DR. CÁSSIA ALVES LIMA-REZENDE (Orcid ID : 0000-0001-7615-7635)

DR. GUSTAVO SEBASTIAN CABANNE (Orcid ID : 0000-0002-4812-2176)

Article type : Original Article

CÁSSIA ALVES LIMA-REZENDE (Orcid ID: 0000-0001-7615-7635)

Article type: Original Article

Running head: A phylogenomic analysis of savanna birds

A comparative phylogenomic analysis of birds reveals heterogeneous differentiation processes among Neotropical savannas

Cássia Alves Lima-Rezende[1*], Gustavo S. Cabanne[1], Amanda Vaz Rocha[2], Martin Carboni[1], Robert M. Zink[3] & Renato Caparroz[2]

[1] División de Ornitología, Museo Argentino de Ciencias Naturales "Bernardino Rivadavia" – CONICET, Buenos Aires, Argentina.

[2] Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, Brazil.

[3] School of Natural Resources, School of Biological Sciences, and Nebraska State Museum, University of Nebraska-Lincoln, Lincoln, Nebraska, United States.

* Corresponding author: calimarezende@gmail.com

Abstract

The main objective of this study is to evaluate biogeographic hypotheses of diversification and connection between isolated savannas north (Amazonian savannas) and south (Cerrado core) of the Amazon River. To achieve our goal, we employed genomic markers (genotyping-by-sequencing) to evaluate the genetic structure, population phylogenetic relationships, and historical range shifts of four Neotropical passerines with peri-Atlantic distributions: the Narrow-billed Woodcreeper (*Lepidocolaptes angustirostris*), the Plain-crested Elaenia (*Elaenia cristata*), the Grassland Sparrow (*Ammodramus humeralis*), and the White-banded Tanager (*Neothraupis fasciata*). The population genetic analyses indicated that landscape (e.g., geographic distance, landscape resistance, and percentage of tree cover) and climate metrics explained divergence among populations in most species, but without indicating a differential role between current and historical factors. Our results did not fully support the hypothesis that isolated populations at Amazonian savannas have been recently derived from the Cerrado core domain. Intraspecific phylogenies and gene flow analyses supported multiple routes of connection between the Cerrado and Amazonian savannas, rejecting the hypothesis that the Atlantic corridor explains the peri-Atlantic distribution. Our results reveal that the biogeographic history of the region is complex and cannot be explained by simple vicariant models.

Keywords: Amazonia, Cerrado, biogeography, landscape genomics, Pleistocene climatic fluctuations, savanna corridors.

Introduction

The origin of the remarkable Neotropical biodiversity is usually related to ecological processes associated with the large variety of habitats of the region (i.e., grasslands, forested savannas, tropical and temperate forests), as well as to different phenomena linked to geological events and glacial cycles (Haffer, 1969, 1985; Rull, 2011; Smith et al., 2014; Turchetto-Zollet, Pinheiro, Salgueiro, & Palma-Silva, 2013; Werneck, 2011). Among the variety of Neotropical biomes, rainforests are one of the best-studied in terms of evolution and diversity, whereas open biomes like grasslands and savannas are less well researched (Turchetto-Zolet et al., 2013; Werneck, 2011), despite covering more than 25% of the region (Eva et al., 2004). The scarcity of studies on the origin and diversification of Neotropical open biomes precludes a full understanding of Neotropical diversification. For instance, fundamental questions still remain unanswered, such as how current and historical landscape configuration impacted diversity, or is there a shared biogeographic history among currently co-distributed species?

Even though it is postulated that open-habitat organisms have high dispersal capacity (Sheard et al., 2020), landscape elements such as rivers, mountains, or regions with unsuitable habitat conditions (e.g., rainforest patches) are expected to modulate gene flow and influence population genetic structure (Cabanne et al., 2016; Moreira, Hernandez-Baños, & Smith, 2020; Vasconcellos, Ortiz, Weber, Cannatella, & Rodrigues, 2019). Furthermore, geomorphological and climatic conditions within biomes are not homogeneous, resulting in a mosaic with different levels of habitat integration, limiting gene flow between populations (McRae, 2006; Taylor, Fahrig, Henein, & Merriam, 1993). Accordingly, it is expected that population genetic differentiation is positively related to lack of connectivity (i.e., landscape resistance). However, the relative roles of current and or historical connectivity (e.g., at the Quaternary) as predictors for population genetic structure of open-habitat organisms is as yet unresolved.

In addition to the question of population connectivity, local adaptation related to climate and or vegetation types is also expected to be an important driver of population genetic variation of savanna organisms. Intraspecific variation due to natural selection in heterogenous environmental conditions has been often observed in plant and animal species (Fitzpatrick & Keller, 2014; Gugger, Fitz-Gibbon, Albarrán-Lara, Wright, & Sork, 2021; Morgan et al., 2020; Schoville et al., 2012). However, the role of local adaptation in driving population genetic differentiation of taxa with postulated high dispersal rates (Sheard et al., 2020) and that inhabit a relatively homogeneous environment, as are Neotropical savanna organisms, is unclear.

The South American savanna-like biomes are distributed in two major blocks located in the north (e.g., Llanos and Amazonian savannas) and south (e.g., Cerrado core domain) of the Amazon River (Figure 1A). Notwithstanding being geographically isolated, these two savanna blocks share several taxa, which denotes a biogeographic connection (Haffer, 1967; Mittermeier, Zyskowski, Stowe, & Lai, 2010; Ratter, Bridgewater, & Ribeiro, 2003; Silva, 1995; Simon & Proença, 2000; Wüster et al., 2005). One hypothesis to explain this biogeographic link is that these regions were continuous in the past, and their current disjunct range is a product of vicariant events promoted by the expansion of the Amazon Forest (Mittermeier et al., 2010; Silva, 1995). As vicariant events should affect all the biota equally, one prediction of this hypothesis is to find congruence among the area cladograms derived from phylogeographic analysis of the species inhabiting both savanna blocks (Brown & Lomolino, 1998; Ridley, 2004; Zink, Blackwell-Rago, & Ronquist, 2000). As an alternative, and based on studies with birds and plants, it has been proposed that the disjunct distribution have been originated by long-distance dispersal from the Cerrado core to the savannas north of the Amazon River (Buzatti et al., 2018; Norambuena & Van Els, 2020; Simon & Proença, 2000). A prediction of such a scenario would imply finding incongruent single-taxon area cladograms for co-distributed organisms (Brown & Lomolino, 1998; Frey, 1993; Ridley, 2004; Zink, 1996; Zink et al., 2000). In addition, if the isolated Amazonian savanna populations were derived from the Cerrado core domain, it is predicted that the new peripheral population will be phylogenetically embedded within its ancestral population (i.e., an apomorphic Amazonian population and a paraphyletic Cerrado, Figure 1A; see further explanation in Frey (1993)). Because the new population should have been founded from a small sample of individuals (founder effect), one would expect lower genetic diversity and a younger age for the new population than in the parental Cerrado population (Merilä, Bjorklund, & Baker, 1997).

Three hypothetical historical connection routes between the Cerrado core domain and the Amazonian savanna enclaves have been proposed (reviewed by Norambuena & Van Els, 2020; Silva & Bates, 2002; Figure 1A): I) West Amazonia corridors (Ribeiro, Werneck, & Machado, 2016; Silva, 1995; Webb, 1978; Werneck et al., 2012), II) a Central Amazonia corridor (Haffer, 1967; Ledo & Colli, 2017; Ribeiro et al., 2016), and III) an Atlantic coast corridor (Silva, 1995; Werneck et al., 2012). Studies with different organisms and methods have reported mixed support for each of these connection routes (Bates, Tello, & Silva, 2003; Lima-Rezende et al., 2019; Norambuena & Van Els, 2020; Ribeiro et al., 2016; Rocha et al., 2020; Silva, 1995), and thus it is not clear whether these routes represent mutually exclusive biogeographic histories or whether each species responded idiosyncratically to the same

biogeographic history. Silva (1995) described distinct distribution patterns of South American savanna birds and proposed that each of them was a result of a historical connection between the Cerrado and the other northern savannas. For instance, species with geographic ranges centered in one or more regions located south and east of the Amazon Forest, with isolated populations in one or more of the savannas located along the Atlantic coast such as Amapá and/or Marajó Island (from now on "peri-Atlantic distribution", Figure 1B), should be connected by the Atlantic coast corridor. Up to now, there is no specific test of these hypotheses, and phylogeographic analyses of species with peri-Atlantic distributions would allow such testing. If the peri-Atlantic distribution pattern is linked to a biogeographic connection through the region, we expected to find the highest gene flow rates through the Atlantic coast and a close phylogenetic relationship between the northern Cerrado and the isolated savannas at northern Amazonia (Figure 1A).

Thus, the general objective of this study is to evaluate biogeographic hypotheses of diversification and connection between Neotropical savannas, with emphasis on bird species from the Cerrado core domain and isolated savanna enclaves in the northern Amazon Forest. To achieve this objective, we conducted a comparative phylogenomic study of four passerines with a peri-Atlantic distribution (Figure 1B), namely: the Narrow-billed Woodcreeper (*Lepidocolaptes angustirostris* (Vieillot, 1818), Furnariidae), the Plain-crested Elaenia (*Elaenia cristata* Pelzeln, 1868, Tyrannidae), the Grassland Sparrow (*Ammodramus humeralis* (Bosc, 1792), Passerellidae) and the White-banded Tanager (*Neothraupis fasciata* (Lichtenstein, 1823), Thraupidae). We collected reduced representation genomic markers (genotyping-by-sequencing; Elshire et al., 2011) to investigate population genetic structure and history, and we modeled range shifts across time, to answer the following questions: (1) how do current and historical landscape characteristics, as well as climate, drive population genetic structure of savanna bird taxa?, (2) are isolated Amazonian savannas (i.e., Marajó Island and savannas of Amapá) derived from the Cerrado core domain? and 3) do species with a peri-Atlantic distribution support a historical Atlantic coast corridor between Neotropical savannas?

Methods

Taxon sampling

We studied four passerine species with peri-Atlantic distribution (Figure 1B): *Lepidocolaptes angustirostris*, which occurs in a variety of woodlands (deciduous forest, gallery forest, and others) and savannas, as well as agricultural and urban areas (Marantz, Aleixo, Bevier, & Patten, 2020; Ridgely & Tudor, 2009); *Elaenia cristata*, which also occurs in a variety of open

vegetation types but always in association with trees and bushes (Cerrado *sensu stricto*, wooded savannas, white-sand forests and adjacent forest edges; Herzog et al., 2016; Hosner, 2020); *Ammodramus humeralis*, which exclusively occurs in grasslands and savannas (Birdlife International, 2018; Ridgely & Tudor, 2009); and *Neothraupis fasciata*, which only occurs in Cerrado *sensu stricto* (Hilty & de Juana, 2017; Ridgely & Tudor, 2009).

For each taxon, we sampled a total of 17 to 36 individuals from five to ten localities located north (hereafter called Amazonian savannas) and south of the Amazon River (hereafter called Cerrado) (Figure 1; Supporting Information Table S1). Amazonian savannas encompass the continent and the Marajó Island located at the mouth of the Amazon River. We neither sampled putative migrant populations of *E. cristata* nor of *A. humeralis* (Hosner, 2020; Jaramillo, 2020).

Genomic data collection and assembly

Total DNA was isolated from blood or muscle samples using the PureLink Genomic DNA Kit. Around 1.8 μg (0.4 – 2.5 μg) of genomic DNA of each sample was used to prepare genomic libraries following the protocol genotyping-by-sequencing (GBS) developed by Elshire et al. (2011). Briefly, genomic DNA was digested with the *PstI* restriction enzyme, and then digested DNA fragments of each sample were tagged with unique barcodes. Tagged fragments were pooled, cleaned, and the libraries were amplified by polymerase chain reaction. The amplified libraries were cleaned, and their quality evaluated on a capillary sizing system. Libraries were sequenced in an Illumina HiSeq 2000 (100 bp fragments, single-end). All steps from library preparation to sequencing and library demultiplexing were conducted at the Cornell Institute of Genomic Diversity (Ithaca, NY, USA).

We obtained an average of 2.5 million raw reads per individual, and we used IPYRAD 0.7.30 (Eaton, 2014) to assemble (*de novo*) the demultiplexed GBS data and to export full GBS loci and single-nucleotide polymorphisms (SNPs). Main IPYRAD parameters were set as follows: maximum low-quality base calls (Q < 20) in a read = 5, minimum Phred quality score = 33, minimum depth for statistical base calling = 6, clustering threshold for de novo assembly = 0.9, minimum length of reads after adapter trim = 35 bp, maximum number of alleles per site = 2, minimum number of samples coverage = all samples, maximum number of heterozygous individuals per locus = 8, and maximum proportion of heterozygous sites per locus = 0.5. The final number of loci for each species ranged from 6,127 to 25,402, and the number of single unlinked biallelic SNPs per locus from 4,542 to 18,564 (Table 1). All SNP datasets have less than 1% missing data. All analyses were performed using datasets with unlinked biallelic

SNPs, except for STRUCTURE analyses, which also included SNPs with more than two alleles, and for G-PHOCS which used full-length loci (Table 1).

Analysis of population genetic structure

We calculated pairwise and global genetic differentiation between localities via F_{ST} in ARLEQUIN 3.5.2.2 (Excoffier & Lischer, 2010). To estimate pairwise F_{ST} values we used a genetic distance matrix based on the mean number of pairwise differences. To calculate global F_{ST} among sampling locations we performed an analysis of molecular variance (AMOVA) based on a locus-by-locus approach. We used a total of 10,000 non-parametric permutations to evaluate significance and F_{ST} confidence intervals.

We also used principal component analysis (PCA) and STRUCTURE 2.3.4 (Pritchard, Stephens, & Donnelly, 2000) to explore population genetic structure. We conducted the PCA with the *snpgdsPCA* function from "SNPRelate" package (Zheng et al., 2012) in R 3.6.2 (R Development Core Team, 2019). A missing rate threshold was not implemented because the datasets have low-levels of missing data. For running STRUCTURE we set a burn-in of 50,000, run length of 500,000, K-values ranging from 1 to the number of sampled sites for each species, and ten iterations for each K. Because Evanno's delta K method and the estimated natural logarithm of the probability of the data given that K – Ln Pr(X|K) – do not always identify correctly the level of structure (Janes et al., 2017), we complemented these methods by inspecting bar plots and geographic distribution of clusters to identify the optimal number of clusters. Bar plots were drawn with "conStruct" R-package (Bradburd, 2019).

Population phylogenetic relationships and gene flow

We reconstructed intraspecific phylogenetic relationships among localities by using species tree approach with SNAPP 1.5.0 (Bryant, Bouckaert, Felsenstein, Rosenberg, & Roychoudhury, 2012) for BEAST 2.6.2 (Bouckaert et al., 2019). SNAPP implements a coalescent method that directly infers the root position and assumes no gene flow between populations; thus, gene heterogeneity is considered a result of incomplete lineage sorting or differences in coalescence times (Bryant et al., 2012). To reduce computation time, we only evaluated two randomly selected individuals per population (Supporting Information Table S1). We fixed both mutation rates (*u* and *v*) at 1, used a gamma prior on population size values (alpha = 10, beta = 100), and a birth rate for the Yule model equal to 0.00765. We performed two independent runs, sampling every 1000 iterations. We inspected effective sample sizes (ESS) and likelihood plots with TRACER 1.7.1 (Rambaut, Drummond, Xie, Baele, & Suchard, 2018), considering stable

analyses with ESS >200. We combined log files with LogCombiner 2.6.3, with a burn-in of 10%, and final analyses had at least 3 million MCMC iterations. Cloud tree diagrams were generated using DensiTree 2.2.7 (Bouckaert & Heled, 2014) and maximum clade credibility trees were obtained with TreeAnnotator 2.6.2.

We used two complementary approaches to investigate historical gene flow among populations. First, we estimated effective migration surfaces (EEMS) to identify regions where historical gene flow was potentially higher or lower-than-average according to a model of isolation by distance (Petkova, Novembre, & Stephens, 2016). EEMS is less sensitive to irregular sampling schemes than other methods like PCA (Petkova et al., 2016). We estimated the identity-by-state similarity index between pairs of individuals using "SNPRelate" R-package and then transformed it to dissimilarity (D_{GEN}) by subtracting each identity-by-state value from 1. We used the *eigen* function in "base" 3.6.2 R-package to test if D_{GEN} met assumptions of Euclidean distances. We used 1,000 nDemes, and tested preliminary different hyperparameter values (results not shown) to tune the variances of proposal distributions and obtain reasonable acceptance proportions (Petkova et al., 2016) The parameters used for final analyses are: mSeedsProposalS2 = 0.9, qSeedsProposalS2 and mEffctProposalS2 = 0.99, qEffctProposalS2 = 0.2, and mrateMuProposalS2 = 0.05 with other priors set as default. We ran three MCMC samplers with a burn-in of 1,000,000 and main chain of 10,000,000 MCMC. Results were plotted using *eems.plots* in the "rEEMSplots" R-package (Petkova et al., 2016).

We also estimated gene flow, total migration, population size parameters (theta), and divergence time using G-PhoCS 1.3 (Gronau, Hubisz, Gulko, Danko, & Siepel, 2011). We estimated the total migration by multiplying the number of new migrants per generation by the divergence time of the youngest lineage in the comparison. We grouped samples into sets according to the results of the PCA and Structure to have populations without strong intra-group genetic differentiation, which is an assumption of any coalescent analysis. We used population topologies from SNAPP as a guide for G-PhoCS analysis and set tau-theta-alpha to 1, mig-rate-alpha to 0.001, and find-finetunes to true. For each dataset, we conducted at least three independent runs with a minimum of 500,000 MCMC iterations (jointly analyzed), sampling every 100 iterations. Burn-in was variable according to each MCMC chain. See Supplementary Information for more details about G-PhoCS analysis.

Ecological niche models

To evaluate historical connection between regions and to obtain landscape resistance metrics (see next section), we constructed ecological niche models of each species for the current time

period and then projected them into three paleoclimate scenarios using MAXENT 3.4.4. (Phillips, Dudík, & Schapire, 2021). Occurrence records were obtained from Global Biodiversity Information Facility (http://www.gbif.org/) and filtered to exclude records that fall outside the species' habitat, duplicated records, or those within the same grid cell. We downloaded 19 current bioclimatic variables (resolution of 2.5 arc-min) from the WorldClim database (http://www.worldclim.org/). Niche models were run with 20 replicates of cross-validation, using 20% of the data for testing, 1,500 maximum iterations, jackknife procedure to measure variable importance. Final niche models were obtained with uncorrelated bioclimatic variables (|Pearson's r| < 0.75) that were selected according to their relative contribution to the model. We projected the models of the present period into paleoclimate scenarios of the mid-Holocene (about 6,000 years ago), Last Glacial Maximum (LGM; about 21,000 years ago), and Last Interglacial (LIG; about 120,000 – 140,000 years ago). We used *calc* function in the "raster" R-package to obtain for each species an averaged historical suitability surface among LIG, mid-Holocene and LGM periods. See Supplementary Information for more details about MAXENT analysis.

Effects of landscape and climate on genetic structure

We conducted multiple matrix regression with randomization analyses (MMRR; Prunier, Colyn, Legendre, Nimon, & Flamand, 2015; Wang, 2013) to investigate the association between genetic differentiation (D_{GEN}) and different predictors (geographic distance, current and historical landscape resistances, and the Amazon River). We used Euclidian geographic distances between individuals (D_{GEO}), estimated with *distm* function of the "geosphere" R-package to test an isolation by distance hypothesis. To test the impact of landscape configuration on genetic differentiation, we estimated landscape resistance matrices among individuals for the current time, and over the past 120,000 years ago (averaged historical suitability) using CIRCUITSCAPE 4.0 (McRae & Beier, 2007). Landscape resistance, which is reciprocal of conductance, describes the degree to which the landscape impedes movement among patches, therefore, the cost of dispersing across landscape elements (McRae, 2006; Taylor et al., 1993). We estimated the present (R_{CUR}) and the past resistance (R_{PAST}) using as conductance surfaces the present and averaged historical habitat suitability, respectively. The Amazon River is the largest river in the study area, and because it is within the Amazon Forest (a putative barrier to the studied organisms), we also evaluated its role on genetic differentiation to disentangle the effects of the forest from the river. For this, we used an indicator matrix "River" that coded whether comparisons were made within the same river

bank (River = 0) or between river banks (River = 1). When sampling was carried out in both Amazonian savannas at Continent and Marajó Island, we coded pairs of individuals from different Amazonian savannas as "1".

To make the coefficients of the predictors comparable, response and independent variables were standardized by subtracting the mean and dividing by the standard deviation. *MMRR* R-function (Wang, 2013) was run using 1,000 permutations to estimate significance. Because multicollinearity may lead to misinterpretation of the MMRR results, we used a variance partitioning procedure to decompose the overall model into unique and common variance components by a commonality analysis (Prunier et al., 2015). Structure coefficients obtained from the commonality analysis are less sensitive to collinearity than beta-weights (partial regression coefficients), and the squared structure coefficient (r^2_s) represents the actual direct contribution of a predictor to model fit (Prunier et al., 2015; Ziglari, 2017). In a commonality analysis, unique effects quantify the amount of variance explained uniquely by a given predictor, whereas common effects quantify the amount of variance in the dependent variable that can be jointly explained by two or more predictors (Prunier et al., 2015). Commonality analysis was run implementing 1,000 bootstrap replicates (without replacement) to estimate 95% coefficient intervals of the commonality coefficients using "yhat" R-package (https://CRAN.R-project.org/package=yhat). MMRR and commonality analysis were performed for each species separately.

Finally, to evaluate the role of landscape and climate predictors in shaping genomic variation we conducted a gradient forest analysis (Ellis, Smith, & Roland Pitcher, 2012; Fitzpatrick & Keller, 2014) using "gradientForest" R-package (Ellis et al., 2012). Gradient forest analysis consists of a machine learning approach based on random forest algorithm to evaluate and model the relationship between response variables (e.g., community or genetic data) and a set of independent factors (e.g., climate and landscape metrics). We used the SNP data as response variables, coding genotypes as 0, 1, and 2, and excluding loci with missing data (Gugger et al., 2021). The evaluated predictors consisted of uncorrelated (|Pearson's r| < 0.75) bioclimatic variables obtained from WorldClim (Bio) and landscape related metrics such as percentage of tree cover (Geospatial Information Authority of Japan, Chiba University and collaborating organization – version 1; https://globalmaps.github.io/ptc.html), maximum green vegetation fraction – MGVF (Broxton, Zeng, Scheftic, & Troch, 2014), elevation (http://www.worldclim.org/), and distance-based Moran's eigenvector map variables (MEMs). Because gradient forest cannot directly accommodate spatial effects (Fitzpatrick & Keller, 2014), we evaluated MEMs, which are metrics representing the decomposition of the spatial

geographic variation among the sampling sites in orthogonal eigenvectors (for further details see Dray, Legendre, & Peres-Neto, 2006). We estimated MEMs from the latitude and longitude data using the *pcnm* function of "vegan" R-package (Oksanen et al., 2020), and used half of the positive eigenvalues for the gradient forest models (Fitzpatrick & Keller, 2014). All spatial datasets were used at a resolution of 30 arcseconds. We ran the models using individual samples as data points following Gugger et al. (2021), set a random forest of 500 trees per SNP, a maximum number of splits equal to $\log_2(0.368N)/2$, N representing sample size, and a correlation threshold of 0.5 for the conditional permutation (Ellis et al., 2012). We followed Fitzpatrick & Keller (2014) to reduce variation in predicted allele composition by PCA. Then, to map genomic geographic variation as modeled by the predictors, we assigned the first three PCs to a red, green, and blue color palette, with resulting color similarity corresponding to the similarity of the expected genomic composition. See Supplementary Information for further details.

Results

Population genetic analysis

Characteristics of genomic datasets of each species used for the different analyses are in Table 1. Global F_{ST} values indicated a weak population genetic structure for all studied species (Global F_{ST} from 0.074 to 0.136; Table 1). Greater pairwise F_{ST} values were obtained between Amazonian savannas and Cerrado than among Cerrado localities (Supporting Information Tables S2 – S5).

Principal components (PC) 1 and 2 of the PCA explained from 5% to 10.8% of the genomic variation of each species (Figure 2). PC3 explained 7.62% of the variation for *L. angustirostris*, 7.27% for *E. cristata*, 6.37% for *A. humeralis*, and 4.21% for *N. fasciata*. For all studied species, and according to the first two PCs, the Amazonian savannas formed separated clusters from the Cerrado localities (Figure 2). For *L. angustirostris, E. cristata*, and *N. fasciata* the genetic clusters of Cerrado samples roughly corresponded with the sampled regions, whereas for *A. humeralis* the Cerrado samples grouped into a single cluster (Figure 2).

In general, STRUCTURE plots are concordant with the PCA results (Figure 2). Regarding Amazonian savanna populations, we observed a tendency for lower admixture levels with western Cerrado than with the other Cerrado localities (except for *A. humeralis*). Despite the observation that Evanno's delta K and the Ln Pr(X|K) methods indicated three clusters for *E. cristata* (K = 3), there is a cluster in this model that is fully sympatric with others, and thus we selected K = 2 as best descriptor of the population genetic structure of the species (Figure 2 and

Supporting Information Figure S10). A similar situation was observed in *N. fasciata*, for which the Evanno's delta K and the Ln Pr(X|K) methods selected K = 4 (Supporting Information Figure S8), whereas K = 3 would better describe the genomic structure of the species because it identifies clusters that are geographically coherent, despite existence of extensive admixture (Figure 2 and Supporting Information Figure S12). Finally, the Evanno's method suggested two genetic clusters in *A. humeralis*. Because all individuals were attributed to the same two genetic clusters (Figure 2) we interpreted this result as indicative of a single population (K = 1, Supporting Information Figure S8).

Intraspecific phylogenetic relationships and gene flow

The studied species presented two distinct phylogeographic patterns involving the Amazonian savannas and Cerrado populations (Figure 3A). For *L. angustirostris* and *E. cristata*, the Cerrado populations formed a paraphyletic group, with a basal split placed between the western Cerrado and the clade formed by Amazonian savannas and the remaining Cerrado populations. For *A. humeralis* and *N. fasciata*, Cerrado populations were placed as a well-supported clade sister to the Amazonian savannas clade; the relationships among Cerrado populations were not well resolved, as indicated by tree fuzziness (Figure 3A), and the Amazonian savannas from the continent and Marajó Island formed a clade. We found low support for the position of the Amazonian Savanna population of *L. angustirostris*, resulting in two alternative topologies: topology I) with north Cerrado as sister to Amazonian savannas (posterior probability PP = 0.60); and topology II) with central Cerrado as a sister lineage to Amazonian savannas (PP = 0.40). Among species the timing of the basal split varied considerably, but occurred during the Middle and Late Pleistocene, as estimated by G-PHOCS (Figure 3A; Supporting Information Table S7). Also, the Amazonian savanna populations have genetic diversities that are half to one order of magnitude lower than most Cerrado populations (but see *E. cristata*; Figure 3A).

The EEMS analyses described areas with lower-than-average migration rates surrounding the Amazonian savannas for all species, except *N. fasciata* (Figure 3B). These results also indicated an area with lower-than-average migration extending from the Amazonian savannas towards the Cerrado, separating the Amazonian savannas and west (*E. cristata*), as well as Amazonian savannas and central (*E. cristata* and *N. fasciata*) Cerrado populations. G-PHOCS analyses supported gene flow between the Amazonian savannas and west (all species), central (at least *A. humeralis*), and east Cerrado (at least *E. cristata*, *A. humeralis*, and *N. fasciata*) populations (Table 2; Supporting Information Table S8).

Ecological niche models

The final niche models performed better than a random model (Test AUC > 0.7; Figure 4; Supporting Information Table S6), indicating that selected variables provided useful information on the presence probability for each studied species (Supporting Information Figures S1-S7 and Table S6). Niche models indicated that climate variation during the last 120,000 years differently affected the studied species (Figure 4; Supporting Information Figures S3 to S7). For *L. angustirostris,* most of the range shifts from the LIG to the present time consisted of shrinkage of suitable areas north of the Amazon Forest. *Neothraupis fasciata* and *E. cristata* probably reached their maximum range during the LGM, whereas all historical predictions for *A. humeralis* indicated greater range extension than the current. Projections also indicated for all studied species range expansions into Amazonia over the last 120,000 years (Figure 4).

Landscape and climate effects on genomic differentiation

The MMRR models indicated that predictors related to landscape resistance (current and historical), geographic distance, and the Amazon River explained from 32.8% to 77.6% of the variation in genetic distance between individuals of the studied species (Table 3; Supporting Information Figure S13 and Table S9). We found positive structure coefficients, some negative beta-weights, and large common effects (common > unique), indicating that predictors are correlated (Table 3; Supporting Information Figure S13 and Table S9). Among species, geographic distance (D_{GEO}) and historical landscape resistance (R_{PAST}) are the significant predictors with the highest r^2_s coefficients, whereas the Amazon River is the predictor with the lowest r^2_s.

The gradient forest analyses indicated that the genomic data for each study species are associated with climate and landscape metrics, supporting the action of natural selection linked to these factors. Among species, the proportion of SNPs with association to climate and landscape metrics ranged from 9.4% to 19.4%, with mean R^2-values varying from 0.20 to 0.27 (Figure 5). Overall, the climate and landscape metrics that are most important in explaining the genomic variation varied among species. However, distance-based Moran's eigenvector map variables (MEMs) showed the highest association to the SNP data, with exception of *A. humeralis*. For the latter species, as well as for *N. fasciata*, elevation was among the metrics with the strongest relationship to the genomic variation. For a single species, one climate metric was among the most important predictors of genomic variation (Figure 5, *E. cristata*).

Maps of predicted genomic variation in response to climate and landscape metrics (Figure 5) revealed a gradual north-south change in allelic composition for all species but *N. fasciata*, for which a notable genomic turnover was observed between Amazonian savannas and Cerrado. Within the Cerrado core domain, we found a heterogeneous change surface in allelic composition related to climate and landscape for all species, with exception of *N. fasciata.* For *L. angustirostris,* the projected genomic composition associated to climate and landscape metrics showed a strikingly differentiated pattern at the Caatinga region (Figure 5).

Discussion

Evolutionary processes associated with Neotropical savannas have been understudied. We investigated the genetic structure, intraspecific phylogenetic relationships, and range shifts across time for four South American passerines to evaluate biogeographic hypotheses of diversification and connection routes between Neotropical savannas. Population genetic analyses revealed that landscape and climate factors explained divergence among populations in most studied species but without indicating a differential role between current and historical landscape configuration. The analyses also indicated an important role of isolation-by-distance in the evolution of these organisms. Our results did not fully support the hypothesis that isolated populations in Amazonian savannas have been recently derived from the Cerrado core domain (Mittermeier et al., 2010; Silva, 1995). Finally, intraspecific phylogenetic and gene flow analyses supported multiple routes of connection between the Cerrado and Amazonian savannas, rejecting the hypothesis that the Atlantic coast corridor explains the peri-Atlantic geographic distribution pattern.

What are the drivers of population genetic structure of savanna birds?

Our findings are consistent with the hypothesis that present and past landscape configuration, as well as climate, are drivers of the genetic structure of South American savanna taxa. First, we found a variable but significant association between genetic differentiation in the study species and present and historical landscape resistance estimated from habitat suitability (Table 3). We have only found a strong impact of historical landscape configuration on genetic divergence of *L. angustirostris*, whereas in the other species the impact of historical landscape configuration was weaker or non-significant. Second, we conducted gradient forest analyses and observed for all the studied species an association between allelic composition and spatial

and climate metrics (Figure 5), which further confirms the importance of landscape and climate factors in explaining patterns of genomic variation. Thus, our analyses reveal inconsistent phylogenomic patterns most likely resulting from species' idiosyncratic responses to landscape and climate.

For *A. humeralis*, the MMRR analysis failed to find significant predictors of pairwise population genetic differentiation. The model was significant but poorly fit and with non-significant individual predictors (Table 3), which we interpret as an effect of multicollinearity. Such lack of association between genetic differentiation and the predictors could be a consequence of lack of equilibrium between drift and gene flow (Hutchison & Templeton, 1999), perhaps due to recent changes in demography. The hypothesis of lack of equilibrium would also be supported by the results of the gradient forest analysis, which indicated that only 9% of the studied SNPs are associated with current conditions such as climate and vegetation cover. In contrast, the other species have a higher percentage of SNPs associated with landscape and climate predictors (Figure 5).

The overall weak impact of historical range shifts on genomic differentiation found in most of the studied species contrasts with previous studies of Cerrado organisms, which indicated a higher relative contribution of historical landscape changes (Diniz-Filho et al., 2016; Vasconcellos et al., 2019). The MMRR result of a weak impact of range shifts on genetic differentiation is compatible with a high impact of isolation by distance in two species (and in a third species as a borderline result, Table 3). Furthermore, we found in the gradient forest analyses strong SNPs-MEMs associations in most studied species, which also can be interpreted as indicators of isolation by distance (Fitzpatrick & Keller, 2014). Thus, our results suggest a strong role of isolation by distance in the evolution of population genetic structures of Cerrado birds.

The relative high importance of isolation by distance found here is not frequently reported for Cerrado bird species (Lima-Rezende et al., 2019; Luna et al., 2017; Savit & Bates, 2015; van Els et al., 2021), although *L. angustirostris* is one exception (Kopuchian et al., 2020; Rocha et al., 2020). These studies rejecting isolation by distance in Cerrado birds differ from our study in respect of the different geographic scale of the study area (e.g., Luna et al., 2017), and in the use of different genetic markers (e.g., Lima-Rezende et al., 2019; Luna et al., 2017; Savit & Bates, 2015; van Els et al., 2021). It is becoming commonly accepted that multilocus studies are more suitable than single locus studies to evaluate population genetic structure and evolutionary phenomena such as isolation by distance (Teske et al., 2018).

Rivers have been shown to be important to explain genetic divergence of Neotropical birds, particularly in the Amazon basin (e.g., Naka & Brumfield, 2018; Ribas, Aleixo, Nogueira, Miyaki, & Cracraft, 2012; Silva et al., 2019). However, assessing the impact of rivers on population genetic structure of open-habitat birds has not been as frequent as in forest birds (Capurucho et al., 2020, 2013; Kopuchian et al., 2020; Naka & Brumfield, 2018). Unexpectedly, the impact of the Amazon River on the population genetic structure of our study species was negligible compared to the other predictors, explaining a maximum of 5.8% of the total determination coefficient (Table 3). Despite that all the four studied species have a genetic split between the Amazonian savannas and the Cerrado core domain (Figure 2 and 3), regions isolated by the Amazon River, the MMRR analyses indicated that this genetic pattern would be best explained by effect of the landscape resistance metrics. Our result implies that large rivers are not a strong gene flow barrier for savanna birds, perhaps because these taxa have high dispersal capacity.

An alternative explanation for the observed weak impact of the Amazon River on genetic differentiation would be related to the fact that we only evaluated individuals sampled near the mouth of the Amazon River (i.e., Marajó Island and Macapá), which is a region containing numerous islands and narrow streams, and therefore it could be assumed a weaker gene flow barrier than wider upstream river portions (Jackson & Austin, 2013; Kopuchian et al., 2020).

Species can respond idiosyncratically to different landscape configurations, and species-specific attributes such as habitat specialization may explain heterogeneous phylogeographic patterns (Khimoun et al., 2016; Kierepka, Anderson, Swihart, & Rhodes, 2016). Among the studied species, *A. humeralis* and *N. fasciata* could be considered open-habitat specialists because they are restricted to Cerrado and other open vegetation areas (Bagno & Marinho-Filho, 2001), whereas *L. angustirostris* and *E. cristata* are forest semi-dependent species (Bagno & Marinho-Filho, 2001). The latter species could also be considered generalist species regarding the lack exclusivity in the use of open-habitats or of a unique forest type. Because specialists tend to be more sensitive to environmental and landscape changes that would either alter or reduce dispersion and gene flow (Borges, Ribeiro, Lopes, & Loyola, 2019; Khimoun et al., 2016; Kierepka et al., 2016), it would be expected: I) that open-habitat specialists will have higher intraspecific genetic differentiation than generalist species, and also II) a higher impact of current and historical landscape features (i.e., landscape resistance) on population genetic structure. Contrary to these expectations, we found that the two generalist species have the highest intraspecific genetic differentiation (Table 1), and that according to the MMRR

analyses the landscape resistance metrics are better predictors of genetic divergence in generalist (models $R^2 > 0.6$) than in specialist taxa (models $R^2 < 0.45$). Thus, these results suggest that the studied open-habitat specialists have higher dispersal capacity than the forest semi-dependent species, which is in agreement with the idea that open-habitat organisms have higher dispersal capacity and lower genetic differentiation than forest organisms (Bates et al., 2003).

A higher dispersal capacity in open-habitat specialists would facilitate gene flow and override dispersal constraints caused by landscape characteristics and geographical distance, which could explain the differences found among species (Table 3). However, according to a morphological metric that is an indicator of dispersal capacity (i.e., Hand Wing Index; Claramunt, Derryberry, Remsen, & Brumfield, 2012; Sheard et al., 2020), there are no clear differences in dispersal and flight capabilities among the studied species once they exhibit similar Hand Wing Index values (Sheard et al., 2020). Thus, an explanation of our results based on dispersal capacity, as evaluated by the Hand Wing Index, is not well supported.

Are isolated Amazonian savannas derived from the Cerrado core domain?

We have not found evidence that the Amazonian savannas and the Cerrado formed a continuous range in the past and that their current disjunct range is a product of vicariance that impacted the whole community. Our results did not confirm the vicariance prediction that organisms shared by these regions should have congruent area cladograms. In contrast, we have found two distinct intraspecific phylogeographic patterns (Figures 2 and 3). In *L. angustirostris* and *E. cristata*, Amazonian savanna samples are nested within the Cerrado core region, being the western Cerrado population sister to all the other samples, whereas in *A. humeralis* and *N. fasciata* the Amazonian savannas form a sister clade to the Cerrado core samples.

As an alternative for the vicariant hypothesis, the avifauna from the isolated Amazonian savannas could have been originated by multiple long-distance dispersal events from the Cerrado core. However, it is not clear if all the studied species have Amazonian savanna populations derived from the Cerrado core. A Cerrado origin was only supported for *L. angustirostris* and *E. cristata*, in which the Cerrado population is paraphyletic with respect to the Amazonia savannas. In the other two species, the phylogenetic pattern of sister clades between Amazonian savannas and the Cerrado populations does not allow identification of the geographic region that acted as a source of immigrants. However, results of the population genetic analyses are consistent with the prediction of the Cerrado as a source because of the

finding of lower genetic diversity in Amazonian savannas (Figure 3A). All studied species have Amazonian savanna populations with theta values (i.e., an indicator of genetic diversity) half to one order of magnitude lower than most Cerrado populations (but see *E. cristata*), supporting a founder effect in the Amazonian savannas. Thus, the possibility of a Cerrado population having been a source for the isolated Amazonian populations is plausible.

A lack of consensus in intraspecific tree topologies of organisms shared between the Cerrado and the Amazonian savannas is in accordance with a previous study (van Els et al., 2021), which supported multiple origins of the isolated Amazonian savannas as well as of the Cerrado populations. For some plant species, genomic and chloroplast DNA data suggest colonization of Amazonian savannas from the Cerrado about 149,000 years ago (Buzatti et al., 2018), whereas in other organisms divergence between the savannas at north and south of the Amazon River predates the Last Interglacial (Buzatti et al., 2018; Resende-Moreira et al., 2019). Thus, we suggest that the biogeographic and phylogeographic history of these regions is complex and that the connection between the Amazonian savannas and the Cerrado might have occurred through processes such as range expansions followed by fragmentation, as well as by long-distance dispersal.

Does the peri-Atlantic biogeographic pattern support the Atlantic coast corridor?

The genomic data and distribution models rejected the hypothesis that species with a peri-Atlantic distribution were historically connected through the Atlantic coast corridor. First, the highest historical gene flow in the sample of four species with a peri-Atlantic distribution did not occur through the Atlantic coast axis. For each species, total migration occurred along multiple axes, and not only through the Atlantic coast route (Figure 3B, Table 2). Although in many species there was evidence of gene flow through the Atlantic coast route (e.g., *E. cristata* and *N. fasciata*), there was considerable gene flow through other routes too. Interestingly, a central Cerrado route appeared as the less likely for most species (Figure 3). Second, one of the predictions of an Atlantic coast gene flow route (i.e., a close phylogenetic relationship between northeastern Cerrado and the isolated savannas at northern Amazonia) was only confirmed in two species (i.e., *L. angustirostris* and *E. cristata*), whereas in the other species the isolated savannas are sister to all the Cerrado, not supporting any specific connection route. These mixed phylogenetic patterns indicate that the connection between regions has been complex, and that the peri-Atlantic distribution does not indicate a specific biogeographic history. Third, historical projections indicated range expansions towards distinct areas of the Amazonia, not

only in the eastern Amazonia region where the Atlantic coast corridor might have occurred (Figure 4).

Even though our research was not oriented to evaluate all the potential corridors between Amazonian savannas and the Cerrado, our results also support the West route between these regions. We have found in all studied species a high total migration between the west Cerrado and the Amazonian savannas populations, which supports a likely biogeographic connection through this route (Figure 1). However, our genetic analysis did not allow us to discriminate between the two geographic routes involving a West connection, namely through the base of the Andes (Silva, 1995; Webb, 1978; Werneck et al., 2012) and/or through the Madeira River basin (Ribeiro et al., 2016). Likewise, the niche models did not differentiate between these two alternatives (Figure 4; Supplementary Material Figure S3 – S7). An exception is *N. fasciata*, for which the LGM projection denoted a corridor through the Madeira River basin with higher climatic suitability than through the Andean region. These results are in agreement with a study of soil carbon isotopes indicating that savanna vegetation expanded in the Madeira River basin during the early to middle Holocene (Freitas et al., 2001), and with pollen records denoting increased dryness in the region during the LGM (Van Der Hammen & Hooghiemstra, 2000), supporting the existence of suitable areas for species of savannas and other types of open areas.

Permits

Fieldworks and sample collection were conducted with the permission of the Ethics Committee for Animal Use of the University of Brasília (UnBDoc no. 75111/2013) and the Brazilian Institute for Biodiversity Conservation — ICMBio (SISBIO no. 27682-1).

Acknowledgements

We are grateful to all graduate and undergraduate students from the Laboratório de Genética e Biodiversidade da Universidade de Brasília for help during the fieldworks. We are thankful to the directors and managers of the Ecological Reserves, National Parks, and farms visited during fieldwork and extended thanks to all the staff for their assistance. We thank museum curators for providing tissue samples and Miguel A. Marini and Gabriela D. Correa for help with the acquisition of the Ecological Station of Águas Emendadas samples. The authors acknowledge funding from CNPq/FNDCT/FAPs/MEC/CAPES/PRO-CENTRO-OESTE (No 031/2010 -

Proc. 564036/2010-2), PPBio - Rede Cerrado (CNPq No 35/2012 - Proc. 457444/2012-6), CNPq Edital Universal (Proc.445025/2014-0), CNPq (#234515/2014-7), as well as from the Fondo para la Investigación Científica y Tecnológica of Argentina PICT 2018 2689 and PUE 0098 (MACN). CAL-R acknowledges the support from the Brazilian Research Council – CNPq (#203013/2019-0), the Consejo Nacional de Investigaciones Científicas y Tecnológicas from Argentina – CONICET, and the Fundação de Amparo à Pesquisa do Distrito Federal – FAPDF. CAL-R is also grateful to the Bell Museum and the Department of Ecology, Evolution and Behavior at University of Minnesota for offering the resources for laboratory activities. AVR received scholarships from CAPES and CNPq. We also give special thanks to Leonardo Campagna (Cornell Lab of Ornithology, USA) for helping us implement the PCA and G-PHOCS analyses. Most of the bioinformatic analyses of this study have been conducted at the facilities of the Museo Argentino de Ciencias Naturales Bernardino Rivadavia, Buenos Aires, Argentina.

Author contribution

This study was conceived and designed by CAL-R, AVR, GSC, RC, and RMZ. Samples were collected by CAL-R, AVR, and RC. CAL-R and AVR conducted laboratory work. CAL-R, GSC, MC, and RC performed statistical analyses. CAL-R drafted the paper with input from all authors.

Data accessibility

Full-length loci, SNP datasets, occurrence points, and outputs from MAXENT are available through Dryad https://doi.org/10.5061/dryad.w9ghx3frh.

References

Bagno, M. A., & Marinho-Filho, J. (2001). A avifauna do Distrito Federal: Uso de ambientes abertos e florestais e ameaças. In J. F. Ribeiro, C. E. L. da Fonseca, & J. C. Sousa-Silva (Eds.), *Cerrado: Caracterização e recuperação de Matas de Galeria.* (pp. 495–516). Planaltina: Embrapa Cerrados. Retrieved from https://www.embrapa.br/cerrados/busca-de-publicacoes/-/publicacao/556930/cerrado-caracterizacao-e-recuperacao-de-matas-de-galeria

Bates, J. M., Tello, J. G., & Silva, J. M. C. (2003). Initial assessment of genetic diversity in ten

bird species of South American Cerrado. *Studies on Neotropical Fauna and Environment*, *38*(2), 87–94.

Birdlife International. (2018). *Ammodramus humeralis*, Grassland Sparrow. The IUCN Red List of Threatened Species 2018: e.T22721147A132007785. Retrieved June 3, 2020, from http://dx.doi.org/10.2305/IUCN.UK.2018- 2.RLTS.T22721147A132007785.en

Borges, F. J. A., Ribeiro, B. R., Lopes, L. E., & Loyola, R. (2019). Bird vulnerability to climate and land use changes in the Brazilian Cerrado. *Biological Conservation*, *236*(November 2018), 347–355. doi: 10.1016/j.biocon.2019.05.055

Bouckaert, R., & Heled, J. (2014). *DensiTree 2: Seeing trees through the forest* (pp. 1–11). pp. 1–11. bioRxiv. doi: https://doi.org/10.1101/012401

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., … Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *15*(4), 1–28. doi: 10.1371/journal.pcbi.1006650

Bradburd, G. (2019). *conStruct: Models Spatially Continuous and Discrete Population Genetic Structure. R package version 1.0.4.* Retrieved from https://cran.r-project.org/package=conStruct

Brown, J. H., & Lomolino, M. V. (1998). *Biogeography* (2nd ed.). Sunderland, Massachusetts: Sinauer Associates.

Broxton, P. D., Zeng, X., Scheftic, W., & Troch, P. A. (2014). A MODIS-based global 1-km maximum green vegetation fraction dataset. *Journal of Applied Meteorology and Climatology*, *53*(8), 1996–2004. doi: 10.1175/JAMC-D-13-0356.1

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & Roychoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, *29*(8), 1917–1932. doi: 10.1093/molbev/mss086

Buzatti, R. S. de O., Pfeilsticker, T. R., Magalhães, R. F. de, Bueno, M. L., Lemos-Filho, J. P., & Lovato, M. B. (2018). Genetic and historical colonization analyses of an endemic savanna tree, *Qualea grandiflora*, reveal ancient connections between Amazonian Savannas and Cerrado. *Frontiers in Plant Science*, *9*(July), 1–16. doi:

10.3389/fpls.2018.00981

Cabanne, G. S., Calderón, L., Trujillo Arias, N., Flores, P., Pessoa, R., d'Horta, F. M., & Miyaki, C. Y. (2016). Effects of Pleistocene climate changes on species ranges and evolutionary processes in the Neotropical Atlantic Forest. *Biological Journal of the Linnean Society*, *119*(4), 856–872. doi: 10.1111/bij.12844

Capurucho, J. M. G., Borges, S. H., Cornelius, C., Vicentini, A., M. B. Prata, E., Costa, F. M., … Ribas, C. C. (2020). Patterns and processes of diversification in Amazonian white sand ecosystems: Insights from birds and plants. In V. A. Rull & A. C. Carnaval (Eds.), *Neotropical diversification: Patterns and processes* (pp. 245–270). Cham, Switzerland: Springer Nature Switzerland AG.

Capurucho, J. M. G., Cornelius, C., Borges, S. H., Cohn-Haft, M., Aleixo, A., Metzger, J. P., & Ribas, C. C. (2013). Combining phylogeography and landscape genetics of *Xenopipo atronitens* (Aves: Pipridae), a white sand campina specialist, to understand Pleistocene landscape evolution in Amazonia. *Biological Journal of the Linnean Society*, *110*(1), 60–76. doi: 10.1111/bij.12102

Claramunt, S., Derryberry, E. P., Remsen, J. V., & Brumfield, R. T. (2012). High dispersal ability inhibits speciation in a continental radiation of passerine birds. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1733), 1567–1574. doi: 10.1098/rspb.2011.1922

Diniz-Filho, J. A. F., Barbosa, A. C. O. F., Collevatti, R. G., Chaves, L. J., Terribile, L. C., Lima-Ribeiro, M. S., & Telles, M. P. C. (2016). Spatial autocorrelation analysis and ecological niche modelling allows inference of range dynamics driving the population genetic structure of a Neotropical savanna tree. *Journal of Biogeography*, *43*(1), 167–177. doi: 10.1111/jbi.12622

Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, *196*(3–4), 483–493. doi: 10.1016/j.ecolmodel.2006.02.015

Eaton, D. A. R. (2014). PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, *30*(13), 1844–1849. doi: 10.1093/bioinformatics/btu121

Ellis, N., Smith, S. J., & Roland Pitcher, C. (2012). Gradient forests: Calculating importance gradients on physical predictors. *Ecology*, *93*(1), 156–168. doi: 10.1890/11-0252.1

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*(5), e19379. doi: 10.1371/journal.pone.0019379

Eva, H. D., Belward, A. S., De Miranda, E. E., Di Bella, C. M., Gond, V., Huber, O., … Fritz, S. (2004). A land cover map of South America. *Global Change Biology*, *10*(5), 731–744. doi: 10.1111/j.1529-8817.2003.00774.x

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567.

Fitzpatrick, M. C., & Keller, S. R. (2014). Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, *18*(1), 1–16. doi: 10.1111/ele.12376

Freitas, H. A., Pessenda, L. C. R., Aravena, R., Marques Gouveia, S. E., de Souza Ribeiro, A., & Boulet, R. (2001). Late Quaternary vegetation dynamics in the southern Amazon Basin inferred from carbon isotopes in soil organic matter. *Quaternary Research*, *55*(1), 39–46. doi: 10.1006/qres.2000.2192

Frey, K. (1993). Modes of peripheral isolate formation and speciation. *Systematic Biology*, *42*(3), 373–381. doi: 10.1093/sysbio/42.3.373

Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G., & Siepel, A. (2011). Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics*, *43*(10), 1031–1035. doi: 10.1038/ng.937

Gugger, P. F., Fitz-Gibbon, S. T., Albarrán-Lara, A., Wright, J. W., & Sork, V. L. (2021). Landscape genomics of *Quercus lobata* reveals genes involved in local climate adaptation at multiple spatial scales. *Molecular Ecology*, *30*(2), 406–423. doi: 10.1111/mec.15731

Haffer, J. (1967). Notas zoogeográficas sobre las avifaunas de las regiones no forestadas de Sudamérica noroccidental. *El Hornero*, *10*(4), 315–333.

Haffer, J. (1969). Speciation in Amazonian forest birds. *Science*, *165*, 131–137.

Haffer, J. (1985). Avian zoogeography of the Neotropical Lowlands. *Ornitological Monograhs*, (36), 113–146.

Herzog, S. K., Terrill, R., Jahn, A. E., Remsen Jr., J. Van, Maillard Z., O., García-Solíz, V. H., … Vidoz, J. Q. (2016). *Birds of Bolivia – Field Guide*. Santa Cruz de La Sierra, Bolivia: Associación Armonía.

Hilty, S., & de Juana, E. (2017). White-banded Tanager (*Neothraupis fasciata*). In J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie, & E. de Juana (Eds.), *Handbook of the Birds of the World Alive*. Barcelona: Lynx Edicions. Retrieved from https://www.hbw.com/node/61574

Hosner, P. (2020). Plain-crested Elaenia (*Elaenia cristata*), version 1.0. In J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie, & E. de Juana (Eds.), *Birds of the World*. Ithaca, NY: Cornell Lab of Ornithology. Retrieved from https://doi.org/10.2173/bow.plcela1.01

Hutchison, D. W., & Templeton, A. R. (1999). Correlation of pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, *53*(6), 1898–1914. doi: 10.1111/j.1558-5646.1999.tb04571.x

Jackson, N. D., & Austin, C. C. (2013). Testing the role of meander cutoff inpromoting gene flow across a riverine barrier in ground skinks (*Scincella lateralis*). *PLoS ONE*, *8*(5). doi: 10.1371/journal.pone.0062812

Janes, J. K., Miller, J. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., Cullingham, C. I., & Andrew, R. L. (2017). The K = 2 conundrum. *Molecular Ecology*, *26*(14), 3594–3602. doi: 10.1111/mec.14187

Jaramillo, A. (2020). Grassland Sparrow (*Ammodramus humeralis*), version 1.0. In J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie, & E. de Juana (Eds.), *Birds of the World*. Ithaca, NY: Cornell Lab of Ornithology. Retrieved from https://doi.org/10.2173/bow.graspa1.01

Khimoun, A., Eraud, C., Ollivier, A., Arnoux, E., Rocheteau, V., Bely, M., … Garnier, S. (2016). Habitat specialization predicts genetic response to fragmentation in tropical birds. *Molecular Ecology*, *25*(16), 3831–3844. doi: 10.1111/mec.13733

Kierepka, E. M., Anderson, S. J., Swihart, R. K., & Rhodes, O. E. (2016). Evaluating the influence of life-history characteristics on genetic structure: A comparison of small mammals inhabiting complex agricultural landscapes. *Ecology and Evolution*, *6*(17), 6376–6396. doi: 10.1002/ece3.2269

Kopuchian, C., Campagna, L., Lijtmaer, D. A., Cabanne, G. S., García, N. C., Lavinia, P. D., … Di Giacomo, A. S. (2020). A test of the riverine barrier hypothesis in the largest subtropical river basin in the Neotropics. In *Molecular Ecology* (Vol. 29). doi: 10.1111/mec.15384

Ledo, R. M. D., & Colli, G. R. (2017). The historical connections between the Amazon and the Atlantic Forest revisited. *Journal of Biogeography*, *44*(11), 2551–2563. doi: 10.1111/jbi.13049

Lima-Rezende, C. A., Rocha, A. V., Couto, A. F., de Souza Martins, É., Vasconcelos, V., & Caparroz, R. (2019). Late Pleistocene climatic changes promoted demographic expansion and population reconnection of a Neotropical savanna-adapted bird, *Neothraupis fasciata* (Aves: Thraupidae). *PLoS ONE*, *14*(3), 1–19. doi: 10.1371/journal.pone.0212876

Luna, L. W., Souza, T. O., Carneiro, L. S., de Girão e Silva, W. A., Schneider, H., Sampaio, I., … Rêgo, P. S. do. (2017). Molecular data and distribution dynamics indicate a recent and incomplete separation of manakins species of the genus *Antilophia* (Aves: Pipridae) in response to Holocene climate change. *Journal of Avian Biology*, *48*(8), 1177–1188. doi: 10.1111/jav.01378

Marantz, C. A., Aleixo, A., Bevier, L. R., & Patten, M. A. (2020). Narrow-billed Woodcreeper (*Lepidocolaptes angustirostris*), version 1.0. In J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie, & E. de Juana (Eds.), *Birds of the World*. Ithaca, NY: Cornell Lab of Ornithology. Retrieved from https://doi.org/10.2173/bow.nabwoo1.01

McRae, B. H. (2006). Isolation by resistance. *Evolution*, *60*(8), 1551. doi: 10.1554/05-321.1

McRae, B. H., & Beier, P. (2007). Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(50), 19885–19890. doi: 10.1073/pnas.0706568104

Merilä, J., Bjorklund, M., & Baker, A. J. (1997). Historical demography and present day population structure of the greenfinch, *Carduelis chloris* - An analysis of mtDNA control region sequence. *Evolution*, *51*, 946–956.

Mittermeier, J. C., Zyskowski, K., Stowe, E. S., & Lai, J. E. (2010). Avifauna of the Sipaliwini savanna (Suriname) with insights into its biogeographic affinities. *Bulletin of the Peabody Museum of Natural History*, *51*(1), 97–122. doi: 10.3374/014.051.0101

Moreira, L. R., Hernandez-Baños, B. E., & Smith, B. T. (2020). Spatial predictors of genomic and phenotypic variation differ in a lowland Middle American bird (*Icterus gularis*). *Molecular Ecology*, *29*(16), 3085–3102. doi: 10.1111/mec.15536

Morgan, K., Mboumba, J. F., Ntie, S., Mickala, P., Miller, C. A., Zhen, Y., … Anthony, N. M. (2020). Precipitation and vegetation shape patterns of genomic and craniometric variation in the central African rodent *Praomys misonnei*. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1930). doi: 10.1098/rspb.2020.0449rspb20200449

Naka, L. N., & Brumfield, R. T. (2018). The dual role of Amazonian rivers in the generation and maintenance of avian diversity. *Science Advances*, *4*(8). doi: 10.1126/sciadv.aar8575

Norambuena, H. V., & Van Els, P. (2020). A general scenario to evaluate evolution of grassland birds in the Neotropics. *Ibis*. doi: 10.1111/ibi.12905

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner, H. (2020). *vegan: Community Ecology Package. R package version 2.5-7.*

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., … Kassem, K. R. (2001). Terrestrial ecoregions of the world: A new map of life on earth. *BioScience*, *51*(11), 933–938.

Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, *48*(1), 94–100. doi: 10.1038/ng.3464

Phillips, S. J., Dudík, M., & Schapire, R. E. (2021). *[Internet] Maxent software for modeling species niches and distributions (Version 3.4.1).* Retrieved from http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on 2021-1-28

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

Prunier, J. G., Colyn, M., Legendre, X., Nimon, K. F., & Flamand, M. C. (2015). Multicollinearity in spatial genetics: Separating the wheat from the chaff using commonality analyses. *Molecular Ecology*, *24*(2), 263–283. doi: 10.1111/mec.13029

R Development Core Team. (2019). *R: A language and environment for statistical computing.* Viena, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. doi: 10.1093/sysbio/syy032

Ratter, J. a., Bridgewater, S., & Ribeiro, J. F. (2003). Analysis of the floristic composition of the Brazilian cerrado vegetation III: Comparison of the woody vegetation of 376 areas. *Edinburgh Journal of Botany*, *60*(01), 153–180. doi: 10.1017/S0960428603000064

Resende-Moreira, L. C., Knowles, L. L., Thomaz, A. T., Prado, J. R., Souto, A. P., Lemos-Filho, J. P., & Lovato, M. B. (2019). Evolving in isolation: Genetic tests reject recent connections of Amazonian savannas with the central Cerrado. *Journal of Biogeography*, *46*(1), 196–211. doi: 10.1111/jbi.13468

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. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1729), 681–689. doi: 10.1098/rspb.2011.1120

Ribeiro, V., Werneck, F. P., & Machado, R. B. (2016). Distribution dynamics of South American savanna birds in response to Quaternary climate change. *Austral Ecology*, *41*(7), 768–777.

Ridgely, R. S., & Tudor, G. (2009). *Field guide to the birds of South America: Passerines* (First Edit). Austin, Texas: University of Texas Press.

Ridley, M. (2004). *Evolution* (3rd ed.). Oxford: Blackwell Publishing.

Rocha, A. V., Cabanne, G. S., Aleixo, A., Silveira, L. F., Tubaro, P., & Caparroz, R. (2020). Pleistocene climatic oscillations associated with landscape heterogeneity of the South American dry diagonal explains the phylogeographic structure of the narrow-billed woodcreeper (*Lepidocolaptes angustirostris*, Dendrocolaptidae). *Journal of Avian Biology*, *51*(9), 1–13. doi: 10.1111/jav.02537

Rull, V. (2011). Neotropical biodiversity: Timing and potential drivers. *Trends in Ecology and Evolution*, *26*(10), 508–513. doi: 10.1016/j.tree.2011.05.011

Savit, A. Z., & Bates, J. M. (2015). Right around the Amazon: The origin of the circum-amazonian distribution in *Tangara cayana*. *Folia Zoologica*, *64*(3), 273–283. doi: 10.25225/fozo.v64.i3.a8.2015

Schoville, S. D., Bonin, A., François, O., Lobreaux, S., Melodelima, C., & Manel, S. (2012). Adaptive genetic variation on the landscape: Methods and cases. *Annual Review of Ecology, Evolution, and Systematics*, *43*(November), 23–43. doi: 10.1146/annurev-ecolsys-110411-160248

Sheard, C., Neate-Clegg, M. H. C., Alioravainen, N., Jones, S. E. I., Vincent, C., MacGregor, H. E. A., … Tobias, J. A. (2020). Ecological drivers of global gradients in avian dispersal inferred from wing morphology. *Nature Communications*, *11*(1). doi: 10.1038/s41467-020-16313-6

Silva, J. M. C. (1995). Biogeographic analysis of the South American Cerrado avifauna. *Steenstrupia*, *21*, 49–67.

Silva, J. M. C., & Bates, J. M. (2002). Biogeographic patterns and conservation in the South American Cerrado: A tropical savanna hotspot. *BioScience*, *52*(3), 225–233.

Silva, S. M., Townsend Peterson, A., Carneiro, L., Burlamaqui, T. C. T., Ribas, C. C., Sousa-Neves, T., … Aleixo, A. (2019). A dynamic continental moisture gradient drove Amazonian bird diversification. *Science Advances*, *5*(7). doi: 10.1126/sciadv.aat5752

Simon, M. F., & Proença, C. (2000). Phytogeographic patterns of *Mimosa* (Mimosoideae, Leguminosae) in the Cerrado biome of Brazil: An indicator genus of high-altitude centers of endemism? *Biological Conservation*, *96*(3), 279–296. doi: 10.1016/S0006-3207(00)00085-9

Smith, B. T., McCormack, J. E., Cuervo, A. M., Hickerson, M. J., Aleixo, A., Cadena, C. D., … Brumfield, R. T. (2014). The drivers of tropical speciation. *Nature*, *515*, 406–413.

Taylor, P. D., Fahrig, L., Henein, K., & Merriam, G. (1993). Connectivity is a vital element of landscape structure. *Oikos*, *68*(3), 571–573. Retrieved from http://www.jstor.org/stable/3544927

Teske, P. R., Golla, T. R., Sandoval-Castillo, J., Emami-Khoyi, A., Van Der Lingen, C. D., Von Der Heyden, S., … Beheregaray, L. B. (2018). Mitochondrial DNA is unsuitable to test for isolation by distance. *Scientific Reports*, *8*(1), 1–9. doi: 10.1038/s41598-018-25138-9

Turchetto-Zolet, A. C., Pinheiro, F., Salgueiro, F., & Palma-Silva, C. (2013). Phylogeographical patterns shed light on evolutionary process in South America.

Molecular Ecology, *22*, 1193–1213.

Van Der Hammen, T., & Hooghiemstra, H. (2000). Neogene and Quaternary history of vegetation, climate, and plant diversity in Amazonia. *Quaternary Science Reviews*, *19*(8), 725–742. doi: 10.1016/S0277-3791(99)00024-4

van Els, P., Zarza, E., Rocha Moreira, L., Gómez-Bahamón, V., Santana, A., Aleixo, A., … Berv, J. (2021). Recent divergence and lack of shared phylogeographic history characterize the diversification of neotropical savanna birds. *Journal of Biogeography*, (March 2020), 1–14. doi: 10.1111/jbi.14065

Vasconcellos, M. M., Ortiz, E. M., Weber, J. N., Cannatella, D. C., & Rodrigues, M. T. (2019). Isolation by instability: Historical climate change shapes population structure and genomic divergence of treefrogs in the Neotropical Cerrado savanna. *Molecular Ecology*, (28), 1748–1764. doi: 10.1111/mec.15045

Wang, I. J. (2013). Examining the full effects of landscape heterogeneity on spatial genetic variation: A multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution*, *67*(12), 3403–3411. doi: 10.1111/evo.12134

Webb, S. D. (1978). A history of savanna vertebrates in the New World. Part II: South America and the great interchange. *Annual Review of Ecology and Systematics*, *9*(1), 393–426. doi: 10.1146/annurev.es.09.110178.002141

Werneck, F. P. (2011). The diversification of eastern South American open vegetation biomes: Historical biogeography and perspectives. *Quaternary Science Reviews*, *30*, 1630–1648.

Werneck, F. P., Nogueira, C., Colli, G. R., Sites, J. W., Costa, G. C., Sites-Jr, J. W., & Costa, G. C. (2012). Climatic stability in the Brazilian Cerrado: Implications for biogeographical connections of South American savannas, species richness and conservation in a biodiversity hotspot. *Journal of Biogeography*, *39*(9), 1695–1706. doi: 10.1111/j.1365-2699.2012.02715.x

Wüster, W., Ferguson, J. E., Quijada-Mascareñas, J. A., Pook, C. E., Salomão, M. D. G., & Thorpe, R. S. (2005). Tracing an invasion: Landbridges, refugia, and the phylogeography of the Neotropical rattlesnake (Serpentes: Viperidae: *Crotalus durissus*). *Molecular Ecology*, *14*(4), 1095–1108. doi: 10.1111/j.1365-294X.2005.02471.x

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-

performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. doi: 10.1093/bioinformatics/bts606

Ziglari, L. (2017). Interpreting multiple regression results: β weights and structure coefficients. *General Linear Model Journal*, *43*(2), 13–22. doi: 10.31523/glmj.043002.002

Zink, R. M. (1996). Comparative phylogeography in North American birds. *Evolution*, *50*(1), 308–317. doi: 10.1111/j.1558-5646.1996.tb04494.x

Zink, R. M., Blackwell-Rago, R. C., & Ronquist, F. (2000). The shifting roles of dispersal and vicariance in biogeography. *Proceedings of the Royal Society B: Biological Sciences*, *267*(1442), 497–503. doi: 10.1098/rspb.2000.1028

Table 1. Characteristics of GBS dataset of four South American savanna birds. Sample size in number of individuals analyzed (N), number of loci in assembly, number of single-nucleotide polymorphisms (SNPs), and global F_{ST} among sampling localities are presented. Ninety-five posterior probability confidence intervals (95% CI) of F_{ST} are also given.

Species	N	Loci[a]	Unlinked SNPs[b]	Unlinked biallelic[c] (excluding SNP positions with missing data[d])	Global F_{ST} (95% CI)
Lepidocolaptes angustirostris	19	12,766	8,320	8,245 (7,483)	0.136 (0.131 – 0.142)
Elaenia cristata	17	25,402	18,735	18,564 (16,856)	0.132 (0.128 – 0.135)
Ammodramus humeralis	19	10,680	8,018	7,897 (7,885)	0.074 (0.069 – 0.080)
Neothraupis fasciata	36	6,127	4,590	4,542 (3754)	0.103 (0.098 – 0.107)

Letters refer to the dataset used in the different analysis: a) G-PHOCS; b) STRUCTURE; c) F_{ST}, PCA, SNAPP, EEMS, and MMRR; and d) gradient forest.

933 **Table 2.** Total migration between populations of South American savanna birds based on GBS genomic data. Gene flow is given as the total
934 migration (Mtot) between Cerrado and Amazonian savannas populations as obtained in G-PHOCS. Confidence intervals are posterior 95%. We
935 present two analyses for *Lepidocolaptes angustirostris* according to two population topologies (Top I and II).

Species		Migration band	Mtot (95% CI)		
			East route	Central route	West route
Lepidocolaptes angustirostris	Top I	Cerrado → Amazonian savanna	49 (0 – 7591)	145 (0 – 7831)	161 (0 – 13584)
		Amazonian savanna → Cerrado	2 (0 – 1005)	68 (0 – 23983)	16752 (0 – 327667)
	Top II	Cerrado → Amazonian savanna	10101 (0 – 23173)	13536 (6875 – 31641)	1249 (0 – 7121)
		Amazonian savanna → Cerrado	18478 (4422 – 67643)	600 (0 – 5324)	1 (0 – 993)
Elaenia cristata		Cerrado → Amazonian savannas	3034 (0 – 14829)	39 (0 – 51456)	7425 (0 – 99381)
		Amazonian savannas → Cerrado	24453 (0 – 220459)	202 (0 – 50859)	2009 (0 – 78190)
Ammodramus humeralis		Cerrado → Amazonian savanna Island	41 (0 – 356635)	57 (0 – 132386)	40 (0 – 126679)
		Amazonian savanna Island → Cerrado	1332 (0 – 1055664)	575 (0 – 667620)	422 (0 – 542472)
		Cerrado → Amazonian savanna Continent	36 (0 – 173496)	40 (0 – 88459)	29 (0 – 136019)
		Amazonian savanna Continent → Cerrado	2058 (0 – 1466654)	1791 (0 – 1339242)	1160 (0 – 921601)
Neothraupis fasciata		Cerrado → Amazonian savanna	690 (0 – 77949)	1 (0 – 5090)	1096 (0 – 45729)

Amazonian savanna → Cerrado	11462 (0 – 45078)	380 (0 – 16961)	6240 (0 – 22721)

936

Table 3. Multiple matrix regression with randomization and commonality analysis of genetic divergence between individuals of South American savanna birds based on a GBS dataset. Response variable was pairwise genetic differentiation (D_{GEN}) and predictors are pairwise geographic distance (D_{GEO}), current (R_{CUR}) and historical (R_{PAST}) landscape resistance, and the Amazon River bank (River). Beta-weights (β), the squared structure coefficient (r^2_s), and the unique, common, and total coefficients are given.

Species and full model fit	Predictors	*p-value*	β	r^2_s	Unique	Common	Total
Lepidocolaptes angustirostris	R_{CUR}	0.525	-0.167	0.194	0.001	0.117	0.118 (19.4%)
$R^2 = 0.608$; **p < 0.001**	R_{PAST}	**< 0.001**	0.834	0.762	0.039	0.424	0.463 (76.2%)
	D_{GEO}	0.065	0.214	0.796	0.008	0.476	0.484 (79.6%)
	River	0.103	-0.273	0.056	0.006	0.028	0.034 (5.6%)
Elaenia cristata	R_{CUR}	0.454	0.419	0.479	0.001	0.371	0.372 (47.9%)
$R^2 = 0.776$; **p < 0.001**	R_{PAST}	0.590	0.261	0.770	0.001	0.597	0.597 (76.9%)
	D_{GEO}	**< 0.001**	0.513	0.877	0.035	0.646	0.681 (87.8%)
	River	**0.018**	-0.490	0.058	0.010	0.035	0.045 (5.8%)
Ammodramus humeralis	R_{CUR}	0.951	-0.035	0.758	0.000	0.249	0.249 (75.9%)
$R^2 = 0.328$; **p < 0.001**	R_{PAST}	0.294	0.522	0.980	0.005	0.317	0.322 (98.2%)
	D_{GEO}	0.508	0.110	0.737	0.002	0.240	0.242 (73.8%)
	River	0.959	-0.013	0.317	0.000	0.104	0.104 (31.7%)
N. fasciata	R_{CUR}	**< 0.001**	-7.834	0.057	0.070	-0.045	0.025 (5.6%)

$R^2 = 0.447$; **$p < 0.001$**	R_{PAST}	**< 0.001**	3.145	0.144	0.069	-0.004	0.064 (14.3%)
	D_{GEO}	**< 0.001**	0.485	0.653	0.059	0.233	0.292 (65.3%)
	River	**< 0.001**	4.604	0.043	0.054	-0.035	0.019 (4.3%)

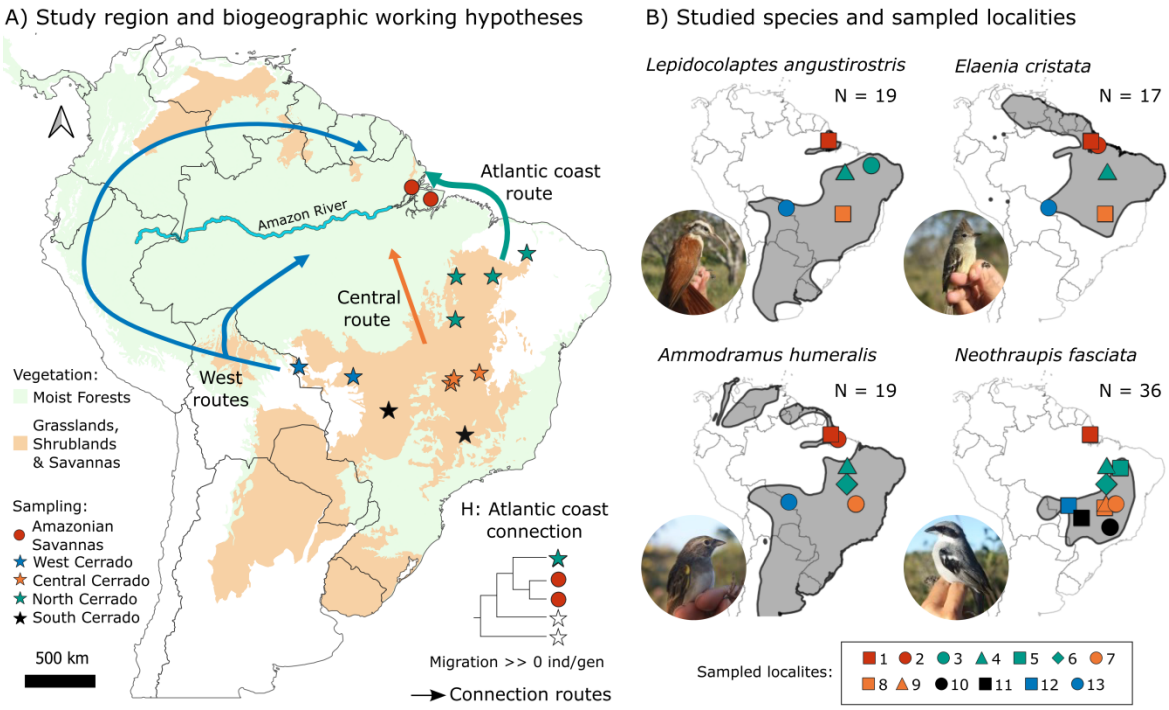
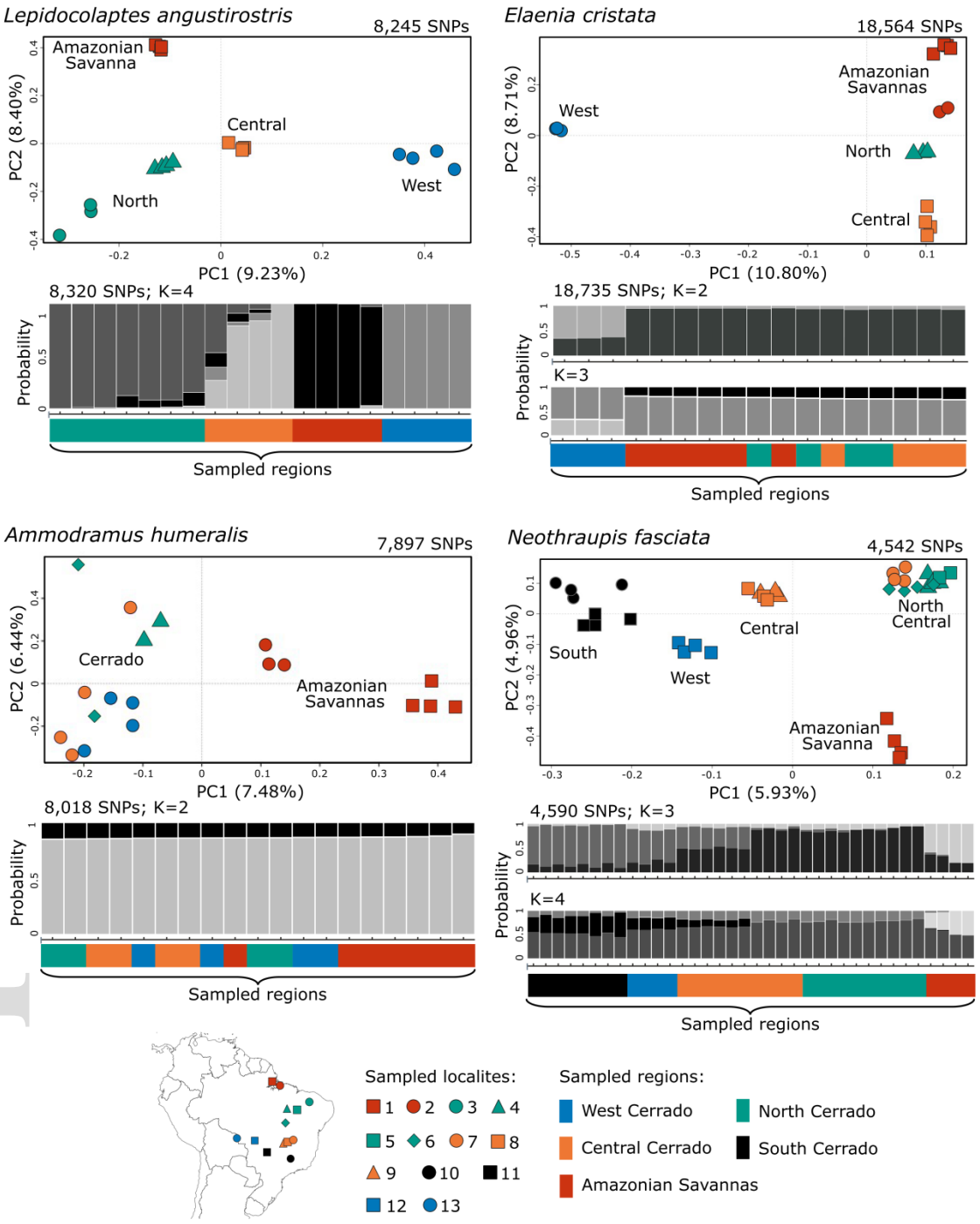


Figure 1. Study region, biogeographic working hypothesis and studied species. A) Main South American open biomes, moist forests, and sampling localities. We show schematic representation of the working hypotheses of connection between the Amazonian savannas and Cerrado: West Amazonia routes connecting the Amazonian savannas and west Cerrado; Central Amazonia route connecting the Amazonian savannas and central Cerrado; and the Atlantic coast route connecting the northern savannas to the northern Cerrado. We also show an example of the expected cladogram under the hypothesis of Atlantic coast connection, and under the case of the north Cerrado acting as a biogeographic source. Vegetation distribution is based on Olson et al. (2001). B) Studied species, tissue sample size (N), sampled localities and its geographic distributions (gray area). See Supplementary Material Table S1 for further details on samples and localities.



954

955 **Figure 2.** Analysis of population genetic structure of four South American savanna birds based
956 on GBS data. For each species we present a PCA and STRUCTURE plots. The proportion of the
957 variance explained by each principal component (PC) is given in parenthesis. Geographic origin
958 of samples is depicted in colors according to the map at the base of the figure. See
959 Supplementary Material Table S1 for further details on samples and localities.

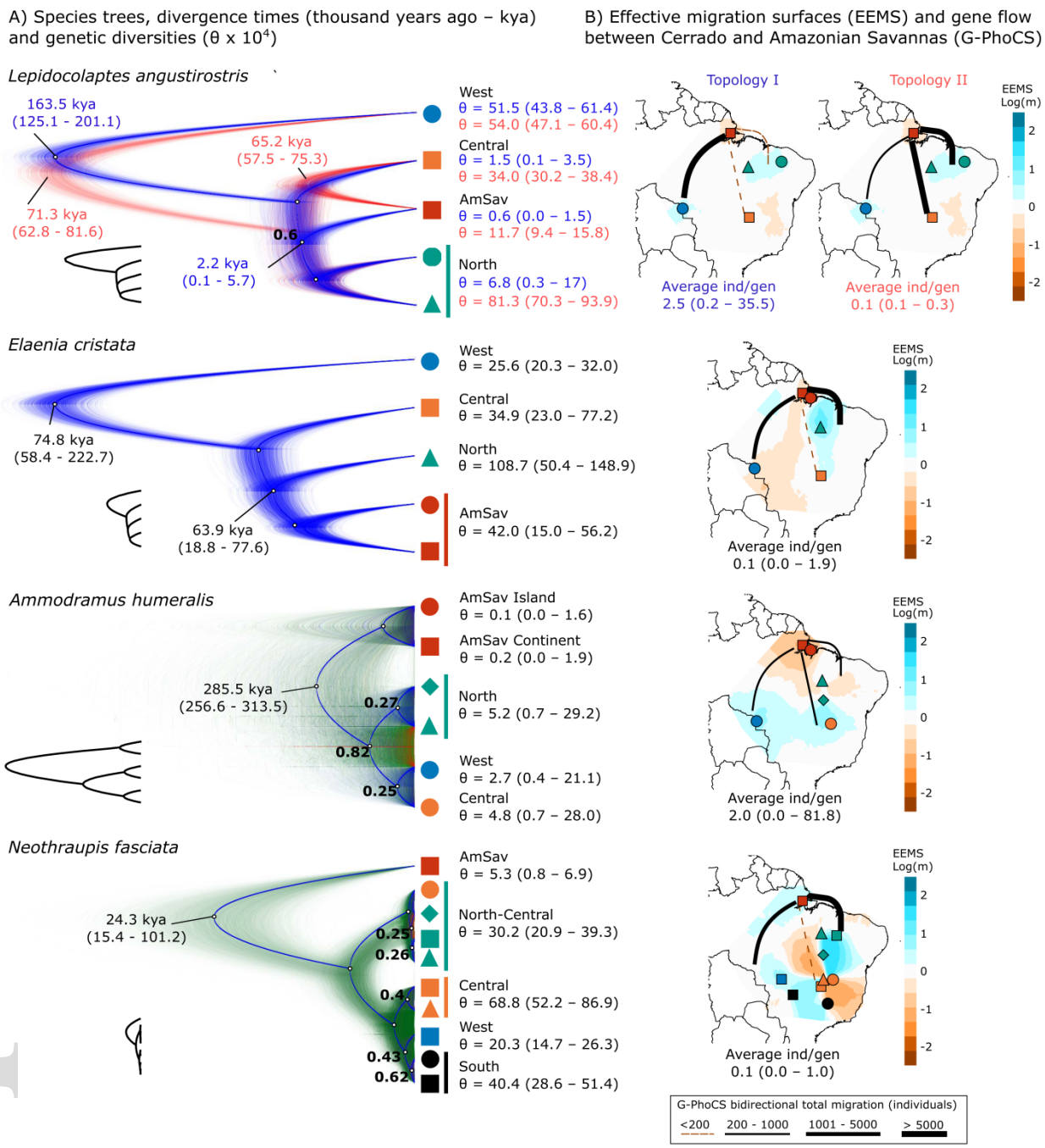
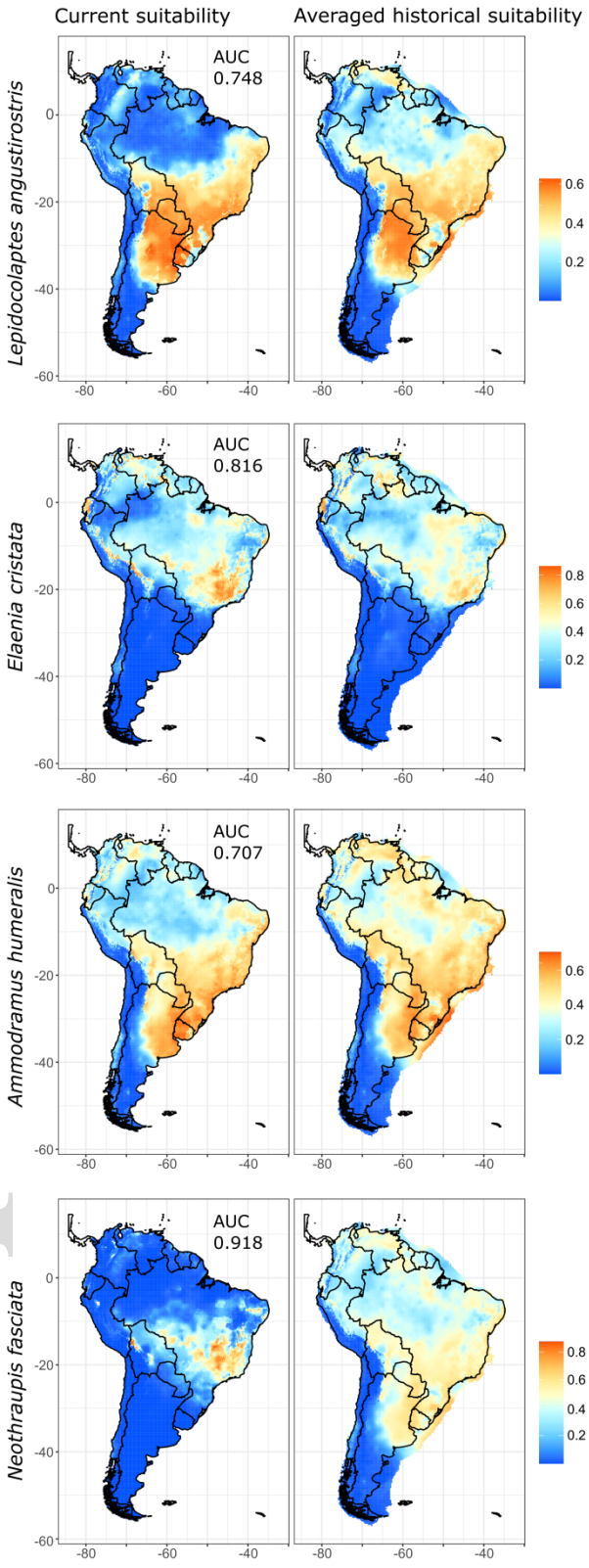


Figure 3. Population phylogenetic relationships, gene flow and divergence time between populations based on GBS markers of four South American savanna birds. A) Species trees obtained in SNAPP and divergence time (thousand years ago – kya) and genetic diversity (theta; θ x 10^4) obtained in G-PHOCS. All posterior probabilities at maximum credibility tree internal nodes are greater than 0.95, unless indicated in bold numbers. For simplicity, we only present divergence times for the basal split and the split between the Cerrado and Amazonian Savannas (AmSav). Cloud tree diagrams are out of scale but inset smaller consensus trees are in scale. B) Estimated effective migration surfaces (EEMS), average gene flow in individuals

per generation (ind/gen) among all G-PHOCS migration bands and average bidirectional total migration (individuals) between Cerrado populations and the studied Amazonas savannas. Population codes correspond to those presented in Figure 1. We present two analyses for *Lepidocolaptes angustirostris,* according to the two topologies obtained in A.



974

975 **Figure 4.** Species distribution models of four South American savanna birds obtained in
976 MAXENT. Historical predictions are based on the average of projections to the mid-Holocene (~
977 6 kya), Last Glacial Maximum (~ 21 kya), and Last Interglacial (~ 120 – 140 kya) periods.

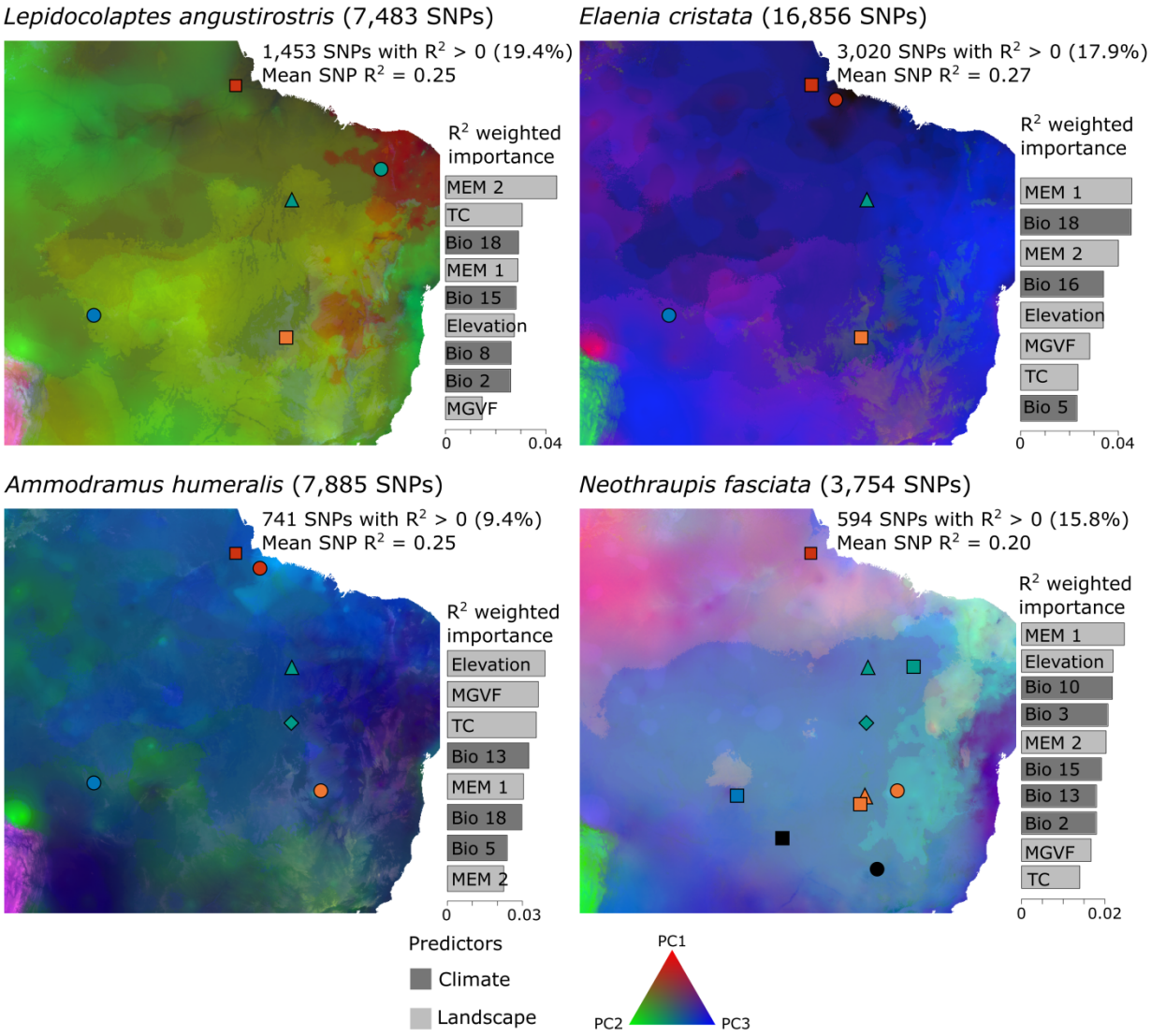


Figure 5. Gradient forest analyses and predicted genomic variation of South American savanna birds in response to climate and landscape predictors. Predictors of genomic variation are ranked according to their importance (R^2 weighted importance) in predicting allelic composition variation. Climate predictors: uncorrelated bioclimate variables (Bio) from WorldClim. Landscape predictors: elevation, distance-based Moran's eigenvector map variables (MEMs), maximum green vegetation factor (MGVF), and percentage of tree cover (TC). Population codes correspond to those presented in Figure 1. Mapped predicted genomic variation has been reduced by PCA to three axes (PC1, PC2, and PC3) and represented by red, green, and blue colors, respectively.