Biological Journal of the Linnean Society, 2021, **132**, 431–450. With 3 figures.

Molecular systematics and phylogeography of a widespread Neotropical avian lineage: evidence for cryptic speciation with protracted gene flow throughout the Late Quaternary

LEONARDO S. MIRANDA[1,2,*], BERNARDO O. PRESTES[1,†] and ALEXANDRE ALEIXO[1,3,*,‡]

[1]*Programa de Pós-graduação em Zoologia – Universidade Federal do Pará / Museu Paraense Emílio Goeldi, CEP 66075-110, Belém, PA, Brazil*
[2]*Coordenação de Ciências da Terra e Ecologia, Museu Paraense Emílio Goeldi, Caixa Postal 399, CEP 66040-170, Belém, PA, Brazil*
[3]*Coordenação de Zoologia, Museu Paraense Emílio Goeldi, Caixa Postal 399, CEP 66040-170, Belém, PA, Brazil*

Received 5 August 2020; revised 21 October 2020; accepted for publication 23 October 2020

Here we use an integrative approach, including coalescent-based methods, isolation–migration and species distribution models, to infer population structure, divergence times and diversification in the two species of the genus *Cymbilaimus* (Aves, Thamnophilidae). Our results support a recent and rapid diversification with both incomplete lineage sorting and gene flow shaping the evolutionary history of *Cymbilaimus*. The spatio-temporal pattern of cladogenesis suggests that *Cymbilaimus* originated in the north/western portion of *cis*-Andean South America and then diversified into the Brazilian Shield and Central America after consolidation of the modern Amazonian drainage and the Andean range. This evolutionary scenario is explained by cycles of range expansion and dispersal, followed by isolation, and recurrent gene flow, during the last 1.2 Myr. Our results agree with those recently reported for other closely related suboscine lineages, whereby the window of introgression between closely related taxa remains open for up to a few million years after their original split. In *Cymbilaimus*, introgression was recurrent between *C. lineatus* and *C. sanctaemariae*, even after they acquired vocal and ecological differentiation, supporting the claim that at least in Neotropical suboscines, full reproductive compatibility may take millions of years to evolve and cannot be interpreted as synonymous with a lack of speciation.

ADDITIONAL KEYWORDS: bamboo forests – biogeography – *Cymbilaimus* – gene flow – incomplete lineage sorting – phylogeography – population structure.

INTRODUCTION

Species delimitation strategies are contentious and often conflict with each other due to the continuous nature of species diversification (De Queiroz, 2007; Zachos, 2018; Collar, 2018). Different criteria tend to emphasize distinct aspects of the speciation process, focusing on measurable proxies such as gene flow, comparative levels of differentiation between phenotypic and genotypic characters, or the evolution of traits thought to be indicators of reproductive compatibility, such as song in birds (Sangster, 2018). Indeed, reproductive compatibility and gene flow are at the centre stage of the species debate, as recognized early on by the 'Evolutionary synthesis' (Mayr, 1942; Harvey *et al.*, 2019). While earlier interpretations have emphasized that the speciation process is only complete when cessation of gene flow and full reproductive

*Corresponding authors. E-mail: miralaba@gmail.com; alexandre.aleixo@helsinki.fi
†Current address: Programa de Pós-graduação em Ecologia – Instituto Nacional de Pesquisas da Amazônia, CEP 69060-000, Manaus, Amazonas, Brazil.
‡Current address: University of Helsinki, Finnish Museum of Natural History, PO Box 17, 00014, Helsinki, Finland.

isolation are attained (Mayr, 1982), recent genomic data suggest much more complex scenarios, where events of reticulation, hybridization and introgression often take place between non-sister lineages that have been diverging from each other for different amounts of time (Ottenburghs *et al.*, 2017; Ottenburghs, 2019).

This seems to be particularly true among suboscine passerines, a highly diverse lineage whose centre of diversification lies in the Neotropics. The advent of next-generation sequencing techniques has demonstrated that new species of suboscines can arise from hybridization (Barrera-Guzmán *et al.*, 2017), and that deeply split non-sister lineages may retain reproductive compatibility (Weir *et al.*, 2015; Pulido-Santacruz *et al.*, 2018) and introgress into each other's genomes (Pulido-Santacruz *et al.*, 2020) even after a few million years of independent evolution. These studies demonstrate that gene flow alone cannot be automatically interpreted as evidence of lack of speciation between any two introgressed taxa, as multiple factors, such as demographics, play important roles in the speciation process (Harvey *et al.*, 2019).

The suboscine genus *Cymbilaimus* (Thamnophilidae) includes two cryptic species that were split relatively recently based on minor morphological, but significant vocal, and ecological differences (Pierpont & Fitzpatrick, 1983). The bamboo antshrike (*Cymbilaimus sanctaemariae*) was originally described as a subspecies of the more broadly distributed fasciated antshrike (*Cymbilaimus lineatus*; Gyldenstolpe, 1941), whose range covers nearly the entire Amazon basin and lowland humid forests in trans-Andean South America and Central America up to Honduras (Zimmer & Isler, 2020a, b). It was later realized that the range of *C. sanctaemariae* was within that of another *C. lineatus* subspecies (*C. l. intermedius*) in south-western Amazonia, but with each taxon replacing each other in different habitats (i.e. bamboo-dominated vs. upland terra firme forests, respectively; Pierpont & Fitzpatrick, 1983; Aleixo & Guilherme, 2010). The sympatry between *C. sanctaemariae* and *C. l. intermedius*, added to their striking vocal differences, was regarded as strong evidence of reproductive isolation between them, supporting the split of the bamboo antshrike as a distinct biological species (Pierpont & Fitzpatrick, 1983), a treatment that has been followed to this day (Dickinson & Christidis, 2014; Gill *et al.*, 2020; Remsen *et al.*, 2020). However, so far, interspecific limits within the genus *Cymbilaimus* have not been investigated with genetic data, which prevents the assessment of whether *C. sanctaemariae* and *C. l. intermedius* are truly reproductively isolated taxa, despite their overlapping distributions. More importantly, the lack of relevant genetic data has prevented an understanding of how phenotypic and ecological differentiation evolved in *Cymbilaimus*, a taxon which provides a useful model to study how sympatric and closely related cryptic species arise and evolve to coexist.

A combination of independent but correlated approaches can be used to assess these processes, such as coalescent-based analyses (Drummond & Rambaut, 2007), and isolation–migration models (Nielsen & Wakeley, 2001; Hey & Nielsen, 2004), which account for the patterns of genetic variability and lineage structuring, while accommodating incomplete lineage sorting (ILS) and gene flow, respectively. Furthermore, reconstructing the history of differentiation and modern sympatry between the two *Cymbilaimus* species and how distinct patterns (i.e. physical barriers or climatic differences) or processes (i.e. adaptation, dispersal and vicariance) might have influenced it is also relevant. Therefore, additional approaches such as ancestral areas and environmental niche reconstructions are necessary.

Here, we present a comprehensive study on the evolutionary history, divergence time, and diversification patterns in the genus *Cymbilaimus* and its constituent species (*C. lineatus* and *C. sanctaemariae*) to address the following questions: (1) What are the evolutionary relationships among *Cymbilaimus* taxa? (2) What are the gene flow dynamics (if any) between *C. sanctaemariae* and *C. lineatus* and among the different subspecies of the fasciated antshrike? (3) What is the spatio-temporal scenario behind the differentiation of *C. sanctaemariae* as a distinct, yet sympatric species with *C. lineatus*? Our hypothesis is that *C. sanctaemariae* evolved as a separate species from *C. lineatus* following the establishment of bamboo-dominated forests in Amazonia, which prompted the genetic, vocal and morphological differentiation and eventual reproductive isolation between them. Alternatively, *C. sanctaemariae* simply represents a weakly differentiated population nested in *C. lineatus* that has recently adapted to explore structurally distinct bamboo forest patches in western Amazonia.

MATERIAL AND METHODS

Genetic data sampling and sequencing analyses

We sequenced 78 *C. lineatus* and seven *C. sanctaemariae* tissue samples collected from throughout both species' ranges, encompassing all recognized subspecies/taxa in the genus *Cymbilaimus* (i.e. *C. l. lineatus*, which occurs on the Guiana shield; *C. l. fasciatus*, from Central America; *C. l. intermedius*, from Venezuela, Colombia and Amazonian Brazil; and *C. sanctaemariae*; Zimmer & Isler, 2020a, b; Fig. 1A). All tissues sequenced had associated voucher specimens housed at different

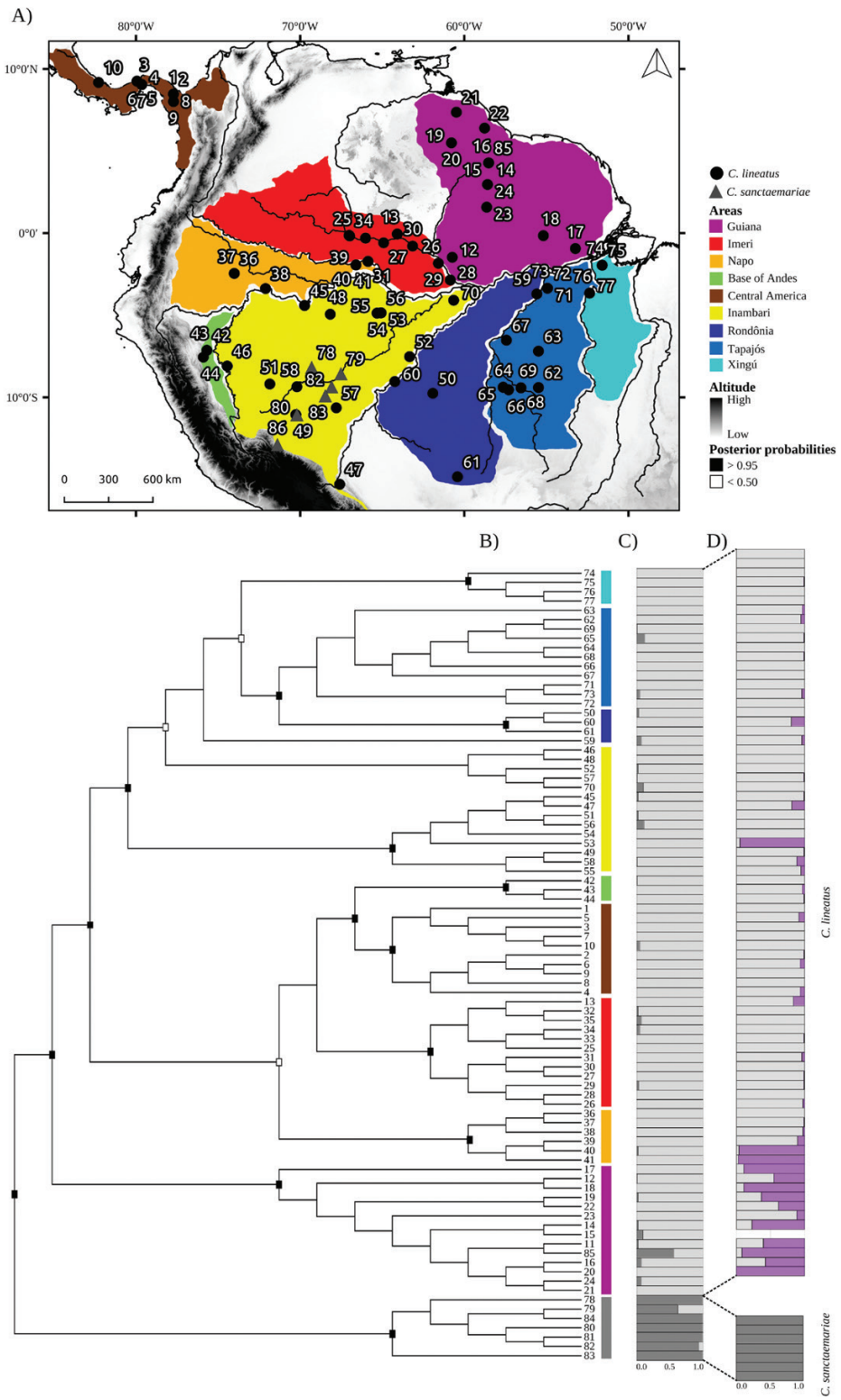

Figure 1. A, map of northern South America showing the sampling localities and their numbers corresponding to those in the phylogenetic tree. Black circles: *Cymbilaimus lineatus*, grey triangles: *C. sanctaemariae*. The gradient of grey colour in the map represents elevation (the darker the higher is the altitude) and the colours between rivers represent hypothesized Amazonian areas of endemism (legend with coloured squares). B, Bayesian inferred maximum clade credibility tree based on

ornithological collections (Supporting Information, Table S1).

Whole genomic DNA was extracted from tissue samples following procedures described by Sambrook *et al.* (1989). Two mitochondrial genes [cytochrome *b* (CYTB) and NADH dehydrogenase subunit 2 (ND2)], three nuclear introns [b-fibrinogen intron 5 (FIB5), glyceraldehyde 3-phosphate dehydrogenase intron 11 (G3PDH) and myoglobin intron 2 (MYO)], and one Z-linked locus [muscle skeletal receptor tyrosine kinase (MUSK)] were amplified via polymerase chain reaction (PCR).

Amplifications were performed in 12.5-µL volumes, containing 1.25 µL 10× reaction buffer, 1.5 mM $MgCl_2$, 0.4 mM each dNTP, 0.2 µM each primer (Supporting Information, Table S2), 1 unit of taq DNA polymerase (Invitrogen) and 10–25 ng of genomic DNA. Thermocycling conditions for mitochondrial DNA (mtDNA) started with an initial denaturation at 95 °C for 5 min, followed by 35 cycles consisting of 30 s denaturation at 95 °C, 30 s annealing at 50 °C (CYTB) or 55 °C (ND2) and 1 min extension at 72 °C, and a final extension of 5 min at 72 °C. Touchdown cycling parameters, differing only in annealing temperatures, as used for nuclear DNA (nDNA) were 50 °C for four cycles, 49 °C for four cycles and 48 °C for 35 cycles. PCR products were purified using 20% polyethylene glycol 8000 (PEG) before sequencing. For each molecular marker, sequences for both forward and reverse strands were performed in the ABI Prism BigDye Terminator Cycle sequencing protocol in an ABI PRISM 3130 XL Genetic Analyser (Applied Biosystems). All DNA sequences generated are available on GenBank (Table S1).

Sequences were visually inspected and aligned using ClustalW (Thompson *et al.*, 1994) as implemented in BioEdit 7.0.5.3 (Hall, 1999). Protein-coding mtDNA was translated into amino acids to verify the absence of stop codons or other anomalous residues. nDNA heterozygous nucleotide positions were identified by double peaks in the electropherograms and heterozygous indels positions were identified by a transition from a single to a series of double peaks in the electropherogram (indels were coded as a single mutational event in analyses that required phased haplotypes). Allelic phases were determined using PHASE 2.1.1 (Stephens *et al.*, 2001; Stephens & Scheet, 2005). Input files were produced with the online software SeqPHASE (Flot, 2010); we applied the algorithm twice with different random seeds, and checked for consistency of results across independent runs. We kept the complete dataset, selecting the haplotype pair with the highest probability among all possible pairs for each individual, as PHASE has been shown to generate a very low number of false positives (Garrick *et al.*, 2010). In addition, the nuclear loci were checked for recombination using the Phi test implemented in SPLITSTREE 4 (Bruen *et al.*, 2006; Huson & Bryant, 2006). DnaSP (Rozas *et al.*, 2017) was used to calculate the diversity statistics and the neutrality test indexes Tajima's *D* (Tajima, 1989) and Fu's *Fs* (Fu, 1997) among phylogroups.

mtDNA phylogenetic inference and nDNA population clustering

The phylogenetic analyses were conducted using BEAST (Drummond & Rambaut, 2007). The mitochondrial genes (CYTB and ND2) were concatenated in a single data matrix to produce an mtDNA Bayesian tree. The evolutionary models were selected with jModelTest 2.1.3 (Darriba *et al.*, 2012) using the Bayesian information criterion (BIC; Posada, 2008; Supporting Information, Table S3). We generated an mtDNA gene tree using an uncorrelated lognormal relaxed clock, a UPGMA topology as a starting tree and a coalescent constant size for the tree prior. Default priors were used except when a uniform prior distribution was involved; in this case we used a lognormal prior distribution, which is more appropriate to describe natural phenomena driven by the accumulation of many small changes (Koch 1966). A mutational rate of 2.1% sequence divergence per million years was applied (Weir & Schluter, 2008). We ran two independent Monte Carlo Markov chains (MCMCs) for 2×10^8 generations (sampling every 10^4 generations and discarding the first 30% as burn-in); and to check the analysis performance (effective sample size values > 200) and convergence of parameters between runs we used TRACER 1.5 (http://beast.bio.ed.ac.uk/Tracer). We combined the tree files from each independent run using LogCombiner, and the maximum clade credibility tree was computed with TreeAnnotator (part of the BEAST package). The consensus tree was visualized in FigTree 1.2.3 (http://tree.bio.ed.ac.uk/software/figtree/).

concatenated mtDNA with posterior probability support for nodes indicated by coded black–white squares. The numbers at tip labels and bar colours on the tree match those in the map. C and D, bar charts to the right denote BAPS results obtained based on nDNA only: C, best $k = 2$, *C. lineatus* (light grey) and *C. sanctaemariae* (dark grey); D, *C. lineatus* (best $k = 2$; purple for Guiana and light grey for the remaining individuals) and *C. sanctaemariae* (best $k = 1$; dark grey). The proportion of different colours in bar charts associated with each individual on the tree depicts the probability of membership to a particular nuclear genetic group.

We performed analyses using phased nDNA only, to assign individuals to populations using the Bayesian clustering program BAPS v.6.0 (Corander *et al.*, 2008). The data were analysed without using previous information on the origin of each individual and as separate partitions with linkage model for sequences. We first performed analyses assuming a mixture model to determine the most probable number of populations (k) using a range for the maximum number of k from 1 to 13 (we ran the analysis ten times to confirm the results). We further used the results from the mixture analysis to perform population assignment and admixture analysis using default settings.

Divergence time estimates and biogeographical reconstruction

For the species tree reconstruction, we used both nuclear and mtDNA sequences and the *BEAST algorithm (Heled & Drummond, 2010). We unlinked substitution model parameters for each gene (Supporting Information, Table S3); trees were also unlinked, except between mtDNA loci. We used the populations that were genetically differentiated in the mtDNA tree (Fig. 1B) and associated with areas bound by main physical barriers (such as large Amazonian rivers and the Andes) to designate the units in *BEAST. This analysis was set with the same parameter settings as the mtDNA Bayesian inference (see above), except that we used published mutational rates for different partitions [CYTB: 2.1% (Weir & Schluter, 2008); ND2: 2.5% (Smith & Klicka, 2010); nuclear introns: 0.27%; and z-linked marker: 0.29% (Ellegren, 2007)] and implemented a Yule process on the species tree prior.

We used the R package BioGeoBEARS (Matzke, 2013) to infer the biogeographical history of *Cymbilaimus*. We defined the following geographical areas for the BioGeoBEARS analysis according to major geological discontinuities and biogeographical provinces recognized within the ranges of *C. lineatus* and *C. sanctaemariae* (see Fig. 2A; Aleixo & Rossetti, 2007; Roddaz *et al.*, 2010; Silva *et al.*, 2019): (1) Northeast, equivalent to the Guiana shield (NE); (2) Northwest, north of the Amazon river and west of the Branco river (NW); (3) Southwest, south of the Amazon river and west of the Madeira river (SW); (4) Central America (CA); and (5) Southeast, encompassing the Brazilian Shield (SE). We allowed the ancestor to occur in a maximum of three areas. To accommodate the degree of uncertainty about the genealogies estimated by the *BEAST species tree, the likelihood and parameter values were estimated with 1000 genealogies sampled from those used to build the maximum clade credibility tree (Sukumaran, 2015). BioGeoBEARS implements models with the 'J' parameter, which takes into account the colonization of new areas located beyond the limits of the most likely ancestral distributions. However, use of the parameter 'J' has been questioned because it may overestimate dispersal in non-insular lineages (Ree & Sanmartín, 2018). That notwithstanding, founder event speciation may be also related to the organism's dispersal ability (Matzke, 2014), which is expected to be higher in *Cymbilaimus*, given its preference for upper strata in the vegetation, than among other related antbird lineages inhabiting the understorey (Burney & Brumfield, 2009). Therefore, because we regard dispersal over already established barriers as realistic biogeographical scenarios for *Cymbilaimus*, we tested and contrasted models with and without the 'J' parameter.

Genetic differentiation between phylogroups was examined using analysis of molecular variance (AMOVA) as implemented in ARLEQUIN v,3.5 (Excoffier & Lischer, 2010) with significance tested using 10 000 random permutations. The phylogroups arrangement was defined according to the *BEAST species tree (Fig. 2C). To assess how spatial variation in genetic diversity can be associated with landscape characteristics (geographical distances and/or environmental differences), pairwise genetic distance matrices [$F_{ST}/(1 - F_{ST})$] were correlated with geographical (ln km) and climatic distances (based on climatic variables that best explained the distribution models, see the Species Distribution Model results), respectively, using the mantel function of the *ecodist* R package (Goslee & Urban, 2007). We used the Mantel test, despite its debated limitations (Guillot & Russet, 2013), as an exploratory analysis and we only attest to significance at $P < 0.001$ (to take into account the multiple testing problem; Legendre & Legendre, 2012).

Gene flow and demographic analyses

As reconstruction methods used above lean on the assumption that no gene flow occurred between a priori defined phylogroups, we used the isolation–migration model (Nielsen & Wakeley, 2001; Hey & Nielsen, 2004) implemented in IMa2 (Hey, 2010) to evaluate the existence and degree of gene flow between pairs of phylogroups. Units of analysis were selected according to specific phylogroups with geographical correspondence (Figs 1B, 2C). For all analyses we used the complete sequences of both mtDNA and nDNA genes together; the HKY model (Hasegawa *et al.*, 1985) was applied to all markers, and inheritance scales were set as 0.25 (mtDNA genes), 0.75 (z-linked nDNA gene) and 1 (remaining nDNA genes). We also used previously published substitution rates for each main category of marker (see above) and a generation time of ~3 years (Bird *et al.*, 2020). Several runs were performed to establish the best priors for effective

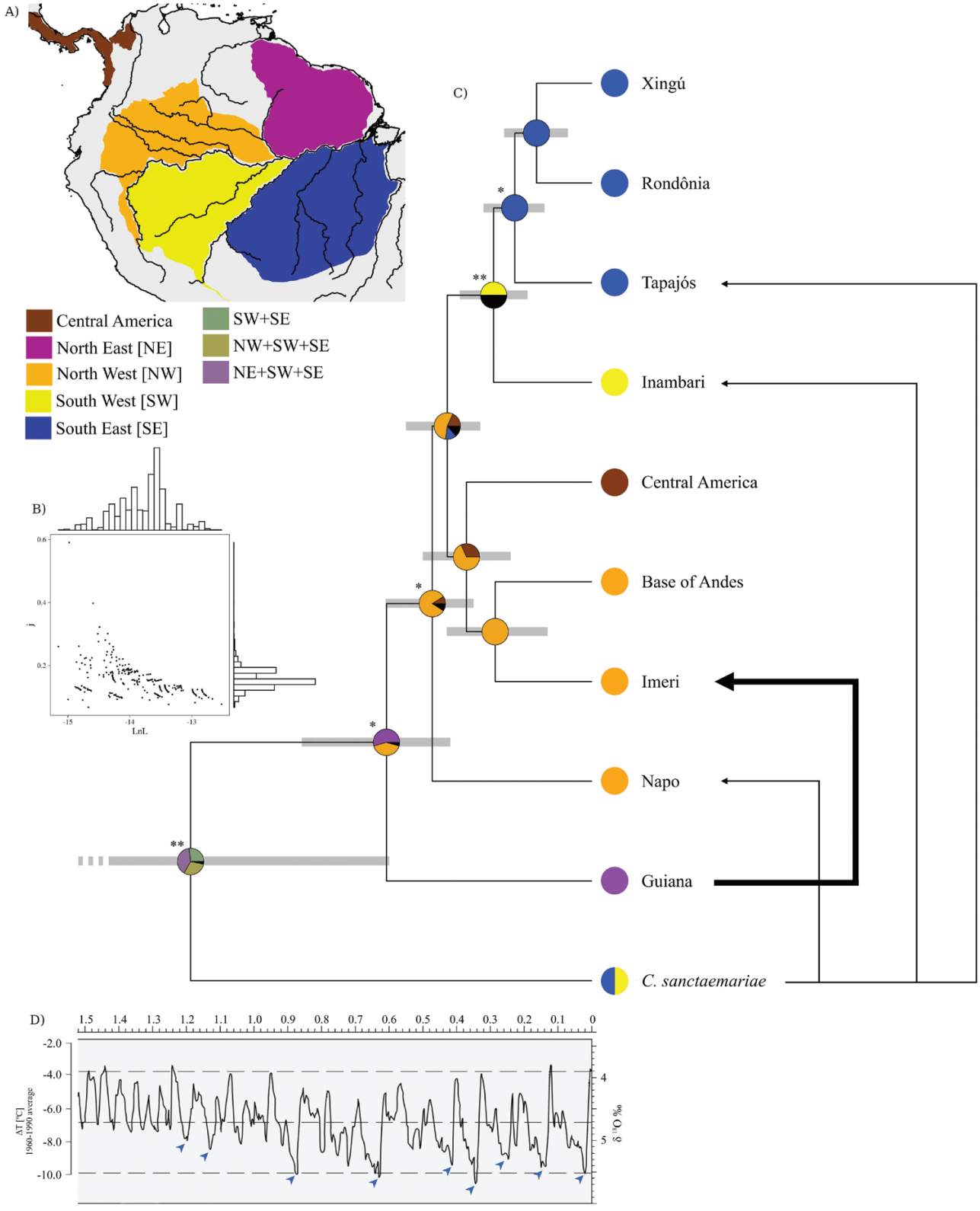

Figure 2. A, map of northern South America showing geographical areas defined for ancestral range reconstructions. B, scatter plot and marginal histogram of paired likelihood and founder event (*J*) parameter from ancestral range estimates of BioGeoBEARS under the DIVALIKE+J model performed over 1000 trees. C, time-calibrated phylogeny estimated with

population sizes, times of divergence and migration parameters. Three final runs were performed using 10^5 generations as burn-in, 10^5 trees sampled during 10^6 generations, and 20 chains (geometric model heating scheme, with 0.96 and 0.9 as the first and second heating parameter, respectively). To test whether a model of isolation without gene flow fitted the data better than a model with gene flow, we used the Nielsen & Wakeley (2001) approach and also the likelihood-ratio test as implemented in IMa2's L mode.

Finally, we used the Extended Bayesian Skyline Plot (EBSP; Heled & Drummond, 2008) method implemented in BEAST to analyse population size dynamics through time for all loci combined. The EBSP analyses were performed for each phylogroup. The best-fit substitution model for each marker, mutation rates, priors and the MCMC run strategy were the same as described above for other Bayesian phylogenetic analyses.

Species distribution modelling

Occurrence records were gathered from the public online database Global Biodiversity Information Facility (GBIF; available from http://gbif.org; accessed 5 July 2017), using the R package *rgbif* (Chamberlain *et al.*, 2015). To reduce spatial autocorrelation, we randomly removed occurrence records that were less than 10 km apart from each other, which resulted in a final data set of 566 records (Supporting Information, Table S4). We separated the occurrence data to generate models for each phylogroup, as we understand that these units can better discern the environmental factors that guide demography and underpin their potential distribution ranges (Marcer *et al.*, 2016). For each phylogroup, we developed present-day species distribution models (SDMs) with the *biomod2* (Thuiller *et al.*, 2009) package for R, using the maximum entropy (MAXENT; Phillips *et al.*, 2006) algorithm. The *C. lineatus* population BA was not included in the SDM due to its small range size (see Results below). The environmental variables were chosen from the 20 least-correlated topo-bioclimatic layers of the dataset available in WorldClim (Hijmans *et al.*, 2005; http://worldclim.org/), with a resolution of 30 arc-seconds (~1 km) and delimited by the following extent: xmin = –87, xmax = –43, ymin = –15, ymax = 15. After performing a pairwise Pearson correlation test to remove those that were highly correlated, we selected eight variables as our predictors: Mean Diurnal Range Temperature, Isothermality, Temperature Seasonality, Mean Temperature of Wettest Quarter, Precipitation Seasonality, Precipitation of Wettest Quarter, Precipitation of Warmest Quarter and Precipitation of Coldest Quarter. General conditions for analyses were: (1) three sets of background data randomly distributed for each phylogroup; (2) data sets were partitioned, i.e. 80% were used to calibrate and 20% were used for model evaluation; and (3) ten replicate runs were performed. The accuracy of the models was evaluated by the True Skill Statistic (TSS). The SDMs obtained with current conditions were projected in the following past periods: last glacial maximum [LGM; ~21 000 years before the present (ybp); general circulation models provided by MIROC-ESM and CCSM4] and last interglacial (LIG; ~120 000 ybp; Otto-Bliesner *et al.*, 2006). Current and past distributions were summarized using the committee average method over the set of single models (i.e. ensemble of three backgrounds per replicate run with TSS > 0.5). We applied the cut-off threshold that maximizes TSS for binarization (Allouche *et al.*, 2006). The *raster* package (Hijmans & Etten, 2012) for R and the Open Source Geospatial Foundation (QGIS) were used for spatial analysis.

RESULTS

Data characteristics

Fragments of 1042 and 1051 bp were obtained for CYTB and ND2, respectively. No indels in unexpected positions, nor stop or nonsense codons were detected in either alignment. Sequence chromatographs of the mitochondrial ND2 marker from *C. sanctaemariae* only (six individuals; 10 shared base pairs in the same position) contained double peaks indicating the presence of multiple different sequence products and possibly heteroplasmy. As heteroplasmy has been little reported in birds and no stop or nonsense codons were involved, we regarded these positions as missing data in all analyses. For nuclear markers, we obtained fragments of 598 bp for FIB5, including two indels in a heterozygous state (ranging between 1 and 3 bp); 409 bp for G3PDH, including six indels in

*BEAST species tree analysis based on all markers. Node support values >90% and >80% are depicted by two asterisks and one asterisk, respectively. Light grey bars represent 95% highest posterior density of divergence times. Pie charts at nodes represent probabilities of the ancestral distributions from BioGEoBEARS with the DIVALIKE+J model, with areas coded by colours (legend with colored squares below the map). Black arrows to the right correspond to gene flow between populations estimated by IMa2 (see Table 2). D, Quaternary climate curve derived from $\delta^{18}O$ (*y*-axis) and scaled against the divergence time analysis (*x*-axis); blue arrowheads correspond to main glaciation events (modified from Head & Gibbard, 2005).

a heterozygous state (ranging between 1 and 20 bp); 531 bp for MYO, including two indels in a heterozygous state (ranging between 1 and 2 bp); and 622 bp for MUSK, including one indel in a heterozygous state. We found no evidence of recombination (all $P < 0.05$). The diversity statistics and the neutrality test indexes Tajima's *D* and Fu's *Fs* are shown in Supporting Information, Table S5.

Phylogroup assignments, divergence times and ancestral distribution

Two main clades were observed in the phylogeny and structure analyses: *C. sanctaemariae* and *C. lineatus* (Fig. 1B, C). In *C. sanctaemariae*, a major split was detected for the mitochondrial genes, which separated sample 83 from the remaining specimens (Fig. 1B). In contrast, nuclear genes pointed to the existence of only one group (Fig. 1D). Within *C. lineatus*, two well-supported major clades were recognized by both inferences as well. One clade consisted of the *C. lineatus* population from the Guiana shield (hereafter GUI phylogroup), while a second one included *C. lineatus* populations from western and southern Amazonia, and Central America (Fig. 1B, D). According to the mtDNA tree only, this latter clade can be divided into two groups. One group has low support for their reciprocal monophyly but contains three clades which coincide with the Napo (NAP phylogroup) and Imeri (IME phylogroup) areas of endemism, plus individuals from Central America (CA phylogroup) together with individuals from the extreme south-west of Amazonia [hereafter called the 'Base of the Andes' (BA) population] that appear as sister taxa. The other group comprises clades of southern Amazonia including individuals from the Inambari, Rondônia, Tapajós and Xingu areas of endemism, which were treated as the Southwest Amazonian phylogroup (SW; including the monophyletic subset of the paraphyletic Inambari population); and the Southeast Amazonian phylogroup (SE; including the Xingú, Tapajós and Rondônia populations; Fig. 1B).

The divergence times in *Cymbilaimus* revealed an initial separation of *C. sanctaemariae* and *C. lineatus* [1.19 (0.6–1.85) Mya, Fig. 2B] and the radiation within *C. lineatus*, with the splitting of the GUI phylogroup from the remaining phylogroups of *C. lineatus* occurring during the late Pleistocene [0.61 (0.42–0.86) Mya, Fig. 2B]. Our species tree was slightly different from the mtDNA tree, whereby the *C. lineatus* NAP phylogroup appears as the sister phylogroup to the remaining *C. lineatus* populations from western and southern Amazonia, and Central America (Fig. 2B), with a splitting time of ~0.47 (0.35–0.61) Mya. Furthermore, the IME phylogroup and BA population were recovered as reciprocally monophyletic sister groups, and the split between them and the CA phylogroup took place around 0.37 (0.24–0.50) Mya (Fig. 2). However, these relationships have low statistical support. In addition, the split between the western and southern Amazonia clades took place around 0.43 (0.33–0.55) Mya. The most recent split was between the phylogroups south of the Amazon River, SW and SE, which was estimated at 0.29 (0.19–0.39) Mya.

DIVALIKE+J was chosen as the most likely ancestral area reconstruction scenarios for the genus *Cymbilaimus* [LnL = −13.782; corrected Akaike Information Criterion (AICc) = 37.56; Supporting Information, Table S6]. It recovered a widely distributed ancestral population inhabiting southern and north-eastern Amazonia (NE+SW+SE), and favoured a vicariant event for the split between *C. sanctaemariae* (SW+SE) and the ancestor of *C. lineatus* (NE; i.e. the ancestor was widespread and the daughter lineages came to occupy a subset of the original ancestral area; Fig. 2C). From this point on, DIVALIKE+J favoured a scenario of sequential colonizations through founder events, starting from the isolated population in the north-east portion, the GUI phylogroup, to the west during the late Pleistocene. The ancestor of the remaining *C. lineatus* phylogroups inhabited north-western Amazonia (NW), from where it colonized south of the Amazon River (SW and SE). Another putative splitting event followed the dispersal of *C. lineatus* from the NW area towards the CA area (Fig. 2C). The uncertainties associated with the topology of the tree apparently have little effect on the likelihood and estimated values of the J parameter, as can be seen in the graph in Figure 2B. Nevertheless, this ancestral area reconstruction should be interpreted with caution, as these uncertainties may be caused by asymmetrical gene flow (Fig. 2B). For comparative purposes, the DIVALIKE model with no founder event speciation (i.e. the best model in our analyses with no 'J' component; Table S6) differed from the one that includes the J parameter in that the ancestor of *C. lineatus* was estimated to occur in a broader region in northern Amazonia (NE+NW) and that, by vicariance, would have given rise to the GUI phylogroup and the ancestor of the remaining phylogroups in *C. lineatus* (Fig. S1). Also according to this model, the next events involved range expansions that provided colonization of CA and southern Amazonia (SW+SE) from NW and then cladogenesis by vicariance (Fig. S1).

AMOVA results showed that most of the genetic variation was found among phylogroups (ranging from 54.15 to 77.96%), followed by the variation found among populations within phylogroups (7.86–47.01%), and the lowest variation due to the differentiation between populations (2.21–22.49%; Table 1). Differentiation due to the variation between populations was significant (F_{ST} ranging from 0.53 to 0.92, all $P < 0.001$), but not significant among the phylogroups (F_{CT}; Table 1).

Table 1. AMOVA results indicating percentage of variation and fixation index (bold numbers are significant); Mantel test correlation scores (and *P*-values) between genetic distance and geographical or climatic distances are given

Marker	Percentage of variation*			Fixation indices			Mantel score†					
	Among groups	Among populations within groups	Within populations	F_{CT}	F_{ST}	F_{SC}	Genetic vs. Geographical	Genetic vs. Preciptation (BIO 18)	Genetic vs. Preciptation/ Geographical	Genetic vs. Temperature (BIO 4)	Genetic vs. Temperature/ Geographical
CYTB	57.19	22.49	20.32	0.57	**0.80**	**0.53**	0.44 (0.01)	0.30 (0.08)	0.21 (0.18)	0.53 (0.05)	0.62 (0.03)
ND2	77.96	14.17	7.86	0.78	**0.92**	**0.64**	0.28 (0.14)	0.26 (0.18)	0.20 (0.24)	0.63 (0.06)	0.68 (0.04)
FIB5	54.15	–1.16	47.01	0.54	**0.53**	**–0.03**	–0.10 (0.67)	0.09 (0.33)	0.12 (0.28)	0.18 (0.24)	0.18 (0.25)
G3PDH	74.65	–2.21	27.56	0.75	**0.72**	–0.09	–0.04 (0.56)	0.45 (0.01)	0.48 (0.01)	0.57 (0.02)	0.57 (0.02)
MYO	58.93	4.71	36.36	0.59	**0.64**	**0.11**	–0.09 (0.63)	0.45 (0.02)	0.49 (0.005)	0.65 (0.02)	0.65 (0.03)
MUSK	62.37	8.79	28.84	0.62	**0.71**	**0.23**	0.40 (0.04)	–0.16 (0.76)	–0.31 (0.92)	–0.19 (0.75)	–0.18 (0.70)

*Tested population structure: [(*C. sanctaemariae*); (Guiana); (Napo); (Imeri, Andes Base, Central America); (Inambari); (Rondônia, Tapajós, Xingu)].
†Only Amazonian lineages of *C. lineatus*.

Mantel tests showed correlations only between genetic variation and environmental variables, namely temperature (BIO4, see SDM resultsbelow), when considering the full data with all species/ phylogroups (although not regarded as significant, $P > 0.001$; Supporting Information, Table S7). In this case, temperature variation explained from 16% (ND2) to 63% (MYO) of the genetic variability (Table S7). However, when comparisons were restricted to the continuously distributed *cis*-Andean Amazonian *C. lineatus* phylogroups (i.e. all *C. lineatus* except the geographically more distant CA phylogroup), Mantel tests revealed a correlation with geographical distance for CYTB and MUSK loci (44 and 40%, respectively), and a correlation with climatic variables for CYTB, ND2, G3PDH and MYO loci (Table 1; Figs S2, S3). In this comparison, precipitation (BIO18) explained 26–30% of the genetic variation in mtDNA and 45% in the autosomal loci; by contrast, temperature (BIO4) had the strongest influence, explaining more than 50% of the genetic differentiation in CYTB, ND2, G3PDH and MYO loci.

Gene flow and demography

In all IMa2 runs, almost all posterior distributions of parameters had a clear peak and the right tails converged to zero (with the exception of the split times estimates involving *C. sanctaemariae*; Supporting Information, Figs S4–S6). IMa2 analyses indicated significant unidirectional and paraphyletic gene flow from *C. sanctamariae* into the NAP, SW and SE phylogroups, with posterior distributions peaking at ~0.1 migrants per generation (Table 2; Fig. S4). Furthermore, the IME phylogroup received migrants from GUI at an average of 1.4 individuals per generation (95% highest posterior density: 0.0–3.62, Table 2; Fig. S4) – the highest migration rate detected. Both rates are significantly different from zero according to the likelihood ratio test of Nielsen & Wakeley (2001). All other pairwise phylogroup comparisons revealed no significant gene flow (Table 2; Fig. S4). IMa2 showed similar effective population sizes (*Ne*) for all phylogroups of *C. lineatus*, except for the CA phylogroup, which had the smallest one; and for *C. sanctaemariae*, with larger values (Table 2). The posterior distributions of ancestral populations showed small sizes for all phylogroups (Fig. S5). Considering a generation time of ~3 years inferred for both *Cymbilaimus* species (Bird *et al.*, 2020), the oldest split time was between the GUI and IME phylogroups and occurred ~0.69 (0.41–2.22) Mya (Table 2; Fig. S6). In addition, between the pairs of phylogroups IME and CA, and NAP and CA, the split occurred between 0.14 and 1.57 Mya. The split between the phylogroup pairs NAP and IME, and NAP and SW may have occurred

Table 2. Parameter estimates from IMa2 including split time, population size and migration rate (modal values; 95% highest posterior density interval in parentheses; bold numbers are significant)

Pairwise analyses (pop1 × pop2)	Split time	N1	N2	2N1M1>2	2N2M2>1
GUI × IME	0.69 (0.41–2.22)	6.13 (3.70–10.19)	2.42 (1.34–4.42)	**1.40 (0.0–3.62)**	0.003 (0.0–0.48)
IME × NAP	0.22 (0.12–0.42)	3.15 (1.37–6.61)	3.25 (1.03–11.33)	0.02 (0.0–2.53)	0.06 (0.0–5.50)
IME × CA	0.28 (0.16–1.33)	3.74 (1.69–7.57)	1.09 (0.44–2.36)	0.01 (0.0–1.83)	0.13 (0.0–1.18)
NAP × CA	0.27 (0.14–1.57)	3.51 (1.01–12.37)	0.99 (0.35–2.35)	0.02 (0.0–7.34)	0.004 (0.00–1.10)
NAP × INA	0.22 (0.13–0.51)	3.76 (1.34–11.54)	7.94 (3.88–16.45)	0.01 (0.0–3.64)	0.01 (0.0–3.97)
INA × TAP	0.14 (0.07–1.24)	6.16 (2.12–17.66)	1.22 (0.38–3.20)	0.004 (0.0–14.80)	0.01 (0.0–1.67)
NAP × sanct	–	2.12 (1.00–3.96)	5.80 (3.48–9.40)	0.008 (0.0–0.13)	**0.08 (0.006–0.31)**
INA × sanct	–	4.70 (2.50–7.70)	6.30 (3.30–10.50)	0.06 (0.0–0.24)	**0.11 (0.01–0.39)**
TAP × sanct	–	1.15 (0.55–2.65)	6.85 (3.85–11.95)	0.009 (0.0–0.11)	**0.11 (0.006–0.44)**

Split time is in milion years, considering a generation time of ~3.01 (Bird *et al.*, 2020).
N1 and N2, population sizes for populations 1 and 2, respectively. The values are $4Ne\mu$, where Ne is the effective population size and μ is the mutation rate.
2N1M1>2, population migration rate into population 2 from population 1 per generation; 2N2M2>1, population migration rate into population 1 from population 2 per generation.
Migration rates that are significantly different from zero at the $P < 0.01$ level in LLR tests (Nielsen & Wakeley, 2001; Hey, 2010) are shown in bold type.

between 0.12 and 0.51 Mya. The most recent split, between the SW and SE phylogroups, occurred ~0.14 (0.07–1.24) Mya.

The EBSPs for *C. sanctaemariae* suggest that it maintained a relatively high and stable population size during the Pleistocene, followed by a gradual decline during the last 0.15 Myr (Fig. 3A). The GUI phylogroup is inferred to have maintained a stable population size (Fig. 3B). Three phylogroups, IME, NAP and SW, also appeared to have maintained stable population sizes during the Pleistocene, but with small sizes and a slight expansion during the last 0.12 Myr (Fig. 3C, D, F). The CA phylogroup is inferred to have maintained a smaller and stable population size and possibly experienced an expansion since the LGM (Fig. 3E). The SE phylogroup has shown a more expressive expansion signal in the last 0.12 Myr (Fig. 3G).

Species distribution modelling

All SDMs showed high TSS values ranging from 0.6 to 0.8, indicating high model fit. In almost all methods, the environmental variables with high explanatory values were Precipitation of Warmest Quarter (BIO18; most important for *C. sanctaemariae*, GUI and SW, but also among the three most important variables for NAP, CA and SE), Precipitation Seasonality (BIO15; for NAP), Precipitation of Coldest Quarter (BIO19. for IME), Temperature Mean Diurnal Range (BIO2; for CA), and Temperature Seasonality (BIO4; most important for SE, but also among the three most important variables for *C. sanctaemariae*, GUI, IME and SW; Supporting Information, Table S8).

The present distribution range was generally well recovered and our models predicted a potential change in the distribution of suitable habitats between the LIG and today, in general agreeing with the results obtained in the EBSP (Fig. 3). Our projections show only two periods, LGM and LIG, when the phylogroups were already established, and we assume that climate change before 0.12 Mya (Fig. 2D) may have driven similar patterns of availability of suitable habitats. The LIG projections for *C. sanctaemariae*, IME, CA and SW phylogroups revealed that suitable habitats could potentially have extended far beyond their current limits (Fig. 3A, C, E, F). However for GUI and SE, our models did not predict suitable habitat for this period (Fig. 3B, G). It is also important to note that the habitats suitable for the IME phylogroup during the LIG were projected for where the GUI occurs today (Fig. 3B, C); similarly, the range of the CA phylogroup was also projected to where NAP occurs today (Fig. 3D, E). From the LIG to the LGM and then to the present, our models suggest that there was displacement (to the south/south-west) and retraction of suitable habitats for *C. sanctaemariae* (Fig. 3A). Suitable habitats for the GUI and SE phylogroups were estimated to have expanded to the current range after the LGM (Fig. 3B, G). Our models also indicate that some phylogroups may have experienced displacements related to suitable habitat contractions followed by expansion, such as IME (from east to west; Fig. 3C) and SW (from the north-west to south/south-east and back to the south-west; Fig. 3F). Finally, the NAP phylogroup did not show significant changes in the range of suitable habitats between these periods (Fig. 3D).

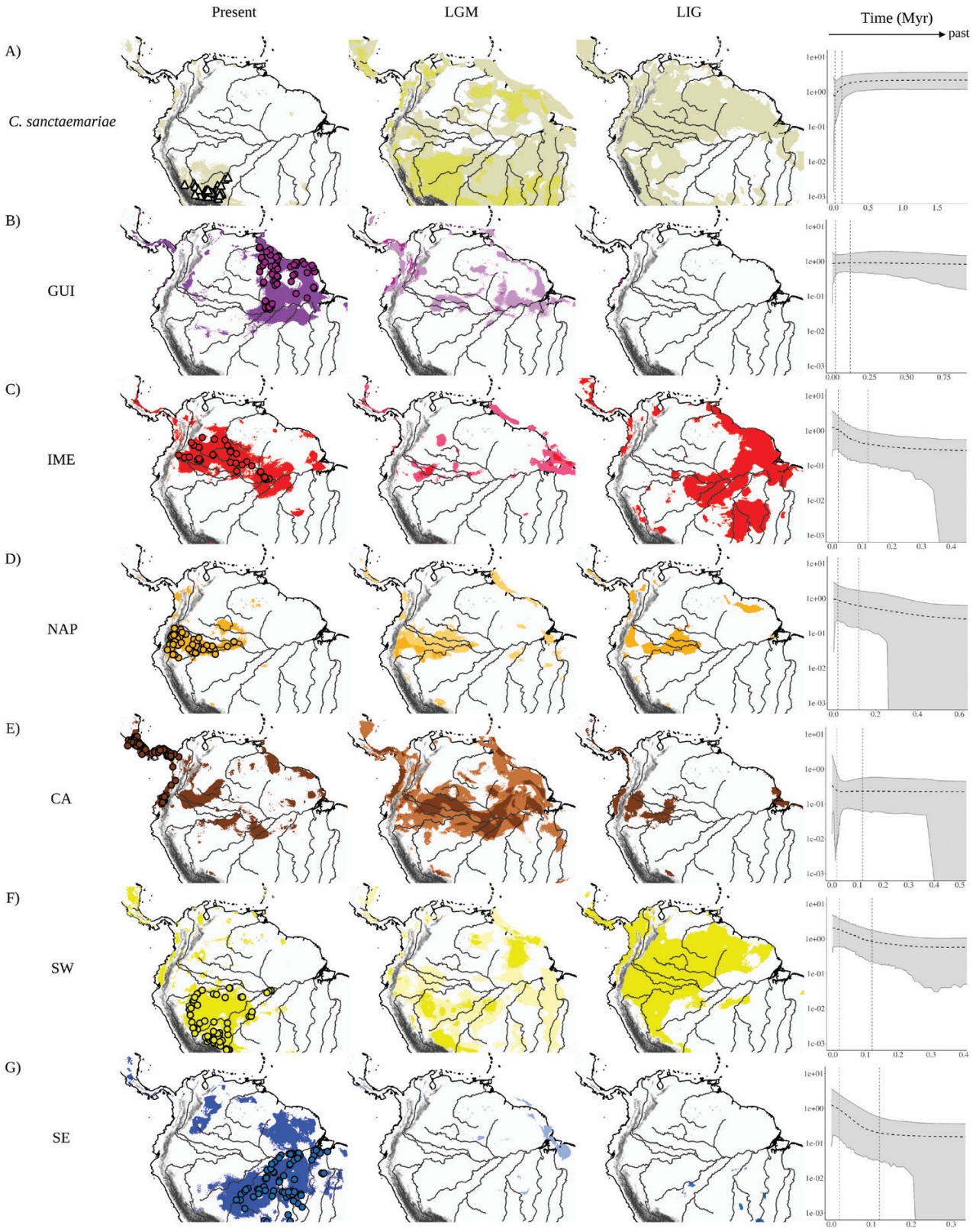

Figure 3. Species distribution models of the phylogroups found in *Cymbilaimus lineatus* and *C. sanctaemariae* obtained with MAXENT, projected for the present, last glacial maximum (LGM) and last interglacial (LIG) periods. Dots in present

DISCUSSION

Our results support the hypothesis that *C. sanctaemariae* split as a separate lineage early during the diversification of *Cymbilaimus*, probably following the establishment of bamboo-dominated forests in western Amazonia. Apparently, *C. sanctaemariae* and *C. lineatus* differentiated largely in allopatry in the mid-Pleistocene, but later came into secondary contact in western Amazonia, where they introgressed multiple times (Fig 2C). Our scenario for the evolution of *C. lineatus* phylogroups agrees with those in previous genetic analyses, but here we were able to show that both ILS and gene flow have been part of the diversification process in this species. Our study also recovered divergence times within the Quaternary, supporting an important role for dispersal in *Cymbilaimus* diversification, and that demographic signatures and range size dynamics coincided with major climatic events (Figs 2C, D, 3). Finally, our results suggest a correlation between genetic and climatic variability for *Cymbilaimus* populations throughout their range (Table 1).

ILS, GENE FLOW AND DIVERGENCE TIMES

Before contrasting the spatio-temporal pattern of diversification of *Cymbilaimus* with the history of the landscape of northern South America, it is important to highlight the evolutionary mechanisms that can generate uncertainties in the inferences made by the methods we have adopted. It is noteworthy that more complex evolutionary scenarios may be poorly described in a tree-like manner, so our inferences should be interpreted with caution for various reasons. ILS, wherein ancestral polymorphisms persist through species divergence, and gene flow among species can generate impacts on phylogenetic inferences (Leaché *et al.*, 2014). First, it is important to distinguish shared ancestral variation from gene flow because both scenarios produce similar patterns. On the one hand, ILS can be accommodated in coalescent approaches such as that implemented in BEAST (Heled & Drummond, 2010), but this approach does not model gene flow. In fact, the impacts of gene flow on phylogenetic inference remain little studied (Pinho & Hey, 2010), with fewer methods that can jointly consider ILS and gene flow (Thom *et al.*, 2018). Because here we explicitly estimated levels of gene flow, our data supported asymmetric gene flow involving four out of nine pairs of populations, i.e. from *C. sanctaemariae* into *C. lineatus* NAP, SW and SE phylogroups (less than one migrant per generation) and from *C. lineatus* phylogroup GUI to IME (more than one migrant per generation). Nevertheless, our study also shows that *C. lineatus* phylogroups IME, NAP, BA, CA, SW and SE have not been completely sorted autosomally, as evidenced by the lack of structure in the BAPS results. Even when analysed using new generation sequencing techniques such as ultraconserved elements (UCEs), a lack of lineage sorting is still verified within *C. lineatus* (see fig. 3 in Smith *et al.*, 2014a). Additional lines of evidence in this direction also include the recent radiation (<1.0 Mya), short internal branches recovered by phylogenetic analyses, and the high degree of shared polymorphic sites and haplotypes among *C. lineatus* phylogroups. Based on predictions from coalescence theory, about four to seven *Ne* generations are necessary to complete the lineage sorting of autosomal genes, which means that species with large population sizes could be even more affected by ILS (Nichols, 2001). Thus, our results show that both ILS and limited gene flow have shaped the evolutionary history of *Cymbilaimus* phylogroups.

The impact of gene flow on species tree inferences has been explicitly demonstrated (e.g. based on decreasing posterior clade probabilities and underestimates of divergence time estimates; Leaché *et al.*, 2014) and great effort is being made to try to explicitly accommodate this source of inconsistency into species tree inference (Than *et al.*, 2008; Solís-Lemus *et al.*, 2017). Our divergence date estimates among phylogroups according to the *BEAST species tree were partially consistent with those inferred with the IM model, given that their confidence intervals overlap (see Fig. 2, Table 2; Supporting Information, Fig. S6). However, these results have some caveats, as we used a general calibration uncorrected for the effects of body mass or generation time (see Nabholz *et al.*, 2016). Our *BEAST estimates may also underestimate actual divergence times, as we observed paraphyletic gene flow from *C. sanctaemariae* into NAP, SW and SE phylogroups and from GUI into IME (see Smith *et al.*, 2014a). Furthermore, it has been shown by using simulation data (Strasburg &

projections represent records used for MAXENT simulations. Different tones in LGM projections correspond to agreements between General Circulation Models (GCMs): darker colours, both GCMs predict habitat suitability; light colours, only one GCM predicts habitat suitability. Graphics on the left represent the demographic trends inferred through Extended Bayesian Skyline plots (EBSP): *x*-axis, time in millions of years (Myr) back to the upper limit of the divergence date inferred by the *BEAST tree; *y*-axis, log population sizes (*Ne*). Black dashed lines represent median values, and while solid lines correspond to 95% confidence intervals. Vertical dashed lines point to interglacial (~0.12 Mya) and glacial (~0.02 Mya) periods.

Rieseberg, 2010) and in terms of the probability of genealogies (Sousa *et al.*, 2011) that the divergence time estimated under a model with migration (IM) is statistically unreliable as the different migration timings could result from misleading genealogies with the same probability. Therefore, regardless of the exact timing and even considering the confidence intervals, most splitting times within *Cymbilaimus* converge on the Late Quaternary, an epoch characterized by great changes in Amazonian landscapes – such as habitat fragmentation – promoted by climatic fluctuations, as well as geo-tectonic processes (Hayakawa & Rossetti, 2015; Sato & Cowling, 2017).

A SCENARIO FOR THE EVOLUTION OF *CYMBILAIMUS* PHYLOGROUPS

Having highlighted the mechanisms that played a role in the diversification process, we can now propose an evolutionary scenario. Our results provide support for diversification of *Cymbilaimus* occurring between 1.2 and 0.3 Mya (Fig. 2), probably due to Pleistocene climatic fluctuations; the mid-point estimates of the main splits appear to coincide with glacial maxima following the Quaternary climate curve derived from $\delta^{18}O$ (modified from Head & Gibbard, 2005, and correlated previously with dated phylogenies; see Ribas *et al.*, 2012; Carneiro *et al.*, 2018). In addition, although some relationships were poorly resolved, some of the phylogroups seem to be bounded by major rivers, highlighting the additional role of rivers as barriers – even if secondary (Burney & Brumfield, 2009).

From this perspective, our DIVALIKE+J model suggested a first vicariant event for the split between *C. sanctaemariae* and the ancestor of *C. lineatus* as being completed between 1.85 and 0.6 Mya (Fig. 2). As a bamboo (mainly *Guadua*) specialist (Pierpont & Fitzpatrick, 1983; Zimmer & Isler, 2020a), the evolution of *C. sanctaemariae* was probably influenced by the distribution of bamboo-dominated forests in Amazonia. The origins of these environments and the potential changes in their range over time remain of debate. Thorny bamboo similar to *Guadua* macrofossils have been found in Peru, suggesting the presence of *Guadua* bamboo-like species in lowland Amazonia since the Plio-Pleistocene (Olivier *et al.*, 2009). Moreover, the two most common species in these south-western Amazonian bamboo-dominated forests, *Guadua sarcocarpa* and *G.weberbaueri*, are endemic and their divergence dates back to ~1.14 Mya (Wu *et al.*, 2015). Our demographic reconstructions inferred population expansions for almost all phylogroups between 0.12 Mya to the present, except for *C. sanctaemariae*, whose effective population sizes seem to have declined during the past 0.15 Myr, suggesting opposite demographic trends around the same time and the potential contraction of the bamboo forest and expansion of closed-canopy humid forests (Fig. 3A). This pattern apparently supports the view that bamboo-dominated forests in the Amazon are relicts of a former more widely distributed formation across the basin (Olivier *et al.*, 2009). *Guadua* species can grow in very different climatic regimes and their distribution has no correlation with soil, topography or human influence (McMichael *et al.*, 2014); however, unlike most species of the genus, *G. sarcocarpa* and *G. weberbaueri* depend on adjacent trees for vertical growth, causing a 'self-perpetuating disturbance cycle' to persist (Griscom & Ashton, 2006). Regardless, whether due to the climatic scenario or some other driver of changes in the range of bamboo-dominated forests, the NAP, SW and SE *C. lineatus* phylogroups or their ancestors potentially secondarily came into contact with *C. sanctaemariae*, and even with limited gene flow, this may have resulted in continuing divergence due to their distinct habitat requirements (Zimmer & Isler, 2020a).

The DIVALIKE+J model also suggests that the ancestor of all phylogroups in *C. lineatus* was in the NE phylogroup, having been isolated there after the split from *C. sanctaemariae*, resulting in the origin of the GUI phylogroup between 0.86 and 0.42 Mya (Fig. 2). The estimated age for this event is closely associated with the so-called 'mid-Pleistocene revolution', when major glacial maxima became progressively more severe, impacting ecosystems more strongly (Fig. 2D; Head & Gibbard, 2005). The interfluve delimited by the upper Negro and Orinoco rivers and the Branco River (Guiana area in Fig. 1) is occupied by the largest patches of savannas and white-sand forests known in Amazonia, which may have experienced significant changes in vegetation cover due to climate change. In addition, geomorphological data suggest that both rivers are relatively recent (<1.2 Mya) and with evidence of drainage capture events as young as thousands of years ago (Schaefer & Vale, 1997; Almeida-Filho & Miranda, 2007; Soares *et al.*, 2010). This demonstrates how dynamic the environment may have been in the Guiana shield. In fact, we have recovered significant changes in the extent of suitable habitats for the GUI phylogroup since the interglacial period (e.g. virtually no suitable habitat was projected for the LIG, but suitable habitats started to expand from the LGM). Nevertheless, there is no sign of demographic fluctuations in the EBSP estimates (Fig. 3B). Interestingly, during the LIG, our results suggest suitable habitats for the IME phylogroup in the area of occurrence that today is occupied by GUI (Fig. 3C). Furthermore, our SDMs also suggest a decrease in the amount of suitable habitat for the IME phylogroup between the LIG and LGM, as well

as an expansion between the LGM and today, but, again, with no signs of demographic fluctuations (Fig. 3C). The GUI and IME phylogroups are the only ones among *C. lineatus* that showed signs of pairwise gene flow greater than one individual per generation (Table 2; Supporting Information, Fig. S4) and these secondary contact events after periods of isolation (caused by climate change that must have led to range overlap) can produce greater genetic diversity (Alcala & Vuilleumier, 2014; Garrick *et al.*, 2019). Recent work with *Pseudopripra pipra* has recovered a reasonably similar pattern with regard to corresponding area and timing (Berv *et al.*, 2019).

Our data also support a pattern of sequential dispersal episodes, almost simultaneous between 0.6 and 0.2 Mya, for the occupancy of the north-west, Central America and the south bank of the Amazon River (Fig. 2). Studies including several avian lineages have identified a wide range of splitting times across the main Amazonian barriers (Smith *et al.*, 2014b). For instance, divergence times estimated for the separation between *cis*- and *trans*-Andean lowland lineages ranged from the late Neogene (>2.58 Mya; ten splitting episodes; Smith *et al.*, 2014b), to the Quaternary (<2.58 Mya; 23 splitting episodes; Smith *et al.*, 2014b). This suggests that changes in the availability of suitable habitats caused by increasingly strong glacial cycles may have driven dispersion pulses (founder events) that produced different phylogroups (IME, CA and SW) in this period from an NAP phylogroup ancestor. In addition to the SDM projections that show a stability of suitable habitats for the NAP phylogroup in the three time periods (Fig. 3D), this scenario is also supported by: (1) the greater genetic diversity of this phylogroup in relation to the others (Supporting Information, Table S5); (2) population expansion patterns recovered by EBSPs (Fig. 3D); and (3) a significant correlation between geographical distance/climatic differences and F_{ST} (Table 1; Figs S2, S3). Furthermore, one of those pulses that led to the colonization of southern Amazonia promoted the secondary contact between two different non-sister phylogroups that diverged in allopatry: the SW *C. lineatus* phylogroup and *C. sanctaemariae*. The presence of phylogroups west of the lower Madeira River (SW phylogroup, Inambari area in Fig. 1) that are nevertheless more closely related to those occurring across the river (i.e. SE phylogroup, Rondônia area in Fig. 1) rather than any co-distributed phylogroups (i.e. *C. sanctaemariae*) on the west bank has been reported for at least five other avian lineages associated with the same habitat as *Cymbilaimus* (upland terra-firme forest; Aleixo, 2004; Patané *et al.*, 2009; Miranda *et al.*, 2013; Fernandes *et al.*, 2014; Ferreira *et al.*, 2017).

It is worth emphasizing that the scenario described above can only be valid assuming jump dispersal as the main initiator of cladogenesis in *Cymbilaimus* and that new biogeographical hypotheses may arise in the light of more robust phylogenies and models. As a counterpoint to a model assuming jump dispersal, our best alternative model that does not include the J parameter (DIVALIKE) supports a shared ancestral range across Amazonia, in addition to range expansion events, favouring splits by vicariance only (Supporting Information, Fig. S1). However, DIVALIKE also points to dispersal as an important factor in the reconstructed scenario of *Cymbilaimus* diversification (Table S6).

Ecology and dispersal ability in the diversification process

Effective population sizes and patterns of gene flow among populations are influenced by the ecology of a species (e.g. species-specific dispersal and environmental tolerance attributes; see Harvey *et al.*, 2019). For example, species with specific habitat requirements, such as *C. sanctaemariae* (a near-obligate bamboo specialist; Zimmer & Isler, 2020a), may have a low ability for moving individuals between populations (hence reducing gene flow), but such difficulty can be surpassed with an increase in dispersal ability (Phillipsen *et al.*, 2014). However, dispersal ability is usually inferred through species attributes associated with habitat, diet and relative abundance (Burney & Brumfield, 2009; Smith *et al.*, 2014a; but see Moore *et al.*, 2008 for experimental evidence). For instance, *Cymbilaimus* species inhabit the mid-storey canopy and forest edges, attributes which are associated with a better ability to cross habitat gaps, and hence functioning as an indirect measure of relatively high dispersal ability (Stotz *et al.*, 1996; Burney & Brumfield, 2009; Zimmer & Isler, 2020a, b). Therefore, our results are in agreement with a growing body of evidence favouring dispersal events as the initiators of geographical isolation and speciation (Burney & Brumfield, 2009; Smith *et al.*, 2014a; Thom *et al.*, 2018) and that past climate fluctuations could have affected the distribution of populations, directing gene flow (sometimes inhibiting flow but at other times promoting flow) and affecting genetic structure (Silva *et al.*, 2019). Given the timing and possible scenario of diversification described in the previous session, landscape change alone (a basic assumption of most diversification hypotheses) is not an exclusive requirement for cladogenesis and speciation among geographically separated populations, but also the level of persistence and dispersive capabilities (Smith *et al.*, 2014b). Our results also reinforce the findings of Silva *et al.* (2019) in which climate-driven distribution dynamics produced a diversification pattern running chronologically from north-western Amazonia (older lineages found in climatically wetter and more stable

areas) to the south-east (younger lineages recovered in climatically drier and more unstable areas). Although more than 90% of the 23 lineages sampled in that study shared this geographically 'anticlockwise' pattern of cladogenesis, the authors did not explicitly test its association with climatic variables. Here, we demonstrate that much of the genetic variability in *cis*-Amazonian *C. lineatus* phylogroups is correlated across a gradient of precipitation and temperature, where the environmentally distant populations were characterized by higher genetic differentiation (e.g. IME, NAP vs. RON, XIN); even after accounting for the effects of isolation by distance (Table 1; Supporting Information, Figs S2, S3). Thus, according to the best model estimated by BioGeoBEARS for *Cymbilaimus*, vicariance and habitat specialization accounted for the earliest split, whereas cladogenesis resulting from jump dispersal (founder events) was favoured for subsequent nodes, indicating that a balance between vicariance and genetic differentiation resulting from dispersal and founder events across pre-existing barriers together explain the pattern of diversification in this antshrike genus. Furthermore, landscape change is not only associated with cladogenesis, but also with promotion of admixture, as inferred for the asymmetric gene flow documented between some phylogroups.

Phylogeographical structure, gene flow and taxonomy

Two species of *Cymbilaimus* and three subspecies in *C. lineatus* are currently recognized on the basis of bioacoustic, morphological and morphometric data (Zimmer & Isler, 2020a, b), although the phenotypic diagnosis of *C. lineatus* subspecies is not straightforward and it is mostly based on the intensity of barring throughout the body (Zimmer, 1932). Genetically, our multi-locus analysis supported the existence of three major groups in *Cymbilaimus*: (1) *C. sanctaemariae*; (2) *C. lineatus lineatus* (nominate *lineatus* is the taxon name applied to the GUI phylogroup; Peters, 1951; Zimmer & Isler, 2020b); and (3) remaining *C. lineatus* NAP, IME, CA, BA, SW and SE phylogroups, which include populations belonging to the taxon names *fasciatus* and *intermedius*. Therefore, current subspecific limits within *C. lineatus* are inconsistent with its evolutionary history, given the recovered paraphyly of *C. lineatus intermedius*, which currently designates all phylogroups, except GUI (*C. l. lineatus*) and CA (to which the name *fasciatus* unequivocally applies; Peters, 1951).

Despite the existence of eight reciprocally monophyletic mitochondrial populations, in the clade grouping phylogroups NAP, IME, CA, BA, SW and SE, analyses based only on the nuclear genes indicated lack of structuring among them, which could be explained either by a lack of lineage sorting or asymmetrical gene flow between mtDNA and nDNA genes or both factors (Tajima, 1983; Maddison & Knowles, 2006; Pinho & Hey, 2010). When pairwise rates of gene flow are estimated among the phylogroups in *C. lineatus*, only one comparison (involving the GUI and IME phylogroups) recovered significant values and a high migration rate. Hence, the evidence indicates that ancestral polymorphism rather than gene flow accounts for most of the comparatively lower differentiation in nDNA than mtDNA in *C. lineatus*, consistent with the notion that at least the CA, SW and SE phylogroups are diverging from each other with little or no gene flow between them. However, our sampling is too sparse to document more localized and potentially high rates of gene flow among the *C. lineatus* phylogroups uncovered by this study [e.g. Rondônia and BA populations that were excluded from the pairwise gene flow estimates due the small sampling (N = 3]).

Given that the highest rates of gene flow among *C. lineatus* populations were detected between two of the most genetically divergent groups (GUI and IME, which are found in parapatry across the Branco River; Naka & Brumfield, 2018), it is likely that other parapatric phylogroups are also reproductively compatible with each other. This supports the treatment of all *C. lineatus* phylogroups as a single biological species, which is also consistent with their lack of any phenotypic diagnosis (Zimmer, 1932; Zimmer & Isler, 2020b). However, as exemplified multiple times, gene flow or lack thereof may not be a consistent criterion for delimiting species, particularly if they are cryptic, as is the case with *C. lineatus* and *C. sanctaemariae* (see Fišer *et al.*, 2018; Pulido-Santacruz *et al.*, 2018). Our estimates indicate that gene flow has taken or is still taking place in western Amazonia between *C. sanctaemariae* and *C. lineatus* phylogroups NAP, SW and SE, even though at a very low rate, which is nevertheless surprising considering the high degree of vocal, ecological and genetic differentiation already attained by these lineages (Zimmer & Isler, 2020a, b). While *C. sanctaemariae* and *C. lineatus* phylogroup NAP are today distributed parapatrically, the current ranges of *C. sanctaemariae* and *C. lineatus* phylogroups SW and SE (Tapajós area of endemism) are sympatric (Zimmer & Isler, 2020a, b), underlining that both recent and even current gene flow might exist between this cryptic species pair. This finding supports the idea that complete reproductive isolation takes a very long time to evolve, even between otherwise vocally divergent cryptic species, particularly in the tropics (Pulido-Santacruz *et al.*, 2018, 2020; Weir & Price, 2019), and that it represents more of an ancestral character that is lost only after

two groups have evolved independently for a long time (Zink, 2004).

Despite this documented mismatch between gene flow and species limits in *Cymbilaimus*, sufficient genetic and phenotypic divergences support the treatment of *C. sanctaemariae* and *C. lineatus* as separate species, while within the latter at least two major genetically divergent groups (at both the mitochondrial and the nuclear levels) can be objectively delimited, but which nevertheless lack any consistent phenotypic diagnosis (Zimmer, 1932). These results are consistent with ranking *C. lineatus lineatus* (including only phylogroup GUI) and *C. lineatus fasciatus* (the name with nomenclatural priority available for all remaining phylogroups, which were consistently recovered as monophyletic) as evolutionarily basal units in both biogeographical studies and conservation assessments.

CONCLUSIONS

The methods used in an integrated manner in this study provide evidence that the evolutionary history of taxa in *Cymbilaimus* developed during the Pleistocene, being influenced by cycles of founder events and secondary contacts caused by climate change. The association of *C. sanctaemariae* with bamboo-dominated forests in western Amazonia is the main evidence supporting the effects of ecological specialization in the diversification of the genus. At least from a vocal point of view, it is known that these environments have sound transmission properties different from terra-firma forests, which may be correlated with differences in vocalizations (Tobias *et al.*, 2010). Although *C. sanctaemariae* and *C. lineatus* are recognized as independent species based on morphological and vocal differences, our work also presented evidence of recent or ongoing gene flow between them. Therefore, our results are in line with recent studies showing that prezygotic barriers (e.g. morphological and vocal differences) are less important as mechanisms of reproductive isolation than postzygotic barriers (i.e. accumulation of genetic incompatibilities) and that detectable recent gene flow can occur even in lineages that diverged just a few million years ago (Pulido-Santacruz *et al.*, 2018, 2020).

ACKNOWLEDGEMENTS

We would like to thank the following institutions, curators and curatorial assistants for providing tissues: Mario Cohn-Haft, Camila Ribas and Ingrid Macedo (Instituto Nacional de Pesquisas da Amazonia, INPA); Cristina Y. Miyaki (Laboratório de Genética e Evolução Molecular de Aves, Universidade de São Paulo, LGEMA); Donna Dittmann and Robb Brumfield (Louisiana State University Museum of Natural History, LSUMZ); John Bates, Ben Marks and Jason Weckstein (Field Museum of Natural History, FMNH); Nate Rice (Academy of Natural Sciences of Drexel, ANSP); and Christopher Milensky (Smithsonian Institution, USNM). The laboratory work was conducted at the Laboratório Multidisciplinar of Museu Paraense Emılio Goeldi (MPEG). Julianna Fernandes helped with editing the figures. Lucas E. Araujo-Silva, Lincoln Carneiro, Tiberio C. T. Burlamaqui, Fabio Raposo, Henrique Batalha-Filho, Sofia Silva and Maria Lucia Harada provided useful and inspiring comments on earlier versions of the manuscript. We also thank three anonymous reviewers for their helpful comments. Field and laboratory work related to this paper was generously funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; grants #310593/2009-3; 'Instituto Nacional de Ciência e Tecnologia em Biodiversidade e Uso da Terra da Amazônia' #574008/2008-0; #563236/2010-8; and #471342/2011-4) and Fundação Amazônia de Amparo a Estudos e Pesquisas (ICAAF 023/2011 and 011/2012). Support of L.S.M.'s graduate research was provided by a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) Doctoral fellowship and CNPq-MPEG 'Programa de Capacitação Institucional' (#302202/2020-0). B.O.P. was supported by a Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) undergraduate (PIBIC) fellowship (grant #16105/2014–7). A.A. was supported by CNPq research productivity fellowships (#310880/2012-2 and 306843/2016-1).

REFERENCES

Alcala N, **Vuilleumier S. 2014.** Turnover and accumulation of genetic diversity across large time-scale cycles of isolation and connection of populations. *Proceedings of the Royal Society B: Biological Sciences* **281:** 20141369.

Aleixo A. 2004. Historical diversification of a terra-firme forest bird superspecies: a phylogeographic perspective on the role of different hypotheses of Amazonian diversification. *Evolution* **58:** 1303–1317.

Aleixo A, **Guilherme E. 2010.** Avifauna da Estação Ecológica do Rio Acre, estado do Acre, na fronteira Brasil/Peru: composição, distribuição ecológica e registros relevantes. *Boletim do Museu Paraense Emílio Goeldi Ciências Naturais* **5:** 279–309.

Aleixo A, **Rossetti DDF. 2007.** Avian gene trees, landscape evolution, and geology: towards a modern synthesis of Amazonian historical biogeography? *Journal of Ornithology* **148:** 443–453.

Allouche O, **Tsoar A**, **Kadmon R. 2006.** Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* **43:** 1223–1232.

Almeida-Filho R, **Miranda FP. 2007.** Mega capture of the Rio Negro and formation of the Anavilhanas Archipelago, Central Amazônia, Brazil: evidences in an SRTM digital elevation model. *Remote Sensing of Environment* **110:** 387–392.

Barrera-Guzmán AO, **Aleixo A**, **Shawkey MD**, **Weir JT. 2017.** Hybrid speciation leads to novel male secondary sexual ornamentation of an Amazonian bird. *Proceedings of the National Academy of Sciences* **115:** E218–E225.

Berv JS, **Campagna L**, **Feo TJ**, **Castro-Astor I**, **Ribas CC**, **Prum RO**, **Lovette IJ. 2019.** Genomic phylogeography of the White Crowned Manakin *Pseudopipra pipra* (Aves: Pipridae) illuminates a continental-scale radiation out of the Andes. *BioRxiv* https://doi.org/10.1101/713081.

Bird JP, **Martin R**, **Akçakaya HR**, **Gilroy J**, **Burfield IJ**, **Garnett S**, **Symes A**, **Taylor J**, **Şekercioğlu Ç**, **Butchart SHM. 2020.** Generation lengths of the world's birds and their implications for extinction risk. *Conservation Biology* **34:** 1252–1261.

Bruen TC, **Philippe H**, **Bryant D. 2006.** A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172:** 2665–2681.

Burney CW, **Brumfield RT. 2009.** Ecology predicts levels of genetic differentiation in neotropical birds. *The American Naturalist* **174:** 358–368.

Carneiro L, **Bravo GA**, **Aristizábal N**, **Cuervo AM**, **Aleixo A. 2018.** Molecular systematics and biogeography of lowland antpittas (Aves, Grallariidae): the role of vicariance and dispersal in the diversification of a widespread Neotropical lineage. *Molecular Phylogenetics and Evolution* **120:** 375–389.

Chamberlain S, **Ram K**, **Barve V**, **Mcglinn D. 2015.** *Package 'rgbif': Interface to the Global 'Biodiversity' Information Facility 'API'*. Available at: https://CRAN.R-project.org/package=rgbif.

Collar NJ. 2018. Taxonomy as tyranny. *Ibis* **160:** 481–484.

Corander J, **Marttinen P**, **Sirén J**, **Tang J. 2008.** Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* **9:** 539.

Darriba D, **Taboada GL**, **Doallo R**, **Posada D. 2012.** JModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9:** 772.

De Queiroz K. 2007. Species concepts and species delimitation. *Systematic Biology* **56:** 879–886.

Dickinson EC, **Christidis L. 2014.** *The howard and moore complete checklist of the birds of the world, 4th edn, Vo. 2: Passeriforms*. Eastbourne: Aves Press.

Drummond AJ, **Rambaut A. 2007.** BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7:** 214.

Ellegren H. 2007. Molecular evolutionary genomics of birds. *Cytogenetic and Genome Research* **117:** 120–130.

Excoffier L, **Lischer HEL. 2010.** Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10:** 564–567.

Fernandes AM, **Wink M**, **Sardelli CH**, **Aleixo A. 2014.** Multiple speciation across the Andes and throughout Amazonia : the case of the spot-backed antbird species complex (*Hylophylax naevius* / *Hylophylax naevioides*). *Journal of Biogeography* **41:** 1094–1104.

Ferreira M, **Aleixo A**, **Ribas CC**, **Santos MPD. 2017.** Biogeography of the Neotropical genus *Malacoptila* (Aves: Bucconidae): the influence of the Andean orogeny, Amazonian drainage evolution and palaeoclimate. *Journal of Biogeography* **44:** 748–759.

Fišer C, **Robinson CT**, **Malard F. 2018.** Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology* **27:** 613–635.

Flot JF. 2010. Seqphase: a web tool for interconverting phase input/output files and fasta sequence alignments. *Molecular Ecology Resources* **10:** 162–166.

Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147:** 915–925.

Garrick RC, **Banusiewicz JD**, **Burgess S**, **Hyseni C**, **Symula RE. 2019.** Extending phylogeography to account for lineage fusion. *Journal of Biogeography* **46:** 268–278.

Garrick RC, **Sunnucks P**, **Dyer RJ. 2010.** Nuclear gene phylogeography using PHASE: dealing with unresolved genotypes, lost alleles, and systematic bias in parameter estimation. *BMC Evolutionary Biology* **10:** 118.

Gill F, **Donsker D**, **Rasmussen P. 2020.** *IOC World Bird List (v10.2)*. doi: 10.14344/IOC.ML.10.2.

Goslee SC, **Urban DL. 2007.** The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* **22:** 1–19.

Griscom B, **Ashton P. 2006.** A self-perpetuating bamboo disturbance cycle in a neotropical forest. *Journal of Tropical Ecology* **22:** 587–597.

Guillot G, **Rousset F. 2013.** Dismantling the Mantel tests. *Methods in Ecology and Evolution* **4**:336–344.

Gyldenstolpe N. 1941. Preliminary diagnosis of some new birds from Bolivia. *Arkiv för Zoologi* **13:** 1–10.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41:** 95–98.

Harvey E, **Harvey MG**, **Singhal S**, **Rabosky DL. 2019.** Beyond reproductive isolation : demographic controls on the speciation process. *Annual Review of Ecology, Evolution and Systematics* **50:** 75–95.

Hasegawa M, **Kishino H**, **Yano T aki. 1985.** Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22:** 160–174.

Hayakawa EH, **Rossetti DF. 2015.** Late quaternary dynamics in the Madeira river basin, southern Amazonia (Brazil), as revealed by paleomorphological analysis. *Anais da Academia Brasileira de Ciências* **87:** 29–50.

Head MJ, **Gibbard PL. 2005.** Early – middle Pleistocene transitions: an overview and recommendation for the defining boundary. *Geological Society of London* **19:** 1–18.

Heled J, **Drummond AJ. 2008.** Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology* **8:** 1–15.

Heled J, **Drummond AJ. 2010.** Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* **27:** 570–80.

Hey J. 2010. Isolation with migration models for more than two populations. *Molecular Biology and Evolution* **27:** 905–920.

Hey J, **Nielsen R. 2004.** Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167:** 747–760.

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

Hijmans RJ, **Etten J. 2012.** *raster: Geographic analysis and modeling with raster data. R package version 2.0–12*. Available at: https://CRAN.R-project.org/package=raster.

Huson DH, **Bryant D. 2006.** Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23:** 254–67.

Koch AL. 1966. The logarithm in biology 1. Mechanisms generating the log-normal distribution exactly. *Journal of Theoretical Biology* **12:** 276–290.

Leaché AD, **Harris RB**, **Rannala B**, **Yang Z. 2014.** The influence of gene flow on species tree estimation: a simulation study. *Systematic Biology* **63:** 17–30.

Legendre P, **Legendre L. 2012.** *Numerical ecology*. Amsterdam: Elsevier.

Maddison W, **Knowles L. 2006.** Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology* **55:** 21–30.

Marcer A, **Méndez-Vigo B**, **Alonso-Blanco C**, **Picó X. 2016.** Tackling intraspecific genetic structure in distribution models better reflects species geographical range. *Ecology and Evolution* **6:** 2084–2097.

Matzke NJ. 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography* **5**. Available at: https://escholarship.org/uc/item/44j7n141.

Matzke NJ. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in Island Clades. *Systematic Biology* **63:** 951–970.

Mayr E. 1942. *Systematics and the origin of species*. New York: Columbia University Press.

Mayr E. 1982. *The growth of biological thought: diversity, evolution, and inheritance*. Cambridge: Harvard University Press.

McMichael CH, **Palace MW**, **Golightly M. 2014.** Bamboo-dominated forests and pre-Columbian earthwork formations in South-Western Amazonia. *Journal of Biogegraphy* **41:** 1733–1745.

Miranda L, **Aleixo A**, **Whitney BM**, **Silveira LF**, **Guilherrne E**, **Dantas MP**, **Schneider MPC. 2013.** Molecular systematics and taxonomic revision of the Iheringi' s Antwren complex (*Myrmotherula iheringi*: Thamnophilidae), with description of a new species from southwestern Amazonia. In: Del Hoyo J, Elliot A, Christie D, eds. *Handbook of the birds of the world special volume*. Barcelona: Lynx Edicions, 268–271.

Moore RP, **Robinson WD**, **Lovette IJ**, **Robinson TR. 2008.** Experimental evidence for extreme dispersal limitation in tropical forest birds. *Ecology Letters* **11:** 960–968.

Nabholz B, **Lanfear R**, **Fuchs J. 2016.** Body mass-corrected molecular rate for bird mitochondrial DNA. *Molecular Ecology* **25:** 4438–4449.

Naka LN, **Brumfield RT. 2018.** The dual role of Amazonian rivers in the generation and maintenance of avian diversity. *Science Advances* **4:** eaar8575.

Nichols R. 2001. Gene trees and species trees are not the same. *Trends in Ecology & Evolution* **16:** 358–364.

Nielsen R, **Wakeley J. 2001.** Distinguishing migration from isolation: a Markov Chain Monte Carlo approach. *Genetics* **158:** 885–896.

Olivier J, **Otto T**, **Roddaz M**, **Antoine PO**, **Londoño X**, **Clark LG. 2009.** First macrofossil evidence of a pre-Holocene thorny bamboo cf. *Guadua* (Poaceae: Bambusoideae: Bambuseae: Guaduinae) in south-western Amazonia (Madre de Dios — Peru). *Review of Palaeobotany and Palynology* **153:** 1–7.

Ottenburghs J. 2019. Multispecies hybridization in birds. *Avian Research* **10:** 20.

Ottenburghs J, **Kraus RHS**, **van Hooft P**, **van Wieren SE**, **Ydenberg RC**, **Prins HHT. 2017.** Avian introgression in the genomic era. *Avian Research* **8:** 30.

Otto-Bliesner BL, **Marshall SJ**, **Overpeck JT**, **Miller GH**, **Hu A. 2006.** Simulating arctic climate warmth and icefield retreat in the last interglaciation. *Science* **311:** 1751–1753.

Patané JSL, **Weckstein JD**, **Aleixo A**, **Bates JM. 2009.** Evolutionary history of Ramphastos toucans: molecular phylogenetics, temporal diversification, and biogeography. *Molecular Phylogenetics and Evolution* **53:** 923–934.

Peters J. 1951. *Check-list of birds of the world*, *Vol. 7*. Cambridge: Museum of Comparative Zoology.

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

Phillipsen IC, **Kirk EH**, **Bogan MT**, **Mims MC**, **Olden JD**, **Lytle DA. 2014.** Dispersal ability and habitat requirements determine landscape-level genetic patterns in desert aquatic insects. *Molecular Ecology* **24:** 54–69.

Pierpont N, **Fitzpatrick JW. 1983.** Specific status and behavior of *Cymbilaimus sanctaemariae*, the Bamboo Antshrike, from southwestern Amazonia. *The Auk* **100:** 645–652.

Pinho C, **Hey J. 2010.** Divergence with gene flow: models and data. *Annual Review of Ecology, Evolution, and Systematics* **41:** 215–230.

Posada D. 2008. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* **25:** 1253–1256.

Pulido-Santacruz P, **Aleixo A**, **Weir JT. 2018.** Morphologically cryptic amazonian bird species pairs exhibit strong postzygotic reproductive isolation. *Proceedings of the Royal Society B: Biological Sciences* **285:** 20172081.

Pulido-Santacruz P, **Aleixo A**, **Weir JT. 2020.** Genomic data reveal a protracted window of introgression during the diversification of a Neotropical woodcreeper radiation. *Evolution* **74:** 842–858.

Ree RH, **Sanmartín I. 2018.** Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection. *Journal of Biogeography* **45:** 741–749.

Remsen JV, **Areta JI**, **Bonaccorso E**, **Claramunt S**, **Jaramillo A**, **Pacheco JF**, **Robbins MB**, **Stiles FG**, **Stotz DF**, **Zimmer KJ. 2020.** *A classification of the bird species of South America, version 1.* American Ornithological Society. Available at: http://www.museum.lsu.edu/~Remsen/SACCBaseline.htm.

Ribas CC, **Aleixo A**, **Nogueira ACR**, **Miyaki CY**, **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:** 681–689.

Roddaz M, **Hermoza W**, **Mora A**, **Baby P**, **Parra M**, **Christophoul F**, **Brusset S**, **Espurt N. 2010.** Cenozoic sedimentary evolution of the Amazonian foreland basin system. In: Hoorn C, Wesselingh FP, eds. *Amazonia: landscape and species evolution – a look into the past*. Oxford: Blackwell Publishing Ltd, 61–88.

Rozas J, **Ferrer-Mata A**, **Sanchez-DelBarrio JC**, **Guirao-Rico S**, **Librado P**, **Ramos-Onsins SE**, **Sanchez-Gracia A. 2017.** DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution* **34:** 3299–3302.

Sambrook J, **Fritschi E**, **Maniatis T. 1989.** *Molecular cloning: a laboratory manual*. New York: Cold Spring Harbor Laboratory Press.

Sangster G. 2018. Integrative taxonomy of birds: the nature and delimitation of species. In: Tietze D, eds. *Bird species. Fascinating life sciences*. Cham: Springer. Available at: https://doi.org/10.1007/978-3-319-91689-7_2.

Sato H, **Cowling SA. 2017.** Glacial Amazonia at the canopy-scale: using a biophysical model to understand forest robustness. *Quaternary Science Reviews* **171:** 38–47.

Schaefer CER, **do Vale JF Jr**. **1997.** Mudanças climáticas e evolução da paisagem em Roraima: uma resenha do Cretáceo ao recente. In: Barbosa R, Ferreira E, Castellón E, eds. *Homem, ambiente e ecologia no estado de Roraima*. Manaus: INPA, 231–265.

Silva SM, **Peterson AT**, **Carneiro L**, **Burlamaqui TC**, **Ribas CC**, **Sousa-Neves T**, **Miranda LS**, **Fernandes AM**, **Dhorta FM**, **Araujo-Silva LE**, **Batista R**, **Bandeira CHMM**, **Dantas SM**, **Ferreira M**, **Martins DM**, **Oliveira J**, **Rocha TC**, **Sardelli CA**, **Thom G**, **Rêgo PS**, **Santos MP**, **Sequeira F**, **Vallinoto M**, **Aleixo A. 2019.** A dynamic continental moisture gradient drove Amazonian bird diversification. *Science Advances* **5:** eaat5752.

Smith BT, **Harvey MG**, **Faircloth BC**, **Glenn TC**, **Brumfield RT. 2014a.** Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Systematic Biology* **63:** 83–95.

Smith BT, **Klicka J. 2010.** The profound influence of the Late Pliocene Panamanian uplift on the exchange, diversification, and distribution of New World birds. *Ecography* **33:** 333–342.

Smith BT, **McCormack JE**, **Cuervo AM**, **Hickerson MJ**, **Aleixo A**, **Cadena CD**, **Pérez-Emán J**, **Burney CW**, **Xie X**, **Harvey MG**, **Faircloth BC**, **Glenn TC**, **Derryberry EP**, **Prejean J**, **Fields S**, **Brumfield RT. 2014b.** The drivers of tropical speciation. *Nature* **515:** 406–409.

Soares EAA, **Tatumi SH**, **Riccomini C. 2010.** OSL age determinations of Pleistocene fluvial deposits in Central Amazonia. *Anais da Academia Brasileira de Ciencias* **82:** 691–699.

Solís-Lemus C, **Bastide P**, **Ané C. 2017.** PhyloNetworks: a package for phylogenetic networks. *Molecular Biology and Evolution* **34:** 3292–3298.

Sousa VC, **Grelaud A**, **Hey J. 2011.** On the nonidentifiability of migration time estimates in isolation with migration models. *Molecular Ecology* **20:** 3956–3962.

Stephens M, **Scheet P. 2005.** Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *American Journal of Human Genetics* **76:** 449–462.

Stephens M, **Smith NJ**, **Donnelly P. 2001.** A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68:** 978–89.

Stotz DF, **Fitzpatrick JW**, **Parker TA**, **Moskovits DK. 1996.** *Neotropical birds: ecology and conservation*. Chicago: University of Chicago Press.

Strasburg JL, **Rieseberg LH. 2010.** How robust are 'isolation with migration' analyses to violations of the im model? A simulation study. *Molecular Biology and Evolution* **27:** 297–310.

Sukumaran J. 2015. *Model-based biogeographical analyses: principles and practice*. Available at: https://github.com/jeetsukumaran/2015-SSB-AnnArbor-Biogeography.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105:** 437–460.

Tajima F. 1989. The effect of change in population size on DNA polymorphism. *Genetics* **123:** 597–601.

Than C, **Ruths D**, **Nakhleh L. 2008.** PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* **9:** 322.

Thom G, **Amaral FR Do**, **Hickerson MJ**, **Aleixo A**, **Araujo-Silva LE**, **Ribas CC**, **Choueri E**, **Miyaki CY. 2018.** Phenotypic and genetic structure support gene flow generating gene tree discordances in an Amazonian floodplain endemic species. *Systematic Biology* **67:** 700–718.

Thompson JD, **Higgins DG**, **Gibson TJ. 1994.** Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22:** 4673–4680.

Thuiller W, **Lafourcade B**, **Engler R**, **Araújo MB. 2009.** BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography* **32:** 369–373.

Tobias JA, **Aben J**, **Brumfield RT**, **Derryberry EP**, **Halfwerk W**, **Slabbekoorn H**, **Seddon N. 2010.** Song divergence by sensory drive in Amazonian birds. *Evolution* **64:** 2820–2839.

Weir JT, **Faccio MS**, **Pulido-Santacruz P**, **Barrera-Guzmán AO**, **Aleixo A. 2015.** Hybridization in headwater regions, and the role of rivers as drivers of speciation in Amazonian birds. *Evolution* **69:** 1823–1834.

Weir JT, **Price TD. 2019.** Song playbacks demonstrate slower evolution of song discrimination in birds from Amazonia than from temperate North America. *PLoS Biology* **17:** e3000478.

Weir JT, **Schluter D. 2008.** Calibrating the avian molecular clock. *Molecular Ecology* **17:** 2321–2328.

Wu M, **Lan S**, **Cai B**, **Chen S**, **Chen H**, **Zhou S. 2015.** The complete chloroplast genome of *Guadua angustifolia* and comparative analyses of Neotropical-Paleotropical Bamboos. *PLoS One* **10:** e0143792.

Zachos FE. 2018. Mammals and meaningful taxonomic units: the debate about species concepts and conservation. *Mammal Review* **48:** 153–159.

Zimmer JT. 1932. *Studies of Peruvian birds. 8, The formicarian genera* Cymbilaimus, Thamnistes, Terenura, Percnostola, Formicarius, Chamaeza*, and* Rhegmatorhina. American Museum Novitates 20.

Zimmer K, **Isler ML. 2020a.** Bamboo Antshrike (*Cymbilaimus sanctaemariae*), version 1.0. In: Del Hoyo J, Elliot A, Christie D, eds. *Birds of the world*. Ithaca: Cornell Lab of Ornithology. Available at: https://doi.org/10.2173/bow.bamant1.01.

Zimmer K, **Isler ML. 2020b.** Fasciated Antshrike (*Cymbilaimus lineatus*), version 1.0. In: Del Hoyo J, Elliot A, Christie D, eds. *Birds of the world*. Ithaca: Cornell Lab of Ornithology. Available at: https://doi.org/10.2173/bow.fasant1.01.

Zink RM. 2004. The role of subspecies in obscuring avian biological diversity and misleading conservation policy. *Proceedings of the Royal Society B: Biological Sciences* **271:** 561–564.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Tissue samples analysed, collection, voucher numbers, localities (map local column is the same as in Fig. 1) and GenBank accession numbers.
Table S2. List of loci and primers used in the study.
Table S3. Evolutionary models and parameter values for each locus.
Table S4. Georeferenced points from Global Biodiversity Information Facility (GBIF) data used in the species distribution models.
Table S5. Summary of population genetic indices for each phylogoup. N: number of sequences; s: number of polymorphic sites; π: nucleotide diversity; #hap: number of haplotypes; Hd: haplotype diversity; θ: population size parameter (=2 Mu) and 95% confidence interval in brackets; *Fs*: Fu's *Fs* test; *D*: Tajima's *D* test, bold numbers are significant.
Table S6. BioGeoBEARS model selection. nparam: number of parameters; LnL: mean value of log-likelihood estimated from 1000 trees; sd: standard deviation of log-likelihood; AICc: corrected Akaike Information Criterion, based on mean value of log-likelihood; AICc_wt: AICc weights; mean values and standard deviations of dispersal (d), extinction (e) and founder (j) parameters.
Table S7. Mantel test correlation scores (and *P*-values) between genetic distance and geographical or climatic distances. A, all *C. lineatus* phylogroups and *C. sanctaemariae*; B, all *C. lineatus* phylogroups; C, *C. lineatus cis*-Andean Amazon phylogroups (i.e. Central America excluded).
Table S8. Values obtained from species distribution model performance tests. True Skill Statistics (TSS) and the relative contributions of each environmental variable for the models.
Figure S1. *BEAST species tree with the results of BioGeoBEARS under the DIVALIKE model. Pie charts at nodes represent probabilities of the ancestral distributions, with areas coded by colours in the pie chart (see map and legend with colored squares on the left).
Figure S2. Scatter plots of genetic distance vs. geographical or climatic distance for pairwise population comparisons. mtDNA and Z-linked loci.
Figure S3. Scatter plots of genetic distance vs. geographical or climatic distance for pairwise population comparisons. nDNA loci.
Figure S4. Posterior distribution of migration estimates from IMa2 analyses. Upper panels, significant values; lower panels, non-significant values.
Figure S5. Posterior distribution of population size estimates from IMa2 analyses.
Figure S6. Posterior distribution of split time estimates from IMa2 analyses.

SHARED DATA

All main raw data, including input and output files, trees and R codes are available on GitHub: https://github.com/miralaba/miranda_et_al_2020