species' traits, environmental variation, and vertebrate diversification. *Proceedings of the National Academy of Sciences*. 113:8041-8048.

Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, et al. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*. 346(6215):1311-1320.