

Geogenomic Predictors of Genetree Heterogeneity Explain Phylogeographic and Introgression History: A Case Study in an Amazonian Bird (*Thamnophilus aethiops*)

LUKAS J. MUSER^{1,2,7}, GLAUCIA DEL-RIO^{3,4}, RAFAEL S. MARCONDES^{5,6}, ROBB T. BRUMFIELD⁵, GUSTAVO A. BRAVO^{7,8}, AND GREGORY THOM⁵

¹Department of Ornithology, The Academy of Natural Sciences of Drexel University, Philadelphia, PA 19103, USA

²Department of Ornithology, American Museum of Natural History, New York, NY 10024, USA

³Cornell Laboratory of Ornithology and Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA

⁴Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA

⁵Department of Biology and Museum of Natural Science, Louisiana State University, Baton Rouge, LA 70803, USA

⁶Department of BioSciences, Rice University, Houston, TX 77005, USA

⁷Sección de Ornitología, Colecciones Biológicas, Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Claustro de San Agustín, Villa de Leyva, Boyacá 111311, Colombia

⁸Museum of Comparative Zoology and Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

*Correspondence to be sent to: Department of Ornithology, The Academy of Natural Sciences of Drexel University, 1900 Benjamin Franklin Blvd., Philadelphia, PA 19103, USA; E-mail: lukas.j.muser@drexel.edu.

Received 20 December 2022; reviews returned 14 September 2023; accepted 4 October 2023

Associate Editor: Frank Burbrink

Abstract.—Can knowledge about genome architecture inform biogeographic and phylogenetic inference? Selection, drift, recombination, and gene flow interact to produce a genomic landscape of divergence wherein patterns of differentiation and genealogy vary nonrandomly across the genomes of diverging populations. For instance, genealogical patterns that arise due to gene flow should be more likely to occur on smaller chromosomes, which experience high recombination, whereas those tracking histories of geographic isolation (reduced gene flow caused by a barrier) and divergence should be more likely to occur on larger and sex chromosomes. In Amazonia, populations of many bird species diverge and introgress across rivers, resulting in reticulated genomic signals. Herein, we used reduced representation genomic data to disentangle the evolutionary history of 4 populations of an Amazonian antbird, *Thamnophilus aethiops*, whose biogeographic history was associated with the dynamic evolution of the Madeira River Basin. Specifically, we evaluate whether a large river capture event ca. 200 Ka, gave rise to reticulated genealogies in the genome by making spatially explicit predictions about isolation and gene flow based on knowledge about genomic processes. We first estimated chromosome-level phylogenies and recovered 2 primary topologies across the genome. The first topology (T1) was most consistent with predictions about population divergence and was recovered for the Z-chromosome. The second (T2), was consistent with predictions about gene flow upon secondary contact. To evaluate support for these topologies, we trained a convolutional neural network to classify our data into alternative diversification models and estimate demographic parameters. The best-fit model was concordant with T1 and included gene flow between non-sister taxa. Finally, we modeled levels of divergence and introgression as functions of chromosome length and found that smaller chromosomes experienced higher gene flow. Given that (1) genetrees supporting T2 were more likely to occur on smaller chromosomes and (2) we found lower levels of introgression on larger chromosomes (and especially the Z-chromosome), we argue that T1 represents the history of population divergence across rivers and T2 the history of secondary contact due to barrier loss. Our results suggest that a significant portion of genomic heterogeneity arises due to extrinsic biogeographic processes such as river capture interacting with intrinsic processes associated with genome architecture. Future phylogeographic studies would benefit from accounting for genomic processes, as different parts of the genome reveal contrasting, albeit complementary histories, all of which are relevant for disentangling the intricate geogenomic mechanisms of biotic diversification. [Amazonia; biogeography; demographic modeling; gene flow; gene tree; genome architecture; geogenomics; introgression; linked selection; neural network; phylogenomic; phylogeography; reproductive isolation; speciation; species tree.]

A key goal of speciation research is to understand the biogeographic mechanisms associated with population divergence and homogenization (Endler 1977). Although the reduction of gene flow between populations caused by biogeographic barriers (henceforth geographic isolation) typically plays a crucial role in speciation, populations may diverge despite high gene flow (Nosil 2008). Such gene flow, especially between non-sister populations, can result in heterogeneous levels of differentiation across the genome (Keller et al. 2013; Gompert et al. 2014; Mallet et al. 2016; Meier et al.

2017; Pulido-Santacruz et al. 2020), a pattern further intensified by interactions with selection and genome architecture (Feder et al. 2012; Cruickshank and Hahn 2014; Irwin et al. 2018; Manthey et al. 2021). For example, genomic regions experiencing strong disruptive selection or low recombination (e.g., large chromosomes; Haenel et al. 2018) may resist the homogenizing effects of gene flow, resulting in elevated peaks of divergence and contrasting genealogies when compared with other parts of the genome. In these cases, recombination rate and introgression should covary when there is baseline

selection against gene flow due to divergent selection or cohesion among co-evolved genes that reduce gene flow. Thus, the genomic landscape of divergence not only reflects disparate levels of differentiation between populations but also a profusion of genealogical relationships (Fontaine et al. 2015; Mallet et al. 2016; Wen et al. 2016). Dealing with and modeling evolutionary history in light of this heterogeneous landscape is crucial for obtaining a detailed understanding of biogeographic processes (Provost et al. 2022).

Although the genomic landscape can be highly heterogeneous, the signals of distinct processes, such as divergence and gene flow, are nonrandomly distributed across the genome (Van Doren et al. 2017). For instance, variation in recombination rates directly impacts levels of gene flow and divergence (Wang et al. 2022). One widely recognized mechanism by which this operates is the breakdown of blocks of linked loci affected by selection. Specifically, linked selection (selection in the genome impacting nearby sites) can significantly diminish genetic variation in blocks of linked loci through genetic hitchhiking of nearby sites (Lohmueller et al. 2011; Feder et al. 2012). When linked selection within populations is strong, measures of relative divergence, such as F_{ST} , which contain a term for within population variation, are expected to increase, but absolute divergence, d_{xy} , may be reduced (barring strong genomic island effects) by the corresponding depletion of allelic diversity (Charlesworth 1998; Nachman and Payseur 2012; Cruickshank and Hahn 2014; Van Doren et al. 2017; Irwin et al. 2018). In regions of the genome where recombination is high, blocks of linked loci are more frequently broken down, lessening the effects of linked selection (Tigano et al. 2022). Where recombination is low, however, the effects of selection are elevated as longer blocks of linked loci are able to persist. Recombination rate varies considerably across the genome and is particularly associated with chromosome size. Because each chromosome must undergo at least one crossing-over event during Meiosis (Mather 1938), smaller chromosomes experience more recombination per base than large chromosomes (Haenel et al. 2018; Tigano et al. 2022). Thus, regions of the genome with higher rates of recombination, such as smaller chromosomes, are also expected to have higher rates of introgression, when population divergence occurs with gene flow (Martin et al. 2019; Manthey et al. 2021). Quantifying the variation and predictability of these genomic processes that are intrinsic to organisms can help shed light on the reticulated history of recent radiations.

In contrast to intrinsic genomic architecture, biogeographic history is an important extrinsic factor influencing reticulation and the genomic landscape (Burbrink and Gehara 2018; Thom et al. 2021; Provost et al. 2022). As levels of isolation associated with physiographic barriers vary through space and time, so too do rates of selection, gene flow, and divergence (Endler 1977; Aguilée et al. 2013; Delmore et al. 2018; He et al. 2019). In many parts of the world,

population isolation and connectivity vary, in part, as functions of spatiotemporal variation in the environment (Flantua et al. 2019; He et al. 2019; Musher et al. 2019). For example, rates of isolation and gene flow among populations that diverged across the Isthmus of Panama may have been affected by the wax and wane of humid and dry forests across that region (David Webb 1991; Vrba 1992; Smith et al. 2012; Musher et al. 2020). Likewise, sea-level fluctuations and rainfall patterns have directly affected the distribution and amount of flooded forest habitat in Amazonia, which also affected levels of gene flow between populations of organisms that occur there (Thom et al. 2020; Sawakuchi et al. 2022; Luna et al. 2023). In Amazonian lowlands, differentiated populations often experience pervasive gene flow, sometimes from multiple non-sister lineages, a factor that complicates inference about their historical relationships, biogeography, and systematics (Pulido-Santacruz et al. 2018; Del-Rio et al. 2022; Musher et al. 2022). This is because gene flow between non-sister taxa results in a network of interpopulation relationships that violates the assumptions of a bifurcating model of evolutionary history (Mallet et al. 2016; Thom et al. 2018). Thus, if biogeography drives opportunities for isolation and contact between non-sister taxa, it should result in distinct predictable signatures of genealogy in the genomes of diverging populations.

Many lowland *terra-firme* (non-flooded forest) Amazonian birds have geographically isolated populations across rivers yet experience high levels of gene flow (Barrera-Guzmán et al. 2022; Del-Rio et al. 2022; Musher et al. 2022). Rivers are key biogeographic barriers for many lowland Amazonian birds, driving population isolation and genetic structure across the landscape (Sick 1967; Capparella 1991; Ribas et al. 2012; Smith et al. 2014; Ferreira et al. 2017). However, 3 well-known features of Amazonian lowlands add complexity to this system. First, the Amazon Basin, especially its southern portion, is characterized by several large tributaries running in quasi-parallel, forming isolated blocks of habitat (interfluvies) wherein a given taxon may be surrounded by 2 or more closely related taxa that occur on opposite river margins. Second, Amazonian rivers get narrower toward their headwaters, which is associated with increased gene flow across their upper portions (Weir et al. 2015). Finally, lowland river basins continuously rearrange via tributary capture (the movement of a tributary from one basin to another) and avulsion (the erosion of channel boundaries, leading to channel migration) (Gascon et al. 2000; Albert et al. 2018). In this geographic configuration, there are opportunities for multiple non-sister taxa to interact, and partially isolated populations can experience gene flow across rivers, leading to highly reticulated patterns of diversification across species' genomes (Musher et al. 2022).

Inferring the history of population isolation and gene flow under these conditions is a major challenge for researchers studying Amazonia because limited knowledge about the historical relationships of taxa

also hampers an understanding of the mechanisms that contribute to the region's high biodiversity (Cracraft et al. 2020). Previous studies of Amazonian birds have greatly advanced our knowledge of the history of the region's taxa, but in many cases, limited data or sparse spatial sampling has resulted in weak resolution of species' complex histories of isolation and gene flow. Genomic approaches, however, are revealing many of the evolutionary and biogeographic mechanisms driving species accumulation in the Neotropics (Thom et al. 2018, 2020; Pulido-Santacruz et al. 2020; Schley et al. 2020). This is especially important given geologists' growing understanding that the Amazonian landscape has been highly dynamic (Bicudo et al. 2019; Pupim et al. 2019; Ruokolainen et al. 2019). Thus, important questions for Amazonian biogeography include, (1) how do we infer population history under conditions of high gene flow among non-sister taxa, especially in the context of a dynamic landscape? (2) What are the consequences of these complex histories of isolation and gene flow for Amazonian organisms at the genomic level? (3) do genealogical patterns vary predictably across the genome in a way that is informative for biogeographic inference? and (4) to what extent can reduced representation genomic data shed light on these questions?

In this study, we address these questions by testing competing hypotheses for the biogeographic history of a passerine bird, the White-shouldered antshrike *Thamnophilus aethiops* (Thamnophilidae), in southern Amazonia. The southern Amazonian lowlands are particularly dynamic and experienced a major riverine restructuring, wherein large tributaries moved among watersheds and across the landscape during the Quaternary (Latrubesse 2002; Rossetti 2014; Ruokolainen et al. 2019; Rossetti et al. 2021). Some of these past movements occurred near the headwaters of the modern Madeira River, a large tributary of the Amazon that is a well-known biogeographic barrier for many bird species (Fernandes 2013; Smith et al. 2014; Silva et al. 2019). For example, paleochannels between the modern upper Madeira and Purus Rivers indicate that the Madeira extended its basin by capturing a large tributary of the Purus ca. 200 Ka or less (Fig. 1) (Ruokolainen et al. 2019). This suggests that the upper portion of the Madeira formed more recently than the lower, likely becoming a barrier for many terrestrial organisms over 2 stages.

The Madeira River capture scenario provides predictions about the history of population divergence and gene flow, wherein populations east of the Upper Madeira (Southern Rondônia; Fig. 1) are expected to be more closely related to populations west of the Madeira (Inambari) than to other populations within Rondônia (Fernandes 2013; Ferreira et al. 2017). Such a scenario also suggests that there will be more genetic differentiation, and therefore potential incompatibilities across the genomes of more deeply diverged populations (e.g., those in Rondônia) than between more recently diverged populations (e.g., populations within Inambari). Previous studies of *T. aethiops* have

been equivocal with respect to these predictions (Thom and Aleixo 2015; Musher et al. 2022). For example, the mitochondrial (mtDNA) phylogeny is consistent with the expectations of river capture, recovering a sister relationship between populations on opposite sides of the Madeira River (but not its downstream portions). Genomic data were more ambiguous, however, recovering relatively strong support for the monophyly of Rondônia and Inambari populations, which conflicts with the expectations of river capture. Thus, if the population in S Rondônia is historically related to Inambari populations, as the river capture scenario predicts, then we expect regions of the genome supporting monophyly of Rondônia to be driven by introgressive hybridization after secondary contact. Herein, we test these alternative genealogical predictions while also utilizing and modeling information about genome architecture. For instance, if gene flow increases on smaller autosomes, genealogies resulting from secondary contact should be less likely on larger autosomes and sex chromosomes. Rather, regions of the genome inferred to have low gene flow should better track phylogeny *sensu stricto* (i.e., the history of population divergences).

METHODS

Study Taxon

Thamnophilus aethiops is well-suited for testing our hypotheses as it exhibits both subspecific variation and genetic structure across southern Amazonian rivers (Thom and Aleixo 2015; Musher et al. 2022). Populations east of the Madeira River belong to a single subspecies, *T. a. punctuliger*. Populations west of the Madeira fall into one of 3 subspecies: *T. a. injunctus* occurs between the Purus and Madeira Rivers, *T. a. juruanus* occurs between the Jurua and Purus Rivers, and *T. a. kapouni* occurs west of the Juruá. Superficially, the phenotypes of *T. a. injunctus* and *T. a. punctuliger* are most similar, with individuals of both taxa being lighter gray overall and marked with white spots on the wing coverts, unlike populations farther west. Moreover, *T. aethiops* is a fairly sedentary understory species, so gene flow expectations across stable rivers are relatively low.

Sampling and Genotyping-by-sequencing Data Assembly

We downloaded genotyping-by-sequencing (GBS) data from a previous study (Musher et al. 2022) and re-assembled it using iPyRAD version 0.9.81 (Elshire et al. 2011; Eaton and Overcast 2020). These data are available on the NIH Sequence Read Archive under project ID PRJNA966941. Raw reads were deposited in NCBI Sequence Read Archive under ID SRR26669395. We were interested in the history and relationships of populations west of the Tapajós River and south of the Amazon River (southwestern Amazon Basin), so we excluded samples falling outside of this region. This dataset totaled 51 samples (Supplementary Tables S1

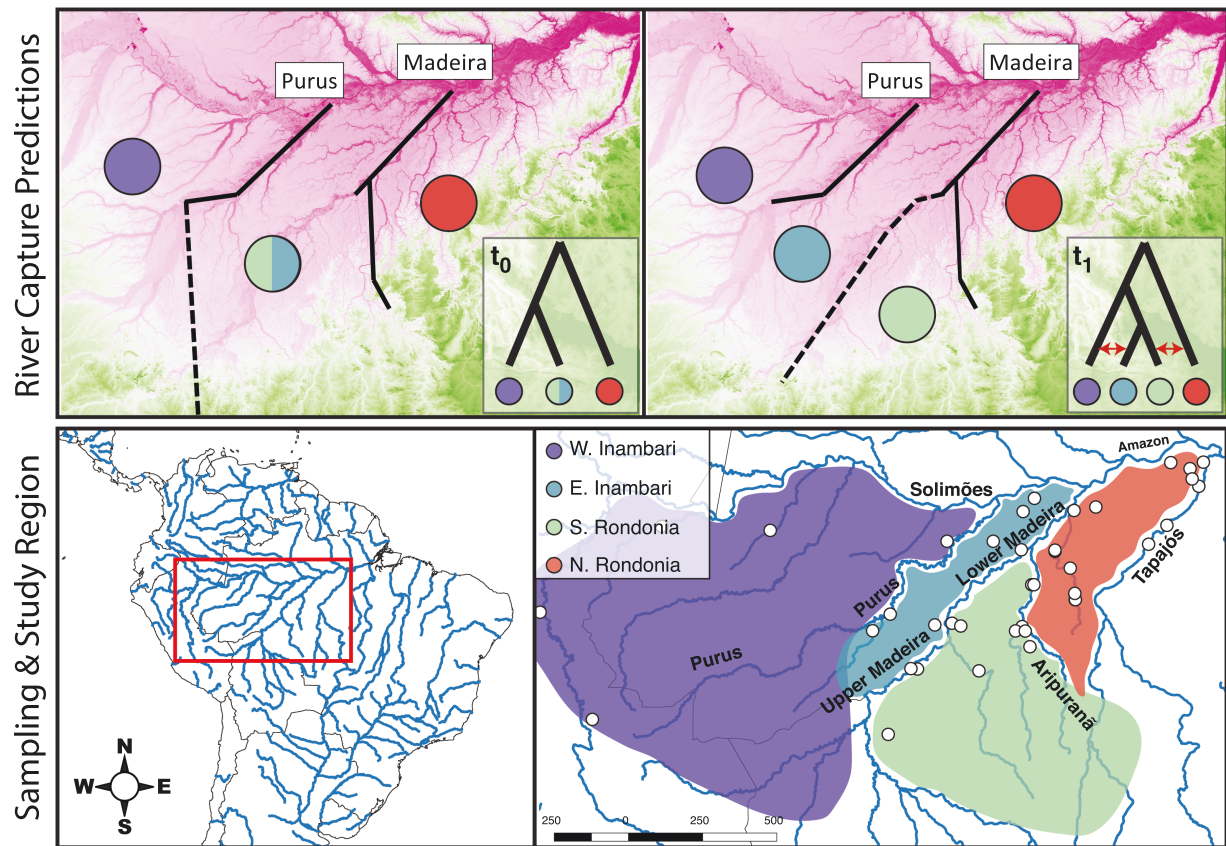


FIGURE 1. Spatially explicit predictions of the river capture scenario (top row). The stippled lines show the movement of a paleo-river, historically draining water via the Purus watershed (time = t_0) and its subsequent capture by the modern Madeira watershed (time = t_1). Phylogenetic predictions assume a set of historical relationships (t_0); the predicted topology (t_1) matches that inferred from mitochondrial DNA in a previous study (Thom and Aleixo 2015). Red arrows indicate opportunities for gene flow between non-sister populations. The maps in the top row are colored by topography, with dark pinks representing lower elevations (e.g., river channels) and darker greens representing uplands. The bottom row shows the study region, sampling (white circles), and key interfluvial areas (from west to east): W Inambari (purple), E Inambari (teal), S Rondônia (pale green), and N Rondônia (red).

and S2). First, we demultiplexed and cleaned raw reads. Specifically, we used TGCA as our restriction overhang (PstI-I, which cleaves DNA at 5'-CTGCA/G-3' sites was the restriction enzyme used in library prep), allowed a maximum of 5 low-quality ($Q < 20$) base calls in each read with a Phred Q-score offset of 33, and discarded any cleaned reads with fewer than 35 bp. We then mapped cleaned reads to a newly assembled reference genome of a closely related species *Thamnophilus caeruleus* (divergence time: ~4 Ma; Harvey et al. (2020); see Supplementary Materials). iPyrad uses vsearch (Rognes et al. 2016) to merge overlapping paired reads and then bwa (Li and Durbin 2009) to map those paired reads to a reference genome and determine locus homology, discarding any unmapped or replicated reads. In sum, paired-end reads are mapped to the reference genome based on sequence similarity, and reads that cluster at the same genomic loci are then aligned using muscle (Edgar 2022). Each of these "aligned clusters" represents a single GBS locus. We applied a minimum statistical coverage of six, which is considered the minimum value for accurate base calling that is widely used in RADSeq studies (Eaton and Overcast 2020; Barreto et

al. 2022; Donoghue et al. 2022; Hanes et al. 2022). When estimating consensus allele sequences from clustered reads, we allowed default settings of a maximum of 5% ambiguous base calls and 5% heterozygous sites. These last 2 filters reduce the risk of incorporating erroneous alignments into the dataset that are likely to have increased heterozygosity and may be caused when many ambiguous bases are included in a consensus.

Population Structure and Ancestry

Characterization of the population structure of *T. aethiops* is crucial for delineating populations for downstream analysis and characterization of patterns of admixture. We first investigated genetic structure among populations using principal components analysis (PCA), implemented in the iPyrad API (Eaton and Overcast 2020). Specifically, we used k -means iterative clustering to assign individuals to populations under a 4-population model ($k = 4$). We iteratively clustered sample single-nucleotide polymorphisms (SNPs) present across 90% of individuals and clustered individuals based on an assumed value for the number of

populations. We repeated this clustering 5 times, allowing more missing data at each successive iteration until reaching a minimum coverage of 75% of individuals for a given SNP. Next, missing genotypes for all samples were imputed by choosing one random genotype from the assigned population based on the k -means assignments. This method allowed imputation without *a priori* geographic bias. We then performed PCA on unlinked genotype calls (one randomly sampled genotype per locus) in this imputed dataset.

We then used sparse non-negative matrix factorization (sNMF), implemented in the R-package LEA3 (Frichot et al. 2014; Gain and François 2021), to calculate admixture coefficients for each population and infer the best-fit number of ancestral populations (k) in the dataset. We tested values of k from one through 5 to infer the optimal number of ancestries. As sNMF results are sometimes sensitive to the regularization parameter, α , we explored our results under multiple values of α (10, 100, 1000, 10,000) using only 10 replicates. However, given our relatively large dataset, our results were stable across all α -values, so we performed 1000 iterations of each k value using $\alpha = 10$ to obtain our final results. The best-fit value of k was determined by choosing the k with the lowest cross-entropy value (Frichot et al. 2014). In addition to the best-fit value of k , we also visualized the results from other values of k , to better characterize additional, biologically relevant population structure present in the data. To assess the stability of population assignments in sNMF, we ran this procedure multiple times, replicating the 1000 iterations of sNMF over multiple runs.

Genetrees, Chromosome Trees, and Species Tree Topologies

We evaluated the phylogenetic relationships for the populations recovered in sNMF and PCA analyses ($K = 4$), which were also consistent with previously published mitochondrial clades (Thom and Aleixo 2015). We estimated genetrees for sliding windows across the genome using TreeSlider, a python program available within the iPyRAD API (Eaton and Overcast 2020). TreeSlider keeps the most common allele for each locus in each population and estimates the maximum likelihood phylogeny at each locus using RAxML (Stamatakis 2014). We specifically compared the results from TreeSlider for 10 kb, 50 kb, 100 kb, and 200 kb sliding windows, first requiring a minimum of 5 SNPs, and then a minimum of 10 SNPs to retain each window. We found that 50 kb windows with a minimum of 5 SNPs retained the most (4858; Supplementary Table S2) loci, so we used the genetrees from this dataset for all downstream phylogenetic analyses.

To understand how phylogenetic history varied across the genome and among chromosomes, we analyzed genetree results at multiple scales using ASTRAL v5.7.8 (Zhang et al. 2018). ASTRAL estimates an unrooted species tree given a set of genetrees by identifying the maximum number of shared induced quartets within the provided genetrees. We used ASTRAL

to estimate the interrelationships of four populations of interest, corresponding to those identified using sNMF and PCA. To understand how these relationships varied among parts of the genome, we first ran ASTRAL for all 4858 genetrees together (genome-wide species tree), second for only autosomal genetrees (autosomal tree; that is, excluding genetrees from sex chromosomes), and finally for genetrees from each chromosome independently (chromosomal trees). As no loci on Chromosome 33 met our criteria (50 kb minimum of 5 SNPs), we excluded this chromosome from all ASTRAL analyses.

To quantify variation in support for competing phylogenetic hypotheses, we calculated unrooted topology weights for alternative topologies across windows of the genome using TWISST (Martin and Van Belleghem 2017). This approach provides an assessment of the relative likelihood of alternative topologies for individual genomic windows. Here, we estimated genetrees from SNPs on sliding windows using PHYML v3.0 (Guindon et al. 2010) following Martin and Van Belleghem (2017). Due to the patchiness of GBS datasets, we analyzed windows of varying lengths with exactly 100 SNPs each. The larger window size and stricter filtering applied here enabled the estimation of relatively high-resolution genetrees, reducing noise in this analysis. We performed 2 independent runs testing the relationships between 4 taxa. First, we calculated topology weights for 3 unrooted topologies using the 4 ingroup populations. This configuration allowed us to test the sister relationship between E and W Inambari populations. Next, we added the reference genome (*T. caerulescens*) to the dataset and removed the W Inambari population, allowing us to test S Rondônia as sister to Inambari populations. Because TWISST quantifies the relative weights of unrooted topologies, running our analyses with 4 taxa enabled us to evaluate support for alternative quartets identified in ASTRAL as well as the mitochondrial topology, without the noise of more complicated 5-taxon statements. Initial tests with our data set revealed that 5-taxon runs had poor resolution given the larger set of alternative topologies.

Given the overall low support for the mtDNA tree across the genome (see “Results” section), we explored whether the few genomic windows with high support for mtDNA topology were linked to nuclear genes associated with mitochondrial activity. We selected genes associated with mitochondrial activity according to Morales et al. (2018) (734 N-mt genes; Gene Ontology term: 0005739 and 0006119) using our annotated reference genome, located within 100 kb of windows with topology weights above 3 standard deviations from the mean for the mtDNA tree.

Demographic Modeling

To estimate genome-wide topology for the 4 populations while accounting for gene flow, we used a combination of coalescent simulations performed with PipeMaster (Gehara et al. 2017), and supervised

machine learning implemented in Keras v2.3 (Arnold 2017; <https://github.com/rstudio/keras>). We first simulated data under 3 demographic models matching alternative phylogenetic hypotheses under the effect of gene flow between selected populations. For the purposes of this section, we abbreviate N Rondônia as “NR,” S Rondônia as “SR,” W Inambari as “WI,” and E Inambari as “EI.” The simulated models were: Model 1) (NR,(SR(EI, WI))) with gene flow between SR and both NR and EI ($M_{SR \leftrightarrow NR}$ and $M_{SR \leftrightarrow EI}$); Model 2) ((NR, SR),(EI, WI)) with gene flow between SR and EI ($M_{SR \leftrightarrow EI}$); and Model 3) (NR,(WI,(EI, SR))) with gene flow between populations on both sides of the Madeira River ($M_{SR \leftrightarrow NR}$ and $M_{WI \leftrightarrow EI}$). Each of the 3 models matches the results obtained with independent phylogenetic analysis: Models 1 and 2 match the alternative topologies obtained with ASTRAL (T1 and T2; see “Results” section); whereas Model 3 matches the mtDNA topology of Thom and Aleixo (2015) (T3; see “Results” section). Symmetric gene flow was allowed in the model between pairs of non-sister populations potentially in contact (e.g., occurring on opposite sides of a river) that could explain the discordance observed in phylogenetic estimates. We only allowed symmetric gene flow between populations to reduce the number of parameters of our models and because our primary goal was to test for alternative phylogenetic topologies while accounting for gene flow; our intent was not to estimate the absolute values of demographic parameters. For all models we set relatively large and uniform priors based on reasonable values for lowland Amazonian species: Ne 100,000–1,000,000 diploid individuals (for all populations); $M_{pop1 \leftrightarrow pop2}$ 0.1–3.0 migrants per generation; Tdiv (Divergence time in generations) Model 1: WI/EI 50,000–800,000, SR/WI+EI 500,000–1,200,000, NR/WI+EI+SR 750,000–1,500,000; Model 2: WI/EI 50,000–800,000, NR/SR 50,000–800,000, NR+SR/WI+EI 750,000–1,500,000; and Model 3: EI/SR 50,000–800,000, WI/EI+SR 500,000–1,200,000, NR/WI+EI+SR 750,000–1,500,000. We assumed a fixed mutation rate of 2.42×10^{-9} mutations per generation and a 1-year generation time (Jarvis et al. 2014; Zhang et al. 2014). Although an assumed 1-year generation time may underestimate the true generation time, limited data about survivorship and reproduction in Amazonian birds precludes a more precise estimate (Saether et al. 2005).

To obtain the observed data on which simulations were based, we initially converted the alleles file produced by iPyrad into individual fasta alignments with the iPyrad.alleles.loci2fasta function of PipeMaster. Since PipeMaster is sensitive to missing data, we also applied multiple filters to the sequence data. First, we removed individuals missing more than 50% of the loci from alignments using Alignment_Refiner_v2.py (Portik et al. 2016) and excluded alignment positions not recovered for at least 80% of the individuals using trimAL (Capella-Gutierrez et al. 2009). Finally, we excluded loci with fewer than 51 individuals, shorter than 100 bp, and missing more than 50% of the sites using AMAS and custom R scripts (Borowiec 2016).

The genetic data for each model was simulated on PipeMaster based on the number of retained loci, matching their length and number of individuals using msABC (Pavlidis et al. 2010). To summarize genetic variation of observed and simulated data, we calculated multiple population genetics summary statistics, including mean and variance across loci: (1) the number of segregating sites, both per population and summed across populations; (2) nucleotide diversity (π), both per population and for all populations combined; (3) Watterson’s theta per population and for all populations combined; (4) pairwise F_{ST} between populations; the number of shared alleles between pairs of populations; (5) the number of private alleles per population and between pairs of populations; and (6) the number of fixed alleles per population and between pairs of populations. These summary statistics were used as feature vectors on a neural network (nnet) approach implemented in Keras designed to estimate the classification probability of the 3 simulated models given our data and associated demographic parameters.

After careful parameter exploration, the final architecture of our neural network had three hidden layers with 32 internal nodes and a “relu” activation function. For model classification, our output layer was composed of 3 nodes and a “softmax” activation function. Three-quarters of the simulations were used as training data, and the remaining 25% were used to test the accuracy of our approach in assigning simulations to the correct model. The neural network training step was run for 1000 epochs using the “adam” optimizer and a batch size of 20,000 using 5% of the data for validation. To track improvements in model classification during training, we calculated the overall accuracy and the sparse_categorical_crossentropy for each epoch. After identifying the most probable model for our observed data, we estimated demographic parameters with a neural network designed to predict continuous variables. Here, we used a similar architecture to the one described above but set an output layer with a single node and a “relu” activation. We used this approach to estimate the effective size of each population, the amount of gene flow between populations, and divergence times. To assess improvements in accuracy during training, we used the mean absolute percentage error (MAE) as an optimizer. We trained the neural network for 3000 epochs with a batch size of 10,000 and a validation split of 0.1. To account for variation in parameter estimation, we ran 10 replicates and summarized the results calculating mean values for each demographic parameter. We also assessed the accuracy of parameter estimation by calculating the coefficient of correlation between estimated and true simulated values of the testing data set.

Testing the Link Between Chromosome Length, Linked Selection, and Introgression

To test for predicted patterns associated with linked selection across the genome, we calculated nucleotide

diversity (π), relative divergence (F_{ST}), and absolute divergence (d_{xy}) across the genome. To calculate π , d_{xy} , and F_{ST} , we used pixy, a command line utility that handles missing data by adjusting the site-level denominators and sequence length (Korunes and Samuk 2021). Summary statistics were estimated based on vcf files with invariant sites included. To examine the robustness of these estimates given the incomplete nature of GBS datasets, we employed pixy for multiple sliding window widths and data filters (see [Supplemental Materials](#)). We specifically examined pixy results for (1) 50 kb windows with an unfiltered vcf; (2) 50 kb windows with a vcf filtered for sites with a minimum depth of 10 \times , and (3) 250 kb windows with a vcf filtered for sites with a minimum depth of 10 \times . As we found that pixy results were robust to these depth filters and window sizes (see [Supplemental Materials](#)), we opted to use the dataset that retained the most variants, the 50 kb windows with a minimum depth of 6 \times .

To better understand how topology and rates of gene flow vary among chromosomes, we fit data to generalized linear models (glms) in R (R Core Team 2019) to test for an association with chromosome length and bootstrapped these models to examine their sensitivity (see [Supplemental Materials](#)). In so doing, we make the assumption that genome structure is relatively conserved across New World suboscine birds (Passeriformes, Tyrannides) since the chromosome lengths of the reference genome are derived from those in *Chiroxiphia lanceolata*, a manakin. This assumption is supported given the high synteny within passerine birds (Dawson et al. 2007; Ellegren 2010; Delmore et al. 2018; Coelho et al. 2019; Peñalba et al. 2020). We first modeled π , d_{xy} , and F_{ST} (unfiltered 50 kb-window dataset) as functions of chromosome length (glm family = "gaussian") based on the results from the pixy analyses outlined above. To do this, we calculated the average value for each statistic across each chromosome by taking the mean value across all windows after eliminating empty windows (i.e., NA values). As the distribution of chromosome lengths was right-skewed, we evaluated these models after \log_{10} transforming the data. This allowed us to track general patterns associated with linked selection and recombination in the data. For example, since prior studies have shown that gene flow is expected to increase on smaller chromosomes (Martin et al. 2019), understanding how patterns of divergence, genealogy, and of introgression change with chromosome length enables us to evaluate whether evidence for certain phylogenetic topologies might be driven by gene flow between non-sister populations. For all Gaussian regressions, we also estimated Pearson's correlation coefficient (r).

Next, we used logistic regression (family = "binomial") to test for an association between chromosome length and phylogenetic topology, scoring genetrees (i.e., RAxML trees from the sliding window analysis above) as monophyletic (1) or non-monophyletic (0) for populations in Rondônia, Inambari, and S Rondônia+E Inambari. These logistic regression analyses enabled

us to test for associations between genealogical relationships across the genome and chromosome length. ASTRAL recovered 2 alternative topologies across the genome corresponding to non-monophyletic (T1) and monophyletic (T2) Rondônia populations, and mtDNA recovered a sister relationship between E Inambari and S Rondônia (T3). Thus, these analyses were aimed at identifying how support for these three topologies varied across the genome to better evaluate whether certain topologies are associated with introgression. For logistic regression, we estimated the coefficient (β) and odds-ratio (ψ) for each model. We ran these logistic regression models on all 4858 genetrees as well as a subset of 1727 genetrees reconstructed from windows with a minimum of 10 SNPs. The latter allowed us to examine the robustness of these models to potential gene tree error. We also replicated this analysis on the chromosome-level ASTRAL phylogenies instead of genetrees.

To obtain direct measures of introgression, we then estimated the introgression proportion (f_{dm}) across the genome using window-based ABBA-BABA tests in non-overlapping windows with exactly 100 SNPs using ABBABABAwindows.py (https://github.com/simonhmartin/genomics_general (Martin et al. 2014)). An initial analysis using windows of fixed width (50 kb and 250 kb) produced highly biased f_{dm} estimates, with high variance in f_{dm} estimates for larger chromosomes. This was due to the incomplete nature of GBS datasets and amplification bias that led to denser sequence coverage on smaller chromosomes and sparser sequence coverage on larger chromosomes (DaCosta and Sorenson 2014) (see [Supplementary materials](#)). Thus, we chose to define windows by the number of SNPs to help reduce this bias and generate windows with equivalent information content. We specifically explored introgression between N and S Rondônia, which were defined by positive values of f_{dm} assuming T1 as the species tree (P1 = WI, P2 = SR, P3 = NR, out = reference; see "Results" section).

These introgression results were then used to model f_{dm} and the proportion of derived alleles shared by N and S Rondônia per window (variants shared by P2 and P3 but not P1 or P4; i.e., ABBA's) as functions of chromosome length to understand how introgression varied across the genome. As positive values of f_{dm} indicate introgression between P2 and P3 (N and S Rondônia), we modeled positive f_{dm} values as a function of chromosome length after excluding negative values from the data frame. To evaluate whether introgression on the Z-chromosome was lower than mean autosomal introgression, we also used an unpaired Bayesian t -test for distributions with unequal variances implemented in the R-package Bolstad (Curran 2013) to statistically test differences in mean values of f_{dm} ($p2 < p3$) across windows on the Z-chromosome versus the mean f_{dm} ($p2 < p3$) for windows on all autosomes combined.

Finally, previous work has shown that recombination not only increases on small chromosomes but also on the periphery of large chromosomes (Haenel

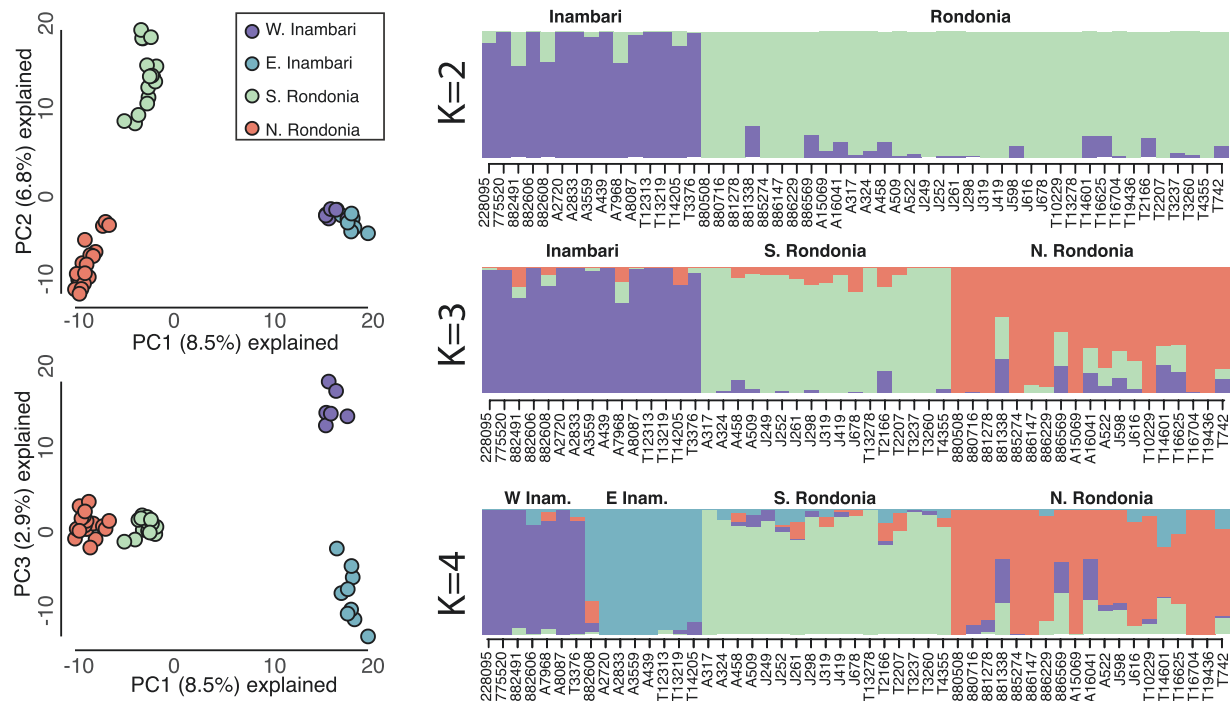


FIGURE 2. Results of PCA and ancestry on the full GBS dataset from samples spanning 4 populations. Colors correspond to shaded regions on the map (Fig. 1): W Inambari (purple), E Inambari (teal), S Rondônia (pale green), and N Rondônia (red). (Left) Results of PCA on all samples. (Right) sNMF results for the optimal K value of $K = 2$ (top), as well as plots for $K = 3$ (center) and $K = 4$ (bottom).

et al. 2018). To test for a potential effect of increased gene flow on the periphery of large chromosomes, we modeled f_{dm} ($p2 < \rightarrow p3$) as a function of percent distance from a chromosome's center (defined as the distance in base pairs of a window from the center of a chromosome divided by half the chromosome's length in base pairs). All statistical analyses were performed in R version 4.1.2 (R Core Team 2019). From all analyses, we excluded the W-chromosome, which does not undergo crossing-over and is subject to error given many males in the dataset. We also removed Chromosomes 31, 32, and 33, for which we recovered relatively little data.

RESULTS

GBS Assembly

Our final GBS assembly included a total of 118,127 loci with a mean and standard deviation of 62,882.65 and 17,506.31 loci assembled per sample, respectively (range = 11,955–85,123 loci). Within this dataset, we recovered a total of 1,277,143 SNPs.

Population Structure and Ancestry

Our results point to between 2 and 4 populations across the region corresponding to W Inambari, E Inambari, S Rondônia, and N Rondônia, including individuals

with a shared coefficient of ancestrality between W Inambari and S Rondônia, as well as between N and S Rondônia (Fig. 2). Although the best-fit value for the number of ancestries was $k = 2$ (Supplementary Fig. S1), corresponding to populations east and west of the Madeira River, higher values of k were consistent with the PCA results, recovering additional distinct ancestries across the Aripuanã ($k = 3$) and Purus ($k = 4$) Rivers. Contrastingly, PCA consistently recovered 4 clusters of samples corresponding to areas delimited by the 3 focal rivers (Fig. 2). These PCA clusters match the spatial boundaries of mitochondrial clades from a previous study (Thom and Aleixo 2015). Still, in consistency with the sNMF results, PC1 (explaining 8.5% of the genomic variation) separated populations from across the Madeira river, PC2 (explaining 6.8% of the genomic variation) separated populations across the Aripuanã river, and PC3 (explaining 2.9% of the genomic variation) separated populations across the Purus river.

Genetrees, Chromosome Trees, Species Tree, and Demographic History

We detected 3 competing phylogenetic topologies across the genome, 2 of which predominated among chromosome trees (Fig. 3): Topology 1 (T1) matched the genome-wide species tree and was recovered for 9 of the 34 chromosomes, including the Z-chromosome; T2 matched the autosomal tree and was recovered for 22 of the autosomes. There were also 3 unique topologies

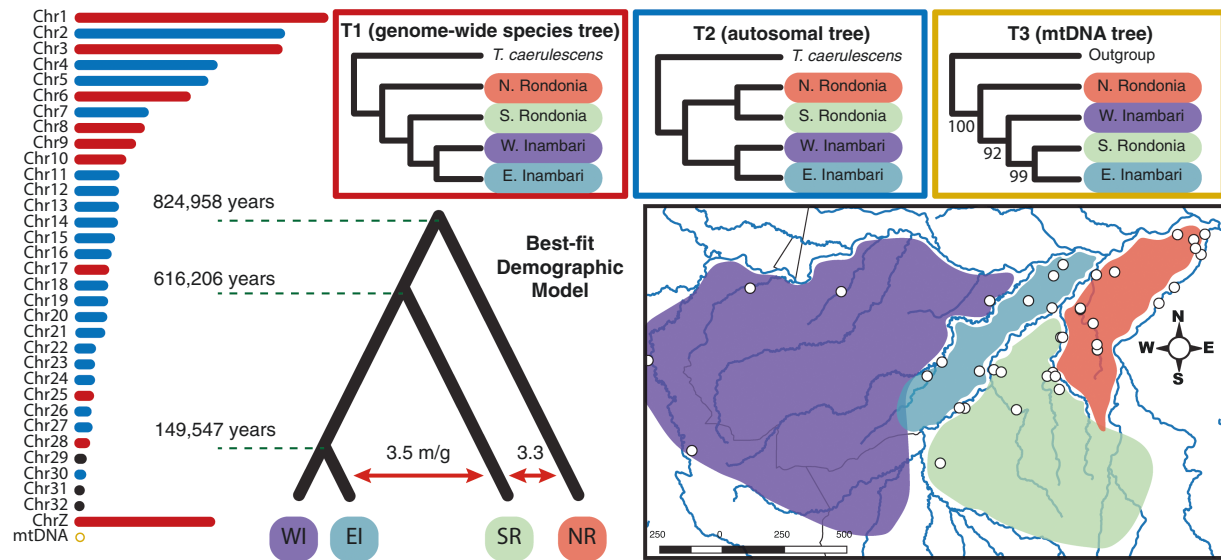


FIGURE 3. Chromosome-level topologies inferred in astral for 50 kb sliding windows across the genome. Two chromosome tree topologies predominated across the genome: T1 (red chromosomes/red box), which is consistent with the genome-wide tree, and T2 (blue chromosomes/blue box), which is consistent with the autosomal tree. The mitochondrial topology is also shown (T3) with bootstrap support, based on the results of a previous study (Thom and Aleixo 2015). The tree at the lower left shows the best-fit demographic model inferred using a convolutional neural network. Bars in the left column are proportional to chromosome length.

recovered for Chromosome 29, Chromosome 31, and Chromosome 32 (Supplementary Fig. S2). None of these were concordant with the mitochondrial topology recovered in a previous study (henceforth, T3) (Thom and Aleixo 2015). Overall, populations from Rondônia were recovered as monophyletic for 22 chromosome trees and Inambari populations were monophyletic for all but 3 chromosomes.

Demographic modeling similarly recovered T1 as the population history (model classification probability: model 2 = 0.99), finding high gene flow between S Rondônia and E Inambari as well as between N and S Rondônia (Fig. 3). Model classification yielded high accuracy, with training data being correctly assigned to the simulated model 99% of the time (accuracy = 0.99). The neural network regression approach designed for demographic parameter estimations produced accurate estimates for most parameters (R^2 simulated ~ estimated > 0.90), except for the 2 oldest divergence times, likely due to the high amount of gene flow between populations (R^2 ~ 0.61; Supplementary Table S4). Effective population sizes were in general consistent with the size of the geographic distribution of the species, with the most restricted E Inambari populations having the smallest size (198,314 individuals; R^2 = 0.93; MAE = 72,272). However, because we only modeled symmetric gene flow between non-sister populations (and not all possible gene flow scenarios), we suggest these N_e be interpreted with caution. The divergence between E and W Inambari populations occurred around 150 Ka (149,546ya; R^2 = 0.95; MAE = 48,492), followed by the split between the ancestor of the Inambari populations from S Rondônia at about 600 Ka (616,206ya; R^2 = 0.61; MAE = 117,941), and the deepest divergence at around 820 Ka (824,957ya; R^2 = 0.61; MAE

= 125,875). Gene flow estimates were high between both migration edges of the model suggesting considerable introgression between E Inambari and S Rondônia, and between Rondônia populations. Parameter estimates were in general contained within simulated priors except for gene flow estimates. Additional runs adjusting priors for gene flow drastically affected the accuracy of model classification, thus we assumed these constrained and conservative estimates.

The support for the 3 alternative topologies varied across the genome (Supplementary Figs. S3 and S4; Table 1). The unrooted T1/T2 topology (T1 and T2 are the same when the ingroup quartet is unrooted) had the highest support across the genome in our analysis for the ingroup populations. However, when including the reference to differentiate between T1 and T2 topologies, both T1/T3 (T1 and T3 are the same when this quartet is unrooted) and T2 had similar support across the genome. There were relatively few outlier loci supporting T3 compared with the other topologies, and we did not find any association between n-mt genes and outlier peaks supporting the mitochondrial topology.

Associations with Chromosome Length

We detected significant associations between chromosome length, genetree topologies, and genome statistics (Fig. 4). First, we found a significant association with F_{ST} (n = 32, 50 kb windows: r = 0.818, P < 0.0001; 250 kb windows: n = 32, r = 0.666, P < 0.0001; Fig. 4a), d_{xy} (50 kb windows: n = 32, r = -0.856, P < 0.0001; 250 kb windows: n = 32, r = -0.931, P < 0.0001; Fig. 4b), and π (S Rondônia: 50 kb windows, n = 32, r = -0.885, P < 0.0001, 250 kb windows, n = 34, r = -0.868, P < 0.0001; 50 kb windows N

TABLE 1. Number of outlier windows supporting 3 alternative topologies (T1/T2, T3, and a third unrooted topology Tx), and the overall number of genes, and number of genes linked to mitochondrial activity less than 100 kb from outlier peaks (N-mt genes) on [Supplementary Fig. S4](#) [ingroup TWISST analysis].

Unrooted topology	Number of outlier windows	Number of genes	Number of N-mt genes	Number of windows with highest weight
T1/T2	43	262	19	1586
Tx	8	26	1	360
T3	10	17	1	441

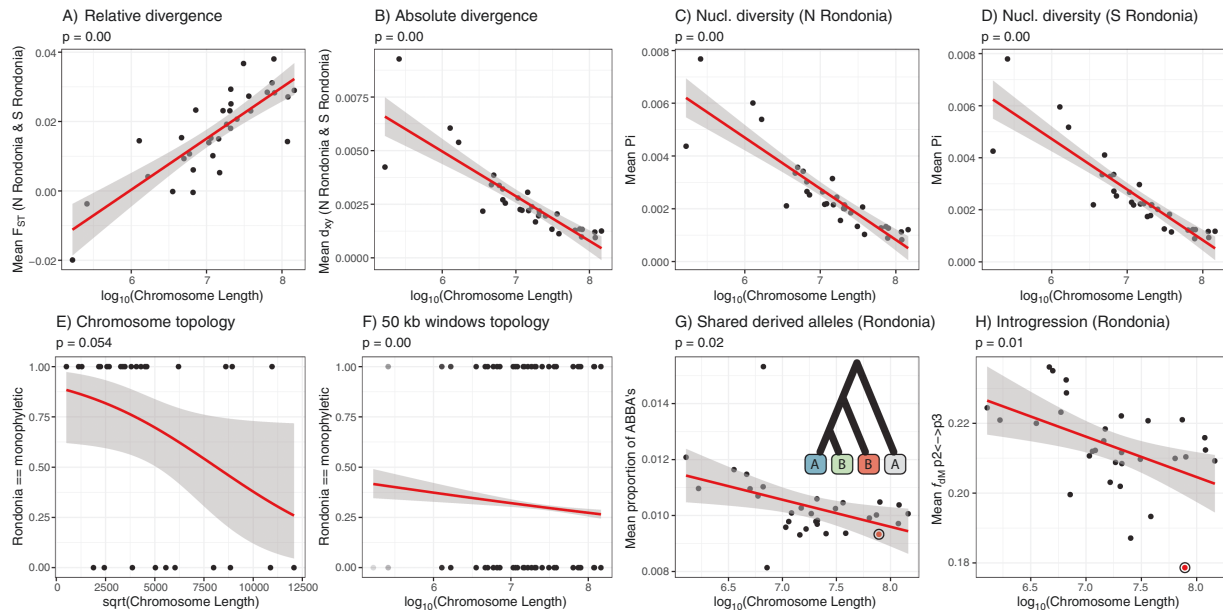


FIGURE 4. Logistic and linear regression tests for associations between genomic characteristics and chromosome length. Y-axis values in the top row reflect chromosome-wide means of 50-kb windows for (a) F_{ST} , (b) d_{xy} , (c) π for N Rondônia, and (d) π for S Rondônia. Plots in the bottom row examine the influence of introgression on genetree variation across the genome showing (e) a nonsignificant association between chromosome topologies and chromosome length despite (f) a significant negative association in genetree topologies consistent with T2 and chromosome size, and (g and h) significant negative associations between introgression and chromosome length. Models for each association are shown as the solid line with the shaded gray area representing the model standard error. Bullseye points in plots g and h represent values for the Z-chromosome. Derived variants shared by Rondônia were defined as SNPs showing an ABBA pattern assuming topology T1, as indicated using the tree at the top right of plot g. The topology tips, from left to right are E Inambari, S Rondônia, N Rondônia, and T. caeruleus (outgroup), with A's and B's representing alternative alleles.

Rondônia: $n = 32$, $r = -0.883$, $P < 0.0001$; 250 kb windows: $n = 34$, $r = -0.846$, $P < 0.0001$; [Fig. 4c,d](#)). Bootstrapping ([Supplementary Fig. S5](#)) and pairwise correlations based on alternate VCF filtering and window sizes showed that the results of our linear regression models were not affected by unequal sequence coverage across windows and chromosomes ([Fig. 5](#), [Supplementary Figs. S6 and S7](#)). We recovered a negative association between chromosome length and genetree topology, wherein genetrees on larger chromosomes have a reduced probability of recovering monophyly of Rondônia ($\beta = -0.231$, $\psi = 0.794$, $P = 0.0005$; [Fig. 4f](#)) and recapitulated this result at the level of chromosomes, though the latter was only weakly supported ($\beta = -0.0003$, $\psi = 0.999$, $P = 0.054$; [Fig. 4e](#)).

We also confirmed that introgression varied across the genome and was negatively correlated with chromosome length ([Fig. 4g,h](#)). Specifically, we found significant negative correlations between f_{DM} ($n = 31$, $r = -0.476$, $P = 0.007$) and the proportion of derived variants shared by N and S Rondônia (ABBA's; $n = 31$, $r = -0.420$, $P = 0.019$) with chromosome length. The Bayesian t -test confirmed ($df =$

237.74, $t = 3.566$, $P = 0.0005$) that the Z-chromosome had lower levels of introgression (mean $f_{DM} = 0.178$) than autosomes (mean $f_{DM} = 0.212$). Overall, f_{DM} values averaged positive across the genome, implying stronger gene flow between N and S Rondônia than between N Rondônia and E Inambari (positive values indicate an excess of gene flow between P2 and P3, whereas negative values indicate an excess between P1 and P3). We found no relationship between f_{DM} and distance from chromosome center ([Supplementary Fig. S9](#)).

DISCUSSION

Nonrandom Variation in Genealogical History Across the Genome

We combine phylogenomics and population genetics to investigate the interplay between genomic architecture and biogeographic processes in generating predictable patterns of genetree variation across the genome

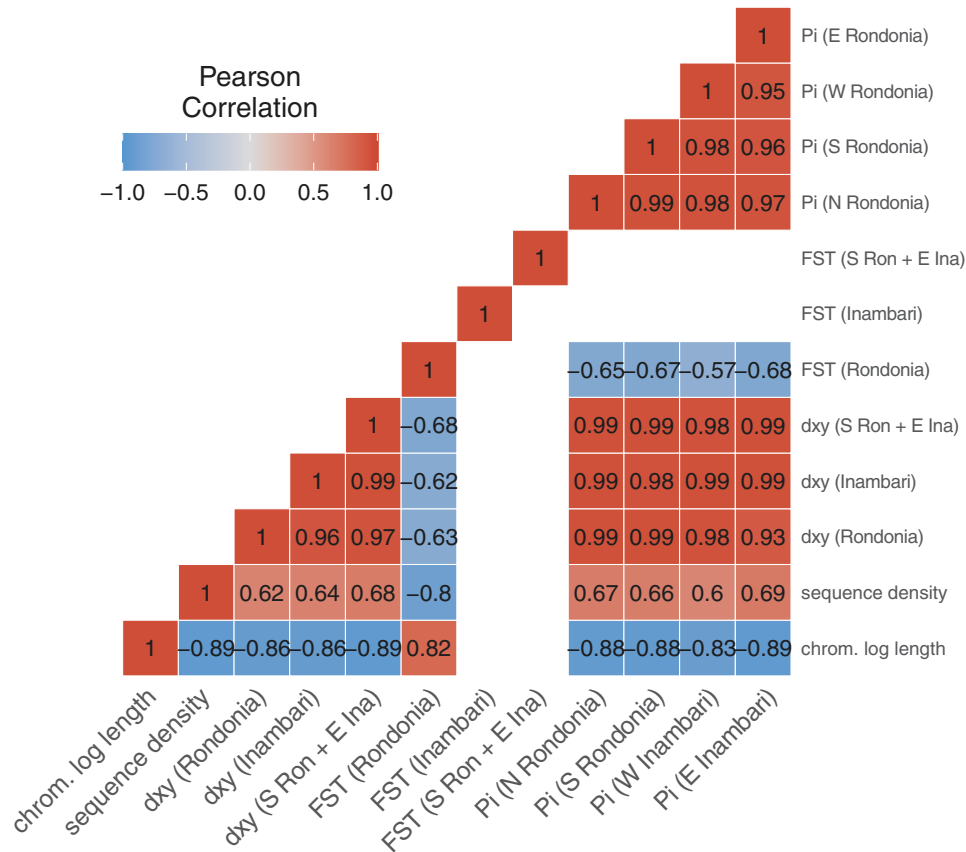


FIGURE 5. Heatmap of Pearson's coefficients for pairwise correlations showing pairwise correlations of within-chromosome means of genetic summary statistics for values of dxy, FST, and pi. Empty (white) boxes represent correlations with nonsignificant P -values ($P > 0.05$). Correlation coefficients are printed on each square.

of an Amazonian antbird, *T. aethiops*. We found that accounting for chromosome length informs phylogenetic and biogeographic inference in cases of high gene flow among non-sister taxa; it may be necessary to examine genome architecture to properly interpret phylogeographic signal when divergence occurs with gene flow. Our results also suggest that reduced representation genomic data such as genotype-by-sequencing can be used with genomic-architecture-aware approaches, recapitulating expected associations between genomic processes and the signal for ancestry and introgression.

We tested 3 competing hypotheses for the relationships of 4 spatially adjacent and genetically differentiated populations that are semi-isolated across Amazonian tributaries. Two of these topologies, T1 and T2, were equally supported across genome-wide sliding windows (Supplementary Fig. S3), which made inferring *T. aethiops*' evolutionary history challenging. We found that genealogical signal for these competing hypotheses was nonrandomly distributed across the genome; areas of low gene flow such as the Z-chromosome and larger autosomes tend to support T1, whereas areas with elevated gene flow such as smaller autosomes tend to recover genetrees consistent with T2 (Fig. 4e–f). The third topology, T3, was not well supported across the nuclear genome (Supplementary

Fig. S4) but was recovered for mitochondrial DNA. Importantly, introgression was negatively correlated with chromosome length and the Z-chromosome exhibited especially lower introgression than autosomes, consistent with expectations (Supplementary Fig. S10). Thus, we suggest that T1 may be seen as representing the initial branching pattern among the 4 taxa, whereas the prevalence of T2 on smaller autosomes probably resulted from introgression between N and S Rondônia. These results suggest that the genome-wide diversification history of *T. aethiops* might be better explained by a complex network of differentiation and introgression between multiple interacting populations.

The idea that the prevalence of T2 is driven by gene flow is supported by theoretical predictions about genome architecture. Smaller chromosomes are expected to exhibit higher levels of gene flow than larger chromosomes due to their higher recombination rates that more effectively break the linkage between introgressed variants (Martin et al. 2019; Tigano et al. 2022). During Meiosis, each chromosome must undergo at least one crossing-over event resulting in fewer recombination events per base on longer chromosomes than shorter ones. This becomes complicated in birds, which may also experience increased selection on smaller chromosomes due to denser gene content

in those regions (Henderson and Brelsford 2020). However, in birds, recombination is also thought to be tied to gene promoter regions (Singhal et al. 2015). Thus, although there is an increase in the density of targets of selection along smaller chromosomes, there is also additional recombination acting to reduce the effects of such linked selection. To test this assumption, we modeled multiple population genetic statistics as functions of chromosome length. We found that F_{ST} decreased on smaller chromosomes, but π and d_{xy} were negatively correlated with chromosome length. These patterns are consistent with expectations associated with a reduction in overall genetic diversity due to linked selection on larger chromosomes, thereby increasing F_{ST} and decreasing d_{xy} . We also found that introgression statistics such as f_{DM}^{xy} and the proportion of derived alleles shared by Rondônia populations (i.e., ABBA's given T1) were negatively correlated with chromosome length. Overall, these results confirm our assumption that rates of gene flow between N and S Rondônia are higher on smaller chromosomes and likely driven by increased recombination.

Biogeographic and Genomic Implications

Understanding how variation in genealogical history is associated with genomic processes enables a closer look at the processes driving population divergence and speciation in *T. aethiops* and can illuminate our understanding of Amazonian biogeography in general. Amazonian biogeography has been at the center of discussions on how landscape evolution leads to allopatric speciation (Haffer 1997; Ribas et al. 2012), but increasingly, researchers are discovering that the histories of taxa across this landscape are marked by high gene flow (Barrera-Guzmán et al. 2022). Recent genomic studies have reported introgression across rivers and considerable phylogenetic conflict, often despite strong genetic and phenotypic structuring (Thom et al. 2021; Del-Rio et al. 2022; Musher et al. 2022). This includes examples of hybrid speciation (Barrera-Guzmán et al. 2018), mitonuclear discordance (Del-Rio et al. 2022), mitochondrial capture (Ferreira et al. 2018), and extensive introgression across river headwaters (Weir et al. 2015). Phylogenetic relationships among interfluvial populations have been used to inform paleogeographic models of landscape evolution and have also helped to generate and test multiple biogeographic hypotheses (Cracraft and Prum 1988; Ribas et al. 2012). However, the unique configuration of the Amazon Basin with massive, unstable tributaries flowing in parallel facilitates episodic or continuous gene flow, which can result in reticulate patterns of differentiation (Barrera-Guzmán et al. 2018; Thom et al. 2018). This process directly affects phylogenetic inference and, if not fully understood, hampers researchers from obtaining a detailed understanding of the region's biogeographic history. Our study highlights how genealogical patterns vary predictably across the genome and inform biogeographic inference (Martin et al. 2019).

We suggest that large-scale river capture events can result in historical signatures of discordant genealogy across the genomes of species that respond to rivers as barriers (Musher et al. 2022). The river capture scenario postulated in prior studies (Fernandes 2013; Weeks et al. 2016; Ruokolainen et al. 2019) predicts a sister relationship between Inambari and S Rondônia with gene flow between E and W Inambari as well as between N and S Rondônia. Thus, either T1 or T3 might match the spatial-phylogenetic expectations under a river-vicariance model. However, our results are consistent with a more nuanced set of expectations associated with barrier change wherein genomic heterogeneity is associated with multiple distinct genealogical histories. Assuming that T1 reflects the history of population isolation (or at least reduced introgression) across the Aripuanã, then T2 results from secondary contact and lineage fusion within the traditionally recognized area of endemism, Rondônia (Cracraft 1985). This gene flow among N and S Rondônia appears to be resulting in autosomal homogenization; sNMF preferred a model of $k = 2$, recovering 2 ancestral populations corresponding to Inambari and Rondônia (Fig. 2 and Supplementary Fig. S1). Within Rondônia, the lack of any apparent plumage variation in *T. aethiops* also supports the notion that there is homogenizing gene flow between the N and S Rondônia populations.

A key objective of this study was to dissect the interplay between genomic and biogeographic processes in generating genomic heterogeneity, which requires some knowledge about landscape history. Recently, researchers have proposed a field of study, called "geogenomics," wherein patterns of genomic differentiation and genomically inferred timings of divergence and gene flow can be used to help test paleogeographic models (Dawson et al. 2022; Ribas et al. 2022). As shown here and elsewhere, spatial diversification patterns within Amazonia are reticulated, and unraveling the evolutionary history of taxa in this system is nontrivial (Dagosta and Pinna 2017; Dagosta and De Pinna 2019). Understanding the geological context of river capture while accounting for intrinsic genomic processes, however, aids in the interpretation of alternative phylogenetic histories. For instance, if the Madeira headwaters were captured within the past few hundred thousand years, the Aripuanã river must predate the upper Madeira, as implied by both T1 and T3. Given that T1 is probably less impacted by introgression and populations within Rondônia are now homogenizing, the Aripuanã probably represented a more important barrier for *T. aethiops* prior to the capture event that is now weakened. In this way, we have a window into the process driving the formation of areas of endemism. As new barriers form on the landscape, old barriers erode; if differentiated taxa are not reproductively isolated, as appears the case for *T. aethiops* populations in Rondônia, they may fuse into a single taxon whose distribution conforms to the boundaries of the new river-barriers, leaving behind only reciprocally monophyletic mitochondrial groups that potentially match the ancestral landscape configuration.

Extensive paleochannels between the Jiparaná and Aripuanã tributaries of the middle Madeira Basin suggest that these rivers have historically been larger and behaved as a dynamic megafan, a hypothesis also supported by biological data (Latrubesse 2002; Wilkinson et al. 2010; Ferreira et al. 2017). We thus postulate a historical river somewhere in the vicinity of these 2 tributaries that could have acted as a historical barrier to dispersal. Under this scenario, the paleo-Madeira River would have been flowing via the Jiparaná or Aripuanã basins or somewhere in between (Hayakawa and Rossetti 2015), acting as a historical barrier for taxa located on either side. Likewise, the Tapajós basin to the east was probably drained via this paleo-Madeira River (Rossetti 2014). If so, the formation of the modern Tapajós would have drawn water away from this basin, reducing the barrier's strength. Once the Purus tributary was captured, the modern Madeira formed by extending its headwaters, which, in turn, generated a new barrier on the landscape. If taken at face value, our best-fit demographic model suggests that this river capture occurred roughly 600 Ka, at least 400 ky prior to the current geological estimate (Ruokolainen et al. 2019), but in line with divergence times of some other taxa in the region (Silva et al. 2019; Musher et al. 2022).

Interpreting the Mitochondrial Topology

Given that T1 and T2 are the primary genealogical signals across the genome, we are left to evaluate the mechanisms that gave rise to the mitochondrial tree (T3). The dispersal ecology of resident Amazonian bird species is poorly understood, but male birds are generally considered philopatric, which means it is unlikely that male-biased dispersal drives mtDNA patterns. T3 was estimated based on 2 loci, cytochrome-*b* and NADH-dehydrogenase subunit 2, and is well supported (Thom and Aleixo 2015). Mitochondrial DNA is known to have faster coalescent times due to its reduced (one quarter) effective population size and has thus traditionally been viewed as an efficacious phylogenetic marker for detecting divergences that occur over short timescales (Zink and Barrowclough 2008). However, mtDNA is a single locus and, therefore, might be expected to disagree with the species tree under certain conditions (Maddison 1997). For example, if we assume T1 to be the "true" species tree, then T3 may have resulted from mitochondrial capture during hybridization between E Inambari and S Rondônia, as documented in other groups (Ferreira et al. 2018; Myers et al. 2022). If this were the case, we should find portions of the genome that are associated with the cell respiratory system to be introgressed due to the same event (Morales et al. 2018). Yet, there is limited support for T3 across the genome (Supplementary Fig. S4), and the windows that do support T3 are not clearly linked to genes associated with mitochondrial activity (Table 1). However, because we used GBS data for this analysis, we could be missing crucial mitonuclear gene clusters in our dataset.

Alternatively, incomplete lineage sorting or biogeographic history could have given rise to the mtDNA

topology. For example, it could have resulted from a deep coalescent event wherein W Inambari haplotypes failed to coalesce with E inambari haplotypes. This is especially likely in instances of very rapid divergence (Degnan and Rosenberg 2006). Given that divergences in our ingroup occurred in under 1 million years, this is certainly plausible, despite the rapid fixation rate of mtDNA. However, it is also possible that T3 represents the history of population isolation across rivers (Fig. 1); it is striking that the distribution of reciprocally monophyletic populations based on mtDNA (Thom and Aleixo 2015) matches the spatial patterns recovered in the PCA and that *T. a. injunctus* (E Inambari) is most phenotypically similar to *T. a. punctuliger* (Rondônia). Therefore, if the T3 topology indicates the signal of population isolation across rivers then, introgression between multiple non-sister lineages has nearly erased the signal of that isolation from the genome since relatively few nuclear loci support T3 (Table 1). In this case, T1 does not exactly show the history of isolation across rivers, but might just reflect fewer genetic incompatibilities between E and W Inambari, which are more recently diverged, than between N and S Rondônia. In other words, T1 may not reflect a lack of gene flow altogether, but instead easier gene flow between E and W Inambari, which lack as many incompatibilities.

The genomes of populations that diversify on dynamic landscapes contain the signatures of multiple histories; that is, they are reticulated. To dissect these histories of isolation and secondary contact, we argue that it is important to understand the biogeographic mechanisms that give rise to predictable genomic patterns. If one's objective is to model the relationships of taxa—that is, reconstruct phylogeny—T1 is the best-supported topology given our phylogenetic and demographic modeling results. However, if the goal is to decipher the complex biogeographic history of these taxa across space and time, we may want to know whether the mtDNA tree resulted from mitochondrial capture, deep coalescence, historical isolation, or even simply phylogenetic error. We suggest that T3 could represent the historical population divergences across rivers as they formed. The divergence time based on mtDNA between E Inambari and S Rondônia occurred at roughly 200 Ka (Thom and Aleixo 2015), which lines up well with geological estimates for the Madeira River capture event (Ruokolainen et al. 2019). However, future studies based on whole-genome resequencing data may be necessary to fully understand the complicated patterns of isolation, reticulation, and homogenization in *T. aethiops*, and the relationships between these processes and genome architecture. It would also enable more detailed models of historical demography and selection not possible with GBS data.

Concluding Remarks

Gene flow may be a creative or destructive force with regard to divergence and speciation. Introgression among divergent populations, wherein standing genetic

variation among populations is episodically reshuffled into novel combinations might be an underappreciated speciation mechanism (Marques et al. 2019). We showed that introgression remains high among taxa of a common understory antbird, *T. aethiops*. Specifically, taxa within *T. aethiops* seem to remain differentiated, despite ongoing and apparently intense introgression. This has been shown in many other Neotropical taxa (Martin et al. 2013; Ebersbach et al. 2020; Musher et al. 2022). Given the potential for high rates of episodic isolation and reconnection due to the movement of large rivers, conditions in Amazonian lowlands, like those in some African lakes (Aguilée et al. 2013; Meier et al. 2017), are potentially ideal for this combinatorial mechanism of adaptation to promote diversification in lowland birds. Still, the process of high introgression can also result in higher rates of extinction (i.e., homogenization) for these young, weakly differentiated taxa, as they fuse with other lineages (Harvey et al. 2017; Barrera-Guzmán et al. 2022). Nonetheless, our results suggest that a nontrivial portion of genealogical heterogeneity across the genome arises due to extrinsic processes—such as river-course rearrangement—interacting with intrinsic processes associated with genome architecture.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: <https://doi.org/10.5281/zenodo.7460835>.

FUNDING

L.J.M. was funded in part by NSF grant DEB-1855812 to J.D. Weckstein.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We are grateful to E. Griffith, A. Del Grosso, J. Weckstein, and J. Merwin for comments on an earlier draft of this manuscript. We also thank B. Faircloth for advice and assistance, as well as the Bauer Core Facility and Cannon/Odyssey HPC system staff at the Harvard University Faculty of Arts and Sciences for support and computational resources needed for sequencing and assembly of the *T. caerulescens* reference genome. We thank S.V. Edwards and the Department of Organismic and Evolutionary Biology at Harvard University for providing financial support to generate the reference genome of *Thamnophilus caerulescens*. L.J.M. was funded in part by NSF grant DEB-1855812 to J.D. Weckstein. We also thank curators and staff at the Instituto Nacional de Pesquisas Amazonica, Museu Paraense Emilio Goeldi, American

Museum of Natural History, and Louisiana Museum of Natural Science for loaning invaluable genetic resources. Reviews from F. Burbrink and two anonymous referees greatly improved the quality of this manuscript.

DATA AVAILABILITY

Supplemental material and scripts are available <https://doi.org/10.5061/dryad.tht76hf3b>. The *T. caerulescens* assembly and raw reads have been submitted to NCBI under project number PRJNA1016420.

REFERENCES

- Aguilée R., Claessen D., Lambert A. 2013. Adaptive radiation driven by the interplay of eco-evolutionary and landscape dynamics. *Evolution* 67:1291–1306.
- Albert J.S., Craig J.M., Tagliacollo V.A., Petry P. 2018. Upland and lowland fishes: a test of the river capture hypothesis. In: Hoorn C., Perrigo A., Antonelli A., editors. *Mountains, climate and biodiversity*. New York (NY): Wiley-Blackwell. p. 273–294.
- Barrera-Guzmán A.O., Aleixo A., Faccio M., Dantas S. de M., Weir J.T. 2022. Gene flow, genomic homogenization and the timeline to speciation in Amazonian manakins. *Mol. Ecol.* 31:4050–4066.
- Barrera-Guzmán A.O., Aleixo A., Shawkey M.D., Weir J.T. 2018. Hybrid speciation leads to novel male secondary sexual ornamentation of an Amazonian bird. *Proc. Natl. Acad. Sci. U.S.A.* 115:E218–E225.
- Barreto S.B., Knowles L.L., Mascarenhas R., Affonso P.R.A. de M., Batalha-Filho H. 2022. Drainage rearrangements and in situ diversification of an endemic freshwater fish genus from north-eastern Brazilian rivers. *Freshw. Biol.* 67:759–773.
- Bicudo T.C., Sacek V., de Almeida R.P., Bates J.M., Ribas C.C. 2019. Andean tectonics and mantle dynamics as a pervasive influence on Amazonian ecosystem. *Sci. Rep.* 9:16879.
- Borowiec M.L. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4:e1660.
- Burbrink F.T., Gehara M. 2018. The biogeography of deep time phylogenetic reticulation. *Syst. Biol.* 67:743–744.
- Capella-Gutierrez S., Silla-Martinez J.M., Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.
- Capparella A.P. 1991. Neotropical avian diversity and riverine barriers. *Acta Congressus Internationalis Ornithologici* 20:307–316.
- Charlesworth B. 1998. Measures of divergence between populations and the effect of forces that reduce variability. *Mol. Biol. Evol.* 15:538–543.
- Coelho L.A., Musher L.J., Cracraft J. 2019. A multireference-based whole genome assembly for the obligate ant-following antbird, *rhegmatorhina melanosticta* (Thamnophilidae). *Diversity* 11:144.
- Cracraft J. 1985. Historical biogeography and patterns of differentiation within the South American avifauna: areas of endemism. *Ornithol. Monogr* 36:49–84.
- Cracraft J., Prum R.O. 1988. Patterns and processes of diversification: speciation and historical congruence in some neotropical birds. *Evolution* 42:603–620.
- Cracraft J., Ribas C.C., d'Horta F.M., Bates J., Almeida R.P., Boubli J.P., Campbell K.E., Cruz F.W., Ferreira M., Fritz S.C., Grohmann C.H., Latruesse E.M., Lohmann L.G., Musher L.J., Nogueira A., Sawakuchi A.O., Baker P. 2020. The origin and evolution of Amazonian species diversity. In: Rull V., Carnaval A.C., editors. *Neotropical diversification: patterns and processes*. Springer. p. 225–244.
- Cruikshank T.E., Hahn M.W. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* 23:3133–3157.

- Curran J.M. 2013. Bolstad2: Bolstad functions. R package version 1.0-28. In: Bolstad W.M., editor. Understanding computational Bayesian statistics. Vol. 644. John Wiley & Sons.
- DaCosta J.M., Sorenson M.D. 2014. Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. *PLoS One* 9:e106713.
- Dagosta F.C.P., De Pinna M. 2019. The fishes of the Amazon: distribution and biogeographical patterns, with a comprehensive list of species. *Bull. Am. Mus. Nat. Hist.* 2019:1.
- Dagosta F.C.P., de Pinna M. 2017. Biogeography of Amazonian fishes: deconstructing river basins as biogeographic units. *Neotrop. Ichthyol* 15:e170034.
- David Webb S. 1991. Ecogeography and the great American interchange. *Paleobiology* 17:266–280.
- Dawson D.A., Akesson M., Burke T., Pemberton J.M., Slate J., Hansson B. 2007. Gene order and recombination rate in homologous chromosome regions of the chicken and a passerine bird. *Mol. Biol. Evol.* 24:1537–1552.
- Dawson M.N., Ribas C.C., Dolby G.A., Fritz S.C. 2022. Geogenomics: toward synthesis. *J. Biogeogr.* 49:1657–1661.
- Degnan J.H., Rosenberg N.A. 2006. Discordance of species trees with their most likely gene trees. *PLoS Genet.* 2:e68.
- Delmore K.E., Lugo Ramos J.S., Van Doren B.M., Lundberg M., Bensch S., Irwin D.E., Liedvogel M. 2018. Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds. *Evol. Lett.* 2:76–87.
- Del-Rio G., Rego M.A., Whitney B.M., Schunck F., Silveira L.F., Faircloth B.C., Brumfield R.T. 2022. Displaced clines in an avian hybrid zone (Thamnophilidae: Rhytmatorhina) within an Amazonian interfluvium. *Evolution* 76:455–475.
- Donoghue M.J., Eaton D.A.R., Maya-Lastra C.A., Landis M.J., Sweeney P.W., Olson M.E., Cacho N.I., Moeglein M.K., Gardner J.R., Heaphy N.M., Castorena M., Rivas A.S., Clement W.L., Edwards E.J. 2022. Replicated radiation of a plant clade along a cloud forest archipelago. *Nat. Ecol. Evol.* 6:1318–1329.
- Eaton D.A.R., Overcast I. 2020. ipyrad: interactive assembly and analysis of RADseq datasets. *Bioinformatics* 36:2592–2594.
- Ebersbach J., Posso-Terranova A., Bogdanowicz S., Gómez-Díaz M., García-González M.X., Bolívar-García W., Andrés J. 2020. Complex patterns of differentiation and gene flow under the divergence of asexual phenotypes in Oophaga poison frogs. *Mol. Ecol.* 29:1944–1956.
- Edgar R.C. 2022. Muscle5: high-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat. Commun.* 13:6968.
- Ellegren H. 2010. Evolutionary stasis: the stable chromosomes of birds. *Trends Ecol. Evol.* 25:283–291.
- Elshire R.J., Glaubitz J.C., Sun Q., Poland J.A., Kawamoto K., Buckler E.S., Mitchell S.E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379.
- Endler J.A. 1977. Geographic variation, speciation, and clines. *Monogr. Popul. Biol.* 10:1–246.
- Feder J.L., Egan S.P., Nosil P. 2012. The genomics of speciation-with-gene-flow. *Trends Genet.* 28:342–350.
- Fernandes A.M. 2013. Fine-scale endemism of Amazonian birds in a threatened landscape. *Biodivers. Conserv.* 22:2683–2694.
- Ferreira M., Aleixo A., Ribas C.C., Santos M.P.D. 2017. Biogeography of the Neotropical genus *Malacoptila* (Aves: Bucconidae): the influence of the Andean orogeny, Amazonian drainage evolution and palaeoclimate. *J. Biogeogr.* 44:748–759.
- Ferreira M., Fernandes A.M., Aleixo A., Antonelli A., Olsson U., Bates J.M., Cracraft J., Ribas C.C. 2018. Evidence for mtDNA capture in the jacamar *Galbula leucogastra*/chalcophorax species-complex and insights on the evolution of white-sand ecosystems in the Amazon basin. *Mol. Phylogenet. Evol.* 129:149–157.
- Flantua S.G.A., O'Dea A., Onstein R.E., Giraldo C., Hooghiemstra H. 2019. The flickering connectivity system of the north Andean páramos. *J. Biogeogr.* 41:1227.
- Fontaine M.C., Pease J.B., Steele A., Waterhouse R.M., Neafsey D.E., Sharakhov I.V., Jiang X., Hall A.B., Catteruccia F., Kakani E., Mitchell S.N., Wu Y.-C., Smith H.A., Love R.R., Lawnczak M.K., Slotman M.A., Emrich S.J., Hahn M.W., Besansky N.J. 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* 347:1258524.
- Frichot E., Mathieu F., Trouillon T., Bouchard G., François O. 2014. Fast and efficient estimation of individual ancestry coefficients. *Genetics* 196:973–983.
- Gain C., François O. 2021. LEA 3: factor models in population genetics and ecological genomics with R. *Mol. Ecol. Resour.* 21:2738–2748.
- Gascon C., Malcolm J.R., Patton J.L., da Silva M.N., Bogart J.P., Loughheed S.C., Peres C.A., Neckel S., Boag P.T. 2000. Riverine barriers and the geographic distribution of Amazonian species. *Proc. Natl. Acad. Sci. U.S.A.* 97:13672–13677.
- Gehara M., Garda A.A., Werneck F.P., Oliveira E.F., da Fonseca E.M., Camurugi F., Magalhães F.M., Lanna F.M., Sites J.W. Jr, Marques R., Silveira-Filho R., São Pedro V.A., Colli G.R., Costa G.C., Burbrink F.T. 2017. Estimating synchronous demographic changes across populations using aABC and its application for a herpetological community from northeastern Brazil. *Mol. Ecol.* 26:4756–4771.
- Gompert Z., Lucas L.K., Buerkle C.A., Forister M.L., Fordyce J.A., Nice C.C. 2014. Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Mol. Ecol.* 23:4555–4573.
- Guindon S., Dufayard J.-F., Lefort V., Anisimova M., Hordijk W., Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59:307–321.
- Haanel 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. *Mol. Ecol.* 27:2477–2497.
- Haffer J. 1997. Alternative models of vertebrate speciation in Amazonia: an overview. *Biodivers. Conserv.* 6:451–476.
- Hanes M.M., Shell S., Shimu T., Crist C., Machkour M., Rabet S. 2022. The phylogeographic history of Megistostegium (Malvaceae) in the dry, spiny thickets of southwestern Madagascar using RAD-seq data and ecological niche modeling. *Ecol. Evol.* 12:e8632.
- Harvey M.G., Seeholzer G.F., Smith B.T., Rabosky D.L., Cuervo A.M., Brumfield R.T. 2017. Positive association between population genetic differentiation and speciation rates in New World birds. *Proc. Natl. Acad. Sci. U.S.A.* 114:6328–6333.
- Harvey M.G., Bravo G.A., Claramunt S., Cuervo A.M., Derryberry G.E., Battilana J., Seeholzer G.F., McKay J.S., O'Meara B.C., Faircloth B.C., Edwards S.V., Perez-Emanm J., Moyle R.G., Sheldon F.H., Aleixo A., Smith B.T., Chesser R.T., Silveira L.F., Cracraft J., Brumfield R.T., Derryberry E. 2020. The evolution of a tropical biodiversity hotspot. *Science* 370:1343–1348.
- Hayakawa E.H., Rossetti D.F. 2015. Late quaternary dynamics in the Madeira river basin, southern Amazonia (Brazil), as revealed by paleomorphological analysis. *An. Acad. Bras. Cienc.* 87:29–49.
- Henderson E.C., Brelsford A. 2020. Genomic differentiation across the speciation continuum in three hummingbird species pairs. *BMC Evol. Biol.* 20:113.
- He Z., Li X., Yang M., Wang X., Zhong C., Duke N.C., Wu C.-I., Shi S. 2019. Speciation with gene flow via cycles of isolation and migration: insights from multiple mangrove taxa. *Natl. Sci. Rev.* 6:275–288.
- Irwin D.E., Milá B., Toews D.P.L., Brelsford A., Kenyon H.L., Porter A.N., Grossen C., Delmore K.E., Alcáide M., Irwin J.H. 2018. A comparison of genomic islands of differentiation across three young avian species pairs. *Mol. Ecol.* 27:4839–4855.
- Jarvis E.D., Mirarab S., Aberer A.J., Li B., Houde P., Li C., Ho S.Y.W., Faircloth B.C., Nabholz B., Howard J.T., Suh A., Weber C.C., da Fonseca R.R., Li J., Zhang F., Li H., Zhou L., Narula N., Liu L., Ganapathy G., Boussau B., Bayzid M.S., Zavidovych V., Subramanian S., Gabaldón T., Capella-Gutiérrez S., Huerta-Cepas J., Rekepalli B., Munch K., Schierup M., Lindow B., Warren W.C., Ray D., Green R.E., Bruford M.W., Zhan X., Dixon A., Li S., Li N., Huang Y., Derryberry E.P., Bertelsen M.F., Sheldon F.H., Brumfield R.T., Mello C.V., Lovell P.V., Wirthlin M., Schneider M.P.C., Prosdocimi F., Samaniego J.A., Vargas Velazquez A.M., Alfaro-Núñez A., Campos P.F., Petersen B., Sicheritz-Ponten T., Pas A., Bailey T., Scofield P., Bunce M., Lambert D.M., Zhou Q., Perelman P., Driskell A.C., Shapiro B., Xiong Z., Zeng Y., Liu S., Li Z., Liu

- B., Wu K., Xiao J., Yinqi X., Zheng Q., Zhang Y., Yang H., Wang J., Smeds L., Rheindt F.E., Braun M., Fjeldsa J., Orlando L., Barker F.K., Jönsson K.A., Johnson W., Koepfli K.-P., O'Brien S., Haussler D., Ryder O.A., Rahbek C., Willerslev E., Graves G.R., Glenn T.C., McCormack J., Burt D., Ellegren H., Alström P., Edwards S.V., Stamatakis A., Mindell D.P., Cracraft J., Braun E.L., Warnow T., Jun W., Gilbert M.T.P., Zhang G. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346:1320–1331.
- Keller I., Wagner C.E., Greuter L., Mwaiko S., Selz O.M., Sivasundar A., Wittwer S., Seehausen O. 2013. Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Mol. Ecol.* 22:2848–2863.
- Korunes K.L., Samuk K. 2021. pixy: unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol. Ecol. Resour.* 21:1359–1368.
- Latrubesse E.M. 2002. Evidence of Quaternary palaeohydrological changes in middle Amazônia: the Aripuanã-Roosevelt and Jiparaná fans". *Z. Geomorphol. Suppl.* 129:61–72.
- Li H., Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
- Lohmueller K.E., Albrechtsen A., Li Y., Kim S.Y., Korneliusen T., Vinckenbosch N., Tian G., Huerta-Sanchez E., Feder A.F., Grarup N., Jørgensen T., Jiang T., Witte D.R., Sandbæk A., Hellmann I., Lauritzen T., Hansen T., Pedersen O., Wang J., Nielsen R. 2011. Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. *PLoS Genet.* 7:e1002326.
- Luna L.W., Naka L.N., Thom G., Knowles L.L., Sawakuchi A.O., Aleixo A., Ribas C.C. 2023. Late Pleistocene landscape changes and habitat specialization as promoters of population genomic divergence in Amazonian floodplain birds. *Mol. Ecol.* 32:214–228.
- Maddison W.P. 1997. Gene trees in species trees. *Syst. Biol.* 46:523–536.
- Mallet J., Besansky N., Hahn M.W. 2016. How reticulated are species? *Bioessays* 38:140–149.
- 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 Biol. Evol.* 13:evab120.
- Marques D.A., Meier J.I., Seehausen O. 2019. A combinatorial view on speciation and adaptive radiation. *Trends Ecol. Evol.* 34:531–544.
- Martin S.H., Dasmahapatra K.K., Nadeau N.J., Salazar C., Walters J.R., Simpson F., Blaxter M., Manica A., Mallet J., Jiggins C.D. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res.* 23:1817–1828.
- Martin S.H., Davey J.W., Jiggins C.D. 2014. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* 32:244–257.
- Martin S.H., Davey J.W., Salazar C., Jiggins C.D. 2019. Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biol.* 17:e2006288.
- Martin S.H., Van Belleghem S.M. 2017. Exploring evolutionary relationships across the genome using topology weighting. *Genetics* 206:429–438.
- Mather K. 1938. Crossing-over. *Biol. Rev. Camb. Philos. Soc.* 13:252–292.
- Meier J.I., Marques D.A., Mwaiko S., Wagner C.E., Excoffier L., Seehausen O. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* 8:14363.
- Morales H.E., Pavlova A., Amos N., Major R., Kilian A., Greening C., Sunnucks P. 2018. Concordant divergence of mitogenomes and a mitonuclear gene cluster in bird lineages inhabiting different climates. *Nat. Ecol. Evol.* 2:1258–1267.
- Musher, Galante P.J., Thom G., Huntley J.W., Blair M.E. 2020. Shifting ecosystem connectivity during the Pleistocene drove diversification and gene-flow in a species complex of Neotropical birds (Tityridae: *Pachyramphus*). *J. Biogeogr.* 66:167.
- Musher L.J., Ferreira M., Auerbach A.L., Cracraft J. 2019. Why is Amazonia a "source" of biodiversity? Climate-mediated dispersal and synchronous speciation across the Andes in an avian group (Tityrinae). *Proc. Roy. Soc. B.* 286:20182343.
- Musher L.J., Giakoumis M., Albert J., Del-Rio G., Rego M., Thom G., Aleixo A., Ribas C.C., Brumfield R.T., Smith B.T., Cracraft J. 2022. River network rearrangements promote speciation in lowland Amazonian birds. *Sci. Adv.* 8:eabn1099.
- Myers E.A., Mulcahy D.G., Falk B., Johnson K., Carbi M., de Queiroz K. 2022. Interspecific gene flow and mitochondrial genome capture during the radiation of Jamaican *Anolis* Lizards (Squamata; Iguanidae). *Syst. Biol.* 71:501–511.
- Nachman M.W., Payseur B.A. 2012. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367:409–421.
- Nosil P. 2008. Speciation with gene flow could be common. *Mol. Ecol.* 17:2103–2106.
- Pavlidis P., Laurent S., Stephan W. 2010. msABC: a modification of Hudson's ms to facilitate multi-locus ABC analysis. *Mol. Ecol. Resour.* 10:723–727.
- Peñalba J.V., Deng Y., Fang Q., Joseph L., Moritz C., Cockburn A. 2020. Genome of an iconic Australian bird: high-quality assembly and linkage map of the superb fairy-wren (*Malurus cyaneus*). *Mol. Ecol. Resour.* 20:560–578.
- Portik D.M., Smith L.L., Bi K. 2016. An evaluation of transcriptome-based exon capture for frog phylogenomics across multiple scales of divergence (Class: Amphibia, Order: Anura). *Mol. Ecol. Resour.* 16:1069–1083.
- Provost K., Shue S.Y., Forcellati M., Smith B.T. 2022. The genomic landscapes of desert birds form over multiple time scales. *Mol. Biol. Evol.* 39:msac200.
- Pulido-Santacruz P., Aleixo A., Weir J.T. 2018. Morphologically cryptic Amazonian bird species pairs exhibit strong postzygotic reproductive isolation. *Proc. Biol. Sci.* 285:20172081.
- Pulido-Santacruz P., Aleixo A., Weir J.T. 2020. Genomic data reveal a protracted window of introgression during the diversification of a neotropical woodcreeper radiation. *Evolution* 74:842–858.
- Pupim F.N., Sawakuchi A.O., Almeida R.P., Ribas C.C., Kern A.K., Hartmann G.A., Chiessi C.M., Tamura L.N., Mineli T.D., Savian J.F., Grohmann C.H., Bertassoli D.J., Stern A.G., Cruz F.W., Cracraft J. 2019. Chronology of Terra Firme formation in Amazonian lowlands reveals a dynamic Quaternary landscape. *Quat. Sci. Rev.* 210:154–163.
- R Core Team. 2019. R: a language and environment for statistical computing.
- Ribas C.C., Aleixo A., Nogueira A.C.R., Miyaki C.Y., Cracraft J. 2012. A palaeobiogeographic model for biotic diversification within Amazonia over the past three million years. *Proc. Biol. Sci.* 279:681–689.
- Ribas C.C., Fritz S.C., Baker P.A. 2022. The challenges and potential of geogenomics for biogeography and conservation in Amazonia. *J. Biogeogr.* 49:1839–1847.
- Rognes T., Flouri T., Nichols B., Quince C., Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584.
- Rossetti D.F. 2014. The role of tectonics in the late Quaternary evolution of Brazil's Amazonian landscape. *Earth-Sci. Rev.* 139:362–389.
- Rossetti D.F., Vasconcelos D.L., Valeriano M.M., Bezerra F.H.R. 2021. Tectonics and drainage development in central Amazonia: the Juruá River. *Catena* 206:105560.
- Ruokolainen K., Moulatlet G.M., Zuquim G., Hoorn C., Tuomisto H. 2019. Geologically recent rearrangements in central Amazonian river network and their importance for the riverine barrier hypothesis. *Front. Biogeogr.* 11:e45046.
- Saether B.-E., Lande R., Engen S., Weimerskirch H., Lillegård M., Altwegg R., Becker P.H., Bregnballe T., Brommer J.E., McCleery R.H., Merilä J., Nyholm E., Rendell W., Robertson R.R., Tryjanowski P., Visser M.E. 2005. Generation time and temporal scaling of bird population dynamics. *Nature* 436:99–102.
- Sawakuchi A.O., Schultz E.D., Pupim F.N., Bertassoli D.J. Jr, Souza D.F., Cunha D.F., Mazoca C.E., Ferreira M.P., Grohmann C.H., Wahnfried I.D., Chiessi C.M., Cruz F.W., Almeida R.P., Ribas C.C. 2022. Rainfall and sea level drove the expansion of seasonally flooded habitats and associated bird populations across Amazonia. *Nat. Commun.* 13:4945.
- Schley R.J., Pennington R.T., Pérez-Escobar O.A., Helmstetter A.J., de la Estrella M., Larridon I., Sabino Kikuchi I.A.B., Barraclough T.G., Forest F., Klitgård B. 2020. Introgression across evolutionary scales suggests reticulation contributes to Amazonian tree diversity. *Mol. Ecol.* 29:4170–4185.

- Sick H. 1967. Rios e enchentes na Amazônia como obstáculo para a avifauna. Atas do simpósio sobre a biota amazônica 5:495–520.
- Silva S.M., Peterson A.T., Carneiro L., Burlamaqui T.C.T., Ribas C.C., Sousa-Neves T., Miranda L.S., Fernandes A.M., d'Horta F.M., Araújo-Silva L.E., Batista R., Bandeira C.H.M.M., Dantas S.M., Ferreira M., Martins D.M., Oliveira J., Rocha T.C., Sardelli C.H., Thom G., Rêgo P.S., Santos M.P., Sequeira F., Vallinoto M., Aleixo A. 2019. A dynamic continental moisture gradient drove Amazonian bird diversification. *Sci. Adv.* 5:eaat5752.
- Singhal S., Leffler E.M., Sannareddy K., Turner I., Venn O., Hooper D.M., Strand A.I., Li Q., Raney B., Balakrishnan C.N., Griffith S.C., McVean G., Przeworski M. 2015. Stable recombination hotspots in birds. *Science* 350:928–932.
- Smith B.T., Amei A., Klicka J. 2012. Evaluating the role of contracting and expanding rainforest in initiating cycles of speciation across the Isthmus of Panama. *Proc. Biol. Sci.* 279:3520–3526.
- Smith B.T., McCormack J.E., Cuervo A.M., Hickerson M.J., Aleixo A., Cadena C.D., Pérez-Emán J., Burney C.W., Xie X., Harvey M.G., Faircloth B.C., Glenn T.C., Derryberry E.P., Prejean J., Fields S., Brumfield R.T. 2014. The drivers of tropical speciation. *Nature* 515:406–409.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
- Thom G., Aleixo A. 2015. Cryptic speciation in the white-shouldered antshrike (*Thamnophilus aethiops*, Aves–Thamnophilidae): the tale of a transcontinental radiation across rivers in lowland *Mol. Phylogenet. Evol.* 82 Pt A:95–110.
- Thom G., Amaral F.R.D., Hickerson M.J., Aleixo A., Araujo-Silva L.E., Ribas C.C., Choueri E., Miyaki C.Y. 2018. Phenotypic and genetic structure support gene flow generating gene tree discordances in an Amazonian Floodplain endemic species. *Syst. Biol.* 67:700–718.
- Thom G., Moreira L.R., Batista R., Gehara M., Aleixo A., Smith B.T. 2021. Genomic architecture controls spatial structuring in Amazonian birds. *bioRxiv*. 2021.12.01.470789. doi:10.1101/2021.12.01.470789
- Thom G., Xue A.T., Sawakuchi A.O., Ribas C.C., Hickerson M.J., Aleixo A., Miyaki C. 2020. Quaternary climate changes as speciation drivers in the Amazon floodplains. *Sci. Adv.* 6:eaax4718.
- Tigano A., Khan R., Omer A.D., Weisz D., Dudchenko O., Multani A.S., Pathak S., Behringer R.R., Aiden E.L., Fisher H., MacManes M.D. 2022. Chromosome size affects sequence divergence between species through the interplay of recombination and selection. *Evolution* 76:782–798.
- Van Doren B.M., Campagna L., Helm B., Illera J.C., Lovette I.J., Liedvogel M. 2017. Correlated patterns of genetic diversity and differentiation across an avian family. *Mol. Ecol.* 26:3982–3997.
- Vrba E.S. 1992. Mammals as a key to evolutionary theory. *J. Mammal.* 73:1–28.
- Wang L., Liu S., Yang Y., Meng Z., Zhuang Z. 2022. Linked selection, differential introgression and recombination rate variation promote heterogeneous divergence in a pair of yellow croakers. *Mol. Ecol.* 31:5729–5744.
- Weeks B.C., Claramunt S., Cracraft J. 2016. Integrating systematics and biogeography to disentangle the roles of history and ecology in biotic assembly. *J. Biogeogr.* 43:1546–1559.
- Weir J.T., Faccio M.S., Pulido-Santacruz P., Barrera-Guzmán A.O., Aleixo A. 2015. Hybridization in headwater regions, and the role of rivers as drivers of speciation in Amazonian birds. *Evolution* 69:1823–1834.
- Wen D., Yu Y., Hahn M.W., Nakhleh L. 2016. Reticulate evolutionary history and extensive introgression in mosquito species revealed by phylogenetic network analysis. *Mol. Ecol.* 25:2361–2372.
- Wilkinson M.J., Marshall L.G., Lundberg J.G., Kreslavsky M.H. 2010. Megafan environments in northern South America and their impact on Amazon Neogene aquatic ecosystems. In: Hoorn C., Wesselingh F.P., editors. Amazonia, landscape and species evolution: a look into the past. 1st edition. pp.162–184.
- Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinf.* 19:153.
- Zhang G., Li C., Li Q., Li B., Larkin D.M., Lee C., Storz J.F., Antunes A., Greenwold M.J., Meredith R.W., Ödeen A., Cui J., Zhou Q., Xu L., Pan H., Wang Z., Jin L., Zhang P., Hu H., Yang W., Hu J., Xiao J., Yang Z., Liu Y., Xie Q., Yu H., Lian J., Wen P., Zhang F., Li H., Zeng Y., Xiong Z., Liu S., Zhou L., Huang Z., An N., Wang J., Zheng Q., Xiong Y., Wang G., Wang B., Wang J., Fan Y., da Fonseca R.R., Alfaro-Núñez A., Schubert M., Orlando L., Mourier T., Howard J.T., Ganapathy G., Pfenning A., Whitney O., Rivas M.V., Hara E., Smith J., Farré M., Narayan J., Slavov G., Romanov M.N., Borges R., Machado J.P., Khan I., Springer M.S., Gatesy J., Hoffmann F.G., Opazo J.C., Håstad O., Sawyer R.H., Kim H., Kim K.-W., Kim H.J., Cho S., Li N., Huang Y., Bruford M.W., Zhan X., Dixon A., Bertelsen M.F., Derryberry E., Warren W., Wilson R.K., Li S., Ray D.A., Green R.E., O'Brien S.J., Griffin D., Johnson W.E., Haussler D., Ryder O.A., Willerslev E., Graves G.R., Alström P., Fjeldsø J., Mindell D.P., Edwards S.V., Braun E.L., Rahbek C., Burt D.W., Houde P., Zhang Y., Yang H., Wang J., Jarvis E.D., Gilbert M.T.P., Wang J.; Avian Genome Consortium. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* 346:1311–1320.
- Zink R.M., Barrowclough G.F. 2008. Mitochondrial DNA under siege in avian phylogeography. *Mol. Ecol.* 17:2107–2121.