



Original Article

Pervasive Genomic Signatures of Local Adaptation to Altitude Across Highland Specialist Andean Hummingbird Populations

Marisa C. W. Lim[✉], Ke Bi, Christopher C. Witt, Catherine H. Graham, and Liliana M. Dávalos[✉]

From the Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY (Lim, Graham, and Dávalos); Museum of Vertebrate Zoology, University of California, Berkeley, CA (Bi); California Institute for Quantitative Biosciences, University of California Berkeley, Berkeley, CA (Bi); Museum of Southwestern Biology and Department of Biology, University of New Mexico, Albuquerque, NM (Witt); Swiss Federal Research Institute (WSL), Birmensdorf, Switzerland (Graham); and Consortium for Inter-Disciplinary Environmental Research, Stony Brook University, Stony Brook, NY (Dávalos).

Address correspondence to M. C. W. Lim at the address above, or e-mail: marisa_lim@pacbell.net.

Received October 28, 2020; First decision November 22, 2020; Accepted February 23, 2021.

Corresponding Editor: Mark Chapman

Abstract

Populations along steep environmental gradients are subject to differentiating selection that can result in local adaptation, despite countervailing gene flow, and genetic drift. In montane systems, where species are often restricted to narrow ranges of elevation, it is unclear whether the selection is strong enough to influence functional differentiation of subpopulations differing by a few hundred meters in elevation. We used targeted capture of 12 501 exons from across the genome, including 271 genes previously implicated in altitude adaptation, to test for adaptation to local elevations for 2 highland hummingbird species, *Coeligena violifer* ($n = 62$) and *Colibri coruscans* ($n = 101$). For each species, we described population genetic structure across the complex geography of the Peruvian Andes and, while accounting for this structure, we tested whether elevational allele frequency clines in single nucleotide polymorphisms (SNPs) showed evidence for local adaptation to elevation. Although the 2 species exhibited contrasting population genetic structures, we found signatures of clinal genetic variation with shifts in elevation in both. The genes with SNP-elevation associations included candidate genes previously discovered for high-elevation adaptation as well as others not previously identified, with cellular functions related to hypoxia response, energy metabolism, and immune function, among others. Despite the homogenizing effects of gene flow and genetic drift, natural selection on parts of the genome evidently optimizes elevation-specific cellular function even within elevation range-restricted montane populations. Consequently, our results suggest local adaptation occurring in narrow elevation bands in tropical mountains, such as the Andes, may effectively make them “taller” biogeographic barriers.

Subject Area: Molecular Adaptation and Selection

Key words: clinal variation, exon capture, gene flow, high-elevation

Steep gradients in environmental conditions create opportunities for local adaptation via a balance between locus-specific (e.g., natural selection) and genome-wide (e.g., gene flow, genetic drift) population genetic processes. In general, selection acts on specific regions of the genome, while demographic processes affect the genome uniformly (Lewontin and Krakauer 1973). Two important population processes—1) migration rate quantified by gene flow and 2) genetic drift—influence whether potentially adaptive alleles are lost or maintained in populations (Lenormand 2002; Yeaman and Whitlock 2011; Blanquart et al. 2012). Locally adapted populations should maintain alleles that confer fitness advantages through positive selection (Savolainen et al. 2013). The genetic composition of populations generally becomes more homogeneous with increasing gene flow, such that adaptive alleles are swamped out by the influx of alleles from other populations (Lenormand 2002). Adaptive alleles can also be lost by chance in small populations due to genetic drift. However, if the strength of natural selection overcomes the effects of migration and drift (migration-drift-selection balance), genetic differentiation can accumulate, resulting in signatures of locus-specific local adaptation (Yeaman and Whitlock 2011; Blanquart et al. 2012). Local adaptation can be measured as clinal variation in the frequency of single nucleotide polymorphisms (SNPs) in genes with biochemical or physiological functions related to adaptation to ecological conditions of the gradient.

Besides creating opportunity for local adaptation, the physical features of landscapes shape spatial population genetic variation (Weir 2009; Turchetto-Zolet et al. 2013). The Andes are outstanding in their rich species diversity and topographical complexity (Chapman 1926; Fjeldsø et al. 1999; Fjeldsø et al. 2012). For Andean animal populations, lowland barriers include rivers and arid inter-mountain valleys, while the high-elevation ridges reduce dispersal above tree-line (Weir 2009; Turchetto-Zolet et al. 2013). In the tropics, mountains are thought to be particularly strong barriers because they both physically impede dispersal and limit movement of tropical species that have evolved narrow physiological tolerances to temperature extremes (Janzen 1967; Ruggiero and Hawkins 2008; Fjeldsø et al. 2012). Specifically, spatial population genetic variation in Andean vertebrate species has been influenced by the high Andean ridge and a few key inter-mountain valleys, including the Marañón, Huallaga, and Apurímac river valleys in Peru (Cracraft 1985; Benham et al. 2015; Hosner et al. 2015; Benham and Witt 2016; Winger 2017; Hazzi et al. 2018; Prieto-Torres et al. 2018). While there is abundant evidence that topography promotes genetic isolation and population structure, the interplay of evolutionary processes driving this separation remain unclear.

It is challenging to disentangle the genetic processes resulting in spatial genetic variation without accounting for both the genetic structure of populations across their geographic range and potential for local adaptation. Targeted high-throughput sequencing methods provide cost-effective ways to obtain genome-wide samples for many individuals to evaluate whether present genetic differentiation was shaped by genome-wide population processes (migration or drift) and/or locus-specific processes (natural selection) (Luikart et al. 2003). A targeted approach can help distinguish highly homogeneous regions of the genome within a population from others that are distinct due to local adaptation to particular environmental conditions (Luikart et al. 2003; Chapman et al. 2013; Hämälä et al. 2018; Herman et al. 2018). Even in wide-ranging species with high dispersal ability and consequently greater potential for

homogenization of the gene pool, there is evidence for locus-specific local adaptation resulting from an equilibrium between migration and natural selection (Cheviron and Brumfield 2009; Galen et al. 2015; Graham et al. 2018).

Species that occur along elevational gradients face variable environmental and physiological extremes along the gradient, including lower oxygen availability (hypobaric hypoxia), greater exposure to UV radiation, more intense desiccation, and lower ambient temperature. For example, the partial pressure of oxygen declines by approximately 10% for every 1000 m increase in elevation and this is critical for small endotherms, like hummingbirds, that sustain high metabolic rates. This makes oxygen uptake difficult and causes a corresponding reduction in arterial blood oxygen saturation. Evidence of a strong gradient in natural selection and consequent local adaptation along elevational gradients could explain the origin and maintenance of narrow elevational ranges themselves. In the past decade, numerous studies have investigated the genetic basis of physiological and biochemical responses to elevation for a variety of taxa across the world's mountains (Storz et al. 2009; McCracken et al. 2010; Muñoz-Fuentes et al. 2013; Projecto-García et al. 2013; Galen et al. 2015; Natarajan et al. 2016). There is also a growing list of candidate genes involved in the hypoxia response, DNA or cell damage, and energy metabolism, among other functions (Yi et al. 2010; Simonson et al. 2012; Huerta-Sánchez et al. 2013; Yanhua Qu et al. 2013; Qu et al. 2015; Zhang et al. 2016; Graham et al. 2018; Lim et al. 2019). Most studies, however, have focused on comparing species or populations at elevation extremes rather than along the gradient.

Here, we tested for local genetic adaptation to elevation within the elevational ranges of 2 Andean hummingbird species (family Trochilidae) that are montane specialists: the violet-throated starfrontlet (*Coeligena violifer*) and the sparkling violetear (*Colibri coruscans*). Both *Coe. violifer* and *Col. coruscans* diverged from their sister species ~5 Ma, when the Peruvian Andes had attained most of their current height (Gregory-Wodzicki 2000; Hoorn et al. 2010; McGuire et al. 2014). Thanks to differences in their geographic distribution and habitat requirements, they may have evolved to highland conditions via different genetic mechanisms. *Coe. violifer* is more restricted to high-elevation edges or clearings in cloud and elfin forest habitats between 2500–3900 m (Schulenberg et al. 2007; del Hoyo et al. 2020). In Peru, there are 3 subspecies defined by plumage differences: *Coe. v. dichrourea* in northern and central Peru, *Coe. v. albicaudata* in the Apurímac drainage region, and *Coe. v. osculans* in southeast Peru (Schuchmann and Zuchner 1997) (Supplementary Figure S1a). In contrast, *Col. coruscans* is found in both natural and urban landscapes, typically from ~2000 to 4500 m, but occasionally as low as sea-level (Schulenberg et al. 2007; Züchner et al. 2020) (Supplementary Figure S1b). There is only one subspecies, *Col. c. coruscans*, in Peru.

Using targeted exon capture sequencing, we generated a dataset of coding sequences sampled across the hummingbird genome, including genes previously implicated in elevation-related adaptation. With this population genomic dataset, we characterized population genetic structure and gene flow, and evaluated SNP frequencies for trends with elevation. Genetic associations between SNPs and elevation, particularly in previously identified candidate genes and for both species, would provide strong evidence for parallels in local adaptation and potentially explain how elevation ranges are themselves maintained. The absence of clinal genetic variation with elevation could indicate gene flow or genetic drift dilutes or removes beneficial alleles from the population.

Materials and Methods

Study System

Our study included 62 *Coe. violifer* and 101 *Col. coruscans* specimens sampled from the Peruvian Andes, encompassing the elevational range for both (Figure 1). *Coe. violifer* individuals were sampled from elevations of 2000 to 4000 m, while *Col. coruscans* individuals were sampled from 1800 to 4200 m. Tissue samples were collected at sites across Peru under an approved Animal Use Protocol from the University of New Mexico Institutional Animal Care and Use Committee (Protocol number 08UNM033-TR-100117; Animal Welfare Assurance number A4023-01) and under permits from the management authority of Peru (76-2006-INRENA-IFFS-DCB, 087-2007-INRENA-IFFS-DCB, 135-2009-AG-DGFFS-DGEFFS, 0377-2010-AG-DGFFS-DGEFFS, 0199-2012-AG-DGFFS-DGEFFS, and 006-2013-MINAGRI-DGFFS/DGEFFS). Complete specimen data are available via the ARCTOS online database (Supplementary Table S1).

Probe Design

We used probes designed by NimbleGen for the exon capture experiment (NimbleGen SeqCap EZ kit; Roche, Pleasanton, CA). We selected exons for capture using the Anna's hummingbird (*Calypte anna*) genome (Gilbert et al. 2014). The genome-wide sample of targeted exons was expected to comprise markers neutral with respect to elevation, thereby allowing for analysis of population genetic structure. To maximize probe hybridization efficiency, we filtered exons by length and GC content. We limited exon lengths to 150–600bp, as shorter loci will have insufficient probe coverage and SNPs on longer loci are effectively linked and counted as a single marker. Probe hybridization efficiency is reduced when GC content is too low or too high, so we retained exons with 35–70% GC content (Bi et al. 2012). We ran RepeatMasker to remove sequences containing low complexity and short repeats using the abblast search engine and chicken as the DNA reference (Smit et al. 2015). Exon sequences with N's were also removed as they lower the probability that probes will hybridize to target DNA. From this filtered set of sequences, we chose the longest coding sequence per protein-coding gene in the *Cal. anna* genome ($n = 16\,000$ annotated genes). As a result, the probe

design included a genome-wide sample of 12 230 exon sequences. To increase enrichment uniformity and coverage of the targets, we added a flanking region of 100bp to the start of each exon.

We targeted 271 genes listed in the high-elevation adaptation literature to ensure they were included in the dataset (Supplementary Table S2). These candidate genes were identified from a range of highland vertebrate taxa (e.g., mammals, birds, amphibians, fish) and have an array of functions potentially related to elevation adaptation, including response to hypoxia, metabolism of carbohydrates and lipids, inflammation and immune response, and roles in cellular respiration. When available, sequences for 40 candidate genes were extracted from transcriptomes generated for the 2 study species to improve probe specificity (Lim et al. 2019). The remaining 231 candidate gene sequences were extracted from the Anna's hummingbird reference genome (Gilbert et al. 2014). We also included 2 mitochondrial genes (*COX1*, *COX3*) in the probe design. Since mitochondrial DNA is more abundant than nuclear DNA, the proportion of probes for mitochondrial to nuclear genes was reduced by NimbleGen to achieve consistent coverage of all targets (1:72).

Exon Capture

Genomic DNA was extracted from liver tissue for 134 individuals using the Zymo Genomic DNA Tissue MiniPrep for Solid Tissue kit (Zymo, Irvine, CA). We also included DNA extractions from the Museum of Southwestern Biology for 29 individuals, bringing the total to 163 individuals. The DNA quality of these extractions was assessed prior to inclusion in our study. Extractions were diluted in 1xTE prior to DNA fragmentation. Extractions were sonicated with the Bioruptor until the DNA was in the 250–400bp range with very few fragments exceeding 600bp. For DNA library preparation, we followed the Meyer and Kircher (2010) protocol and used unique dual indexes. To reduce PCR bias, we combined products from 2 sets of indexing PCR reactions for each sample. Libraries were then pooled in groups of 48 for the exon capture procedure.

For the capture experiment, we followed the NimbleGen SeqCap EZ protocol. We used 1.1µg of DNA from the pooled genomic libraries and chicken COT DNA (1:10 COT to library) to block nonspecific hybridization in the capture reaction. We verified that the capture reaction DNA fragment size range was between 150–500bp

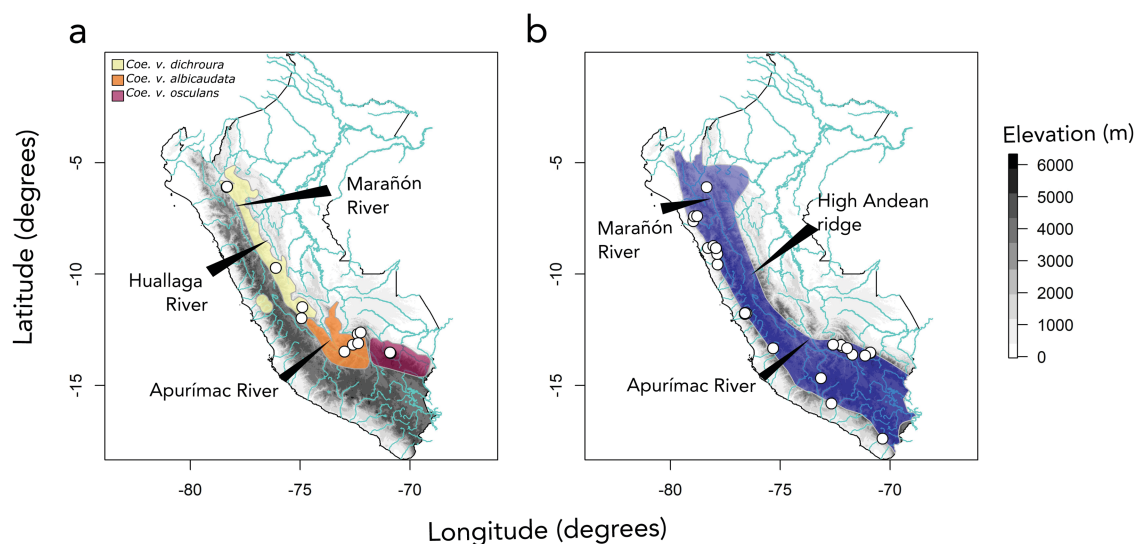


Figure 1. Specimen localities plotted on an elevation map with major waterways and range maps for (a) *Coe. violifer* subspecies and (b) *Col. coruscans*.

by comparing pre- and post-capture DNA Bioanalyzer traces. We also compared qPCR results for 6 positive (exon sequence targeted) and 6 negative (exon sequence not targeted) control primers between pre- and post-capture reactions. Finally, post-capture enrichment libraries derived from each hybridization were quantified and combined for sequencing. To minimize bias from different sequencing lanes and to maximize sequence coverage, the pooled capture reaction was split equally across 2 lanes on an Illumina HiSeq 4000 with 150bp paired-end sequencing.

Bioinformatics

After raw sequence reads were de-multiplexed, we used a custom-designed pipeline to filter reads, build species-specific reference assemblies, align reads to the assemblies, evaluate alignment quality, and call SNPs for population genetic analyses. We used Trimmomatic v 0.36 (Bolger et al. 2014) to remove adapter sequences and low-quality reads shorter than 36bp; SuperDeduper to remove PCR duplicates (<https://ibest.github.io/HTStream/#https://github.com/ibest/SuperDeduper>); and alignment to the *Escherichia coli* genome with Bowtie2 to remove potential bacterial contamination sequences (Langmead and Salzberg 2012). Overlapping paired-end reads were merged with Flash v1.2.11 (Magoč and Salzberg 2011). We created *de novo* species-specific assemblies as references for mapping reads, rather than mapping to the *Cal. anna* genome. We chose this species-specific reference approach to maximize the percentage of aligned reads for downstream analysis, even though it meant we could not directly compare gene sequences for the 2 species because assembly length and number of exons captured differed per species. Efforts to generate a hybrid *de novo* assembly with sequences from both species failed to increase the percentage of aligned reads (~40–50%) to an acceptable threshold (>60%).

To build the species-specific references, we chose six individuals per species that had the most exon capture data and generated *de novo* assemblies with SPAdes using default k-mer sizes (Bankevich et al. 2012). The SPAdes assemblies were combined to make one reference assembly per species. We mapped unique reads to the species-specific references with NovoAlign (<http://www.novocraft.com/products/novoalign/>), added read groups with Picard v2.1.1 (<http://broadinstitute.github.io/picard/>), and realigned reads around indels with GATK v3.6 (McKenna et al. 2010). After alignment, we evaluated the data with the following statistics: number of raw, filtered, and mapped nucleotides; percentage of reads retained after sequence quality filtering; percentage of reads aligned to exon target region (specificity), percentage of exon targeted regions covered by at least one read (sensitivity), average read coverage, variation in read coverage, and percentage of nucleotide sites retained at multiple read coverage depths (2×, 5×, 10×, 20×) (Supplementary Table S4). We used SAMtools/bcftools v1.3.1 (Li et al. 2009) to generate raw variant call format (VCF) files containing all possible variant and non-variant sites. We filtered these sites with SNPcleaner v.225 (Bi et al. 2013) to remove loci with too much missing data prior to downstream analyses. For this site filtering step, we retained sites for which at least 70% of the samples had a minimum of 3× coverage to reduce the effects of missing data. We also removed sites that were within 10bp of indels, were not biallelic, or had excessive heterozygosity (removed sites with P -value < 0.0001).

For calculating allele frequency, genotype likelihood estimation, and genotype calling, we used the analysis of next-generation sequencing data package, ANGSD (Korneliussen et al. 2014). ANGSD incorporates uncertainty from base calling and alignment errors into

genotype likelihoods with posterior probabilities for genotype calls, which is especially useful for low-to-medium coverage sequencing datasets. The scripts for our pipeline and downstream analyses are available at <https://github.com/marisalim/HbirdSeqCap> and <https://github.com/CGRL-QB3-UCBerkeley/seqCapture>.

Population Genetic Structure

We estimated 2 measures of within-locality diversity: the proportion of segregating sites (Watterson's θ , θ_w) (Watterson 1975) and nucleotide diversity (θ_π) (Nei 1987). For each species, individuals were divided into geographic clusters based on specimen localities (Figure 1, Supplementary Table S1; referred to as sites A-M in Figure 2 and text). For average θ_w and θ_π , we used ANGSD to generate per locality site frequency spectra (with respect to minor allele; folded) and custom scripts to calculate these statistics.

We conducted several analyses to assess genetic structure between populations: PCA, test for relatedness, test for admixture, and between-locality differentiation. PCA results help detect potential outlier samples, which may exaggerate downstream measures of genetic differentiation between samples. For PCA analysis, we computed a genotype covariance matrix based on genotype likelihoods with ngsCovar from the ngsTools package (Fumagalli et al. 2014). The inclusion of closely related individuals can bias analyses that estimate population genetic structure, as family groups are likely to cluster together, forming their own “population” (Anderson and Dunham 2008). To determine whether our samples included closely related individuals, we estimated the coefficient of relatedness (θ ; siblings: $\theta > 0.50$, half-siblings: $\theta > 0.25$) with NgsRelate using genotype likelihoods from ANGSD (Korneliussen and Moltke 2015). To test for admixture, we used NgsAdmix, which relies on genotype likelihoods (Skotte et al. 2013). For each species, we tested an estimated number of subpopulations (K) from 1 to 5 with 10 iterations each.

We calculated F_{ST} and D_{xy} as 2 independent measures of genetic differentiation between populations based on geographic locality because each is subject to calculation biases depending on levels of current (F_{ST}) or ancestral sequence variation (D_{xy}) (Cruickshank and Hahn 2014). $F_{ST} = 0$ indicates complete mixing and $F_{ST} = 1$ indicates complete isolation between populations. D_{xy} measures genetic sequence divergence between populations and indicates whether populations are more or less diverse compared to each other. These measures of genetic differentiation between localities served as independent estimations of population genetic structure for comparison to PCA and admixture results, which have no *a priori* population assignment. F_{ST} was calculated in ANGSD from the shared site frequency spectra for each pairwise combination of localities (2dSFS; (Korneliussen et al. 2014). For D_{xy} , we identified variant and invariant sites for each locality with ANGSD and calculated D_{xy} with a custom script (<https://github.com/CGRL-QB3-UCBerkeley/seqCapture/blob/master/scripts/popStats.pl>). We compared spatial genetic variation to species and subspecies range boundaries with IUCN Red List maps (IUCN 2016).

Isolation by Distance

To incorporate information about the geography and distribution of study species in the Peruvian Andes with the population genetic structure patterns, we tested for isolation by distance. We calculated geographic distance in 2 ways. First, we calculated pairwise geodesic distances, which measure the shortest distance along a curved path as opposed to a straight line Euclidean distance using the R package lmap (Wallace 2012). Second, we calculated cost distances

between individuals weighted by species distribution using the R package *gdistance* (van Etten 2017), and species distribution models for each species (Fajardo et al. 2014). Paths through regions with lower species suitability cost more than regions where the species is predicted to occur. We compared the 2 geographic distances to the pairwise distance between genotypes using genetic distances calculated from genotype likelihoods in *ngsDist* (Korneliussen et al. 2014; Vieira et al. 2016). We conducted Mantel tests with 10 000 permutations to estimate the correlation between genetic and geographic distances.

To visualize isolation by distance patterns on a map, we used the Estimated Effective Migration Surfaces (EEMS) method (Petkova et al. 2016). EEMS uses a stepping-stone model to estimate migration rates along every edge of a user-defined grid of subpopulations (demes) and interpolates across the region of interest. These maps provide a means to test for, and identify, barriers to gene flow without prior knowledge of landscape or geography. Since migration rates from unsampled regions of the map are interpolated from regions with samples, we focus our interpretation on results in sampled areas. This method uses the same pairwise genetic distance matrix as calculated above for the isolation by distance analysis. We ran EEMS with 100 demes, 4×10^6 MCMC iterations, and 10^6 burn-in iterations sampled every 9999 iterations. We checked for convergence by looking for stationarity in MCMC chain trace plots. Following Petkova et al. (2016), we adjusted proposal variance values on migration and diversity rate parameters to achieve proposal acceptance rates between ~20–30%.

Latent Factor Mixed Models

For selection test analyses, genotype calls (as opposed to genotype likelihoods) are required to determine which allele each individual carries at SNPs associated with an environmental variable. We used ANGSD to call genotypes (Korneliussen et al. 2014). Genotype calling is sensitive to errors from nucleotide base assignment and alignment. Therefore, we conducted sensitivity analyses for a combination of several SNP filter thresholds to maximize the number of SNPs retained for genotype calling. We used the “hetBias_pval” filter to minimize over-calling heterozygote genotypes, measured as deviations from Hardy-Weinberg equilibrium (test: default/not used, 10^{-3} , 10^{-4}). “SNP_pval” filters out polymorphic sites that may result from sequencing error (test: 10^{-3} , 10^{-4} , 10^{-5} , 10^{-6}). Finally, we tested different posterior distribution thresholds for calling a genotype with the “postCutOff” filter (test: 0.75, 0.85, and 0.95). After assessing broad parameter behavior, we conducted a reduced sensitivity analysis of the filters for the study species. Next, we filtered the genotype call dataset to remove all SNPs with minor allele frequencies below 5%, since rare variants are unlikely to be associated with an environmental variable.

We tested for associations between SNPs and elevation using the genotype call dataset with the Latent Factor Mixed Models (LFMM) program (v1.5) (Frichot et al. 2013). We tested other outlier approaches, but these methods resulted in either too many false positives (PCAdapt), skewed F_{ST} distributions for genetically structured populations (OutFLANK), or low statistical power (Pearson’s correlation). LFMM is statistically robust and infers both the gene-environment association, as well as the underlying population structure from the data, which is modeled with unobserved latent factors (K). LFMM can be used to test for associations between environmental measures and allele frequency inputs.

To calculate allele frequencies, we grouped samples into bins based on the elevation at which they were collected and their genetic population assignment (Supplementary Table S1). We tested increments of 400 m for *Coe. violifer* and 300 m or 400 m for *Col. coruscans*. These increments were chosen to spread individuals as evenly as possible across bins. Fourteen *Coe. violifer* samples were dropped from analysis to ensure bins did not mix samples from genetically distant populations (Supplementary Table S1). Bins were standardized by the mean and standard deviation elevation for all individuals per species. We used discrete elevation bins instead of continuous elevation values in order to calculate allele frequencies from genotype calls for individuals in each bin as follows: $\frac{\text{sum of genotypes}}{\text{total number of genotypes} * 2}$. We tested parameter sensitivity with allele frequency inputs and set the latent factor to $K = 1, 2$, or 3 with 5 iterations per test. Per author recommendation, the allele frequency mode works better for lower latent factor values (E. Frichot, personal communication). Each iteration was sampled 250 000 times with 25 000 burn-in cycles from the Gibbs Sampler algorithm. Z-score outputs from the 5 iterations were combined following LFMM documentation protocol. To control for confounding effects on the number of significant SNPs identified by LFMM, we followed author guidelines by evaluating the effect of different genomic inflation factor (λ) thresholds on P -value distributions. The P -value distributions are expected to be flat with a peak at 0 and λ close to 1. LFMM authors recommend choosing the latent factor K based on P -value distributions for λ values, rather than genetic structure methods. For the final LFMM results, we used the R package “qvalue” to control the false discovery rate at the 0.05 level (Storey 2015). We calculated a Pearson’s correlation coefficient to quantify the proportion of significant SNPs with positive ($r > 0$) or negative ($r < 0$) clinal shifts with elevation.

Clinal SNP Analysis

For the LFMM SNPs with significant associations to elevation (clinal SNPs), we next extracted gene and Gene Ontology (GO) biological process information and compared the results between our study species. To extract GO terms, we conducted a statistical test for overrepresented biological process GO categories in the sets of genes with clinal SNPs for each species from the Panther database (Mi et al. 2019). There were 3 sets of genes with clinal SNPs: 1) shared by both species, 2) only *Coe. violifer*, and 3) only *Col. coruscans*. For Panther’s overrepresentation test tool, we selected Fisher’s Exact test with false discovery rate correction ($P < 0.05$).

Results

Bioinformatics

By using a *de novo* assembly approach, ~70–88% of reads mapped to species-specific references; in contrast, only ~30% of reads mapped to the *Cal. anna* genome. The *de novo* assemblies produced 11 843 gene contigs (assembly size: 17.2Mb) for *Coe. violifer* and 11 815 gene contigs (assembly size: 19.1Mb) for *Col. coruscans*. All of the targeted candidate genes from previous studies were captured. Additional Illumina sequence quality statistics are summarized in Supplementary Table S3. On average, 74% of reads were aligned to a targeted region (specificity) and 99.47% of the targeted regions had at least 1X coverage (sensitivity). At least 92.19% of nucleotide sites had 20X coverage. Further exon capture experiment statistics are in Supplementary Table S4.

The final dataset for downstream population genetic analyses included 156 individuals: 59 *Coe. violifer* and 97 *Col. coruscans*. We removed one sample per pair of closely related *Coe. violifer* individuals that had $\theta > 0.25$ ($n = 3$; average $\theta = 0.32$) and 4 *Col. coruscans* samples that were extreme outliers in initial PCA analyses based on component scores (Supplementary Table S5).

Population Genetic Structure

The relative genetic diversity within localities was comparable for θ_w and θ_π for both species (Supplementary Table S6). For *Coe. violifer*, within-locality diversity was highest for the samples near the Huallaga River (B) ($\theta_w = 3.21 \times 10^{-3}$, $\theta_\pi = 3.09 \times 10^{-3}$) and lowest for samples near the Marañón River (A) ($\theta_w = 1.87 \times 10^{-3}$, $\theta_\pi = 2.03 \times 10^{-3}$). There was little variation in within-locality genetic diversity at all *Col. coruscans* localities ($\theta_w = 1.01 \times 10^{-2} - 1.25 \times 10^{-2}$, $\theta_\pi = 8.51 \times 10^{-3} - 8.70 \times 10^{-3}$).

The PCA and admixture analyses ($K = 3$) identified 3 distinct genetic clusters for *Coe. violifer* (PC1: 12.2%; PC2: 3.2%; PC3: 2.9%; Figure 2a), however the groupings differed slightly. The admixture results match the PC1 $\times$ PC3 groups, while PC1 $\times$ PC2 groups the samples south of the Huallaga River (B, C) in one cluster separate from those near the Marañón River (A). The F_{ST} results support the PC1 $\times$ PC2 results, as there is gene flow between samples south of the Huallaga River (B, C) ($F_{ST} = 6.20 \times 10^{-2}$), while the D_{xy} results show very similar values for all 3 northern localities (A, B, C) ($D_{xy} = 5.60 \times 10^{-2} - 5.80 \times 10^{-2}$) (Supplementary Table S7). PC1 $\times$ PC2 grouped southern samples (D, E, F) together and D_{xy} was similar ($D_{xy} = 5.40 \times 10^{-3} - 5.60 \times 10^{-3}$). In contrast, samples east of the Apurímac River (E) were distinct in admixture and PC1 $\times$ PC3 assignment, and gene flow was lower between them (E) and nearby localities (D: $F_{ST} = 0.11$; F: $F_{ST} = 0.12$). These analyses collectively support a primary divide between northern (A, B, C) and southern (D, E, F) localities. Based on subspecies range maps, this divide roughly matches the distribution break of *Coe. v. dichroua* in the north from the 2 in the south (Supplementary Figure S1a), however genetic differentiation between *Coe. v. albicaudata* and *Coe. v. osculans* was unclear: none of the genetic structure patterns match the distribution of samples based on subspecies range maps.

Our sampling for *Col. coruscans* was distributed across the entire length of Peru and, surprisingly, there was no apparent population structure (Figure 2b). In contrast to the *Coe. violifer* PCA, the first 2 principal components explained very little of the variation (PC1: 1.72%; PC2: 1.19%) and therefore support a lack of population genetic structure (Figure 2b). Along PC1, the apparent 3 vertical bands were not explained by variation in sampling conditions (locality, elevation, season), sex, or sequencing design or quality (Supplementary Figure S2). Thus, the “pattern” may not be real as the proportion of variance explained by PC1 was quite small (1.72%). Furthermore, the admixture analysis with the lowest maximum likelihood score was $K = 1$ and between-locality genetic differentiation was low for all pairwise comparisons ($F_{ST} = 1.41 \times 10^{-2} - 4.39 \times 10^{-2}$ and $D_{xy} = 1.52 \times 10^{-2} - 1.55 \times 10^{-2}$) (Supplementary Table S8). Together, these analyses support a single large population with high gene flow throughout the range in Peru. However, gene flow across the sampled range may not be random. For illustration only, we show $K = 2$ admixture results in Figure 2b, which suggests slight divisions in population assignment pointing to gene flow around the high Andes in northern Peru linking populations on the Western and Eastern Cordilleras.

Isolation by Distance

Based on Mantel test results for both species, the geodesic distance and cost distance showed similar trends with genetic distance (Supplementary Figure S3). The role of isolation by distance was evident for *Coe. violifer* ($r = 0.75$, $P < 0.01$; $r = 70$, $P < 0.01$) (Supplementary Figure S3a). In contrast, there was no isolation by distance between *Col. coruscans* individuals ($r = 0.05$, $P = 0.15$; $r = 0.03$, $P = 0.23$) (Supplementary Figure S3b).

EEMS analysis maps showed regions of high and low migration between localities. For *Coe. violifer*, the primary region of low migration splits the northern (A, B, C) from the southern (D, E, F) localities (Figure 2a). There was slightly lower migration between the most northern samples (A) and the others. The estimated region of low effective migration was striking for *Col. coruscans* and likely corresponds to the high ridge of the Peruvian Andes, echoing support for nonrandom movement (Figure 2b). The EEMS map estimated higher migration rates around the Andes with a possible route in the Marañón river depression to connect the Western and Eastern Cordilleras.

Latent Factor Mixed Models

The final parameters for calling genotypes from all captured genes excluded the “hetBias_pval” filter, as it was too stringent. The “postCutOff” filter was the least sensitive parameter, so we chose the more conservative value (0.95). Smaller values of “SNP_pval” removed too many SNPs, so we set it to 10^{-3} . With these parameters, we obtained 198 092 SNPs for *Coe. violifer* and 690 274 SNPs for *Col. coruscans* (Supplementary Table S9). After filtering out rare variants, we retained 91 127 SNPs for *Coe. violifer* and 215 694 SNPs for *Col. coruscans* for analysis with LFMM (Supplementary Table S10). We had to remove hemoglobin markers (*HBB* and *HBE1*) from the analysis, as alleles were difficult to resolve between members of this gene family. These genes of interest should be pursued in future study but require refined sequencing to accurately differentiate SNPs from gene duplicate sequences.

After running LFMM, we evaluated the resulting number of statistically significant SNPs associated with elevation by examining P -value frequency distribution shape for several inflation factor (λ) values (Supplementary Table S11, Supplementary Figure S4). For both species, the P -value frequency distributions for $\lambda > 1$ shifted up, indicating too few SNPs detected (false negatives). The distributions for $\lambda < 1$ shifted down slightly, indicating too many SNPs were detected (false positives). Thus, we chose the SNP results from $\lambda = 1$ and $K = 2$ to define clinal SNPs for final analysis. We used the elevation bins spaced every 400 m for the final analysis of both species datasets.

Clinal SNP Analysis

In total, we identified 3495 SNPs (3.84%) on 1903 genes for *Coe. violifer* and 3962 SNPs (1.84%) on 2137 genes for *Col. coruscans* that exhibited significant elevational clines from LFMM analysis (Supplementary Table S11). A set of 567 genes had clinal SNPs for both species. For *Coe. violifer*, 1980 SNPs had positive trends (average $r = 0.79 \pm 0.12$) and 1515 SNPs had negative trends (average $r = -0.67 \pm 0.19$). For *Col. coruscans*, 1674 SNPs had positive trends (average $r = 0.80 \pm 0.11$) and 2288 SNPs had negative trends (average $r = -0.76 \pm 0.12$). The distribution of Pearson correlation coefficients for these trends are shown in Supplementary Figure S5.

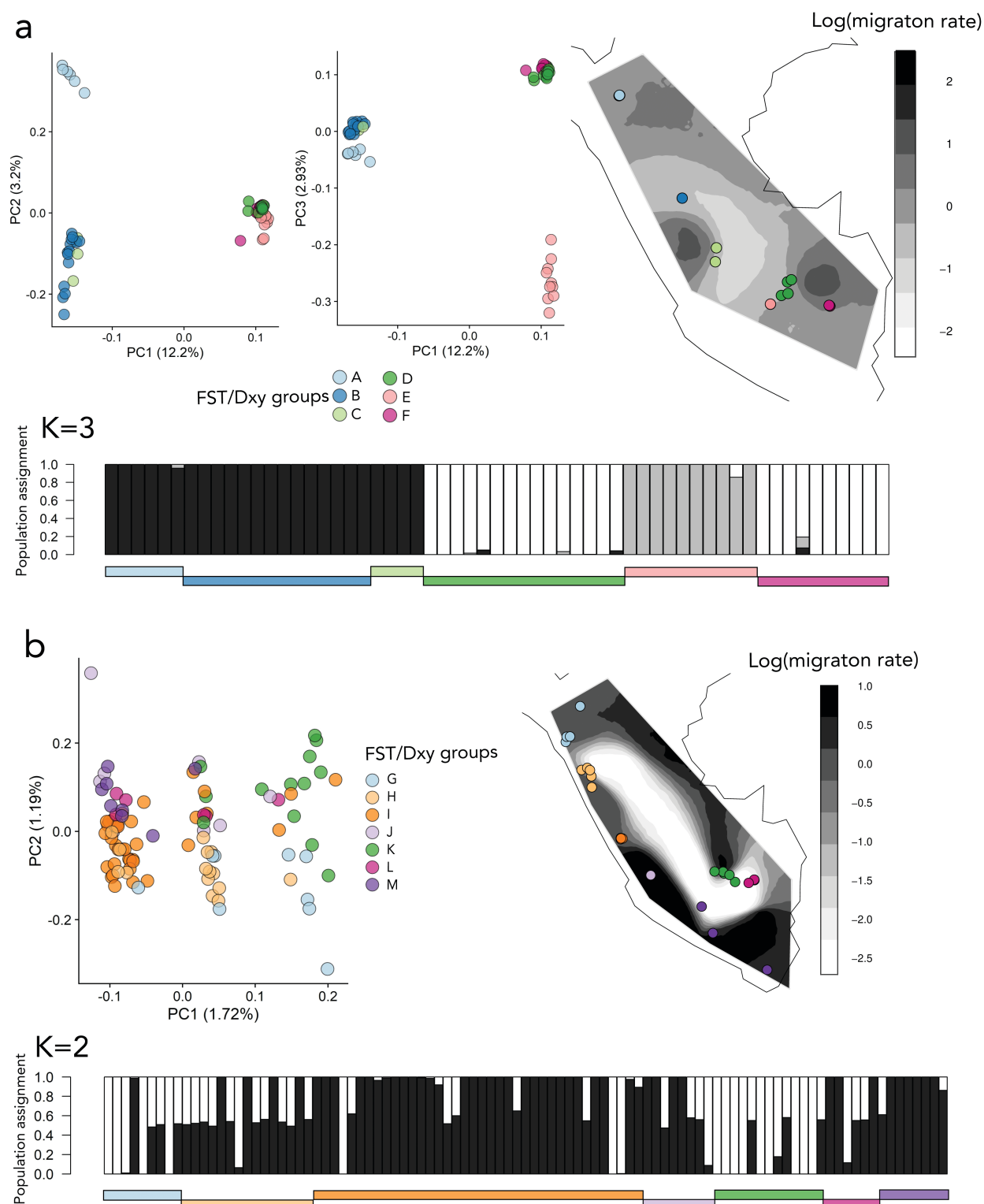


Figure 2. Three types of genetic structure plots each for (a) *Coe. violifer* (n = 59) and (b) *Col. coruscans* (n = 97). For reference, the point or horizontal line colors refer to sample groups used to calculate F_{ST} and Dxy (see online version for sample group colors). The PCA plots show the percentage of variation explained by principal components. The estimated effective migration surface maps show log migration rate in geographic space overlaid on a map of Peru (black = higher gene flow, white = lower gene flow). The vertical bars of the admixture plots show the proportion of individuals assigned to K = 3 populations (black, light gray, white) for *Coe. violifer* and, shown for illustration purposes only, K = 2 populations (black, white) for *Col. coruscans*.

The full list of candidate and novel genes with clinal SNPs recovered in our study is available in [Supplementary Table S12](#). The Panther overrepresentation test did not isolate statistically

significant overrepresented GO biological process categories, however the analysis did serve to characterize the categories represented in our dataset, many of which relate to findings in the high-elevation

literature (Figure 3, Supplementary Table S13). While it is difficult to draw further conclusions about the exact mechanisms connecting these functions to elevation adaptation, we examined 1) how the GO categories compared between our 2 study species and 2) whether genes with clinal SNPs have been previously identified in the high-elevation adaptation literature.

For a broader investigation of the GO categories, we compared the number of genes per GO category between the sets of unique genes per species or shared by both (Figure 3). Across all sets, the majority of genes function in metabolism and cellular processes, though numbers vary. For example, 257 metabolic-related genes were shared by both species, while there were 1216 and 2242 different metabolic-related genes for *Coe. violifer* and *Col. coruscans*, respectively. There were similar patterns for other gene functions, illustrating the redundancy of function across different genes with clinal SNPs in these highland species. There were some differences in the number or presence of genes for other categories. For example, 24 muscle-related genes were shared by both species, while *Coe. violifer* had 41 and *Col. coruscans* had 7 different muscle-related genes. There were 16 immune response-related genes shared by both species, while *Coe. violifer* had 326 and *Col. coruscans* had 54 different immune response-related genes. Two categories, embryo development and regulation of response to oxidative stress, were only present for *Coe. violifer*, highlighting that genetic adaptation to elevation is complex and impacts both repeated and unique gene functions for different species.

Of the genes with clinal SNPs, 154 genes for *Coe. violifer*, 166 genes for *Col. coruscans*, and 116 shared by both species are candidate genes in high-elevation literature (Supplementary Table S12). The proportion of clinal SNPs in candidate (*Coe. violifer*: 0.77%, *Col. coruscans*: 0.34%) and in other genes (*Coe. violifer*: 3.1%, *Col. coruscans*: 1.5%) was comparable between species. For example, we identified clinal SNPs on *ASH2L*, *WNT7B*, and *EPAS1* in the *Coe. violifer* analysis (Supplementary Figure S6a). *ASH2L* plays a role in DNA repair and response to radiation (Tibetan highland chicken: Zhang et al. 2016). *WNT7B* regulates hypoxia-induced development of vascular network in the lungs and heart, and is involved in oxygen homeostasis (Tibetan ground tit: Yanhua Qu et al. 2013). *EPAS1* (*HIF2α*), a component of the hypoxia-inducible factor pathway that regulates changing oxygen levels, has been studied extensively in the elevation adaptation literature for human populations (Beall et al. 2010; Yi et al. 2010) and identified for other highland species (Wang et al. 2014; Zhang et al. 2014; Graham and McCracken 2019), including Andean hummingbirds (Lim et al. 2019). For *Col. coruscans* examples, we identified clinal SNPs in *SEN1* and *COX1* (Supplementary Figure S6b). *SEN1* is associated with chronic mountain sickness (Andean human populations: Zhou et al. 2013). *COX1* plays an important role in the production of ATP during respiratory electron transport and oxidative phosphorylation (bar-headed goose: Scott et al. 2011; polar bears: Welch et al. 2014). *SMURF2* and *RYR2* were among the shared candidate genes with

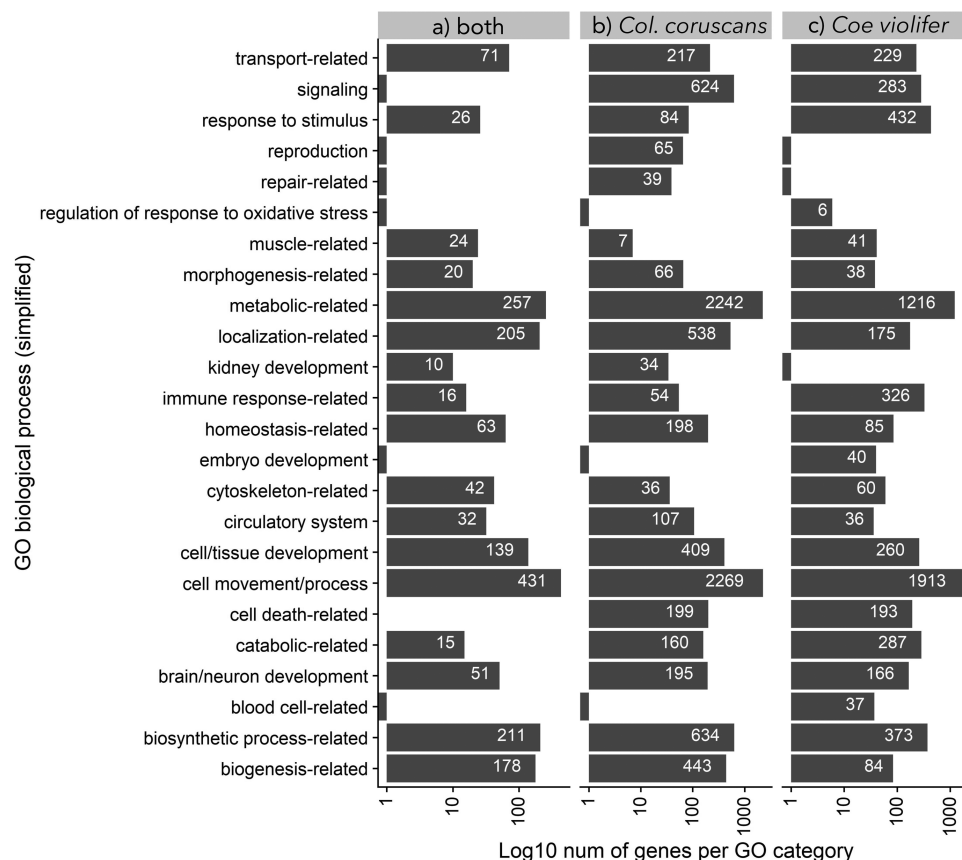


Figure 3. The bar plots show the number of genes categorized per simplified Gene Ontology (GO) biological process (y-axis log10 scaled). Each column of plots represents a different and unique set of genes for which clinal SNPs were identified as significantly associated with elevation by LFMM. The columns differ by whether clinal SNPs occur on (a) the same genes for both highland species, or different genes for (b) *Col. coruscans* only or (c) *Coe. violifer* only. The full list of genes and their associated GO biological process (level 6) category is available in Supplementary Table S13.

clinal SNPs for both species (Supplementary Figure S6). *SMURF2* plays a role in the vascular inflammatory response of endothelial cells in hypoxic conditions (highland Ethiopian human populations: Huerta-Sánchez et al. 2013). *RYR2* is involved in cardiac muscle contraction (snub-nosed monkey: Yu et al. 2016; Qinghai-Tibet Plateau gray wolf: Zhang et al. 2014). While the examples above represent a small snapshot, the presence of any clinal SNPs on candidate genes again highlights their likely importance for high-elevation adaptation across taxa (Supplementary Table S12).

Discussion

Migration, genetic drift, and natural selection determine the fate of beneficial alleles in populations, but resulting patterns can be indistinct. By quantifying population genetic demography, we could present a more nuanced interpretation of clinal variation in SNPs detected across elevational gradients in the Peruvian Andes for 2 highland hummingbird species. Since population genetic structure patterns differed between the 2 study species, the presence of genetic clines from LFMM analysis suggests putatively adaptive alleles are maintained in different ways. As these genetic clines encompass elevation changes of a few hundred meters, our genome-wide surveys highlight how natural selection along gradients makes tropical mountains “taller” by promoting local adaptation. In addition to finding clinal SNPs on previously identified candidate genes, we detected clinal SNPs on genes novel to our study with functions potentially relevant to responding to high-elevation conditions. Concordant with previous research, including our own, these results point to natural selection and repeated evolution on genes, especially in their cellular functions, instead of specific point substitutions (Arendt and Reznick 2008; Natarajan et al. 2016; Lim et al. 2019).

Our analysis of the genetic structure for *Coe. violifer* revealed geographically structured populations, isolation by distance, and some gene flow. Previous research on the evolutionary history of the *Coe. violifer* subspecies suggested diversification occurred from north to south in Peru based on the disjointed distribution of plumage color differences (Remsen 1984; Schuchmann and Zuchner 1997). Our results suggest the subspecies are not only phenotypically different (Schuchmann and Zuchner 1997), but *Coe. v. dichroua* is also genetically distinct from the southern subspecies (Figure 2a, Supplementary Figure S1a). The PCA and admixture results differed slightly in how individuals were assigned to genetic clusters, but there was consensus on the division between northern and southern localities approximately divided by the Apurímac drainage region (Figure 2a). There is a transition in vegetation from dry forest in the west to humid, montane forest in the east in the Apurímac River valley, a prominent barrier to dispersal for other avian species (Hosner et al. 2015; Benham and Witt 2016). The distribution mismatch of phenotypic and genotypic variation for *Coe. v. albicaudata* and *Coe. v. osculans* has been observed for other *Coeligena* species (Parra 2008) and other phenotypic traits (e.g., Benham and Witt 2016), suggesting these subspecies have not diverged genetically yet as ecological pressures rather than geographic ones influence plumage evolution (Winger 2017).

In contrast, the widespread geographic distribution, broad habitat requirements, and degree of gene flow show that *Col. coruscans* has few dispersal limitations. Without exception, all analyses identified a single, large panmictic population. *Col. coruscans* is one of the few well-documented neotropical hummingbird species with some migratory subpopulations for which migration includes

elevational shifts. The extensive gene flow we observed may reflect the inclusion of migratory individuals during the non-breeding or migration seasons (only 12 of 97 specimens were sampled during the core breeding season), but nonetheless illustrate the distances *Col. coruscans* individuals are capable of traversing. In addition, the θ_w and θ_π values were higher for *Col. coruscans* compared to *Coe. violifer*. Higher proportions of segregating sites and nucleotide diversity could result from large effective population size during the evolutionary history of the species, or recent population expansion (Fuchs et al. 2017). Despite the lack of population genetic structure, the Estimated Effective Migration Surface map indicated that movement was reduced across the high Andean ridge. Birds appear more likely to cross the Andes in the lower elevation Marañón river depression region compared to the Central Andes which start south of the depression (Figure 2b). This pattern of movement could explain the PCA and admixture results where northern and eastern individuals (e.g., G, K) were slightly more genetically similar compared to specimens collected in southwestern localities (e.g., M).

Given the population genetic structure of these species, several lines of evidence support our interpretation that clinal SNPs relate to local adaptation to elevation. First, migration-drift-selection balance and local adaptation along elevational gradients have been observed for other Andean bird species, such as the rufous-collared sparrow (Chevion and Brumfield 2009), house wren (Galen et al. 2015), and speckled teal (Graham et al. 2018). Genes with clinal SNPs in our study have similar functions to those in these studies. Andean avian research has also shown that genetic differentiation can result in morphological clines along broad (e.g., torrent ducks across the entire Andean mountain chain; Gutiérrez-Pinto et al. 2014) and narrow elevation gradients (e.g., wedge-billed woodcreepers along a 1500 m elevational gradient in Ecuador; Milá et al. 2009), suggesting that different migration-drift-selection dynamics can result in local adaptation.

Second, clinal SNPs were found for both hummingbird species despite demographic and ecological differences. These patterns illustrate that natural selection can shape locus-specific variation, even as gene flow and genetic drift shape genome-wide population genetic variation. Given their genetic structure and reduced levels of gene flow, clinal variation in *Coe. violifer* populations could have arisen from the addition of new genetic material via small amounts of gene flow between populations. Simulation studies have demonstrated that drift and moderate migration rates decrease the potential for adaptation, with migration adding new variants to the population, counteracting the loss of diversity from drift (Blanquart et al. 2012). This migration-drift balance implies genetic variation at adaptive loci could subsequently be maintained in a given population by low gene flow with little influence from drift. In contrast, high migration rates can homogenize genetic differentiation (Blanquart et al. 2012). The widespread distribution and lack of population subdivision in *Col. coruscans* suggests the observed locus-specific clinal variation is maintained by selection strong enough to supersede the homogenizing effect of gene flow. In both cases, the presence of clinal SNPs provides evidence that local adaptation can maintain higher genetic variation at specific SNPs, independent of genome-wide population genetic processes.

Finally, we found similar categories of cellular gene function across all sets of genes with clinal SNPs (Figure 3). Some of the strongest evidence for adaptation across global highland taxa relies on identifying selection for the same biochemical pathway, gene, or amino acid substitution (Rosenblum et al. 2014; Bailey et al. 2015; Lim et al. 2019). Our results suggest that redundancy

in cellular gene function allows for multiple mechanisms to underlie genetic adaptation to high elevation. For example, there were strong clines in SNP allele frequencies in genes with cellular functions that could be related to high-elevation conditions stressors, including hypoxia response, energy metabolism, and immune response. For both *Coe. violifer* and *Col. coruscans*, previously identified candidate genes, and those novel to our study, shared these key functions (Figure 3, Supplementary Table S13). Still, the lack of complete overlap in candidate genes with significant SNP-elevation associations between our hummingbird species, as well as the presence of the many candidate genes novel to our study reinforce prior findings that there are many ways to adapt (Arendt and Reznick 2008; Natarajan et al. 2016; Lim et al. 2019) and much more to discover about the mechanisms of genetic adaptation.

Our study shows natural selection can operate in similar ways despite differences in demographic history. Population structure resulting from isolation across valleys has been found for many species (Weir 2009; Turchetto-Zolet et al. 2013; Winger 2017). Our findings provide genome-wide evidence connecting turnover in environmental conditions of a few hundred meters in elevation to shifts in potentially adaptive allele frequencies. That narrow niche tolerance in tropical species effectively makes mountains “taller” in the tropics is a longstanding hypothesis for explaining the extraordinary biodiversity of the tropical Andes (Janzen 1967; Fjeldsø et al. 2012). By controlling for drift and migration to identify signals of natural selection, our results reveal more localized putative adaptations to environmental conditions along altitudinal gradients than previously suspected. The parallels in genes with clinal SNPs between *Coe. violifer* and *Col. coruscans*, despite their unique population dynamics, reinforces evidence for strong selective gradients with elevation adaptation to the ecological conditions of the Andes. The presence of elevation-associated SNPs points to local adaptation as a likely mechanism making tropical mountains “taller” as elevational gradients have structured variation across hundreds of genes despite migration.

Future research is needed to investigate whether and how many elevation-associated SNPs identified here confer fitness advantages. Additional research may also be directed at the role of small-effect loci, as well as features of genome architecture that could explain regions of SNP abundance, such as linked loci or mutation hotspots (e.g., Galen et al. 2015). Our results reflect the complexity of tracing the genetic underpinnings of adaptation and reveal predominantly repeated selection for cellular functions rather than specific genetic markers. These results may further shed light on how narrow elevational ranges are maintained in hummingbirds, and likely other montane taxa distributed within sharply defined elevational limits.

Supplementary Material

Supplementary material can be found at *Journal of Heredity* online.

Funding

This work was supported by a National Science Foundation (NSF) Doctoral Dissertation Improvement Grant (NSF DEB-1601477); American Genetic Association Evolutionary, Ecological, or Conservation Genomics Research Award; American Museum of Natural History Chapman Memorial Fund; and American Society

of Naturalists Student Research Award. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley (NIH S10 OD018174 Instrumentation Grant). For data analysis, we used XSEDE (NSF ACI-1548562), the University of Copenhagen's High Performance Computer Centre, and SeaWulf from Stony Brook Research Computing and Cyberinfrastructure and the Institute for Advanced Computational Science (NSF 1531492). L.M.D. was supported, in part, by NSF DEB-1442142. C.W. was supported by NSF DEB-1146491.

Acknowledgments

We thank Lydia Smith, Mark Phuong, and the UC Berkeley Evolutionary Genetics Laboratory for lab support. We also thank M. Thomas P. Gilbert's lab group, Walter Eanes, Douglas Futuyma, Elise Lauterbur, and Fanny Cornejo for helpful discussions. Samples were provided by the Museum of Southwestern Biology (MSB) (University of New Mexico (UNM)) and Museum of Vertebrate Zoology (University of California, Berkeley). Sampling was assisted by Emil Bautista, Andrew B. Johnson, Thomas Valqui, and many students from CORBIDI and the MSB at UNM.

Data Availability

De-multiplexed raw read data is available on the National Center for Biotechnology Information Sequence Read Archive ([dataset] Lim, M.C.W., Bi, K., Witt, C., Graham, C., & Dávalos, L.; 2020; Hummingbird exon capture dataset; NCBI SRA (BioProject: PRJNA688518, accessions: SAMN17186358-SAMN17186513)). All scripts used for analyses are available on Github (<https://github.com/marisalim/HbirdSeqCap>).

References

- Anderson EC, Dunham KK. 2008. The influence of family groups on inferences made with the program structure. *Mol Ecol Resour.* 8:1219–1229.
- Arendt J, Reznick D. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol Evol.* 23:26–32.
- Bailey SF, Rodrigue N, Kassen R. 2015. The effect of selection environment on the probability of parallel evolution. *Mol Biol Evol.* 32:1436–1448.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 19:455–477.
- Beall CM, Cavalleri GL, Deng L, Elston RC, Gao Y, Knight J, Li C, Li JC, Liang Y, McCormack M, et al. 2010. Natural selection on EPAS1 (HIF2α) associated with low hemoglobin concentration in Tibetan highlanders. *Proc Natl Acad Sci USA.* 107:11459–11464.
- Benham PM, Cuervo AM, McGuire JA, Witt CC. 2015. Biogeography of the Andean metaltail hummingbirds: contrasting evolutionary histories of tree line and habitat-generalist clades. *J Biogeogr.* 42:763–777.
- Benham PM, Witt CC. 2016. The dual role of Andean topography in primary divergence: functional and neutral variation among populations of the hummingbird, *Metallura tyrianthina*. *BMC Evol Biol.* 16:22.
- Bi K, Linderot T, Vanderpool D, Good JM, Nielsen R, Moritz C. 2013. Unlocking the vault: next-generation museum population genomics. *Mol Ecol.* 22:6018–6032.
- Bi K, Vanderpool D, Singhal S, Linderot T, Moritz C, Good JM. 2012. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genomics.* 13:403.
- Blanquart F, Gandon S, Nuismer SL. 2012. The effects of migration and drift on local adaptation to a heterogeneous environment. *J Evol Biol.* 25:1351–1363.

- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30:2114–2120.
- Chapman FM. 1926. The distribution of bird-life in Ecuador: a contribution to a study of the origin of Andean bird-life. *Bull. Amer. Mus. Nat. Hist.* 55:1–784.
- Chapman MA, Hiscok SJ, Filatov DA. 2013. Genomic divergence during speciation driven by adaptation to altitude. *Mol Biol Evol.* 30:2553–2567.
- Cheviron ZA, Brumfield RT. 2009. Migration-selection balance and local adaptation of mitochondrial haplotypes in rufous-collared sparrows (*Zonotrichia capensis*) along an elevational gradient. *Evolution*. 63:1593–1605.
- Cracraft J. 1985. Historical biogeography and patterns of differentiation within the South American avifauna: Areas of endemism. *Ornithological Monographs*. 49–84.
- Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol.* 23:3133–3157.
- del Hoyo J, Züchner T, Collar N, Boesman PFD, Kirwan GM. 2020. Violet-throated Starfrontlet (*Coeligena violifer*), version 1.0. In Billerman SM, Keeney BK, Rodewald PG, Schulenberg TS, editors. *Birds of the world*. Ithaca, NY: Cornell Lab of Ornithology. doi:10.2173/bow.vitsta1.01.
- Fajardo J, Lessmann J, Bonaccorso E, Devenish C, Muñoz J. 2014. Combined use of systematic conservation planning, species distribution modelling, and connectivity analysis reveals severe conservation gaps in a megadiverse country (Peru). *PLoS One*. 9:e114367.
- Fjeldsø J, Bowie RCK, Rahbek C. 2012. The role of mountain ranges in the diversification of birds. *Annu Rev Ecol Evol Syst.* 43:249–265.
- Fjeldsø J, Lambin E, Mertens B. 1999. Correlation between endemism and local ecoclimatic stability documented by comparing Andean bird distributions and remotely sensed land surface data. *Ecography*. 22:63–78.
- Frichot E, Schoville SD, Bouchard G, François O. 2013. Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol.* 30:1687–1699.
- Fuchs J, Fjeldsø J, Bowie RCK. 2017. Diversification across major biogeographic breaks in the African Shining/Square-tailed Drongos complex (Passeriformes: Dicuridae). *Zoologica Scripta*. 46:27–41.
- Fumagalli M, Vieira FG, Linderot T, Nielsen R. 2014. ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics*. 30:1486–1487.
- Galen SC, Natarajan C, Moriyama H, Weber RE, Fago A, Benham PM, Chavez AN, Cheviron ZA, Storz JF, Witt CC. 2015. Contribution of a mutational hot spot to hemoglobin adaptation in high-altitude Andean house wrens. *Proc Natl Acad Sci U S A*. 112:13958–13963.
- Gilbert M, Jarvis E, Li B, Li C, Mello C, Wang J, Zhang G. 2014. Genomic data of the Anna's Hummingbird (*Calypte anna*). *GigaScience Database*.
- Graham AM, Lavretsky P, Muñoz-Fuentes V, Green AJ, Wilson RE, McCracken KG. 2018. Migration-selection balance drives genetic differentiation in genes associated with high-altitude function in the speckled teal (*Anas flavirostris*) in the Andes. *Genome Biol Evol.* 10:14–32.
- Graham AM, McCracken KG. 2019. Convergent evolution on the hypoxia-inducible factor (HIF) pathway genes EGLN1 and EPAS1 in high-altitude ducks. *Heredity (Edinb)*. 122:819–832.
- Gregory-Wodzicki KM. 2000. Uplift history of the Central and Northern Andes: a review. *Geol Soc Am Bull.* 112:1091–1105.
- Gutiérrez-Pinto N, McCracken KG, Alza L, Tubaro P, Kopuchian C, Astie A, Cadena CD. 2014. The validity of ecogeographical rules is context-dependent: testing for Bergmann's and Allen's rules by latitude and elevation in a widespread Andean duck. *Biol J Linn Soc.* 111:850–862.
- Hämälä T, Marttilä TM, Savolainen O. 2018. Local adaptation and ecological differentiation under selection, migration, and drift in *Arabidopsis lyrata*. *Evolution*. 72:1373–1386.
- Hazzi NA, Moreno JS, Ortiz-Movliav C, Palacio RD. 2018. Biogeographic regions and events of isolation and diversification of the endemic biota of the tropical Andes. *Proc Natl Acad Sci U S A*. 115:7985–7990.
- Herman A, Brandvain Y, Weagley J, Jeffery WR, Keene AC, Kono TJY, Bilandžija H, Borowsky R, Espinosa L, O'Quin K, et al. 2018. The role of gene flow in rapid and repeated evolution of cave-related traits in Mexican tetra, *Astyanax mexicanus*. *Mol Ecol.* 27:4397–4416.
- Hoorn C, Wesselingh FP, ter Steege H, Bermudez MA, Mora A, Sevink J, Sanmartín I, Sanchez-Meseguer A, Anderson CL, Figueiredo JP, et al. 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*. 330:927–931.
- Hosner PA, Andersen MJ, Robbins MB, Urbay-Tello A, Cueto-Aparicio L, Verde-Guerra K, Sánchez-González LA, Navarro-Sigüenza AG, Boyd RL, Núñez J, et al. 2015. Avifaunal surveys of the upper Apurímac River Valley, Ayacucho and Cuzco Departments, Peru: new distributional records and biogeographic, taxonomic, and conservation implications. *The Wilson Journal of Ornithology*. 127:563–581.
- Huerta-Sánchez E, Degiorgio M, Pagani L, Tarekgn A, Ekong R, Antao T, Cardona A, Montgomery HE, Cavalleri GL, Robbins PA, et al. 2013. Genetic signatures reveal high-altitude adaptation in a set of Ethiopian populations. *Mol Biol Evol.* 30:1877–1888.
- IUCN. 2016. The IUCN Red List of Threatened Species. In Version 2016-3. Available from: <http://www.iucnredlist.org>
- Janzen DH. 1967. Why mountain passes are higher in the tropics. *Am Naturalist*. 101:233–249.
- Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinf.* 15:356.
- Korneliussen TS, Moltke I. 2015. NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. *Bioinformatics*. 31:4009–4011.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 9:357–359.
- Lenormand T. 2002. Gene flow and the limits to natural selection. *Trends Ecol Evol.* 17:183–189.
- Lewontin RC, Krakauer J. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*. 74:175–195.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. 25:2078–2079.
- Lim MCW, Witt CC, Graham CH, Dávalos LM. 2019. Parallel molecular evolution in pathways, genes, and sites in high-elevation hummingbirds revealed by comparative transcriptomics. *Genome Biol Evol.* 11:1552–1572.
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet.* 4:981–994.
- Magoč T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*. 27:2957–2963.
- McCracken KG, Barger CP, Sorenson MD. 2010. Phylogenetic and structural analysis of the HbA (alphaA/betaA) and HbD (alphaD/betaA) hemoglobin genes in two high-altitude waterfowl from the Himalayas and the Andes: Bar-headed goose (*Anser indicus*) and Andean goose (*Chloephaga melanoptera*). *Mol Phylogenet Evol.* 56:649–658.
- McGuire JA, Witt CC, Remsen JV Jr, Corl A, Rabosky DL, Altshuler DL, Dudley R. 2014. Molecular phylogenetics and the diversification of hummingbirds. *Curr Biol.* 24:910–916.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297–1303.
- Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc.* 2010:prot5448.
- Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 47:D419–D426.
- Milá B, Wayne RK, Fitz P, Smith TB. 2009. Divergence with gene flow and fine-scale phylogeographical structure in the wedge-billed woodcreeper, *Glyphorynchus spirurus*, a Neotropical rainforest bird. *Mol Ecol.* 18:2979–2995.

- Muñoz-Fuentes V, Cortázar-Chinarro M, Lozano-Jaramillo M, McCracken KG. 2013. Stepwise colonization of the Andes by ruddy ducks and the evolution of novel β -globin variants. *Mol Ecol*. 22:1231–1249.
- Natarajan C, Hoffmann FG, Weber RE, Fago A, Witt CC, Storz JF. 2016. Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. *Science*. 354:336–339.
- Nei M. 1987. *Molecular evolutionary genetics*. New York, NY: Columbia University Press.
- Parra JL. 2008. *Color evolution in Andean hummingbirds (dissertation)*. Berkeley, CA: University of California.
- Petkova D, Novembre J, Stephens M. 2016. Visualizing spatial population structure with estimated effective migration surfaces. *Nat Genet*. 48:94–100.
- Prieto-Torres DA, Cuervo AM, Bonaccorso E. 2018. On geographic barriers and Pleistocene glaciations: tracing the diversification of the Russet-crowned Warbler (*Myiothlypis coronata*) along the Andes. *PLoS One*. 13:e0191598.
- Projecto-García J, Natarajan C, Moriyama H, Weber RE, Fago A, Cheviron ZA, Dudley R, McGuire JA, Witt CC, Storz JF. 2013. Repeated elevational transitions in hemoglobin function during the evolution of Andean hummingbirds. *Proc Natl Acad Sci U S A*. 110:20669–20674.
- Qu Y, Tian S, Han N, Zhao H, Gao B, Fu J, Cheng Y, Song G, Ericson PG, Zhang YE, et al. 2015. Genetic responses to seasonal variation in altitudinal stress: whole-genome resequencing of great tit in Eastern Himalayas. *Sci Rep*. 5:14256.
- Qu Y, Zhao H, Han N, Zhou G, Song G, Gao B, Tian S, Zhang J, Zhang R, Meng X, et al. 2013. Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau. *Nat Commun*. 4:2071.
- Remsen JV Jr. 1984. High incidence of “leapfrog” pattern of geographic variation in andean birds: implications for the speciation process. *Science*. 224:171–173.
- Rosenblum EB, Parent CE, Brandt EE. 2014. The molecular basis of phenotypic convergence. *Annu Rev Ecol Syst*. 45:203–226.
- Ruggiero A, Hawkins BA. 2008. Why do mountains support so many species of birds? *Ecography*. 31:306–315.
- Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation. *Nat Rev Genet*. 14:807–820.
- Schuchmann KL, Zuchner T. 1997. *Coeligena violifer albicaudata* (Aves, Trochilidae): a new hummingbird subspecies from the Southern Peruvian Andes. *Ornitol Neotrop*. 24:247–253.
- Schulenberg TS, Stotz DF, Lane DE, O'Neill JP, Parker TA. 2007. *Birds of Peru*. Princeton, NJ: Princeton University Press.
- Scott GR, Schulte PM, Egginton S, Scott AL, Richards JG, Milsom WK. 2011. Molecular evolution of cytochrome C oxidase underlies high-altitude adaptation in the bar-headed goose. *Mol Biol Evol*. 28:351–363.
- Simonson TS, McClain DA, Jorde LB, Prchal JT. 2012. Genetic determinants of Tibetan high-altitude adaptation. *Hum Genet*. 131:527–533.
- Skotte L, Korneliussen TS, Albrechtsen A. 2013. Estimating individual admixture proportions from next generation sequencing data. *Genetics*. 195:693–702.
- Smit A, Hubley R, Green P. 2015. RepeatMasker Open-4.0. Available from: <http://www.repeatmasker.org>
- Storey JD. 2015. *qvalue: Q-value estimation for false discovery rate control*.
- Storz JF, Runck AM, Sabatino SJ, Kelly JK, Ferrand N, Moriyama H, Weber RE, Fago A. 2009. Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. *Proc Natl Acad Sci U S A*. 106:14450–14455.
- Turchetto-Zolet AC, Pinheiro F, Salgueiro F, Palma-Silva C. 2013. Phylogeographical patterns shed light on evolutionary process in South America. *Mol Ecol*. 22:1193–1213.
- van Etten J. 2017. R Package gdistance: distances and routes on geographical grids. *J Stat Soft*. 76:1–21.
- Vieira FG, Lassalle F, Korneliussen TS, Fumagalli M. 2016. Improving the estimation of genetic distances from Next-Generation Sequencing data. *Biol J Linn Soc*. 117:139–149.
- Wallace JR. 2012. Imap: Interactive Mapping. R package version, 1.
- Wang GD, Fan RX, Zhai W, Liu F, Wang L, Zhong L, Wu H, Yang HC, Wu SF, Zhu CL, et al. 2014. Genetic convergence in the adaptation of dogs and humans to the high-altitude environment of the Tibetan plateau. *Genome Biol Evol*. 6:2122–2128.
- Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol*. 7:256–276.
- Weir JT. 2009. Implications of genetic differentiation in Neotropical montane birds. *Ann Mo Bot Gard*. 96:410–433.
- Welch AJ, Bedoya-Reina OC, Carretero-Paulet L, Miller W, Rode KD, Lindqvist C. 2014. Polar bears exhibit genome-wide signatures of bioenergetic adaptation to life in the arctic environment. *Genome Biol Evol*. 6:433–450.
- Winger BM. 2017. Consequences of divergence and introgression for speciation in Andean cloud forest birds. *Evolution*. 71:1815–1831.
- Yeaman S, Whitlock MC. 2011. The genetic architecture of adaptation under migration-selection balance. *Evolution*. 65:1897–1911.
- Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, et al. 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*. 329:75–78.
- Yu L, Wang GD, Ruan J, Chen YB, Yang CP, Cao X, Wu H, Liu YH, Du ZL, Wang XP, et al. 2016. Genomic analysis of snub-nosed monkeys (*Rhinopithecus*) identifies genes and processes related to high-altitude adaptation. *Nat Genet*. 48:947–952.
- Zhang Q, Gou W, Wang X, Zhang Y, Ma J, Zhang H, Zhang Y, Zhang H. 2016. Genome resequencing identifies unique adaptations of Tibetan chickens to hypoxia and high-dose ultraviolet radiation in high-altitude environments. *Genome Biol Evol*. 8:765–776.
- Zhang W, Fan Z, Han E, Hou R, Zhang L, Galaverni M, Huang J, Liu H, Silva P, Li P, et al. 2014. Hypoxia adaptations in the grey wolf (*Canis lupus chanco*) from Qinghai-Tibet plateau. *PLoS Genet*. 10:e1004466.
- Zhou D, Udpa N, Ronen R, Stobdan T, Liang J, Appenzeller O, Zhao HW, Yin Y, Du Y, Guo L, et al. 2013. Whole-genome sequencing uncovers the genetic basis of chronic mountain sickness in Andean highlanders. *Am J Hum Genet*. 93:452–462.
- Züchner T, Boesman PFD, Kirwan GM. 2020. Sparkling Violetear (*Colibri coruscans*), version 1.0. In del Hoyo J, Elliott A, Sargatal J, Christie DA, de Juana E, editors. *Handbook of the birds of the world alive*. Ithaca, NY: Cornell Lab of Ornithology. doi:10.2173/bow.spvear1.01