Recombination rate variation shapes genomic variability of phylogeographic structure in a widespread North American songbird (Aves: *Certhia americana*)

Joseph D. Manthey[1]*, Garth M. Spellman[2]

[1]Department of Biological Sciences, Texas Tech University. Lubbock, Texas, USA.

[2]Department of Zoology, Denver Museum of Nature & Science, Denver, Colorado, USA.

*Corresponding author email: jdmanthey@gmail.com

Manuscript Type: Original Article

Running head: Recombination rate variation shapes phylogeographic structure variation

Keywords: phylogeography, population genomics, genetic architecture, demography

ABSTRACT

The nonrandom distribution of chromosomal characteristics and functional elements—genomic architecture—impacts the relative strengths and impacts of population genetic processes across the genome. Due to this relationship, genomic architecture has the potential to shape variation in population genetic structure across the genome. Population genetic structure has been shown to vary across the genome in a variety of taxa, but this body of work has largely focused on pairwise population genomic comparisons between closely related taxa. Here, we used whole genome sequencing of seven phylogeographically structured populations of a North American songbird, the Brown Creeper (*Certhia americana*), to determine the impacts of genomic architecture on phylogeographic structure variation across the genome. Using multiple methods to infer phylogeographic structure—ordination, clustering, and phylogenetic methods—we found that recombination rate variation explained a large proportion of phylogeographic structure variation. Genomic regions with low recombination showed phylogeographic structure consistent with the genome-wide pattern. In regions with high recombination, we found strong phylogeographic structure, but with discordant patterns relative to the genome-wide pattern. In regions with high recombination rate, we found that populations with small effective population sizes evolve more rapidly than larger populations, leading to discordant signatures of phylogeographic structure. These results suggest that the interplay between recombination rate variation and effective population sizes shape the relative impacts of linked selection and genetic drift in different parts of the genome. Overall, the combined interactions of population genetic processes, genomic architecture, and effective population sizes shape patterns of variability in phylogeographic structure across the genome.

INTRODUCTION

The formation of population structure is implicit in the speciation process (Allmon, 1992; Mayr et al., 1963). Speciation is a continuum; early differentiating populations may have few genomic regions with evidence for genetic differentiation or barriers to gene flow (Ellegren et al., 2012; Nosil & Feder, 2012; Toews et al., 2016; Wu, 2001) while later in the speciation process two diverging populations may exhibit strong genetic differentiation across their entire genomes (Nosil & Feder, 2012; Ravinet et al., 2018; Wu, 2001). Across the genome, variation in local evolutionary history is shaped by the relative strengths of the population genomic processes of natural selection, gene flow, and genetic drift. For example, population genetic structure may differ in genomic regions with strong natural selection or differential patterns of gene flow relative to the rest of the genome (Louis et al., 2020; Whiting et al., 2021).

Variable impacts of selection, drift, and gene flow across the genome may be partially attributed to genomic architecture—the nonrandom distribution of chromosomal characteristics and functional elements (Koonin, 2009)—whereby variation in gene content, recombination rates, or other genomic characteristics impact the relative effects of population genomic processes. For example, we expect that genomic regions with lower recombination rates will be relatively depleted of genetic diversity and genetic differentiation will accumulate faster, due to relatively increased effects of linked selection (Cruickshank & Hahn, 2014; Haenel et al., 2018; Hey & Kliman, 2002; Nachman & Payseur, 2012). Indeed, speciation and population genomic studies have often found a negative relationship between recombination rates and genetic differentiation across the genome (Kulathinal et al., 2008; J. D. Manthey, Klicka, & Spellman, 2021; Roesti, Moser, & Berner, 2013; Stankowski et al., 2019; Vijay et al., 2016).

If genomic architecture impacts population genomic patterns within populations and between pairs of taxa, we may naturally extend this idea to phylogeographic patterns in taxa with multiple lineages at different stages of evolutionary distinctiveness. As such, we may

predict that estimates of phylogeographic structure will vary across the genome and be shaped by several factors and their interactions: (1) population size variation of differentiating populations, (2) time since divergence began, and (3) genomic architecture. Exemplar taxa to better understand the impacts of genomic architecture on population genetic structure are species (or species complexes) with multiple lineages on distinct evolutionary trajectories, populations with different effective sizes, and genomes exhibiting heterogeneous genomic architecture.

The Brown Creeper (*Certhia americana*) is an excellent study organism to understand the impacts of genomic architecture on phylogeographic structure; its genome is highly heterogeneous in both structure and content, and it has seven distinct phylogeographic clades. The genome of the Brown Creeper, as like other birds, has chromosomes that (1) span two orders of magnitude in size, (2) vary in their gene and repetitive element content, and (3) exhibit local effective recombination rates that can vary by an order of magnitude or more (Dutoit, Burri, et al., 2017; Ellegren, 2010; Kapusta & Suh, 2017; Kawakami et al., 2014; J. D. Manthey et al., 2021). Additionally, the Brown Creeper has strong phylogeographic structure, with two main lineages that have been diverging around one million years (Fig. 1A), multiple phylogeographically structured clades within each of these two main lineages, and populations with different effective sizes (J. D. Manthey, Klicka, & Spellman, 2011a, 2011b, 2015; J. D. Manthey et al., 2021).

Using the Brown Creeper as our study taxon, we used whole-genome sequencing data to assess how genome architecture influences phylogeographic structure; we used phylogenetic, ordination, and population genetic clustering methods to characterize phylogeographic structure across the genome. Our aims here were twofold. First, we looked to decipher how genomic architecture and effective population sizes shape phylogeographic structure variation across the genome. Second, because phylogeographic structure may be

estimated in many ways, we wanted to identify whether phylogeographic signal variation was consistent using different methodologies.

Based on previous population genomic research, we hypothesized that genomic regions with low recombination rates will have faster lineage sorting. Based on this hypothesis, we'd predict that phylogeographic estimates consistent with the true evolutionary history of the group will be concentrated in genomic regions with low recombination. Additionally, we may predict a couple alternative patterns in high recombination regions. We may identify mixed or weak phylogeographic structure in high recombination regions. Alternatively, we may identify strong—but divergent—patterns of phylogeographic structure in high recombination regions linked with varied evolutionary rates among sampled populations.

METHODS

Study organism, sampling, and sequencing. The Brown Creeper is a songbird found in forested regions of the Americas; it has a widespread distribution ranging from Honduras in the south to Alaska in the north (Fig. 1A) (Poulin et al., 2020). The Brown Creeper is aptly named; it creeps up trees in search of its largely invertebrate diet and has dark coloration on its back and sides (Fig. 1A inset) that makes for excellent camouflage on trees with dark bark. In previous work using tens to thousands of genetic markers, we have identified phylogeographic structure in the Brown Creeper, with up to seven distinct lineages (J. D. Manthey et al., 2011a, 2011b, 2015). Here, we aimed to obtain population genomic sequencing data for seven populations with three individuals sampled per population (Fig 1, Table S1). We downloaded previously published sequencing data from two of these populations and one outgroup sample of *C. familiaris* (Table S1) (J. D. Manthey et al., 2021). For the other five populations, we obtained representative tissue samples from natural history museums and extracted genomic DNA using QIAGEN DNeasy blood and tissue kits following manufacturer guidelines. DNA from extractions

was used to create standard Illumina sequencing libraries, followed by sequencing on an Illumina NovaSeq6000 at the Texas Tech University Center for Biotechnology and Genomics with the goal of obtaining ~15-30× coverage per individual.

Reference genome and genomic architecture. For the reference genome, we used the chromosome-scale *Certhia americana* genome we assembled for a prior study (NCBI assembly ASM1869719v1; Manthey et al. 2021). From data reported in this previous study (Manthey et al. 2021), we summarized mean effective recombination rates for each of the two main Brown Creeper lineages, transposable element (TE) and gene density, and GC content across the genome in 50kbp and 100 kbp non-overlapping sliding windows (Fig. 2). Briefly, these estimates were initially obtained with the following methods: effective recombination rate was estimated using LDhat (McVean & Auton, 2007), TEs were annotated with RepeatMasker v4.08 (Smit, Hubley, & Green, 2015), and gene content was annotated with MAKER v2.31.10 (Cantarel et al., 2008).

Sequencing data filtering and genotyping. We used the program bbduk (Bushnell, 2014) to quality filter raw sequencing data. We then use BWA v0.7.17 (Heng Li & Durbin, 2009) to align filtered reads to the *Certhia americana* reference genome. Next, we used samtools v1.4.1 (Heng Li et al., 2009) to convert the BWA output SAM file to BAM format, followed by using the Genome Analysis Toolkit (GATK) v4.1.0.0 (McKenna et al., 2010) to clean, sort, add read groups to, and remove duplicates from BAM files. We used the samtools ‘depth’ command to measure sequencing coverage across the reference genome for each individual (Fig. S1, Table S1).

We then used GATK in three steps to genotype all individuals: we used the ‘HaplotypeCaller’ function to call preliminary genotypes for each individual, built a database for all samples using the ‘GenomicsDBImport’ function, and lastly used the ‘GenotypeGVCFs’

function to group genotype all individuals together for both variant and invariant sites. We filtered the genotype data using VCFtools v0.1.14 (Danecek et al., 2011) with the following restrictions: (1) minimum site quality of 20, (2) minimum genotype quality of 20, (3) minimum depth of coverage of 6, (4) maximum mean depth of coverage of 50, and (5) removal of indels. For downstream analyses, we used various settings for (1) amount of missing data allowed, (2) whether the outgroup was included, (3) inclusion or exclusion of invariant sites, and (4) thinning between sites (see Table 1 for exact dataset characteristics used for all seven datasets used for analyses). Although it has been suggested that variation in minor allele frequency (MAF) filtering may impact estimates of genetic structure (Linck & Battey, 2019), we did not filter for a minimum MAF because we previously showed that changing MAF did not significantly impact pairwise F_{ST} calculations in *C. americana* (J. D. Manthey et al., 2021).

Phylogenomics. We estimated "gene trees" in non-overlapping windows (window size 50 kbp and 100 kbp; datasets C and D in Table 1) using RAxML v8.2.12 (Stamatakis, 2014) with the GTRCAT model of sequence evolution. For each phylogeny, we used 100 rapid bootstraps to assess support and determined a best-supported tree for each window. For a window to be included, we required a minimum of 10 kbp in the alignment following filtering. This resulted in 19,639 phylogenies for the 50 kbp sliding windows and 9846 phylogenies for the 100 kbp sliding windows. We summarized the best-supported trees using the sumtrees.py script, part of the DendroPy Python package (Sukumaran & Holder, 2010). Lastly, we used ASTRAL III v 5.6.3 (Zhang et al., 2018) to calculate a species tree from all gene trees, using quartet frequencies as a measure of local support (Sayyari & Mirarab, 2016).

For each gene tree, we estimated the genealogical sorting index (GSI) (Cummings, Neel, & Shaw, 2008) for (1) each of the seven populations, and (2) each of the two main lineages (i.e., north and south shaded blue and red, respectively, in Fig. 1). The GSI statistic uses rooted gene trees to quantify lineage sorting; the GSI quantifies exclusive ancestry of

individuals in labelled groups by dividing the minimum number of nodes to produce monophyly for a defined group divided by the observed number of nodes necessary to connect all individuals of a group in a phylogeny. The maximum GSI value is one and occurs when a defined group is monophyletic. We calculated the GSI statistic for each sampled population and for each of the two main lineages (Fig. 1). We expect that each sampled population will show evidence of phylogenetic clustering, regardless of position in the genome. As such, GSI at the population level may have little variation tied to genomic architecture. In contrast, which populations group with which other populations are likely to be influenced by rates of lineage sorting and are therefore likely influenced by genomic architecture. As such, we may expect GSI measures at the lineage level (i.e., two main clades) to be impacted by genomic architecture.

For each individual across every gene tree, we calculated the root-to-tip distance (RTD) using the R package ape (Paradis, Claude, & Strimmer, 2004). As a summary of these RTDs, we calculated the mean RTD per population relative to the maximum RTD value for each gene tree (i.e., this relative statistic's range = 0 – 1). We then measured the relationships between harmonic mean effective population sizes (mean per population) and relative RTD in regions with low recombination and high recombination, as well as the shift in values between high and low recombination regions. We also estimated relationships between these characteristics while accounting for the evolutionary relationships of the samples using phylogenetic independent contrasts (PICs) calculated in the R package ape (Paradis et al., 2004).

Genetic structure. We estimated genetic structure using the program ADMIXTURE v1.3.0 (Alexander, Novembre, & Lange, 2009). Here, we used genome-wide SNPs thinned to either a minimum of 50 kbp or 100 kbp distance between SNPs (datasets A and B in Table 1). We ran ADMIXTURE with an assumed number of genetic clusters equal to two or seven (i.e., k = 2 and k = 7) based on our knowledge of phylogeographic structure in this species from previous studies (J. D. Manthey et al., 2011a, 2011b, 2015). To assess how genetic structure estimation

varied across the genome, we also ran ADMIXTURE for SNP datasets in non-overlapping sliding windows of 50 kbp and 100 kbp (datasets E and F in Table 1). We compared the genome-wide ADMIXTURE results to those in each window by summing the differences in group membership for each individual. In other words, this deviation value ranges from zero to nearly one, where a value of zero is identical group assignment for all individuals with the same admixture coefficients for each genetic cluster, and higher values indicate more dissimilar matrices.

We also estimated variation in genetic structure across the genome using LOSTRUCT (Han Li & Ralph, 2019). This method runs in three steps: (1) reducing the dimensionality of the data using principal component analysis (PCA) for each genomic window, (2) find distances between PCA maps, and (3) using multidimensional scaling (MDS) to display variation in PCA-based genetic structure across the genome. We used LOSTRUCT in sliding windows of 50 kbp and 100 kbp (datasets E and F in Table 1). We used variation in MDS axis one (hereafter MDS1) to describe variation in these PCA-based estimates of genetic structure across the genome.

We expect that genomic regions with low recombination will exhibit strong phylogeographic structure similar to genome-wide patterns (e.g., Fig. 1). In contrast, we expect that genomic regions with high recombination will either exhibit little genetic structure or alternative patterns of genetic structure. As such, we expect that variation in ADMIXTURE and PCA results will covary with genomic architecture.

Summary statistics. Across 50 kbp and 100 kbp sliding windows (datasets C and D in Table 1), we estimated population genomic summary statistics to summarize patterns of genetic variation across the genome both within and among populations. We measured observed heterozygosity per individual and estimated nucleotide diversity per population as measures of genetic diversity. We calculated genetic differentiation (F_{ST}) and genetic divergence (D_{XY})

between geographically neighboring populations as measures of distinct evolutionary trajectories among populations. We used the Reich et al. (2009) F_{ST} estimator because it is unbiased when using small sample sizes, unlike some other common F_{ST} estimators (Willing, Dreyer, & Van Oosterhout, 2012). Lastly, we measured Tajima's D (Tajima, 1989) as an estimator of whether populations tend to evolve neutrally across the genome.

Population genomic correlations. We used the R package Hmisc (Harrell & Dupont, 2020) to estimate both Pearson and Spearman correlation coefficients between population genomic summary statistics and characteristics of genomic architecture in sliding windows. We used the R package corrplot (Wei et al., 2017) to visualize correlations between all summary statistics. Lastly, we used variance partitioning, implemented in the R package vegan (Oksanen et al., 2007), to assess the proportion of variance in phylogeographic structure estimates explained by genomic architecture.

Population demographic history. We used the program MSMC2 v1.1.0 (Schiffels & Durbin, 2014) to estimate demographic history for each individual. For use in MSMC, we masked regions of the genome not genotyped due to low coverage or low genotype quality scores. Additionally, we did not include the sex chromosomes in demography calculations. We ran MSMC for each individual allowing up to 20 iterations (default setting) and used up to 23 inferred distinct time segments because this setting worked well to reduce spurious or inconsistent results in other songbirds that we have studied (J. Manthey et al., 2022). We performed 10 bootstraps for each demography estimate, using 1 Mbp bootstrapped segments of the genome, to assess how demographic signal varies when subsetting parts of the genome. Because the output of MSMC is interpreted relative to assumed mutation rates and generation times, we used the Brown Creeper genome mutation rate estimate of 2.506×10^{-9} substitutions per site per year (J. D. Manthey et al., 2021). Because there are no published estimates of

Brown Creeper generation times, we used a proxy generation time of double the age of sexual maturity (Nadachowska-Brzyska et al., 2015) using maturity estimates from the Animal Aging and Longevity Database (age of maturity = one year) (Tacutu et al., 2017). With the MSMC demographic model output, we estimated the harmonic mean population sizes for each population by using the mean MSMC population size estimates for each individual in 1000-year discrete time intervals over the past 200 kya.

Potential mutational load. Because we identified different relative rates of molecular evolution in differently sized populations, we wanted to identify whether smaller populations were accumulating a relatively greater potential mutational load. First, we used SnpEff (Cingolani et al., 2012) to annotate estimated functional effects of SNPs in coding regions based on a previous functional annotation of the *C. americana* reference genome (J. D. Manthey et al., 2021). Next, as a relative estimate of potential mutational load, we measured the proportion of polymorphisms in each population that changed the amino acid sequence (i.e., missense and nonsense mutations). Specifically, we calculated the number of nonsynonymous polymorphisms (P_N) and synonymous polymorphisms (P_S) and measured the ratio $\frac{P_N}{P_N + P_S}$ for each population in genomic quintiles of recombination rate variation.

Code availability. All computer code used for analyses and figure creation (applicable for some figures) for this project is available here: github.com/jdmanthey/certhia_phylogeography. We used the following R packages for figure creation and file manipulation that were not cited in other parts of the methods section: Biostrings (Pagès et al., 2017), ggplot2 (Wickham, 2011), palettetown (Lucas, 2016), phytools (Revell, 2012), and RColorBrewer (Neuwirth, 2014).

RESULTS

Genome-wide phylogeographic structure and demography. We used genomic sequencing data from 21 ingroup individuals sequenced at ~18-29× genomic coverage to estimate genome-wide patterns of phylogeographic structure. We corroborated previous genetic work (that used few to thousands of genetic markers) by identifying hierarchical and strong phylogeographic structure among sampled populations (J. D. Manthey et al., 2011a, 2011b, 2015). The deepest phylogenomic split separates northern from southern populations and there is additional support for distinctiveness of each of the regionally sampled populations (Fig. 1C). Genome-wide ADMIXTURE results were consistent across different thinning strategies (i.e., 50 kbp and 100 kbp thinning; only 50 kbp results plotted) and showed patterns of genetic structure that aligned well with the species tree (Fig. 1C). Demographic history estimates for each individual were consistent within populations, but each population exhibited a distinct demographic history (Fig. 1D). In the Pacific and Central American localities, population sizes have fluctuated somewhat, but generally stay below an N_E of 200,000 over the past 100 ky (Fig. 1D). The Sierra Madre Oriental population showed evidence for a sharp decline in N_E over the past 50 ky, shifting from an N_E ~ 200,000 at about 50 kya to an N_E ~ 10,000 in the past ten kya (Fig. 1D). The Rocky Mountains population has remained relatively stable over the past 100 ky with an N_E ~ 200,000 (Fig. 1D). Lastly, the Eastern North America and Central Mexico populations have exhibited fluctuations and much larger N_E than all other populations over the past 100 ky (Fig. 1D). Harmonic mean effective population sizes over the past 200 ky ranged from ~85,000 (Central America South) to ~320,000 (Central Mexico) (Fig. 1D). Harmonic mean effective population size estimates over the past 200 ky are highly correlated with observed heterozygosity estimates for each individual (r = 0.962).

Summary Statistics. Genetic diversity in each population varied widely across the genome, with mean nucleotide diversity values for each population between ~0.002 and 0.004 (Fig 1B;

Fig. S2; Fig. S3). Estimates of Tajima's D for each population appeared to be somewhat consistent across the genome in three populations (Central Mexico, Eastern North America, and Rocky Mountain), but highly variable across the genome in the other four populations (Fig. S4).

Pairwise values of genetic divergence and differentiation between geographically proximate populations varied widely across the genome (Fig. 2; Fig. S2; Fig. S5; Fig. S6, Fig. S7). F_{ST} and D_{XY} estimates between the two main lineages (Rocky Mountain × Central Mexico comparison) were the highest, as expected; for all other comparisons, genetic divergence (D_{XY}) had pairwise mean values between 0.0025 and 0.0050, and genetic differentiation (F_{ST}) had pairwise mean values between 0.15 and 0.25 (Fig. 2; Fig. S2). Notably, only two pairwise comparisons of populations—Rocky Mountain × Central Mexico and Rocky Mountain × Eastern North America—showed considerable variation in F_{ST} across the genome, while the other pairwise comparisons showed relatively consistent estimates of F_{ST} across the genome (Fig. S5). Notably, while there was a negative relationship between F_{ST} and D_{XY} when using the mean values across all pairwise comparisons, this relationship also varied by pairwise population comparison (Fig. S7); the Rocky Mountain × Central Mexico showed a strong positive correlation between F_{ST} and D_{XY}, the Rocky Mountain × Eastern North America showed a strong negative correlation, and the other pairwise comparisons showed a weaker relationship between the two summary statistics (Fig. S7).

Phylogeographic structure variation across the genome. We estimated phylogeographic structure across the genome using PCA, ADMIXTURE, and the GSI metric from phylogenies; all estimates of phylogeographic structure varied across the genome, with the outliers deviating the most from genome-wide patterns clustering mostly on small chromosomes and the ends of large chromosomes (Fig. 2; Fig. S2). The most extreme deviations from genome-wide phylogeographic structure appear to cluster the Central Mexico population with the northern lineage (Fig. S8), and in some cases also cluster the Sierra Madre Oriental population with the

northern lineage (Fig. S8; Fig. 3). As an example, we plot the phylogeographic structure for one of these outlier windows in Fig. 3, demonstrating the clustering of the Sierra Madre Oriental and Central Mexico populations with the northern lineage.

Correlations between phylogeographic structure estimates and genomic architecture. Almost all characteristics of the genome, population genomic summary statistics, and statistics characterizing variation in phylogeographic structure were strongly correlated in sliding windows across the genome (Fig. 4; Fig. S9). Notably, the relationship between population pairwise genetic divergence (D_{XY}) and recombination rate variation varied with different population comparisons (Fig. S7); D_{XY} was positively correlated in all the comparisons except the Rocky Mountain × Central Mexico population, which exhibited a strong negative correlation between D_{XY} and recombination rate (Fig. S7).

We found a strong correlation between different measures of phylogeographic structure, including GSI at the lineage level, MDS1 of the PCA, and ADMIXTURE deviations (Fig. 4; Fig. S9). Phylogeographic structure deviations (i.e., most different from genome-wide patterns) were strongly positively correlated with recombination rate, GC content, and genetic diversity (Fig. 4; Fig. S9). In contrast, TE content, Tajima's D, and GSI calculations at the population level had no or weak correlations with all other statistics that we calculated (Fig. 4; Fig. S9). Correlations were consistent between 50 kbp and 100 kbp estimates, with no correlations deviating more than ~0.12 when estimated with different window sizes (Fig. S9).

Using variance partitioning, recombination rate variation most strongly explained deviations in phylogeographic structure, but interactions of recombination rate, GC content, CDS content, and TE content also explained some portions of the variance (Table 2). Most notably, genomic architecture explained approximately 60% of the variance in both (1) the MDS1 axis explaining differences in PCA-based estimates of phylogeographic structure and (2) ADMIXTURE results at the lineage level with an assumed k = 2 (Table 2). The relationship

between recombination rate and phylogeographic structure is exemplified when comparing the genomic windows with the highest or lowest values of recombination rate; estimates of phylogeographic structure and genetic diversity have little to no overlap in statistic distributions in high recombination regions versus low recombination regions (Fig. 5). Additionally, when comparing the genomic windows with relatively higher or lower recombination rates, all the populations exhibit shifts in their relative RTD distributions (Fig. S10).

Because the effective population sizes for each locality varied substantially (Fig. 1; Fig. S11), we looked to see if there was a relationship between relative RTD and population size in regions with very different recombination rates. Using both the raw measurements and PICs, we found a negative association between relative rates of evolution and harmonic mean effective population sizes (Fig. 6). Notably, there was also a strong positive association between effective population sizes and shifts in relative rates of evolution between high and low recombination regions (Fig. 6). We also looked at the relationship between population sizes and potential mutational load; we found no statistically significant associations but did find a slight, non-significant, negative association between population sizes and potential mutational load in high recombination regions (Fig. S12).

DISCUSSION

Using whole-genome sequencing for seven populations of the Brown Creeper, we found variation in genomic diversity and differentiation was associated with local recombination rate, consistent with a large body of previous population genomics research in both model and non-model organisms. For our main research question, we aimed to decipher how genome architecture influences phylogeographic structure variation across the genome. We found that recombination rate and other characteristics of the genome explain 6-68% of the variance in

phylogeographic structure inference using different metrics and genomic sliding window sizes (Table 2).

Genomic architecture shapes genomic diversity, divergence, and differentiation.

While not the focus of our study, we identified correlations between population genomic statistics and characteristics of genomic architecture; for most population genomic statistics, including measures of genetic diversity, divergence, and differentiation, we found significant associations with multiple measures of genomic characteristics including recombination rate and GC, gene, and repetitive DNA content (Fig. 4; Fig. S9). The strongest correlations were those with local recombination rate, so we focus our brief discussion here on the relationships between population genomics summary statistics and recombination rate and consistency with previous studies.

Genetic diversity—as measured by nucleotide diversity and observed heterozygosity—was positively associated with recombination rate across the genome (Fig. 4; Fig. S9). We may expect increased diversity in high recombination regions due to decreased effects of linked selection and genetic drift as well as a higher propensity for gene flow. This positive association between genetic diversity and recombination rate has been observed in multiple species of animals, fungi, and plants (Branca et al., 2011; Cutter & Payseur, 2003, 2013; Dutoit, Vijay, et al., 2017; Kraft et al., 1998; Kulathinal et al., 2008; Moreira, Klicka, & Smith, 2023; Roesti et al., 2013; Takahashi, Liu, & Saitou, 2004; Thom et al., 2021; Vijay et al., 2016; Vijay et al., 2017).

In sliding windows across the genome, we found a strong negative correlation between genetic differentiation (F_{ST}) and genetic diversity, as expected in the presence of linked selection (Charlesworth, Nordborg, & Charlesworth, 1997). As such, F_{ST} also had a strong negative correlation with local recombination rates (Fig. 4). Notably, this relationship was largely driven by population pairwise comparisons with strong genetic differentiation (Fig. S7);

population comparisons with weak differentiation showed little relationship between genetic differentiation and recombination rate (Fig. S7). A negative relationship between local genomic differentiation and recombination rate has been identified in many other species, including birds, fruit flies, fishes, and monkeyflowers (Kulathinal et al., 2008; J. D. Manthey et al., 2021; Roesti et al., 2013; Stankowski et al., 2019; Vijay et al., 2016). We may expect this relationship due to increased effects of linked selection and genetic drift in genomic regions with low recombination rates (Nordborg, Charlesworth, & Charlesworth, 1996), particularly when gene flow is not strong enough to homogenize large genomic portions between differentiating taxa. As may be expected, genetic differentiation negatively covaries with genetic diversity across the genome (Fig. 4).

In our dataset, genetic divergence (D_{XY}) has a complicated relationship with recombination rate; when using the mean genetic divergence across all population pairwise comparisons measured, we identified a positive association between recombination rate and genetic divergence (Fig. 4; Fig. S9). In recent population splits, we may expect D_{XY} to reflect the ancestral population genetic diversity; consistent with these expectations, in each of our comparisons with recent population divergence, we found D_{XY} to be positively associated with genetic diversity, and both these measures positively associated with recombination rate (Fig. 4.; Fig. S7). In contrast, when we compared the two most divergent lineages (Central Mexico × Rocky Mountains) we found a negative association between recombination rate and genetic divergence (Fig. S7). In this comparison between more divergent lineages, there are varied correlations between diversity, divergence, and recombination rates across the genome, indicative of varied strengths of different evolutionary processes across the genome (discussed in Manthey et al., 2021).

Phylogeographic structure shaped by genomic architecture.

We used genetic clustering, ordination, and phylogenetic methods to assess phylogeographic structure in genomic sliding windows. We found that recombination rate variation was the strongest predictor of phylogeographic structure variation across the genome (Table 2). Consistent with our expectation, phylogeographic signal best representing the evolutionary history of the group was identified in genomic regions with relatively low recombination rate; this was consistent across all methods used (Fig. 2). Notably, the genomic windows with the lowest recombination rates deviated little from the genome-wide patterns (Fig. 5).

Previous work in insects, birds, and mammals has shown that phylogenomic signal varies across the genome (Edelman et al., 2019; Fontaine et al., 2015; G. Li et al., 2019; Martin et al., 2019; Thom et al., 2021). Generally, these studies have found or suggested that the background species tree is best represented by phylogenomic patterns exhibited in genomic regions with low recombination, and that gene flow among taxa is more prevalent in genomic regions with high recombination rate (Edelman et al., 2019; G. Li et al., 2019; Martin et al., 2019; Thom et al., 2021). Our results in the Brown Creeper echo these findings *at a shallow evolutionary timescale*, whereby we find highest support for the genome-wide supported phylogeographic relationships in genomic regions with low recombination rate (Fig. 2; Fig. S2). We may interpret these patterns in the context of lineage sorting during the speciation process, whereby genomic regions with low introgression exhibit quicker polymorphism fixation due to relatively increased effects of linked selection and genetic drift and reduced homogenization among lineages due to gene flow.

Genomic studies investigating intraspecific population and phylogeographic structure have more commonly used PCA-based methods (e.g., LOSTRUCT) to investigate variation in genomic structure across the genome. In two species of woodpeckers, a small portion of the variance in PCA-based phylogeographic structure was found to be associated with local recombination rate (Moreira et al., 2023). We found that variation in phylogeographic structure across the genome of the Brown Creeper—inferred using both ordination and genetic clustering

methods—was strongly associated with recombination rate variation (Fig. 2; Fig. S2; Table 2). We may infer that the same interactions between recombination rate variation and population genomics processes shaping variance in phylogenomic signal are also shaping variance in clustering- and ordination-based estimates of phylogeographic structure.

In contrast to generalizations about low recombination regions often reflecting a taxon's background evolutionary history, in some scenarios we may expect regions of extremely low recombination to greatly differ from genome-wide signatures of population genetic structure. In several taxa, including sunflowers, flies, fishes, and birds, the greatest deviations from the genome-wide pattern of intraspecific population structure (inferred with PCA) have been identified in genomic regions with extremely low recombination rates that are often inferred to be polymorphic inversions within populations or species (Hale, Campbell, & McKinney, 2021; Huang et al., 2020; Han Li & Ralph, 2019; Mérot et al., 2021; Perrier, Rougemont, & Charmantier, 2020; Shi et al., 2021; Todesco et al., 2020; Whiting et al., 2021). In these cases, if a taxon has relatively little population structure but also exhibits polymorphic inversions, we may expect the biggest deviations in population genetic structure from genome-wide patterns to reflect these polymorphic inversions that exhibit suppressed recombination. These cases present interesting patterns to think about when interpreting genomic phylogeography data. Genomic regions with low recombination may best represent a taxon's evolutionary history, excepting times when these regions are associated with large structural variants.

Interaction of recombination rate, population demography, and rate of evolution.

While we may expect that the interactions between recombination rate and population genomics processes will impact the rate of evolution across the genome, we did not simply observe little phylogeographic structure in genomic regions with high recombination rates and strong phylogeographic structure in genomic regions with low recombination rates. Conversely, we

found strong—and differing—patterns of phylogeographic structure across the genome (Fig. 1; Fig. 3; Fig. S8). Because these differing patterns of phylogeographic structure generally separated the Central American populations and sometimes the Sierra Madre Oriental population from all others, we hypothesized that populations with relatively smaller N_E exhibit relatively higher evolutionary rates in genomic regions with high recombination. Indeed, using relative RTD measures as a relative rate of evolution for each population, we found that larger populations exhibited reduced rates of molecular evolution in genomic windows with high recombination rates (Fig. 6; Fig. S10) and that shifts in relative rates of molecular evolution in high versus low recombination genomic regions were positively associated with effective population sizes (Fig. 6; Fig. S10). This is consistent with simulations by Tigano and colleagues (2021) that showed the interplay between recombination rate and selection causes greater genomic variability in evolutionary rates in larger populations relative to more consistent evolutionary rates across the genome in smaller populations.

We suspect that molecular evolution in genomic regions with high recombination rates is a product of complex interactions between population demography and the relative strengths of linked selection and genetic drift. In the Brown Creeper genome, recombination rate variation is positively associated with gene density (Fig. 2; Fig. 4) and we may interpret gene-dense genomic regions as targets for natural selection. Indeed, in a previous study in the Brown Creeper (J. D. Manthey et al., 2021), we showed that the two main Brown Creeper lineages exhibited relatively less neutral evolution in the gene-dense microchromosomes relative to the larger macrochromosomes. In gene-dense and high recombination regions under the effects of purifying selection, we may expect populations with relatively smaller N_E to accumulate substitutions faster than populations with higher N_E (Lanfear, Kokko, & Eyre-Walker, 2014), consistent with the trends we identified here in the Brown Creeper (Fig. 6; Fig. S11). Additionally, when we looked at potential mutational load, we found a weak but nonsignificant negative association between effective population size and mutational load accumulation in

genomic regions with high recombination rates (Fig. S12); this weak association also lends support to the idea that smaller populations have relatively less efficient purifying selection than larger populations in high recombination regions.

CONCLUSIONS

We used whole genome sequencing to estimate how genomic architecture shapes variation in phylogeographic structure across the genome of the Brown Creeper. We found that phylogeographic structure—as measured using ordination, phylogenetic, and clustering methods—was strongly associated with regional genomic variation in recombination rates. In low recombination regions, we recovered phylogeographic structure concordant with genome-wide patterns. The most divergent phylogeographic patterns were in high recombination regions, wherein populations with small effective population sizes were distinct from all other populations due to relatively faster evolution than larger populations in these high recombination regions. Because these high recombination regions are rich in coding genes, we suspect that large populations have higher relative effects of purifying selection in these regions, and overall slower relative evolutionary rates compared to smaller populations. Overall, our results show that phylogeographic structure may vary widely across the genome, and that effective population sizes of sampled populations and genomic architecture and their interactions will impact regional genomic variation in phylogeographic structure.

ACKNOWLEDGEMENTS

We would like to thank the collection managers, curators, and contributors for generous tissue loans for all individuals used in this study from the following museums: Burke Museum of Natural History and Culture, and University of Kansas Biodiversity Institute. Sequencing was

supported by Texas Tech University start-up funding to JDM and a generous donation from the Ferguson family to GMS and the Denver Museum of Nature and Science. Support was also provided by NSF awards #1953688 to JDM and #0814841 to GMS. The High-Performance Computing Center at TTU supported computational analyses.

DATA ACCESSIBILITY

All raw resequencing data are currently being uploaded to NCBI's SRA database. All code used for analysis on this project is available on GitHub: (github.com/jdmanthey/certhia_phylogeography).

AUTHOR CONTRIBUTIONS

JDM and GMS designed the study. JDM performed laboratory and bioinformatic work and wrote the first draft of the manuscript. JDM and GMS contributed to revising and improving the final draft of the manuscript.

TABLES AND FIGURES

Figure 1. Sampling map, genetic diversity, phylogeographic structure, and demography. (A) Sampling localities for this study. Colors for localities are consistent across plots. (B) Violin plots of nucleotide diversity across all 50 kbp sliding windows. White circles indicate means. (C) Consensus phylogeny using 19,639 gene trees from the 50 kbp windows. Support values are labelled for all branches showing the highest support from all estimation methods (N = 4), including consensus and ASTRAL phylogenies for both the 50 kbp and 100 kbp window datasets. At tips of phylogeny are ADMIXTURE results from 19,602 genome-wide SNPs (thinned every 50 kbp) for either two or seven assumed genetic clusters (k = 2 or k = 7, respectively). (D) Demographic history estimated with MSMC2. Different individuals are represented with differently colored lines and bootstraps are shown with thin lines. In (D), HM = harmonic mean. Brown Creeper photo in (A) by JDM from the Chiricahua Mountains in southern Arizona.

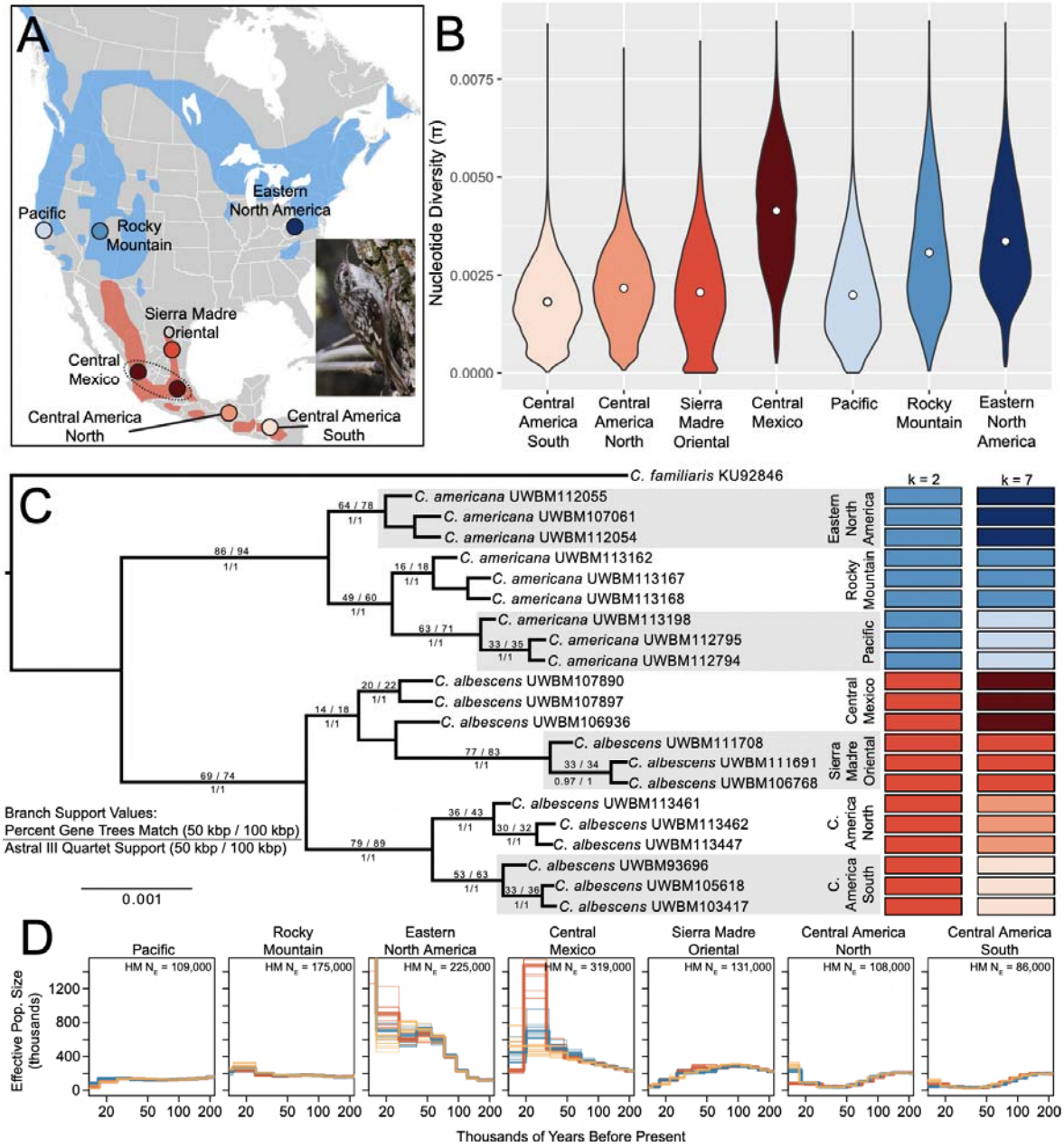


Figure 2. Genomic architecture and phylogeographic structure variation across the genome in 50kbp sliding windows. Gray points indicate window values and dark gray lines indicate means across 20 windows (i.e., 1 Mbp windows). Blue indicates top 2.5% of outliers and orange indicates bottom 2.5% of outliers for each statistic. Bottom outliers not shown for CDS, ADMIX., and GSI as these each have large proportions of values equal to zero. Abbreviations: recombination rate (ρ), gene content (CDS), transposable element and repetitive DNA content (TE), multidimensional scaling axis one from principal components analyses (-MDS1), mean genealogical sorting index for the two main *Certhia* lineages (1 - GSI), ADMIXTURE deviation relative to genome-wide analysis (ADMIX.), and mean genetic diversity across individuals as measured using observed heterozygosity (H_O). MDS1 is plotted as negative and GSI is plotted as (1 - GSI) so that deviations from expected phylogeographic structure for MDS1, GSI, and ADMIXTURE are all represented as higher values.

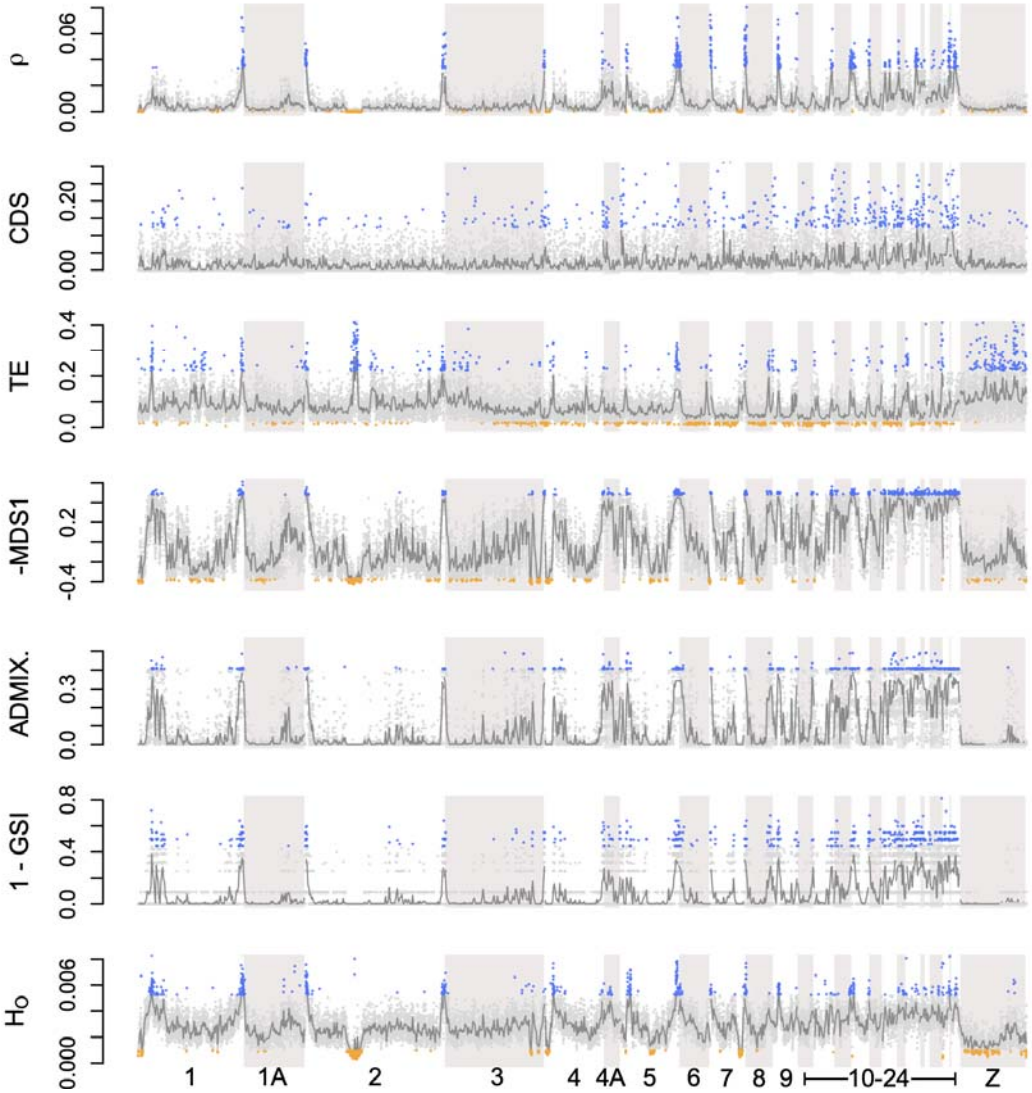


Figure 3. Example phylogeographic structure in a window that deviates from genome-wide patterns. This window had the 15th lowest MDS1 value of all windows from the principal components analysis (PCA) across the genome. (A) PCA, (B) RAxML phylogeny, and (C) ADMIXTURE results for a window on chromosome 4.

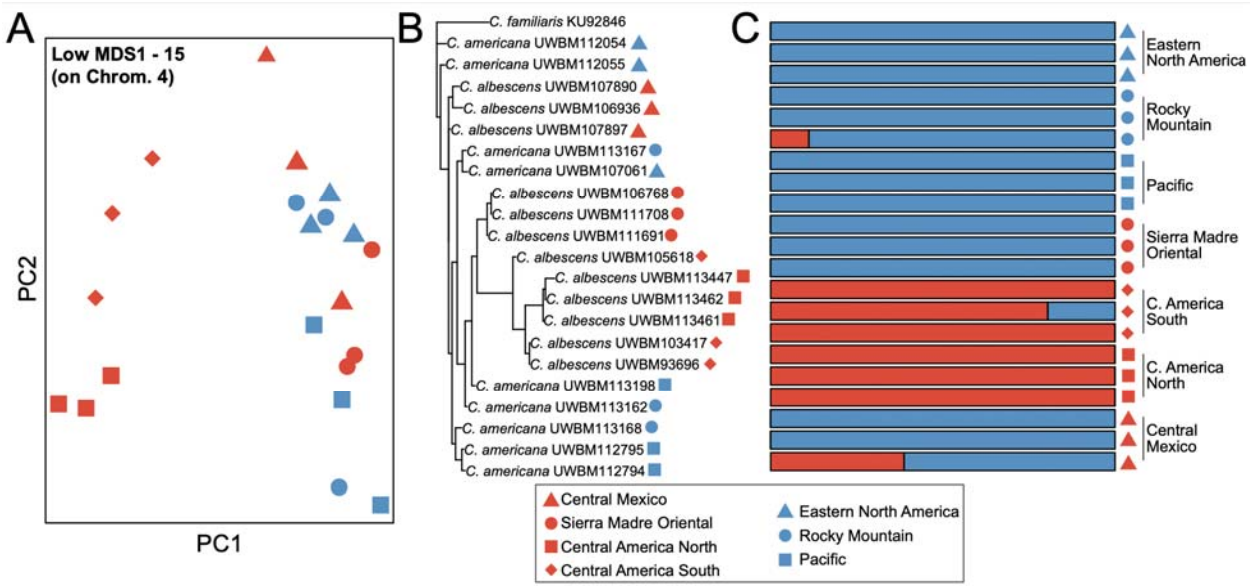


Figure 4. Correlation plot of genomic architecture, genomic diversity, and phylogeographic structure for 50 kbp windowed statistics. Pearson's and Spearman's r coefficients are plotted above and below the diagonal, respectively. Abbreviations shown in plot: recombination rate (ρ); GC content (GC%); gene content (CDS%); transposable element content (TE%); mean observed heterozygosity (H_O); mean nucleotide diversity (π); mean pairwise genetic divergence (D_{XY}); mean pairwise genetic differentiation (F_{ST}); genealogical sorting index at the population [1 - GSI (Pop.)] and lineage [1 - GSI (Lin.)] levels; negative multidimensional scaling axis one from principal components analyses (-MDS1); ADMIXTURE deviation relative to genome-wide analysis for two [ADMIX. (k = 2)] or seven [ADMIX. (k = 7)] assumed genetic clusters.

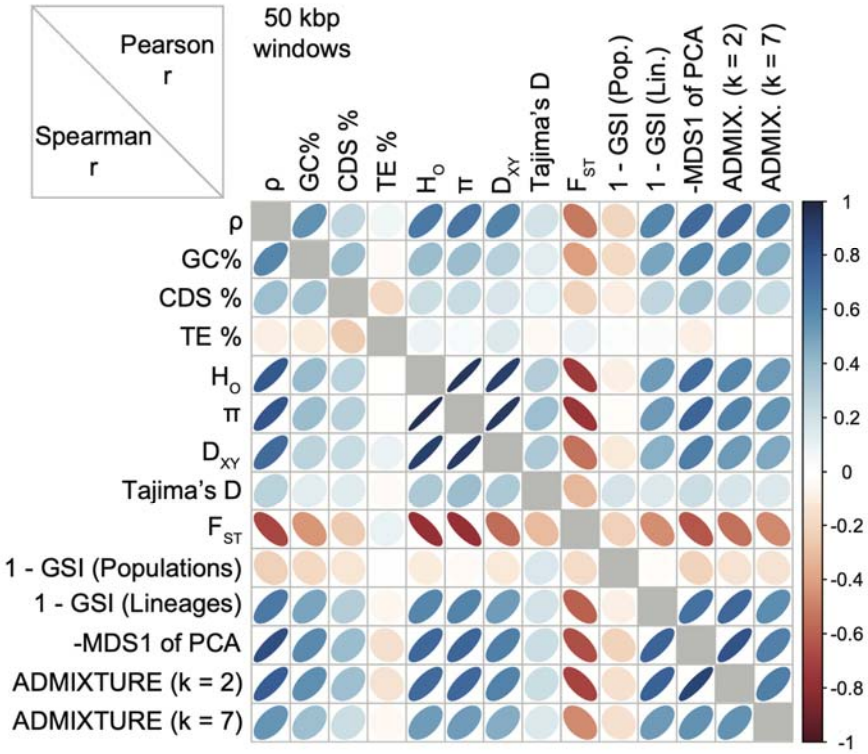


Figure 5. Violin plots of phylogeographic structure and genetic diversity variation in genomic regions with very high or low recombination rates. 50 kbp windows with the highest or lowest 2.5% of recombination rate values are included. Higher values for ADMIXTURE deviations, 1 - GSI, and -MDS1 indicate values deviating relatively more from the genome-wide patterns of phylogeographic structure. For nucleotide diversity (π), D_{XY}, and F_{ST}, values for each window are the mean across all populations (π) or population pairwise comparisons (D_{XY} and F_{ST}).

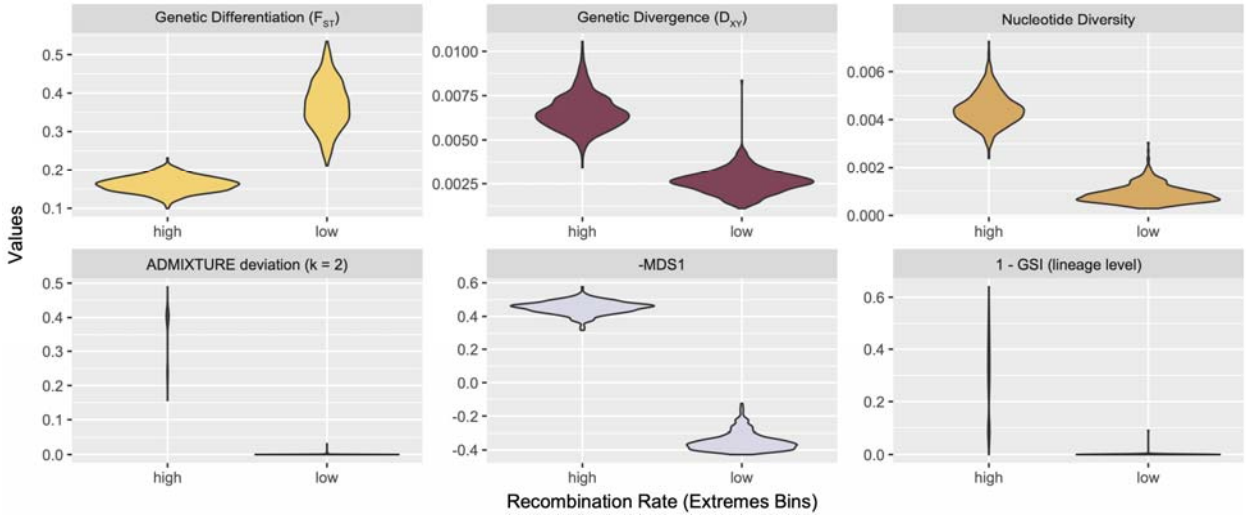


Figure 6. Relationship between effective population sizes and rate of evolution (measured by relative root-to-tip distance for each population) in genomic regions that vary in recombination rate. Extreme low and high recombination (rho) regions represent the quintiles of the genome with the lowest or highest recombination rates (i.e., highest 20% and lowest 20% of windows). Results for raw data are shown in (A-C) and for phylogenetic independent contrasts (PIC) in (D-F). The right panels in each row (C + F) show the shift in relative rates of evolution between high and low recombination genomic regions.

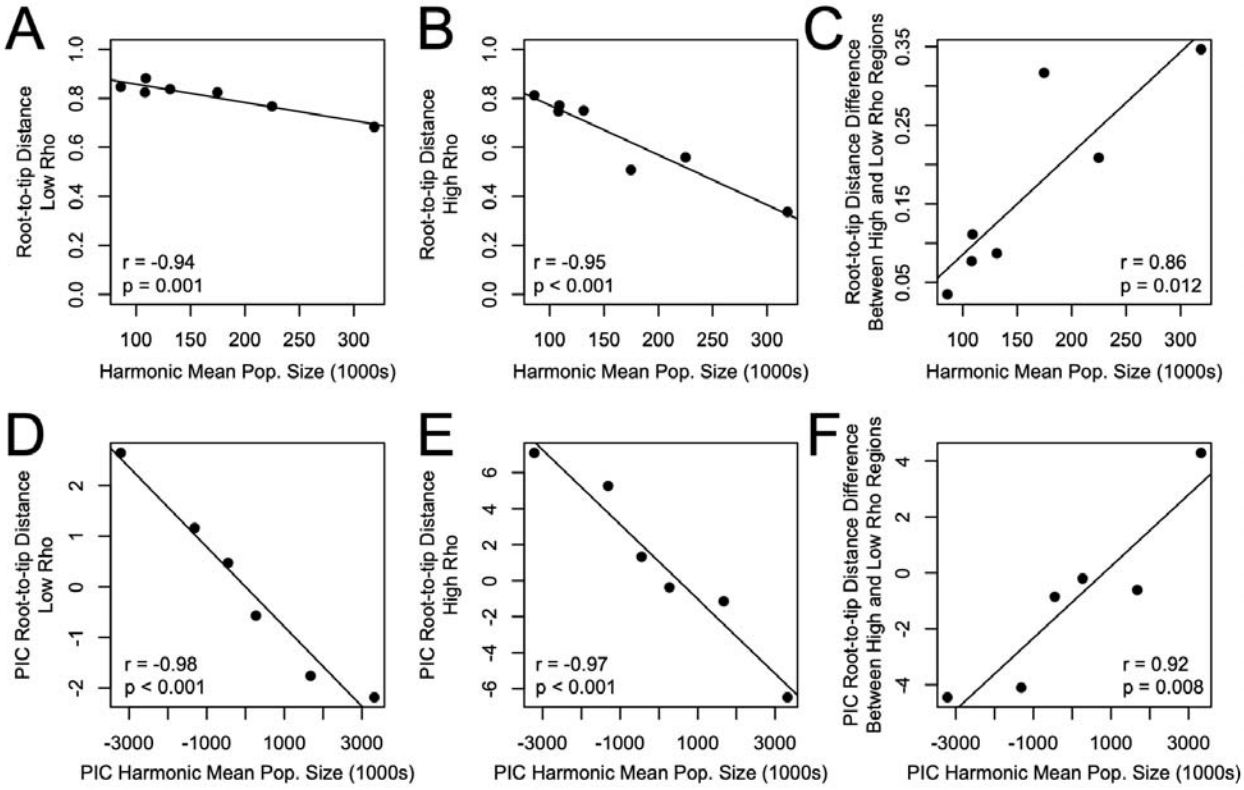


Table 1. Different datasets used in this study and their characteristics. For inclusion, all datasets required sites to have a minimum quality score of 20, minimum of 20 genotype quality, a minimum depth per individual for a site to be called of 6, and a max mean depth across individuals of 50.

Set	Missing Allowed	OG[1]	Window size	# Windows	Min. Sep. per SNP[2]	# SNPs	# Sites	Analyses
A	0%		-	-	50 kbp	19,602	19,602	Genome-wide ADMIXTURE
B	0%		-	-	100 kbp	9,838	9,838	Genome-wide ADMIXTURE
C	20%	X	50 kbp	19,713	-	35,463,408	890,335,872	RAxML, GSI, H_O, F_{ST}, D_{XY}, Tajima's D
D	20%	X	100 kbp	9,862	-	35,463,408	890,335,872	RaxML, GSI, H_O, F_{ST}, D_{XY}, Tajima's D
E	0%		-	-	-	21,657,220	21,657,220	LOSTRUCT
F	0%		50 kbp	19,697	-	21,657,220	21,657,220	ADMIXTURE per window
G	0%		100 kbp	9,860	-	21,657,220	21,657,220	ADMIXTURE per window
H	0%		-	-	-	*	*	MSMC2

[1]OG = outgroup included

[2]Min. Sep. per SNP = minimum separation between SNPs (i.e., thinning)

*SNPs and sites not labeled for MSMC2 analyses because these values varied per individual. For each individual, no missing data was allowed for that particular individual's MSMC2 demographic analyses input files. The MSMC2 dataset did not include the Z chromosome.

Table 2. Variance partitioning where the explanatory variables are characteristics of genomic architecture and response variables are different measures of population genetic structure. Values indicative of variance proportion explained by each explanatory variable or their interactions.

50 kbp windows						
Measure of Genetic Structure	ρ	CDS	TE	GC	Interactions	Total
MDS1	0.23	0.00	0.01	0.04	0.33	0.61
GSI (Lineages)	0.16	0.00	0.00	0.02	0.22	0.40
GSI (Populations)	0.02	0.00	0.00	0.01	0.03	0.06
ADMIXTURE (k = 2)	0.24	0.00	0.00	0.03	0.30	0.57
ADMIXTURE (k = 7)	0.19	0.00	0.00	0.01	0.18	0.38
Mean F_{ST}	0.13	0.00	0.01	0.01	0.15	0.30
100 kbp windows						
Measure of Genetic Structure	ρ	CDS	TE	GC	Interactions	Total
MDS1	0.22	0.01	0.01	0.03	0.41	0.68
GSI (Lineages)	0.16	0.00	0.00	0.01	0.23	0.40
GSI (Populations)	0.02	0.00	0.00	0.01	0.05	0.08
ADMIXTURE (k = 2)	0.25	0.00	0.00	0.02	0.35	0.62
ADMIXTURE (k = 7)	0.20	0.00	0.00	0.01	0.22	0.43
Mean F_{ST}	0.15	0.00	0.02	0.01	0.17	0.35

SUPPLEMENTAL TABLES AND FIGURES

Table S1. Sampling metadata and characteristics of sequencing amount and coverage when aligned to the reference genome. Samples from Utah, Morelos, Jalisco, and the outgroup were retrieved from the NCBI SRA and originally sequenced for a previous study by Manthey and colleagues (2021).

Museum #	**Lineage**	**Locality**	**Lat.**	**Long.**	**Raw bp**	**Filtered bp**	**Coverage**
UWBM103417	South	Copan, HN	14.866	-89.050	40,501,485,458	34,714,897,849	22.79
UWBM105618	South	Copan, HN	14.866	-89.050	30,149,924,172	26,420,616,801	18.55
UWBM93696	South	Copan, HN	14.866	-89.050	40,591,948,048	36,089,400,965	26.06
UWBM113447	South	Chiapas, MX	16.691	-92.606	37,741,499,908	33,063,229,986	23.32
UWBM113461	South	Chiapas, MX	16.691	-92.606	32,696,721,580	27,488,157,228	18.51
UWBM113462	South	Chiapas, MX	16.691	-92.606	48,519,014,074	42,841,474,607	27.09
UWBM106936	South	Morelos, MX	19.0873	-99.194	40,136,994,108	35,563,542,819	26.35
UWBM107890	South	Jalisco, MX	21.882	-103.865	43,601,103,228	38,563,138,625	26.14
UWBM107897	South	Jalisco, MX	21.882	-103.865	30,536,607,690	27,038,872,945	18.55
UWBM106768	South	Nuevo Leon, MX	24.872	-100.224	42,330,853,397	37,109,282,807	27.62
UWBM111691	South	Nuevo Leon, MX	24.872	-100.224	41,523,724,956	36,804,470,356	26.21
UWBM111708	South	Nuevo Leon, MX	24.872	-100.224	43,392,483,440	36,820,385,840	27.04
UWBM113198	North	California, US	36.245	-121.700	40,117,812,880	35,379,027,102	26.26
UWBM112794	North	California, US	36.245	-121.553	37,692,762,846	33,128,089,974	23.33
UWBM112795	North	California, US	36.245	-121.553	44,376,620,806	38,809,177,747	28.71
UWBM113162	North	Utah, US	37.317	-113.455	40,934,172,636	36,217,919,293	26.25
UWBM113168	North	Utah, US	37.317	-113.455	39,295,062,200	34,925,676,137	25.72
UWBM113167	North	Utah, US	37.377	-113.467	42,535,860,138	37,425,027,332	27.55
UWBM107061	North	West Virginia, US	38.289	-79.937	43,886,484,470	38,699,509,454	28.65
UWBM112055	North	West Virginia, US	38.289	-79.937	47,972,794,224	40,458,960,409	29.36
UWBM112054	North	West Virginia, US	38.641	-79.838	39,431,412,482	34,774,499,050	25.96
KU92846	Outgroup	England			21,417,629,506	19,195,682,239	13.58

Figure S1. Sequencing coverage aligned to the reference genome for each individual. The vertical red lines indicate the mean coverage per individual.

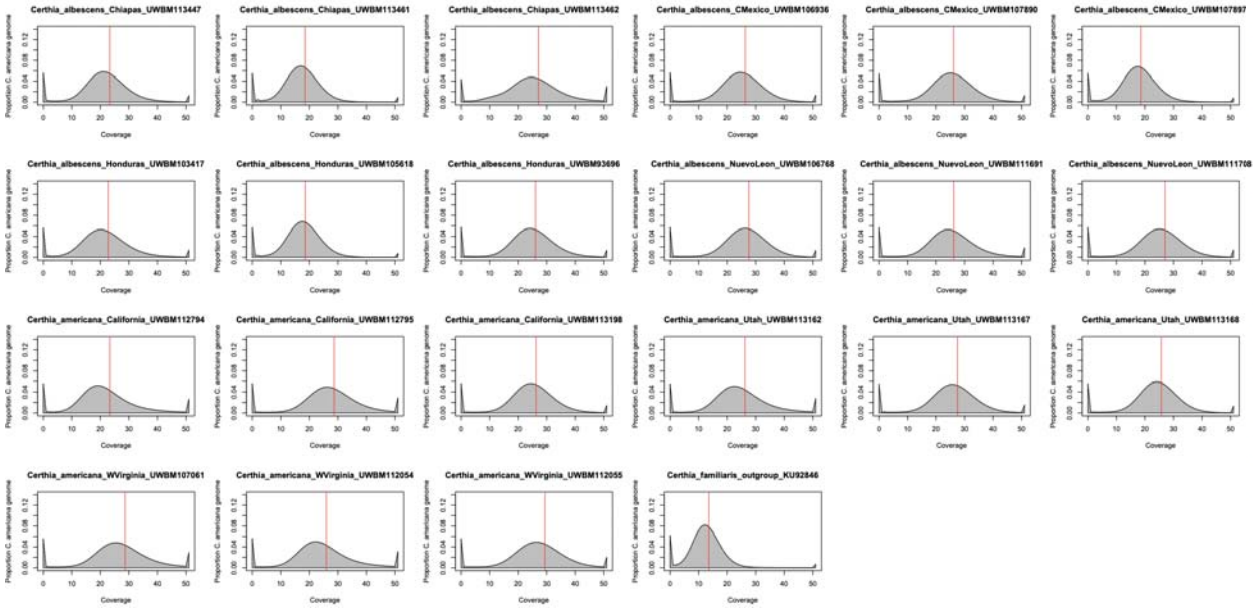


Figure S2. Genomic architecture and phylogeographic structure variation across the genome in 100kbp sliding windows. Gray points indicate window values and dark gray lines indicate means across ten windows (i.e., 1 Mbp windows). Blue indicates top 2.5% of outliers and orange indicates bottom 2.5% of outliers for each statistic. Bottom outliers not shown for CDS, ADMIX., and GSI as these each have large proportions of values equal to zero. Abbreviations: recombination rate (ρ), gene content (CDS), transposable element and repetitive DNA content (TE), multidimensional scaling axis one from principal components analyses (-MDS1), mean genealogical sorting index for the two main *Certhia* lineages (1 - GSI), ADMIXTURE deviation relative to genome-wide analysis (ADMIX.), and mean genetic diversity across individuals as measured using observed heterozygosity (H_O). MDS1 is plotted as negative and GSI is plotted as (1 - GSI) so that deviations from expected phylogeographic structure for MDS1, GSI, and ADMIXTURE are all represented as higher values.

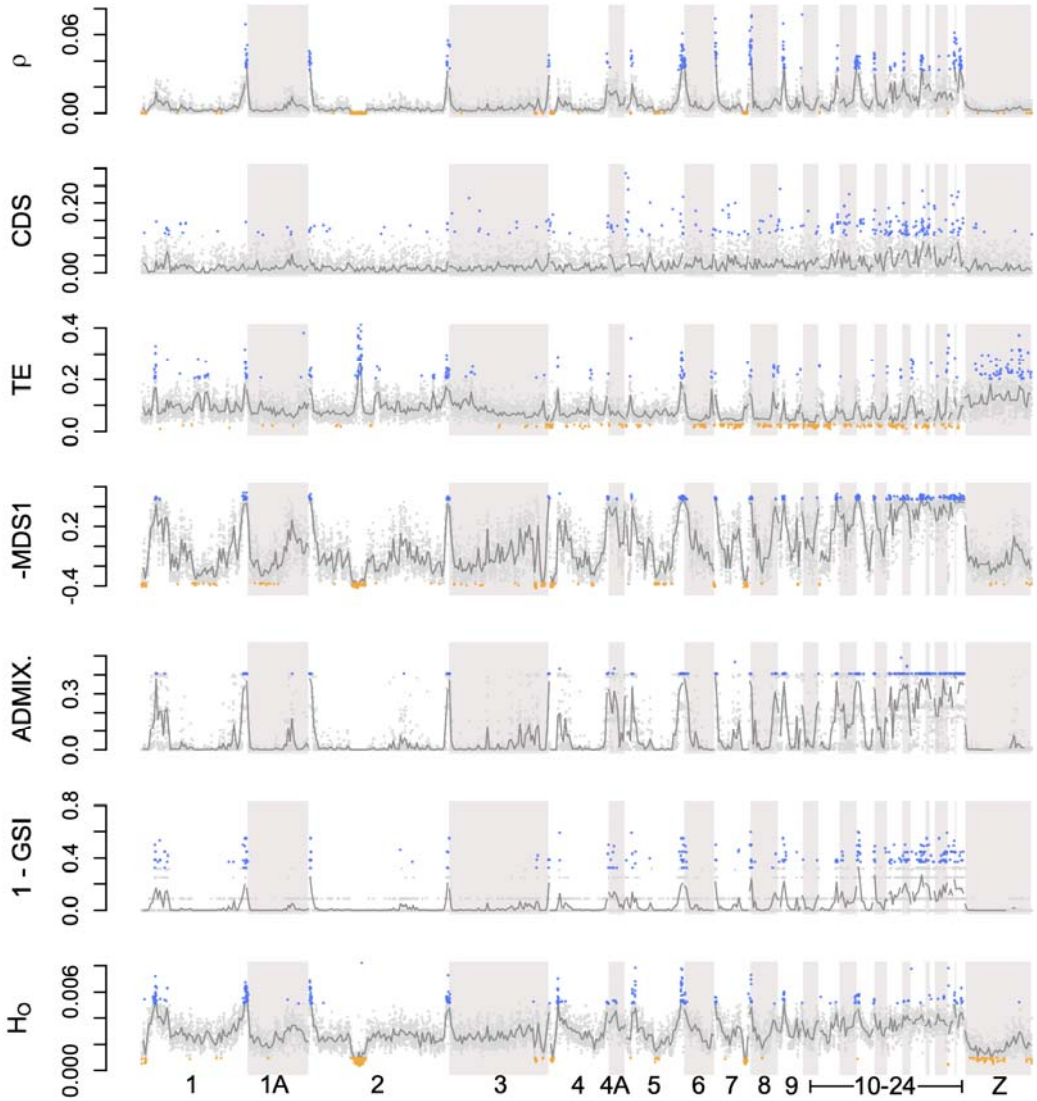


Figure S3. Nucleotide diversity variation across the genome in 100kbp sliding windows. Gray points indicate window values and dark gray lines indicate means across ten windows (i.e., 1 Mbp windows).

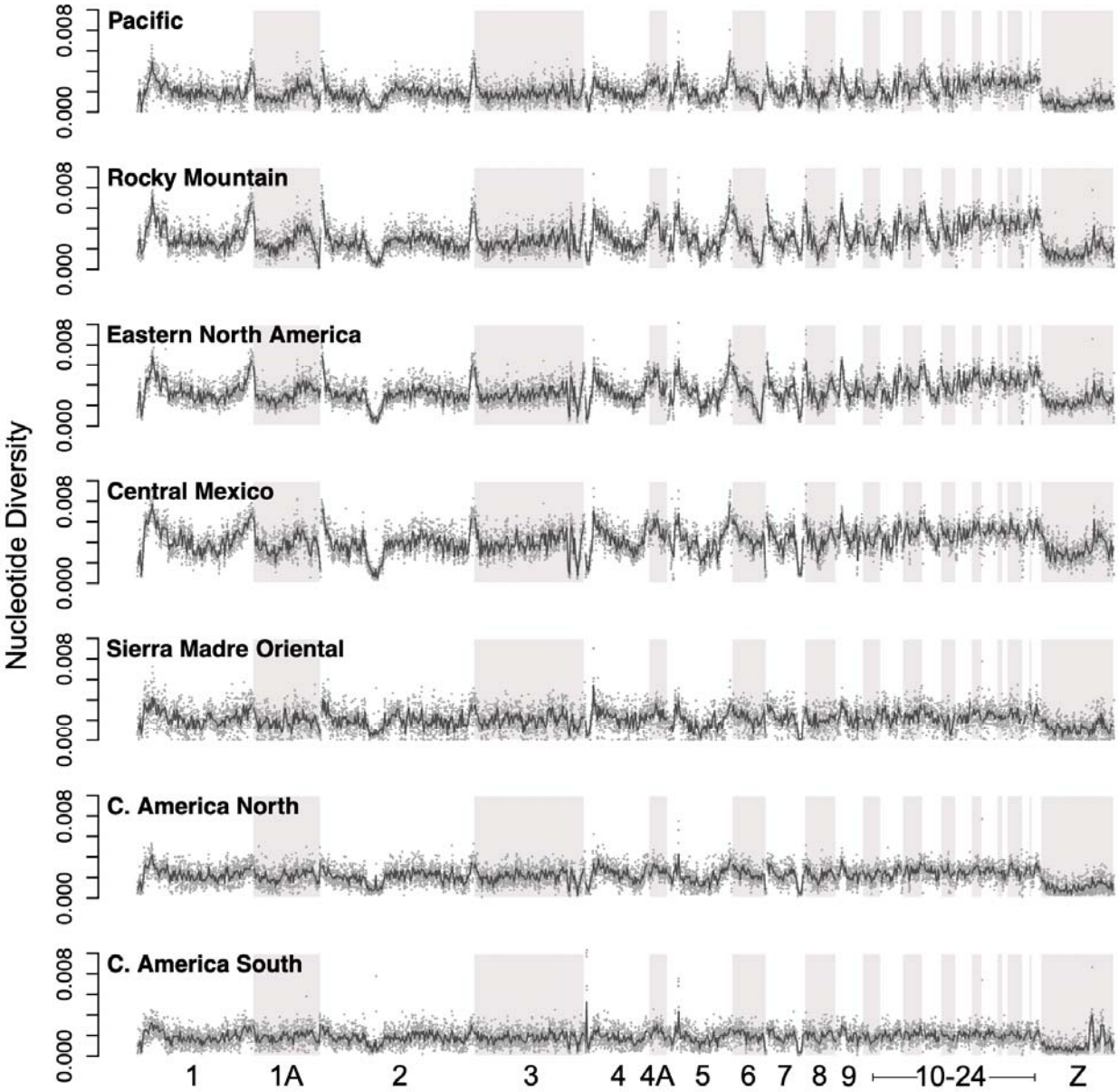


Figure S4. Tajima's D variation across the genome in 100kbp sliding windows. Gray points indicate window values and dark gray lines indicate means across ten windows (i.e., 1 Mbp windows).

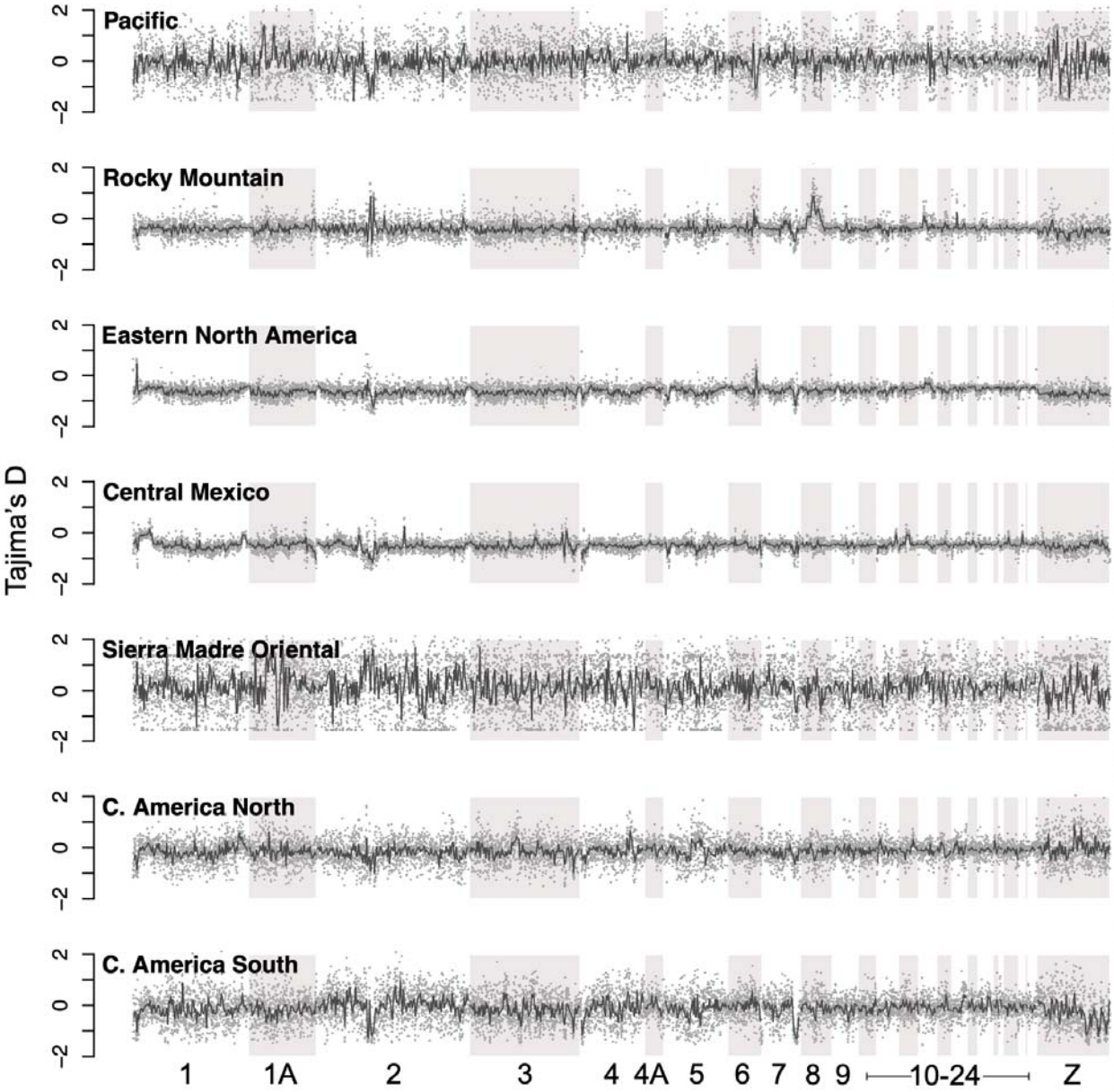


Figure S5. Genetic differentiation (F_{ST}) variation across the genome in 100kbp sliding windows. Gray points indicate window values and dark gray lines indicate means across ten windows (i.e., 1 Mbp windows).

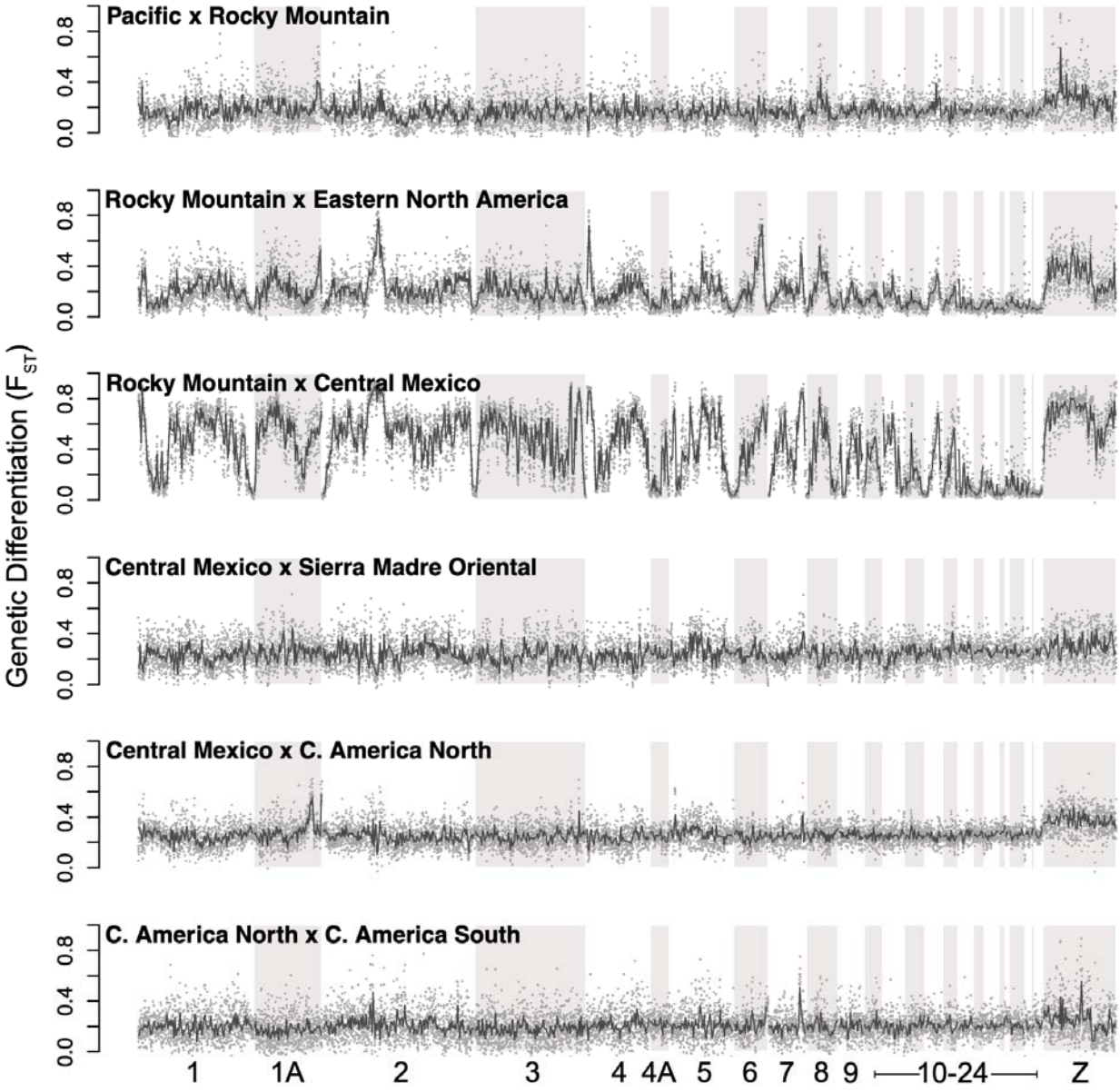


Figure S6. Genetic divergence (D_{XY}) variation across the genome in 100kbp sliding windows. Gray points indicate window values and dark gray lines indicate means across ten windows (i.e., 1 Mbp windows).

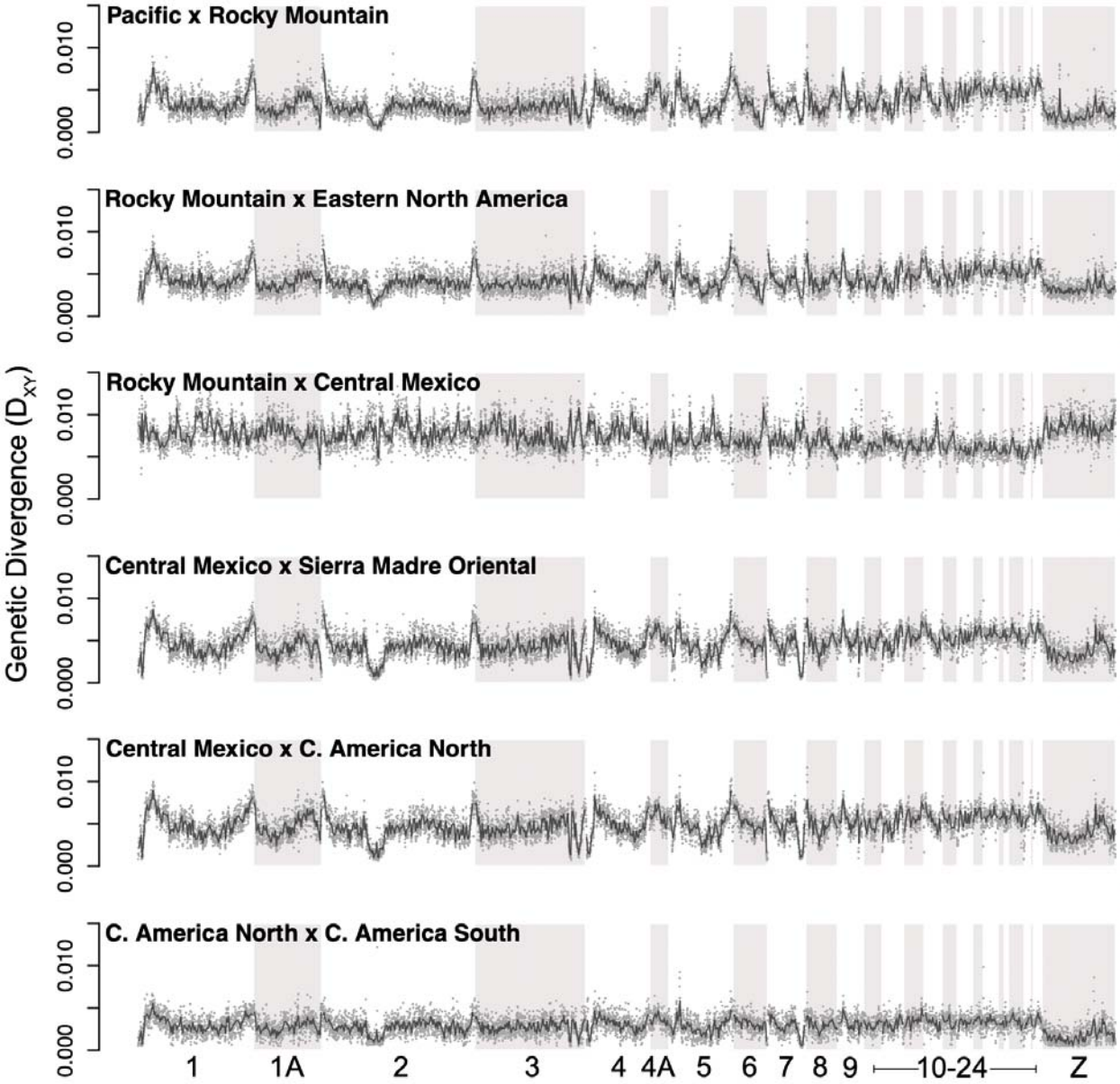


Figure S7. Correlations between genetic divergence (D_{XY}), genetic differentiation (F_{ST}), and recombination rate variation in 100 kbp sliding windows.

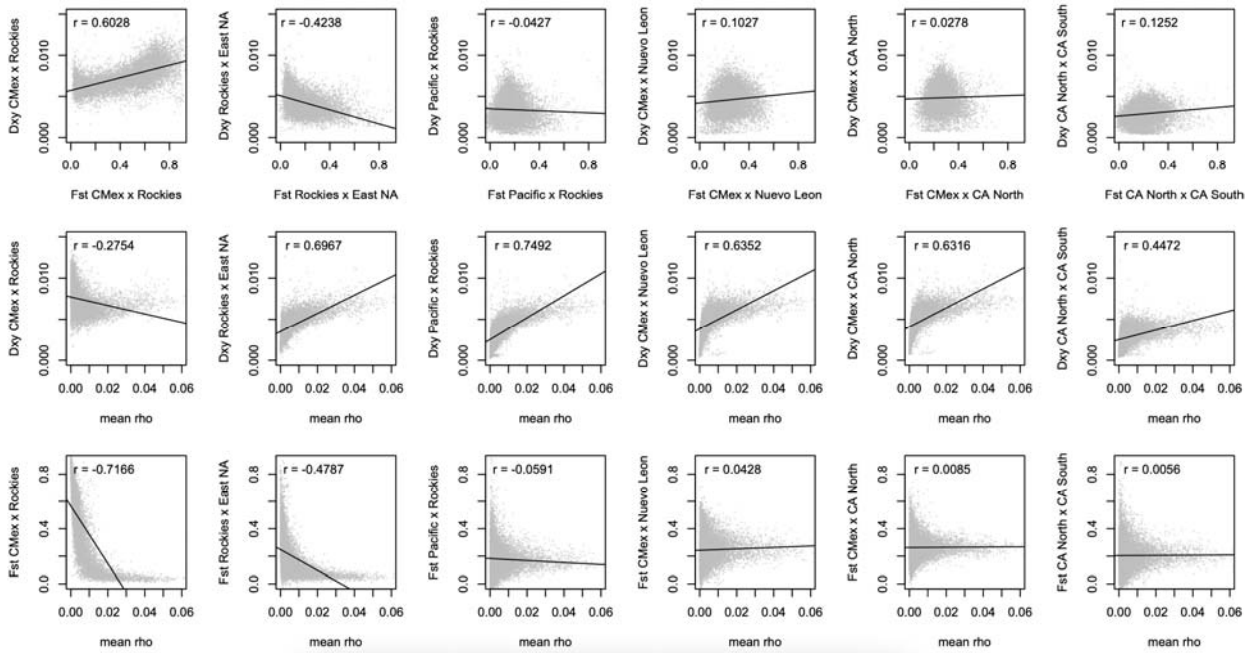


Figure S8. The 15 most extreme outlier principal components analysis (PCA) windows identified from those windows with low MDS dimension one values.

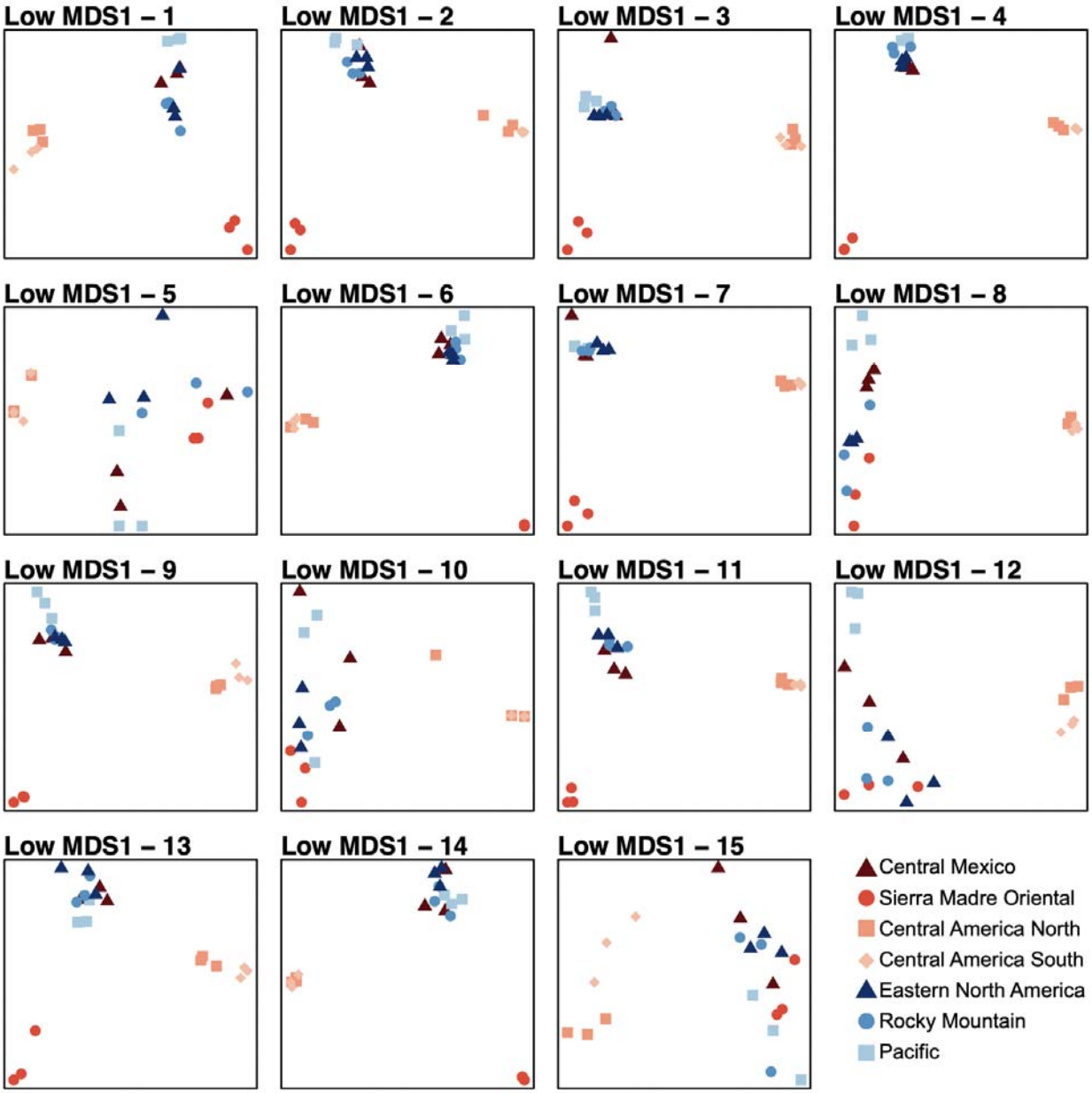


Figure S9. Correlation plot of genomic architecture, genomic diversity, and phylogeographic structure for (A) 50 kbp and (B) 100 kbp windowed statistics. Pearson's and Spearman's r coefficients are plotted above and below the diagonal, respectively. In (C), the absolute value difference between correlations in 50 kbp and 100 kbp windows. Abbreviations shown in plot: recombination rate (ρ); GC content (GC%); gene content (CDS%); transposable element content (TE%); mean observed heterozygosity (H_O); mean nucleotide diversity (π); mean pairwise genetic divergence (D_{XY}); mean pairwise genetic differentiation (F_{ST}); genealogical sorting index at the population [1 - GSI (Pop.)] and lineage [1 - GSI (Lin.)] levels; negative multidimensional scaling axis one from principal components analyses (-MDS1); ADMIXTURE deviation relative to genome-wide analysis for two [ADMIX. (k = 2)] or seven [ADMIX. (k = 7)] assumed genetic clusters.

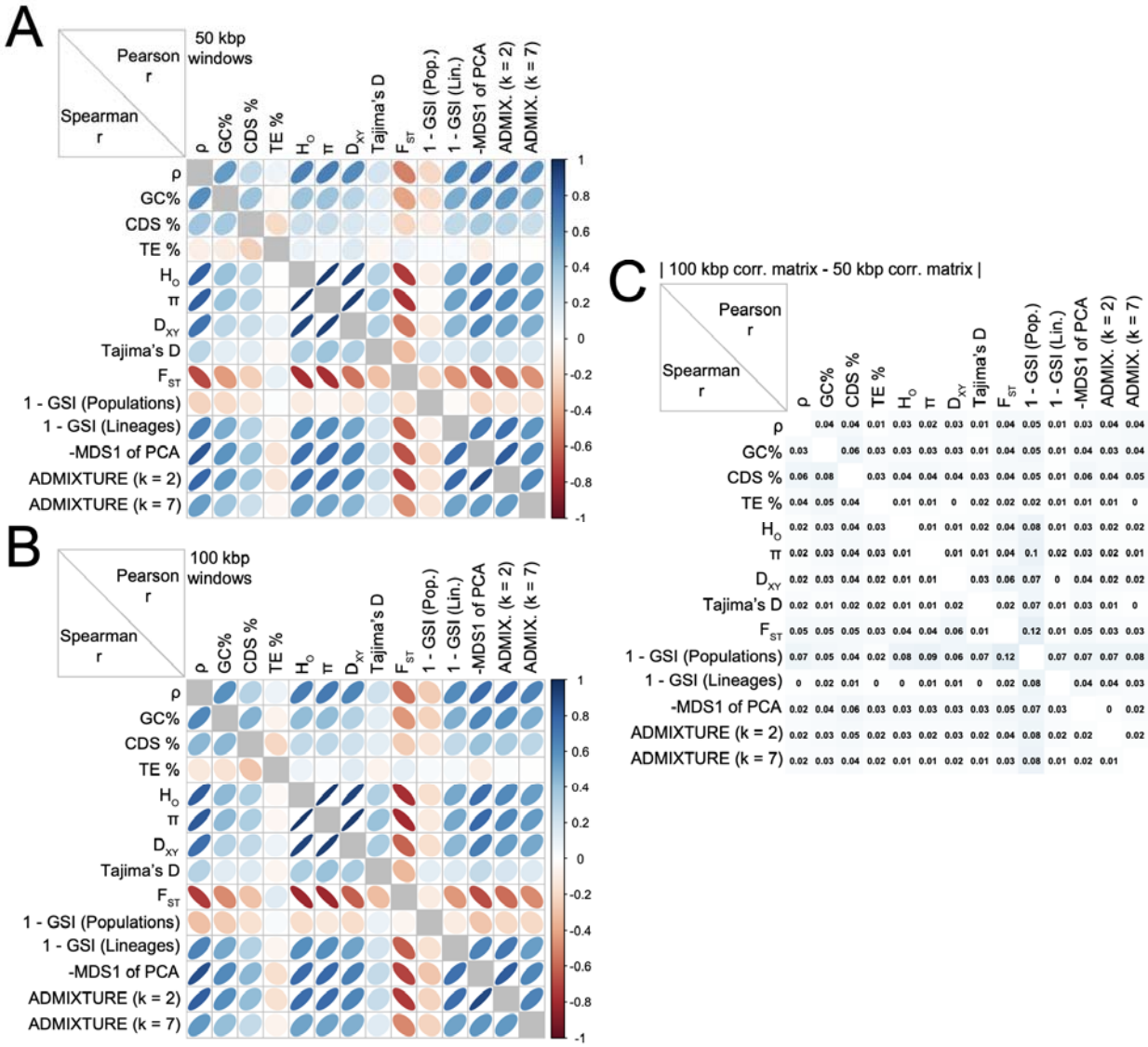


Figure S10. Violin plots of evolutionary rates (as measured by relative root-to-tip distances) for each population in the upper and lower genomic quintiles of high or low recombination (rho) values.

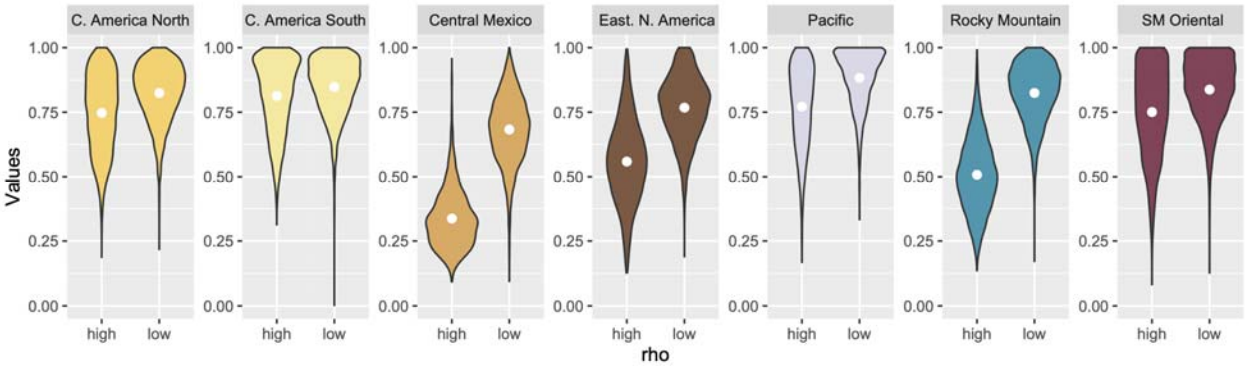


Figure S11. Demographic history estimated with MSMC2 for each individual. Thick lines indicate the whole genome estimate and thin lines indicate results from ten bootstrap replicates. Sample numbers correspond with those reported in Table S1.

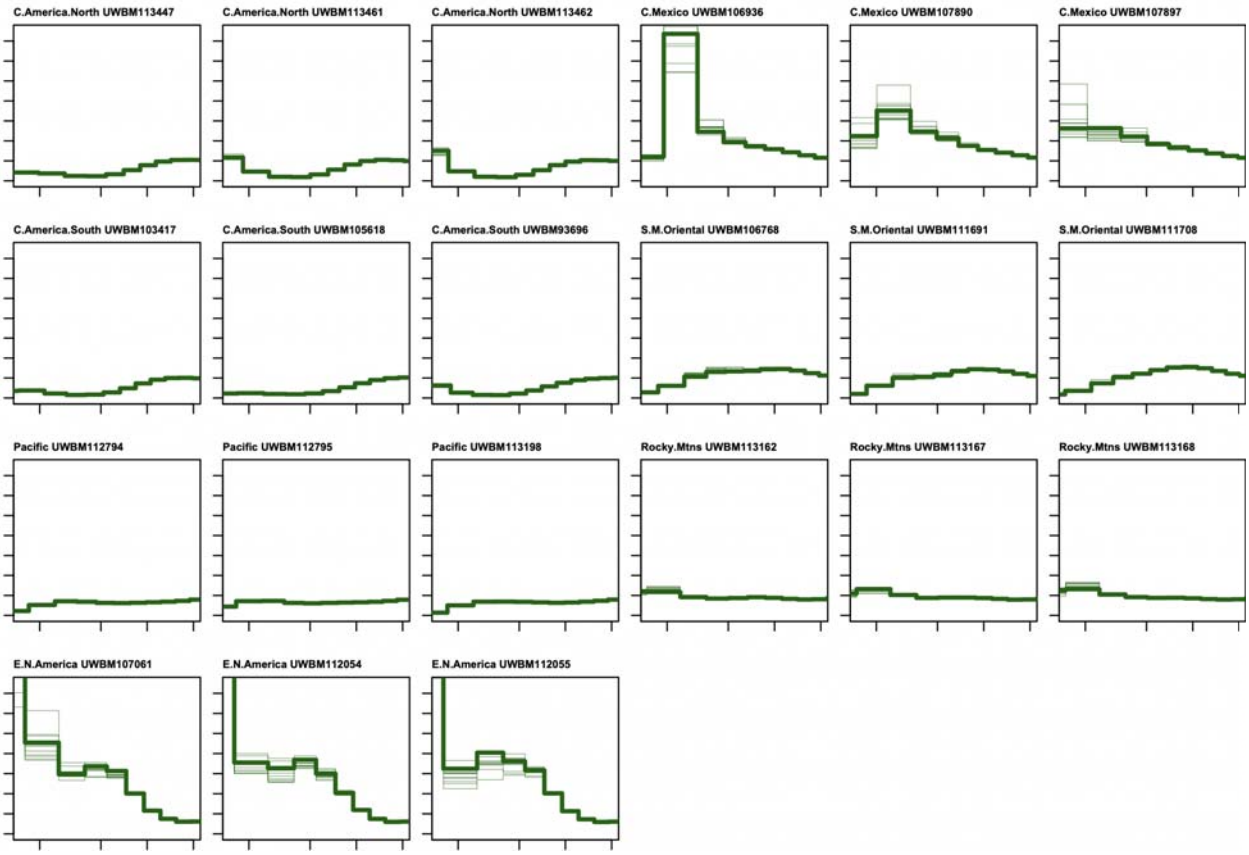
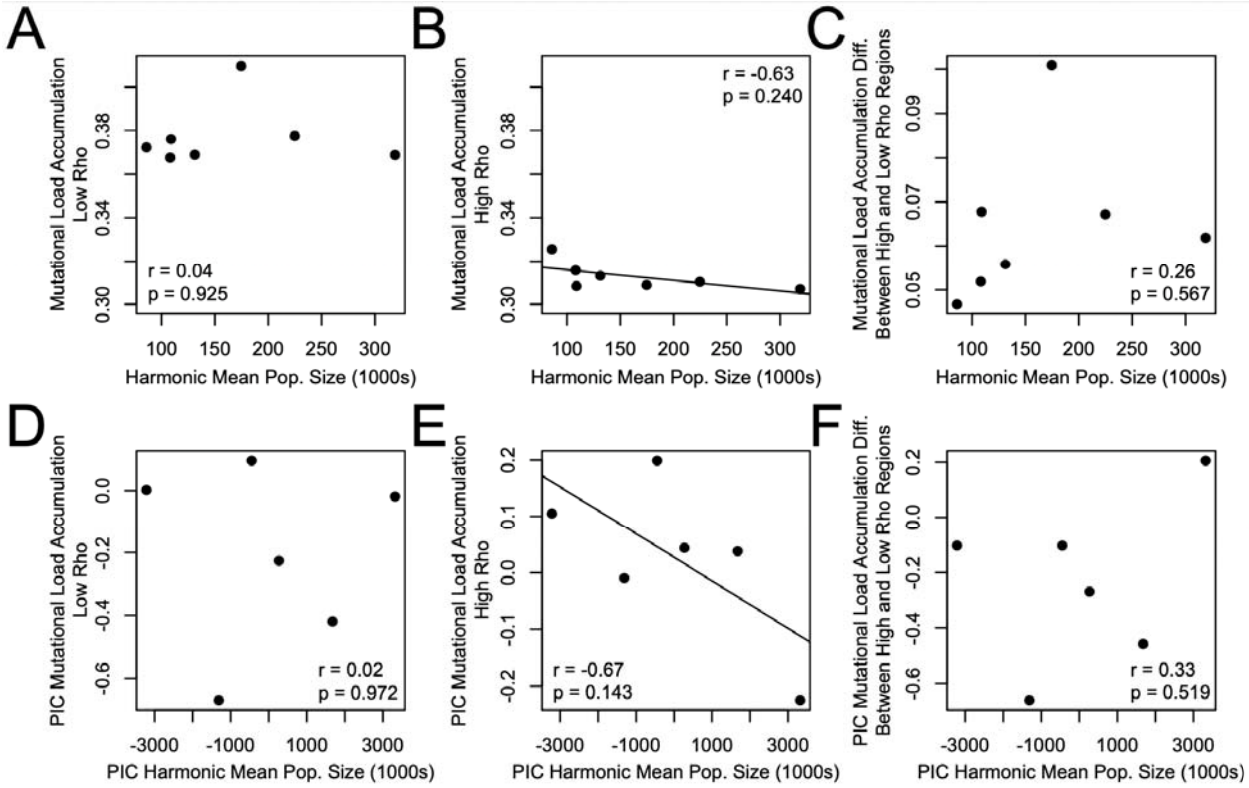


Figure S12. Relationship between effective population sizes and potential mutational load in genomic regions that vary in recombination rate. Extreme low and high recombination (rho) regions represent the quintiles of the genome with the lowest or highest recombination rates (i.e., highest 20% and lowest 20% of windows). Results for raw data are shown in (A-C) and for phylogenetic independent contrasts (PIC) in (D-F). The right panels in each row (C + F) show the shift in potential mutational load between high and low recombination genomic regions.



LITERATURE CITED

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research, 19*(9), 1655-1664.

Allmon, W. D. (1992). A causal analysis of stages in allopatric speciation. *Oxford surveys in evolutionary biology, 8*, 219-219.

Branca, A., Paape, T. D., Zhou, P., Briskine, R., Farmer, A. D., Mudge, J., . . . Gentzbittel, L. (2011). Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume Medicago truncatula. *Proceedings of the National Academy of Sciences, USA, 108*(42), E864-E870.

Bushnell, B. (2014). *BBMap: A Fast, Accurate, Splice-Aware Aligner*. Retrieved from

Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., . . . Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research, 18*(1), 188-196.

Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetics Research, 70*(2), 155-174.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., . . . Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly, 6*(2), 80-92.

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology, 23*(13), 3133-3157.

Cummings, M. P., Neel, M. C., & Shaw, K. L. (2008). A genealogical approach to quantifying lineage divergence. *Evolution, 62*(9), 2411-2422.

Cutter, A. D., & Payseur, B. A. (2003). Selection at linked sites in the partial selfer *Caenorhabditis elegans*. *Molecular Biology and Evolution, 20*(5), 665-673.

Cutter, A. D., & Payseur, B. A. (2013). Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature reviews genetics, 14*(4), 262-274.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Sherry, S. T. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158.

Dutoit, L., Burri, R., Nater, A., Mugal, C. F., & Ellegren, H. (2017). Genomic distribution and estimation of nucleotide diversity in natural populations: perspectives from the collared flycatcher (Ficedula albicollis) genome. *Molecular Ecology Resources, 17*(4), 586-597.

Dutoit, L., Vijay, N., Mugal, C. F., Bossu, C. M., Burri, R., Wolf, J., & Ellegren, H. (2017). Covariation in levels of nucleotide diversity in homologous regions of the avian genome long after completion of lineage sorting. *Proceedings of the Royal Society B: Biological Sciences, 284*(1849), 20162756.

Edelman, N. B., Frandsen, P. B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R. B., . . . Neafsey, D. E. (2019). Genomic architecture and introgression shape a butterfly radiation. *Science, 366*(6465), 594-599.

Ellegren, H. (2010). Evolutionary stasis: the stable chromosomes of birds. *Trends in Ecology & Evolution, 25*(5), 283-291.

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., . . . Qvarnström, A. (2012). The genomic landscape of species divergence in Ficedula flycatchers. *Nature, 491*(7426), 756.

Fontaine, M. C., Pease, J. B., Steele, A., Waterhouse, R. M., Neafsey, D. E., Sharakhov, I. V., . . . Kakani, E. (2015). Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science, 347*(6217), 1258524.

Haenel, Q., Laurentino, T. G., Roesti, M., & Berner, D. (2018). Meta□analysis of chromosome□scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. *Molecular Ecology, 27*(11), 2477-2497.

Hale, M. C., Campbell, M. A., & McKinney, G. J. (2021). A candidate chromosome inversion in Arctic charr (*Salvelinus alpinus*) identified by population genetic analysis techniques. *G3, 11*(10), jkab267.

Harrell, F. E., & Dupont, C. (2020). Hmisc: Harrell Miscellaneous (Version 4.3-1). R: R.

Hey, J., & Kliman, R. M. (2002). Interactions between natural selection, recombination and gene density in the genes of Drosophila. *Genetics, 160*(2), 595-608.

Huang, K., Andrew, R. L., Owens, G. L., Ostevik, K. L., & Rieseberg, L. H. (2020). Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Molecular Ecology, 29*(14), 2535-2549.

Kapusta, A., & Suh, A. (2017). Evolution of bird genomes—a transposon's□eye view. *Annals of the New York Academy of Sciences, 1389*(1), 164-185.

Kawakami, T., Smeds, L., Backström, N., Husby, A., Qvarnström, A., Mugal, C. F., . . . Ellegren, H. (2014). A high□density linkage map enables a second□generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Molecular Ecology, 23*(16), 4035-4058.

Koonin, E. V. (2009). Evolution of genome architecture. *The international journal of biochemistry & cell biology, 41*(2), 298-306.

Kraft, T., Säll, T., Magnusson-Rading, I., Nilsson, N.-O., & Halldén, C. (1998). Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (Beta vulgaris subsp. maritima). *Genetics, 150*(3), 1239-1244.

Kulathinal, R. J., Bennett, S. M., Fitzpatrick, C. L., & Noor, M. A. (2008). Fine-scale mapping of recombination rate in *Drosophila* refines its correlation to diversity and divergence. *Proceedings of the National Academy of Sciences, USA, 105*(29), 10051-10056.

Lanfear, R., Kokko, H., & Eyre-Walker, A. (2014). Population size and the rate of evolution. *Trends in Ecology & Evolution, 29*(1), 33-41.

Li, G., Figueiró, H. V., Eizirik, E., & Murphy, W. J. (2019). Recombination-aware phylogenomics reveals the structured genomic landscape of hybridizing cat species. *Molecular Biology and Evolution, 36*(10), 2111-2126.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics, 25*(14), 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079.

Li, H., & Ralph, P. (2019). Local PCA shows how the effect of population structure differs along the genome. *Genetics, 211*(1), 289-304.

Linck, E., & Battey, C. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources, 19*(3), 639-647.

Louis, M., Galimberti, M., Archer, F., Berrow, S., Brownlow, A., Fallon, R., . . . Rosel, P. E. (2020). Selection on ancestral genetic variation fuels parallel ecotype formation in bottlenose dolphins. *hal-03020428*.

Lucas, T. (2016). palettetown: Use Pokemon Inspired Colour Palettes. R package version 0.1.1. https://CRAN.R-project.org/package=palettetown.

Manthey, J., Bourgeois, Y., Meheretu, Y., & Boissinot, S. (2022). Varied diversification patterns and distinct demographic trajectories in Ethiopian montane forest bird (Aves: Passeriformes) populations separated by the Great Rift Valley. doi: 10.1111/mec.16417. *Molecular Ecology*.

Manthey, J. D., Klicka, J., & Spellman, G. M. (2011a). Cryptic diversity in a widespread North American songbird: phylogeography of the Brown Creeper (Certhia americana). *Molecular Phylogenetics and Evolution, 58*(3), 502-512.

Manthey, J. D., Klicka, J., & Spellman, G. M. (2011b). Isolation□driven divergence: speciation in a widespread North American songbird (Aves: Certhiidae). *Molecular Ecology, 20*(20), 4371-4384.

Manthey, J. D., Klicka, J., & Spellman, G. M. (2015). Chromosomal patterns of diversity and differentiation in creepers: a next-gen phylogeographic investigation of *Certhia americana*. *Heredity, 115*(2), 165-172.

Manthey, J. D., Klicka, J., & Spellman, G. M. (2021). The genomic signature of allopatric speciation in a songbird is shaped by genome architecture (Aves: *Certhia americana*). *Genome Biology and Evolution, 13*(8), evab120.

Martin, S. H., Davey, J. W., Salazar, C., & Jiggins, C. D. (2019). Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biology, 17*(2), e2006288.

Mayr, E., Mayr, E., Mayr, E., & Mayr, E. (1963). Animal species and evolution. In: Belknap Press of Harvard University Press Cambridge, Massachusetts.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., . . . Daly, M. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research, 20*(9), 1297-1303.

McVean, G., & Auton, A. (2007). LDhat 2.1: A package for the population genetic analysis of recombination. *Department of Statistics, Oxford, OX1 3TG, UK*.

Mérot, C., Berdan, E. L., Cayuela, H., Djambazian, H., Ferchaud, A.-L., Laporte, M., . . . Bernatchez, L. (2021). Locally adaptive inversions modulate genetic variation at different geographic scales in a seaweed fly. *Molecular Biology and Evolution, 38*(9), 3953-3971.

Moreira, L. R., Klicka, J., & Smith, B. T. (2023). Demography and linked selection interact to shape the genomic landscape of codistributed woodpeckers during the Ice Age. *Molecular Ecology*, DOI: 10.1111/mec.16841.

Nachman, M. W., & Payseur, B. A. (2012). Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1587), 409-421.

Nadachowska-Brzyska, K., Li, C., Smeds, L., Zhang, G., & Ellegren, H. (2015). Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. *Current Biology, 25*(10), 1375-1380.

Neuwirth, E. (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. https://CRAN.R-project.org/package=RColorBrewer.

Nordborg, M., Charlesworth, B., & Charlesworth, D. (1996). The effect of recombination on background selection. *Genetics Research, 67*(2), 159-174.

Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 367*, 332-342.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M. H. H., Oksanen, M. J., & Suggests, M. (2007). The vegan package. *Community ecology package, 10*, 631-637.

Pagès, H., Aboyoun, P., Gentleman, R., & DebRoy, S. (2017). Biostrings: Efficient manipulation of biological strings. *R Package Version 2.0*.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics, 20*(2), 289-290.

Perrier, C., Rougemont, Q., & Charmantier, A. (2020). Demographic history and genomics of local adaptation in blue tit populations. *Evolutionary applications, 13*(6), 1145-1165.

Poulin, J., D'Astous, É., Villard, M., Hejl, S., Newlon, K., McFadzen, M., . . . Ghalambor, C. (2020). *Brown Creeper (Certhia americana), version 1.0. In Birds of the World.* . In A. Poole (Ed.). doi:https://doi.org/10.2173/bow.brncre.01

Ravinet, M., Yoshida, K., Shigenobu, S., Toyoda, A., Fujiyama, A., & Kitano, J. (2018). The genomic landscape at a late stage of stickleback speciation: High genomic divergence interspersed by small localized regions of introgression. *PLoS Genetics, 14*(5), e1007358.

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature, 461*(7263), 489.

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution, 3*(2), 217-223.

Roesti, M., Moser, D., & Berner, D. (2013). Recombination in the threespine stickleback genome—patterns and consequences. *Molecular Ecology, 22*(11), 3014-3027.

Sayyari, E., & Mirarab, S. (2016). Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution, 33*(7), 1654-1668.

Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics, 46*(8), 919.

Shi, Y., Bouska, K. L., McKinney, G. J., Dokai, W., Bartels, A., McPhee, M. V., & Larson, W. A. (2021). Gene flow influences the genomic architecture of local adaptation in six riverine fish species. *bioRxiv*, DOI: https://doi.org/10.1101/2021.1105.1118.444736.

Smit, A., Hubley, R., & Green, P. (2015). RepeatMasker Open-4.0. 2013–2015. *Institute for Systems Biology*. *http://repeatmasker*. *org*.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*(9), 1312-1313.

Stankowski, S., Chase, M. A., Fuiten, A. M., Rodrigues, M. F., Ralph, P. L., & Streisfeld, M. A. (2019). Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers. *PLoS Biology, 17*(7), e3000391.

Sukumaran, J., & Holder, M. T. (2010). DendroPy: a Python library for phylogenetic computing. *Bioinformatics, 26*(12), 1569-1571.

Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., . . . Wang, J. (2017). Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Research, 46*(D1), D1083-D1090.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics, 123*(3), 585-595.

Takahashi, A., Liu, Y.-H., & Saitou, N. (2004). Genetic variation versus recombination rate in a structured population of mice. *Molecular Biology and Evolution*, *21*(2), 404-409.

Thom, G., Moreira, L. R., Batista, R., Gehara, M., Aleixo, A., & Smith, B. T. (2021). Genomic architecture controls spatial structuring in Amazonian birds. *bioRxiv*, DOI: https://doi.org/10.1101/2021.1112.1101.470789.

Tigano, A., Khan, R., Omer, A. D., Weisz, D., Dudchenko, O., Multani, A. S., . . . Fisher, H. (2021). Chromosome size affects sequence divergence between species through the interplay of recombination and selection. *bioRxiv*.

Todesco, M., Owens, G. L., Bercovich, N., Légaré, J.-S., Soudi, S., Burge, D. O., . . . Imerovski, I. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*, *584*(7822), 602-607.

Toews, D. P., Taylor, S. A., Vallender, R., Brelsford, A., Butcher, B. G., Messer, P. W., & Lovette, I. J. (2016). Plumage genes and little else distinguish the genomes of hybridizing warblers. *Current Biology*, *26*(17), 2313-2318.

Vijay, N., Bossu, C. M., Poelstra, J. W., Weissensteiner, M. H., Suh, A., Kryukov, A. P., & Wolf, J. B. (2016). Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature Communications*, *7*(1), 1-10.

Vijay, N., Weissensteiner, M., Burri, R., Kawakami, T., Ellegren, H., & Wolf, J. B. (2017). Genomewide patterns of variation in genetic diversity are shared among populations, species and higher□order taxa. *Molecular Ecology*, *26*(16), 4284-4295.

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., & Zemla, J. (2017). Package 'corrplot'. *Statistician*, *56*(316), e24.

Whiting, J. R., Paris, J. R., van der Zee, M. J., Parsons, P. J., Weigel, D., & Fraser, B. A. (2021). Drainage-structuring of ancestral variation and a common functional pathway shape limited genomic convergence in natural high-and low-predation guppies. *PLoS Genetics*, *17*(5), e1009566.

Wickham, H. (2011). ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*(2), 180-185.

Willing, E.-M., Dreyer, C., & Van Oosterhout, C. (2012). Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers. *PLoS One*, *7*(8), e42649.

Wu, C. I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, *14*(6), 851-865.

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, *19*(6), 153.