Phenotypic and Genetic Structure Support Gene Flow Generating Gene Tree Discordances in an Amazonian Floodplain Endemic Species

GREGORY THOM[1,*], FABIO RAPOSO DO AMARAL[2], MICHAEL J. HICKERSON[3], ALEXANDRE ALEIXO[4],
LUCAS E. ARAUJO-SILVA[4], CAMILA C. RIBAS[5], ERIK CHOUERI[5], AND CRISTINA Y. MIYAKI[1]

[1]*Departamento de Genética e Biologia Evolutiva, Universidade de São Paulo, Rua do Matão, 277, Cidade Universitária, São Paulo, SP 05508-090, Brazil;*
[2]*Instituto de Ciências Ambientais, Químicas e Farmacêuticas, Laboratório de Genética Evolutiva, Universidade Federal de São Paulo, Rua Professor Artur Riedel, 275, Diadema, SP 09972–270, Brazil;* [3]*Department of Biology, Marshak Science Building, City College of New York, 160, Convent Avenue, 10031 New York, NY, USA;* [4]*Departamento de Ornitologia, Museu Paraense Emílio Goeldi (MPEG), Caixa Postal 399, Belém, PA 66040-170, Brazil; and*
[5]*Coordenação de biodiversidade, Instituto Nacional de Pesquisas da Amazônia (INPA), Av. André Araújo 2936, Manaus, AM 69060-001, Brazil*
Correspondence to be sent to: Departamento de Genética e Biologia Evolutiva, Universidade de São Paulo, Rua do Matão, 277,
sala 230 Cidade Universitária, São Paulo, SP 05508-090, Brazil;
E-mail: biogrego@yahoo.com.br.

Abstract.—Before populations become independent evolutionary lineages, the effects of micro evolutionary processes tend to generate complex scenarios of diversification that may affect phylogenetic reconstruction. Not accounting for gene flow in species tree estimates can directly impact topology, effective population sizes and branch lengths, and the resulting estimation errors are still poorly understood in wild populations. In this study, we used an integrative approach, including sequence capture of ultra-conserved elements (UCEs), mtDNA Sanger sequencing and morphological data to investigate species limits and phylogenetic relationships in face of gene flow in an Amazonian endemic species (*Myrmoborus lugubris*: Aves). We used commonly implemented species tree and model-based approaches to understand the potential effects of gene flow in phylogenetic reconstructions. The genetic structure observed was congruent with the four recognized subspecies of *M. lugubris.* Morphological and UCEs data supported the presence of a wide hybrid zone between *M. l. femininus* from the Madeira river and *M. l. lugubris* from the Middle and lower Amazon river, which were recovered as sister taxa by species tree methods. When fitting gene flow into simulated demographic models with different topologies, the best-fit model indicated these two taxa as non-sister lineages, a finding that is in agreement with the results of mitochondrial and morphological analyses. Our results demonstrated that failing to account for gene flow when estimating phylogenies at shallow divergence levels can generate topological uncertainty, which can nevertheless be statistically well supported, and that model testing approaches using simulated data can be useful tools to test alternative phylogenetic hypotheses. [Antbird; hybridization; migration; multispecies coalescent; *Myrmoborus lugubris*.]

In an evolutionary perspective, the speciation process is concluded when genetic differentiation promotes complete reproductive isolation between diverging populations (Harrison 1990). However, before populations become independent evolutionary lineages, the effects of micro evolutionary processes tend to generate complex scenarios of diversification that may affect phylogenetic reconstruction (Hickerson et al. 2006; Leaché et al. 2014a,b; Poelstra et al. 2014; Nater et al. 2015; Edwards et al. 2016; Mallet et al. 2016; Barley et al. 2017).

Incomplete lineage sorting (ILS) and gene flow between taxa are commonly reported in studies involving recent diversification scenarios in the gray zone of phylogeography (Edwards et al. 2016; Meyer et al. 2016; Oswald et al. 2017). Around 10% of animal and 25% of plant taxa exchange genes with other taxa (Mallet 2005, 2007). In phylogenetic reconstructions the effects of ILS and gene flow can be misleading, since both tend to produce inconsistencies between gene trees and the species tree and the outcomes of each process are often difficult to be distinguished and modeled by the coalescent process (Kubatko and Degnan 2007; Kubatko 2009).

The multispecies coalescent (MSC) methods have become one of the most used approaches to perform species tree estimation by using a statistically standardized methodology to test alternative hypotheses of isolation (Liu et al. 2009a; Leaché and Fujita 2010; Bryant et al. 2012; Leaché et al. 2014a,b). Such methods can incorporate the uncertainty of gene trees given the coalescent stochasticity as well as parameters (e.g., effective population sizes and times of divergence) during the estimation of the species tree (Yang and Rannala 2010; Zhang et al. 2014). However, gene flow is not taken into account by most of the species tree methods available (Liu et al. 2009a; Liu and Yu 2011; Bryant et al. 2012; Sousa and Hey 2013; Mirarab et al. 2014), whereby ILS is assumed as the only source of gene tree discordance in a given dataset.

The presence of gene flow between taxa violates the assumption of bifurcating branches arising from a common ancestor, and not properly accounting for it in species tree estimates can directly impact parameter inferences such as topology, effective population sizes, and branch lengths (Eckert and Carstens 2008; Leaché et al. 2014a,b; Solís-Lemus et al. 2016). Notwithstanding the fact that some species tree methods have proven to be robust in the presence of low levels of gene flow, such as ASTRAL and NJst (Solís-Lemus et al. 2016), the explicit accommodation of migration in phylogenetic estimations is critical, specially due to the large amount of recent evidence for reticulated evolution

(Lamichhaney et al. 2015; Mallet et al. 2016; Edwards et al. 2016), including cases of speciation with gene flow (Feder et al. 2012), ephemeral speciation (Rosenblum et al. 2012), and hybrid speciation (Mallet 2007).

Temporally dynamic environments, such as areas affected by climatic oscillations during the Quaternary when patches of specific habitats were connected and disconnected as predicted by the refugia hypothesis (Haffer 1969), or riverine environments constantly changed by river dynamics (Bagley et al. 2011), are prone to produce complex scenarios of diversification. The combination of recent history of divergence and dynamic secondary contact between populations can potentially result on paraphyly and introgression of diverging lineages (Hewitt 2004). This scenario demands the application of methods that explicitly model ILS and gene flow to explore phylogenetic relationships among populations or species (Nater et al. 2015; Morales et al. 2017).

The Amazon Basin comprises the largest fluvial system in the world and its floodplains extend for over 550 000 km^2, with the major tributaries representing approximately 8% of the area of this biome (Wittmann et al. 2010; Junk et al. 2011). Floodplain environments in the Amazon are strongly influenced by annual flooding cycles and were historically shaped by paleoclimatic changes affecting the river dynamics, yielding the formation of a plethora of specific and ephemeral habitats heterogeneously distributed over the basin (Irion et al. 2009; Junk et al. 2011; Wittmann et al. 2012; Rossetti et al. 2015). As a result, a high level of species endemicity is reported, such as 10% of tree species and 15% of non-aquatic bird species (Remsen and Parcker 1983; Wittmann et al. 2012). The high levels of endemicity contrast with the presence of widely distributed species with low levels of genetic diversity suggesting that the linear distribution and connectivity of floodplains enable high levels of gene flow or that the ephemeral environments could be prone to high extinction rates (Aleixo 2006; Cadena et al. 2011; but see Choueri et al. 2017).

The Ash-breasted Antbird (*Myrmoborus lugubris*) inhabits the floodplain forests of large Amazonian rivers, being restricted to a narrow strip of river edge forests, mainly found on river-created islands (Zimmer and Isler 2003). *M. lugubris* as currently recognized, is a single species with four subspecies that are recognized based on minor to moderate plumage differences: *M. l. lugubris* occurs in the east of the Madeira river to the Amazon river mouth; *M. l. berlepschi* is restricted to the Solimões river; *M. l. femininus* is restricted to the Madeira river; and *M. l. stictopterus* occurs in the Negro river (Fig. 1a). Despite the apparent continuity of river created environments along the entire extension of the main channel of the Amazon river, there is no evidence—historical or recent records—that any taxa of this group occurs in the lower portion of the Solimões and Negro rivers (Fig. 1a; Petermann 1997; Cohn-Haft et al. 2007). It is worth mentioning that these are thoroughly surveyed areas by ornithologists and birdwatchers. Thus, there is strong evidence that *M. l. berlepschi* and *M. l. stictopterus* are allopatric taxa. Haffer and Fitzpatrick (1985) suggested that the central Amazonian taxa (*M. l. femininus* and *M. l. stictopterus*) could be intermediate forms between the two extreme plumage types (*M. l. berlepschi* and *M. l. lugubris*), as a result of secondary contact and intergradation between the western black-faced (*M. l. berlepschi*) and eastern reddish-faced (*M. l. lugubris*) taxa. However, no study has been conducted so far to describe the relationships among these taxa nor whether phenotypic variation mirrors neutral genetic structure. Thus, *M. lugubris* is an interesting model to test the robustness of distinct phylogenetic inference methods in face of a complex scenario of diversification that may present ILS and gene flow.

To test this, we used a model-based integrative approach, that combined sequence capture of ultra-conserved elements (UCEs), mtDNA Sanger sequencing and morphological data to investigate the phylogenetic relationships in the Amazon endemic Ash-breasted Antbird complex. We applied empirical and model-based approaches to explore hypotheses regarding species tree estimation. In addition, we tested the robustness of typically used species tree methods in the presence of gene flow between non-sister taxa. Our results indicated that species tree methods, including full likelihood and summary-based approaches, which do not account for migration, supported distinct topologies or did not have phylogenetic resolution. Our results supported that phylogenetic network methods and model-testing approaches based on simulation with thousands of SNPs agreed with the mtDNA phylogeny and phenotypic analyses. Also, diversification occurred around the Middle and Late Pleistocene and secondary contact between non-sister species seem to be related to the dynamics of Amazon floodplains.

MATERIALS AND METHODS

Morphological Analyses

We examined 160 skins belonging to all taxa of *Myrmoborus lugubris* (*M. l. lugubris*, *M. l. berlepschi*, *M. l. femininus*, and *M. l. stictopterus*) housed at various institutions (see Supplementary Table S1 available on Dryad at http://dx.doi.org/10.5061/dryad.13b88). The phenotypic variation among *M. lugubris* subspecies is a case of heterogynism (Hellmayr 1929), in which female plumage from different taxa differ more conspicuously than that of males. In general, the front and sides of the head are light ferruginous in females from the lower Amazon river (*M. l. lugubris*), whereas those body parts are black or blackish in western Amazonian subspecies (*M. l. berlepschii, M. l. femininus, M. l. stictopterus*; Zimmer and Isler 2003). Plumage variation was described observing discrete characters over the distribution of the species and comparing with the original description of each taxon. We used
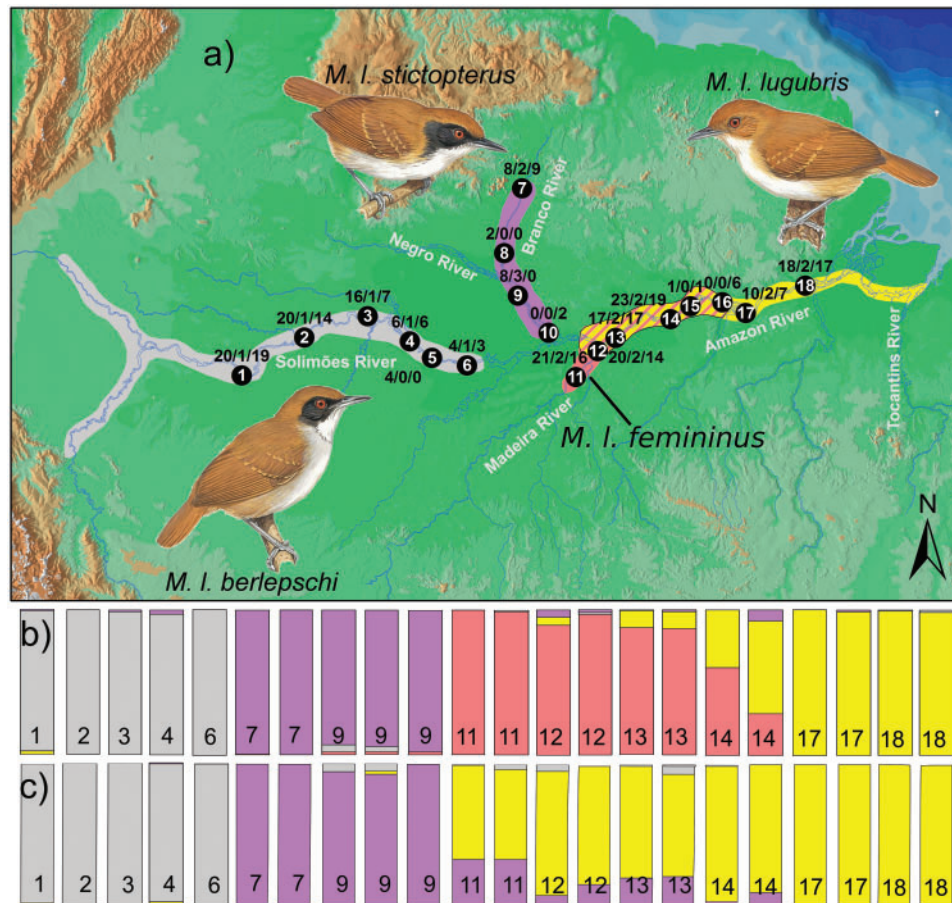
FIGURE 1. a) Map of sampling localities (black circles, see Supplementary Table S1 available on Dryad) of the *Myrmoborus lugubris* species complex and numbers of samples used for mtDNA, sequence capture of UCEs and morphological analyses, respectively. Colors represent the geographic distribution of the four subspecies: gray—*M. l. berlepschi*, purple—*M. l. stictopterus*, pink—*M. l. femininus*, and yellow—*M. l. lugubris*. b) Model with the best value of cross-entropy ($K = 4$) for the population genetic structure inferred in sNMF based on 1664 SNPs. Numbers in the bars correspond to the localities in the map. c) Model with the second best value of cross-entropy ($K = 3$) for the population genetic structure inferred in sNMF. Illustrations retrieved from del Hoyo et al. (2017) (Handbook of the birds of the world alive).

Smithe (1975, 1981) color catalog to describe color variation and plumage tones. Phenotypically intermediate individuals between *M. l. lugubris* and *M. l. femininus* included females from a contact zone between these taxa with a mixed combination of plumage characters diagnostic of each taxa, whereas males from the same localities were assigned by default as intermediates as well. Measurements of the following morphometric characters were taken by G.T. to the nearest 0.1 mm with an Mitutoyo electronic caliper: wing length (from the wrist to the tip of the longest primary), tail length (from the pygostyle insertion to the tip of the longest rectrices), tarsus length, culmen, bill length from the distal points of the nostril to the tip of the bill, bill depth and width at the distal point of the nostril. Wing and tarsus measurements were always taken from the specimens' right side. The normality of the data was assessed with the Kolmogorov–Smirnov tests. We used discriminant function analysis (DFA) to test for differences in the morphometric space among recognized taxa of *M. lugubris*. Missing morphometric

values for any given character were replaced with the corresponding average value obtained for the taxon to which the individual was assigned and did not exceeded >5% of the total number of measurements. Intermediate individuals (based on plumage patterns) between *M. l. lugubris* and *M. l. femininus* were considered as a distinct group. All statistical tests were performed with STATISTICA version 8.0. We adopted a statistical significance of $P \leqslant 0.05$.

mtDNA—DNA Extraction, Sequencing, and Data Analyses

A total of 198 individuals from 25 localities were sampled throughout the entire distribution of *M. lugubris* (see Supplementary Table S1 available on Dryad; Fig. 1a). Subspecific identification was based on the plumage of vouchered individuals. We used *M. leucophrys*, *M. myotherinus*, and *Percnostola lophotes* as outgroups due their close phylogenetic relationship to *M. lugubris* (Isler et al. 2013). Total DNA was

extracted from approximately 20 mg of muscle tissue following a standard phenol/chloroform protocol (Sambrook et al. 1989). We amplified the mitochondrial cytochrome b gene (cyt b, ~1040 bp; Primers L14841: CCATCCAACATCTCAGCATGATGAAA and H16065: AACTGCAGTCATCTCCGGTTTACAAGAC) for all sampled individuals. Polymerase chain reaction (PCR) was performed in 25 µL of final volume, and approximately 50 ng of genomic DNA, 1.5–2.5 mM of MgCl$_2$, 200 mM of dNTPs and 0.1 U of *Taq* DNA polymerase (Promega, Madison, WI, USA). An initial denaturation step was performed at 94°C for 5 min, followed by 35 cycles of: (i) 94°C for 1 min; (ii) 50°C for 1 min; and (iii) 72°C for 1 min. The final extension was at 72°C for 5 min. PCR products were purified with Exo-Fap enzymes. Sequencing was carried out on an ABI 3130 or 3730 automated capillary sequencer (Applied Biosystems, Foster City, CA, USA) with the ABI Prism Big Dye terminator Kit. Both DNA strands were sequenced for all samples. The DNA sequences were edited and aligned in CodonCode aligner 3.7.1 (CodonCode Inc.).

Phylogenetic analyses were performed using the Bayesian inference (BI) in MrBAYES 3.1.2 (Ronquist et al. 2012). The best-fit evolutionary model was selected in JMODELTEST 0.1.1 (Posada 2008), using the Bayesian information criterion (BIC). BI analysis was performed by two independent runs of 10 million generations, each with four Markov chains. The parameters of the chains were sampled every 1000 generations and the first 1000 trees were discarded as burn-in. We used random seed for starting tree and default priors and initial searching values. The posterior probability for each estimated node was obtained through a majority rule consensus of the remaining MCMC samples.

We performed a Bayesian population (*k*) clustering analyses without prior information of the sampling locations and the number of taxa in BAPS 6.0 (Corander and Marttinen 2006), using the option "clustering with linked loci," which accounts for dependence between sites within the aligned sequences. We tested multiple *k* values (1–8) and performed ten independent runs in order to assess the consistency of the results.

Sequence Capture of UCEs

We selected 22 tissue samples of *M. lugubris* representing all four taxa, from 13 localities covering the entire geographic distribution of the species for capture and sequencing UCEs. Given the close phylogenetic relationship to *M. lugubris*, one sample of *M. leucophrys*, and one sample of *M. myotherinus* were used as outgroups (Isler et al. 2013). Genomic DNA of all samples were extracted with QIAGEN DNeasy tissue and Blood kit (Valencia, CA, USA) according to the manufacturer's protocol with a minimum amount of 2 µg of DNA eluted in ~50 µL of AE buffer. All samples were treated with QIAGEN RNAse during extraction.

Library preparation and sequencing of UCEs were performed by RAPiD Genomics (Gainsville, FL, USA),

following most of the protocol described by Faircloth et al. (2012). Modifications of the original protocol concern the use of a probe set containing 2312 probes targeting UCEs (ultraconserved.org) plus 97 probes targeting exons typically used in avian phylogenetic studies (Hackett et al. 2008; Kimball et al. 2009), and sequencing of 150 bp paired-ends in Illumina Hiseq 2500. Samples were sequenced in a multiplexed batch of 96 samples, including individuals not analyzed in this study.

Raw read files were separated per individual using Illumina's Casava software and read quality of each individual was evaluated in FastQC 0.11.4 (Andrews 2014). We filtered raw reads for each individual using Illumiprocessor, which trims off adapter sequences and excludes low-quality reads (Faircloth et al. 2012). We conducted *de novo* assembly across all samples using Trinity 2.4 (Grabherr et al. 2011). Contigs were aligned to UCE probe reference sequences. Contigs that did not align or that matched multiple UCEs were removed with LASTZ using the "match_contigs_to_probes.py" script, available in PHYLUCE 1.4 package (Faircloth et al. 2012; https://github.com/faircloth-lab/phyluce). After generating a SQL database of matches for all individuals, we generated alignments in fasta format and trimmed off long ragged-ends. A matrix of concatenated sequences with 15% of missing data was used for the phylogenetic analyses. For the remaining analyses we generated a sequence matrix with all *M. lugubris* individuals without filtering for missing data nor trimming, which was used as reference for SNP calling (see below). The longest sequence without indels per loci was selected as reference. Sequence alignments for each locus were performed using MAFFT (Katoh and Standley 2013). Finally, additional scripts available in the PHYLUCE package were used to obtain summary statistics of all loci and to convert file formats for phylogenetic analyses.

We aligned the cleaned and synchronized reads of each individual with the generated reference using BWA (Li and Durbin 2009) and called SNPs using the Genome Analyses Tool Kit (GATK; McKenna et al. 2010), hard-masking low-quality bases (<Q30) and keeping sites with a minimum read depth of >8. In order to obtain an unlinked SNP matrix, we randomly selected one SNP per locus, excluding sites with missing data. We also blasted (Blast+) the reference sequence used to call the SNPs against the zebra finch genome (available at https://www.ncbi.nlm.nih.gov/books/NBK279690/) to select for autosomal loci (excluding UCEs in sexual chromosomes). We used VCFTOOLS (Danecek et al. 2011) and custom scripts to generate input files for the programs described below. Finally, to obtain phased alleles for each locus we used the function "ReadBackedPhasing" of GATK (McKenna et al. 2010) on the final VCF with a phasing quality threshold of 20. Phased alleles were incorporated to the reference sequences using the "add_phased_snps_to_seqs_filter.py" from the seqcap_pop package (https://github.com/mgharvey/seqcap_pop; Harvey et al. 2016). Sequence alignments

for each locus were performed using MAFFT (Katoh and Standley 2013). No threshold for missing data was applied.

UCEs: Genetic Structure

We used sNMF (Frichot et al. 2014) to infer the best-fit number of ancestral populations (*k*) within *M. lugubris* and to assign individuals to populations. The sNMF algorithm applies a sparse non-negative matrix factorization to compute least-squares estimates of ancestry coefficients. When compared with more commonly used programs, such as STRUCTURE (Pritchard et al. 2000) and ADMIXTURE (Alexander et al. 2009), sNMF can analyze large bi-allelic datasets more efficiently without loss of accuracy. Also, it seems to perform well even in scenarios of departure from the Hardy Weinberg and linkage equilibrium assumptions (Frichot et al. 2014; Harris et al. 2015).

We tested multiple *k* values (1–8), with 100 replicate runs for each *k* value. The robustness of the results was assessed by testing four values of the alpha regularization parameter (1, 10, 100, 1000). Additionally, we used the *k*-means (find.clusters) algorithm available in the package ADEGENET 2.0 (Jombart and Ahmed 2011) to test for the number of genetic clusters in the sample. *K*-means reduces the number of variables (in our case, SNPs) through a principal component analyses (PCA), maximizes the variation between groups and indicates the optimal number of groups by comparing different clustering solutions with BIC. In order to avoid information loss, all components of the PCA were retained. To check the concordance between both clustering methods we plotted the sNMF results as a function of the *k*-means classification. Clusters obtained in *k*-means analysis were graphically described using a discriminant analysis of principal components (DAPCs; Jombart et al. 2010) available in ADEGENET 2.0 (Jombart and Ahmed 2011).

UCEs: Phylogenetic Estimations and Species Tree

We analyzed the concatenated loci data set using two phylogenetic methods. First, we performed a maximum likelihood (ML) search in RAxML v8.1.18 (Stamatakis 2014) with the GTR+G model of sequence evolution. Node support values were accessed based on 1000 rapid bootstrap replicates. Second, we implemented a BI analysis in MrBayes v3.2 (Ronquist et al. 2012). We performed two independent runs of 2 million generations, each with four chains, sampling every 1000 generations, excluding the first 15% of generations as burn-in. Best-fit models of sequence evolution were determined in CloudForest (github.com/ngcrawford/CloudForest). We used random seed for starting tree and default priors and initial search values.

We also used species tree methods under the MSC framework. To test the robustness of alternative species tree methods in face of potential gene flow among *M. lugubris* taxa, we implemented three methods that sample individual gene trees to infer the species tree under the coalescent: (i) maximum pseudo-likelihood to estimating species tree (MP-EST; Liu et al. 2010); (ii) species tree from average ranks of coalescence (STAR; Liu et al. 2009b); and (iii) neighbor joining species tree (NJst; Liu and Yu 2011), using the STRAW online platform (Shaw et al. 2013; http://bioinformatics.publichealth.uga.edu/SpeciesTreeAnalysis/index.php). We identified the best-fit molecular substitution model and estimated ML gene trees and 100 bootstrap replicates for each gene tree based on sequences of phased alleles with Cloudforest, which runs the Phyml (Guindon et al. 2010) in background (github.com/ngcrawford/CloudForest). The node support values were obtained by calculating an extended majority-rule consensus tree for the 100 species tree obtained by each of the three methods. Despite the potential ascertainment bias introduced by sub sampling loci (Knowles 2010), previous studies suggested that informative loci with limited variation—as typically observed in UCEs datasets—can negatively affect results of summary-based species tree methods, given the assumption of bifurcating trees (Lanier et al. 2014; Hosner et al. 2015; Manthey et al. 2016). Thus for these analyses we selected loci with more than five informative sites that represent approximately 25% of the most variable loci, which is assumed to be a good threshold following Hosner et al. (2015). Given that the inclusion of hybrid individuals explicitly violates the MSC model, for all three methods we performed analyses with and without individuals that are morphological intermediates between *M. l. femininus* and *M. l. lugubris*.

Comparatively, we implemented the full likelihood MSC model of Bayesian phylogenetics and phylogeography (BP&P; Yang and Rannala 2010; Rannala and Yang 2013; Yang and Rannala 2014) to estimate species tree and the probability of speciation at each node, testing our hypothesized taxa, with and without intermediate individuals between *M. l. femininus* and *M. l. lugubris*. In BP&P we ran 2×10^6 generations sampling every five generations with a burn-in of 10%. To estimate divergence time (Tau) and population size (Theta) we set a gamma distribution $[G(\alpha,\beta)]$ as $G(2,1000)$, with the prior mean $= \alpha/\beta$; and prior variance $= \alpha/\beta$. All analyses were performed twice to check for consistency between independent runs. As for the summary-based species tree methods, we used the data set containing loci with five or more informative sites.

Additionally, we applied the MSC method implemented in SNAPP (Bryant et al. 2012) that is executable in BEAST v.2 (Bouckaert et al. 2014). SNAPP infers a likelihood species tree using allele frequency of unlinked SNPs bypassing the need to integrate the probabilities of gene trees as a function of a given species tree. We used gamma rate priors for alpha and beta parameters, with all other priors with default values. Two replicates of 2,5 million MCMC generations were run with 100 000 burn-in interactions. Estimated parameters were sampled every 500 generations.

TABLE 1. Bayes factor delimitation (BFD) results for alternative arrangements of individuals for the four *Myrmoborus lugubris* subspecies

Model	Rank	ML	BF	Ln(BF)
(be),(st),(fe),(lu)	1	−1770.50	***	***
(be),(st),(fe,lu)	3	−18069.44	735.861380	6.60
(be),(st,fe),(lu)	2	−18044.73	686.447462	6.53

Clusters of individuals are represented between parentheses; be = *M. l. berlepschi*; st = *M. l. stictopterus*; fe = *M. l. femininus*; lu = *M. l. lugubris*. *** = not applicable.

TABLE 2. Bayes factor delimitation (BFD) results for alternative arrangements of individuals for the four *Myrmoborus lugubris* subspecies and morphologically intermediate individuals between *M. l. lugubris* and *M l. femininus*

Model	Rank	ML	BF	Ln(BF)
(be),(st),(fe),(hy),(lu)	1	−22095.83	***	***
(be),(st),(fe,hy,lu)	4	−22483.43	775.18983	6.65
(be),(st,fe),(hy),(lu)	5	−22484.03	776.394753	6.65
(be,(st,fe),(hy,lu)	6	−22665.59	1139.50468	7.04
(be),(st,fe,hy),(lu)	7	−22689.10	1186.54008	7.08
(be),(st),(fe),(hy,lu)	2	−22281.99	372.312179	5.92
(be),(st),(fe,hy),(lu)	3	−22352.77	513.873197	6.24

Clusters of individuals are represented between parenthesis; be = *M. l. berlepschi*; st = *M. l. stictopterus*; fe = *M. l. femininus*; lu = *M. l. lugubris*; hy = Intermediate individuals between *M. l. lugubris* and *M. l. femininus*. *** = not applicable.

Burn-in values for the MCMC chains were accessed in Tracer v1.4 (Rambaut and Drummond 2007). Finally, using SNAPP we implemented the Bayes factor delimitation (BFD) of species to test which arrangement of individuals as potential species had the best Bayes factor value based on Marginal likelihoods (Leaché et al. 2014a,b). Models tested the potential effects of gene flow in species tree analyses and the results obtained based on mtDNA (Tables 1 and 2). We ran 50 steps of 100 000 MCMC generations with a pre burn-in of 10 000 MCMC generations for each arrangement of individuals. For both methods, SNAPP and BFD, we performed analyses with and without intermediate individuals between *M. l. femininus* and *M. l. lugubris*. To visualize the posterior distribution of sampled species trees we used DENSITREE v2.2 (Bouckaert 2010).

UCEs: Gene Flow and Demographic Models

In order to account for gene flow while inferring the relationships among *M. lugubris* taxa we used TREEMIX 1.12 (Pickrell and Pritchard 2012). This method infers patterns of population splitting and mixing accessing the covariance structure of allele frequencies between populations and performing a Gaussian approximation for genetic drift, producing a ML graph linking sampled populations to their most common ancestor, improving the fit of the inferred model given the observed data by enabling migration edges. We performed the analyses varying from zero to four migration edges assessing the significance of each added migration edge and the residue covariance matrix to access the fit of the model to the data. Additionally, the $f3$ statistic was calculated

using the three-pop test for admixture, available in the TREEMIX package (Reich et al. 2009). This test detects correlations in allele frequencies that are not compatible with population evolution following a bifurcating tree (Reich et al. 2009). All possible clusters of populations were tested.

Additionally, we estimated phylogenetic networks including branch length and inheritance probabilities with the software PhyloNet 3.6.1 (Than et al. 2008). Given the relative large number of terminal taxa and number of sampled individuals we implemented the Maximum Pseudo Likelihood algorithm (Yu and Nakhleh 2015), "InferNetwork_MPL," allowing for none and two reticulations and performing 1000 independent searches to avoid sampling in local optimums, with the other parameters set as default. For this analysis we used the bootstrapped (100 replicates) gene trees of loci with at least five informative sites. To bypass uncertain nodes in gene trees, we applied a bootstrap support threshold of 75 in all PhyloNet analyses with the -b flag.

In order to compare the topologies obtained based on mtDNA and ST/TREEMIX/PhyloNet analyses based on UCEs and to infer the potential effects of gene flow in the diversification of *M. lugubris*, we implemented a coalescent model-based approach using Fastsimcoal2 (Excoffier et al. 2013). Fastsimcoal2 estimates the composite likelihood of a specific scenario of diversification given the observed data as well as population genetic parameters such as divergence time, effective population size and gene flow using the joint site frequency spectrum (jSFS) as input, to summarize the complexity of the data. The jSFS for each pairwise population ($n = 10$; five groups) were generated in $\partial a \partial I$ 1.7 (Gutenkunst et al. 2009). For parameter estimation and model comparison we ran 50 replicates per model retaining the parameters that maximized the composite likelihood across all interactions. Parameter optimization was performed through 50 cycles of the Brent algorithm and the composite likelihood calculated using 100 000 simulations per replicate. Using the replicate that optimized the composite likelihood for each tested model, we calculated the Akaike information criterion (AIC), AIC = $2k$-2ln(L), where k is the number of parameters estimated in the model and L the composite likelihood value. To obtain confidence intervals (CIs), we performed 100 parametric bootstraps using the parameters of the best model to simulate 100 new sets of jSFS and re-estimated these parameters. We performed 30 independent runs for each new simulated data set running 50 cycles of the Brent algorithm and 100 000 simulations. We applied a mutation rate of 2.5×10^{-9} substitution per site per generation (Nadachowska-Brzyska et al. 2015) and assumed a generation time of 2.33 year (Maldonado-Coelho 2012).

RESULTS

Morphology

Plumage variation (see Supplementary SI1 available on Dryad) was congruent with the original taxonomic

descriptions of each taxon, with subtle differences regarding population polymorphisms detected due to the large number of specimens analyzed in this study. In general, male plumage was conserved along the entire distribution of *M. lugubris,* despite a tendency for darker neutral gray individuals in *M. l. berlepschi* (see Supplementary Fig. S1 available on Dryad). The plumage of females was more variable and geographically diagnosable than that of males, enabling the distinction of four groups congruent with the current taxonomy (see Materials and Methods; Supplementary Figs. S2, S3 and SI1 available on Dryad). In general, the front and sides of the head were light ferruginous in females from the lower Amazon river (*M. l. lugubris*), whereas those body parts were black or blackish in western Amazonian subspecies (*M. l. berlepschii, M. l. femininus, M. l. stictopterus;* Zimmer and Isler 2003). Besides, females of *M. l. lugubris* from Almerim (extreme of its distribution) presented throat and central portions of the chest pure white. *M. l. femininus* females from Borba and Autazes (extreme of its distribution) showed white throat with small and subtle black spots (scale) and a subtle black spotted collar delimiting the throat and chest. Females of *M. l. berlepschi* had white throat with small and subtle black spots (scale) as observed in *M. l. femininus* and an evident collar of black spots with variable intensity delimiting the throat and chest. Finally, *M. l. stictopterus* females were very similar to *M. l. femininus* and diagnosable only by the color of the facial mask, which was larger and pure black without signs of ferruginous feathers. Despite the clear diagnosis of the females of *M. l. lugubris* and *M. l. femininus* in the extreme of their distributions (Almeirim for *M. l. lugubris* and Borba for *M. l. femininus*, localities 18 and 11 in Fig. 1a, respectively), these diagnostic characters were blurred in geographically intermediate localities forming a gradual transition between these two forms (Fig. 2).

We detected significant ($P \leqslant 0.05$) differences between sexes in bill width and wing length in all four taxa of *M. lugubris*. To extract the variation in the data not explained by sexual dimorphism, we performed a linear regression of each character measured as a function of sex, retaining the residues of these analyses. The DFA resulted in significant differences in the morphometric space among *M. lugubris* taxa, with the first two canonical discriminant variables accounting for 95% of the total variation (see Supplementary Fig. S4 available on Dryad). The variables that most contributed to this differentiation were tail length, wing length, and exposed culmem (Wilks's lambda = 0.1149, P >0.0001; Supplementary Fig. S4 available on Dryad; Table 3). The classification matrix revealed that 68% (11 specimens) of *M. l. lugubris* individuals were correctly identified, with all five misclassified specimens placed in the morphologically intermediate group (hereafter named as intermediate group). A similar pattern was observed in *M. l. femininus*, with 80% of the individuals (24 specimens) correctly classified and six specimens placed in the intermediate group (Tables 1 and 2). Among the *M. l. stictopterus* individuals, 75%
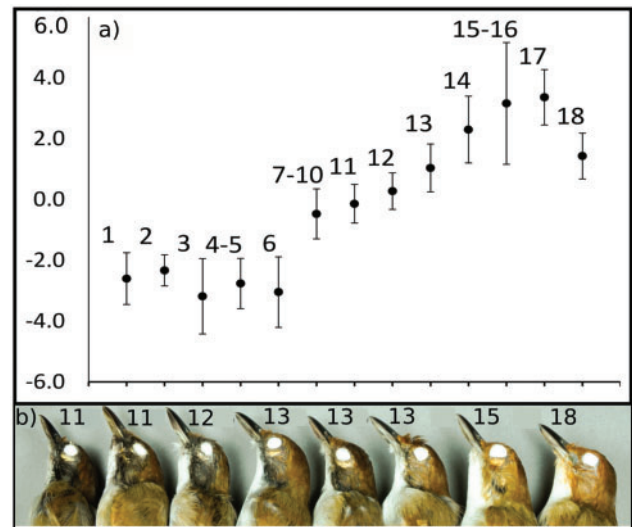

FIGURE 2.   a) Score values of the first canonical axis of the discriminant functional analysis performed on specimens of the four *Myrmoborus lugubris* subspecies and morphologically intermediate individuals between *M. l. lugubris* and *M l. femininus* based on measurements of seven morphometric characters plotted as a function of the 18 localities sampled longitudinally and displayed from west to east (see locality numbers in Fig. 1a. b) Picture of the left side of the head of females collected between localities 11 and 18 (see Fig. 1a) representing the cline in plumage between *M. l. lugubris* and *M l. femininus.*

(6 specimens) were correctly classified and two were misclassified as belonging to the intermediate group. All the *M. l. berlepschi* individuals (100%; 48 specimens) were correctly classified. The intermediate group (53 specimens) had the second highest correct classification index (88.67%), with 47 individuals correctly identified and only six individuals misclassified in other groups (Table 3). Thus, individuals from the intermediate group that were inferred to be hybrids between *M. l. lugubris* and *M. l. femininus* (see below) can be diagnosed by morphometric characters. The geographic distribution (from west to east) of the first canonical axis (that explains 85.4% of the total variation and expresses the size of individuals) and the three measured characters that most contributed to the observed pattern (wing length, tail length, and exposed culmem), revealed that intermediate individuals in plumage are larger than individual with "pure" plumage patterns (Fig. 2, Supplementary Fig. S5 available on Dryad). Additionally, most of the morphometric variables of the intermediate group was statistically different (the Mann–Whitney test) from those of "pure" *M. l. femininus* and *M. l. lugubris* (see Supplementary Table S3 available on Dryad). Considering only the plumage pattern of non-introgressed individuals, *M. l. lugubris* had significantly larger wings, tail, and bill length than any of the other taxa. The opposite occurred with *M. l. berlepschi*, which was significantly smaller for the same characters than the other taxa, with *M. l. stictopterus* and *M. l. femininus* widely overlapping in measurements of these characters (see Supplementary Tables S2 and S3 available on Dryad).

TABLE 3.   Percentage (% correct) and number of individuals correctly identified by a discriminant function analysis based on measurements of seven morphometric characters expressing the individual size of *Myrmoborus lugubris* taxa

Taxon	% correct	M. l. berlepschi	M. l. femininus	M. l. lugubris	M. l. lugubris/femininus	M. l. stictopterus
M. l. berlepschi (48)	100.0	48	0	0	0	0
M. l. femininus (30)	80.0	0	24	0	6	0
M. l. lugubris (16)	68.7	0	1	11	4	0
M. l. lugubris/femininus (53)	88.7	0	3	1	47	2
M. l. stictopterus (8)	75.0	0	0	0	2	6
Total	87.7	48	28	12	59	8

Numbers in parenthesis represent the total number of individuals analyzed per taxon.

mtDNA Phylogeography

Mitochondrial phylogenetic analyses supported three main reciprocally monophyletic groups within *M. lugubris*, which were congruent with three described subspecies that present distinct plumage. However, it showed paraphyly between *M. l. stictopterus* and *M. l. femininus* (Fig. 3a), which could be distinguished by plumage characters. The first split supported the separation of *M. l. berlepschi* from the ancestral population of the remaining taxa, whereas a second event split *M. l. lugubris* from *M. l. femininus* and *M. l. stictopterus.* Within *M. l. berlepschi* an internal clade with some individuals from localities 1 and 2 was observed. Within *M. l. lugubris*, two groups had high support partially clustering individuals from westernmost (locality 14) and easternmost (localities 17 and 18) localities. The paraphyly between *M. l. stictopterus* and *M. l. femininus* was due to two individuals from locality 9 (*M. l. stictopterus*) that were more closely related to *M. l. femininus* individuals than to the reciprocally monophyletic clade formed by the remaining individuals of *M. l. stictopterus*.

Clustering analyses considering linked sites performed in BAPS (Corander and Marttinen 2006) supported $K = 5$ as the best model (likelihood $= -699.0878$; probability for five populations $= 0.9999$), clustering individuals congruently with the phylogenetic analyses and splitting *M. l. berlepschi* in two clusters with partial geographic structure (Fig. 3b). The gradual transition between the phenotypes of *M. l. femininus* and *M. l. lugubris* detected in morphometric and plumage analyses was not observed in BAPS results, suggesting an abrupt transition between these two taxa. Only specimens from Parintins (locality n°14 in Figs. 1a and 3b) included individuals of the two clades occurring in sympatry. However, the individual of locality 15, eastwards from Parintins, clustered with *M. l. femininus* individuals.

UCEs—Summary Information and Genetic Structure

Data processing in PHYLUCE produced a concatenated matrix with 15% missing data with 2151 loci. The average locus size was 549 bp ranging
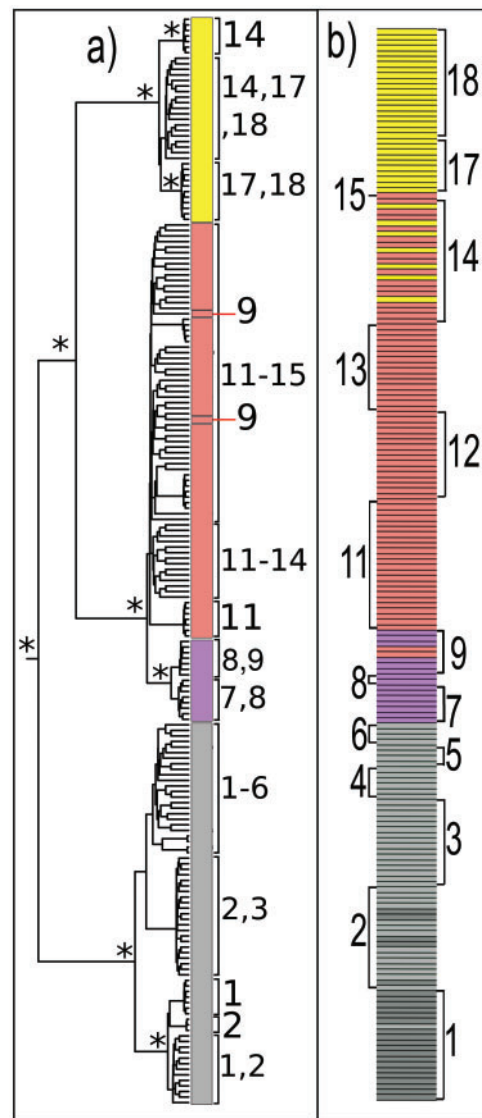

FIGURE 3.   a) Bayesian phylogenetic inference based on mtDNA (cyt b, 1040 bp). *posterior probability >0.95. b) Population genetic structure inferred with BAPS ($k = 5$) using the mtDNA data (cyt b, 1040 bp). Note that *M. l. berlepschi* individuals were split into two groups (light and dark gray). Numbers represent localities and colors, taxa in Fig. 1a.

from 259 to 763 bp. For the SNP calling procedure, we were able to obtain a reference for 2378 loci without cleaning for missing or trimming long edges. Data filtering for loci with five or more informative sites between all individuals of *M. lugubris* produced a data set with 473 loci with average size of 813 bp ranging from 330 (ARNTL-exon13) to 2092 bp (RAG2) with an average of 6.57 informative sites per locus (range of 5–21 informative sites per locus). The selection of a single variant per locus without missing data resulted in a matrix with 1664 SNPs. Overall mean sequencing coverage of SNPs among individuals was 29.3 X varying from 20.9 to 46.3 X.

The population structure obtained in sNMF was concordant with the plumage variation, with the best value of cross entropy (alpha parameter = 10; cross-entropy = 0.513584) supporting the presence of four ancestral populations ($k=4$) that matched the geographical distributions of *M. l. berlepschi*, *M. l. stictopterus*, *M. l. femininus*, and *M. l. lugubris* (Fig. 1b). The pattern of individuals with intermediate morphology between *M. l. femininus* and *M. l. lugubris* was congruent with the result that individuals of localities 13 (Itacoatiara) and 14 (Parintins) presented shared ancestry (Fig. 1a,b). The k-means and DAPC analyses recovered the same number of clusters as sNMF ($k=4$; Supplementary Fig. S6 available on dryad). The only discordance between these two analyses was in the classification of the morphologically intermediate individuals from Parintins (locality 14 in Fig. 1a) as the sNMF analyses showed a higher coefficient of ancestry with *M. l. lugubris* while in the classification matrix of the DAPC analysis they were classified as *M. l. femininus*. The second best model estimated in sNMF supported the presence of three ancestral populations ($K=3$), placing *M. l. femininus* individuals as intermediates between *M. l. lugubris* and *M. l. stictopterus* (alpha parameter = 10; cross-entropy = 0.521805; Fig. 1c).

UCEs—Phylogenetic Estimates, Species Tree, and Gene Flow

The RAxML and MrBayes phylogenies based on concatenated loci supported similar topologies that suggested the monophyly of *M. lugubris* and *M. leucophrys* as its sister species, and revealed at least three main clades within *M. lugubris* (see Supplementary Fig. S7 available on dryad). These relationships were partially congruent with currently recognized subspecies. The first split within *M. lugubris* separated *M. l. berlepschi*, which occurs in western Amazon (Solimões river) from the remaining populations. The second split separated *M. l. stictopterus* (Negro-Branco rivers) from a clade including individuals assigned to *M. l. femininus* and *M. l. lugubris* (Madeira and Amazon rivers). This latter clade was not recovered by mtDNA results (Fig. 3). Furthermore, internal relationships within this clade were generally poorly supported, with two exceptions: (i) easternmost individuals of *M. l. lugubris* (localities 17 and 18; Fig. 1a) clustered in a well-supported clade and
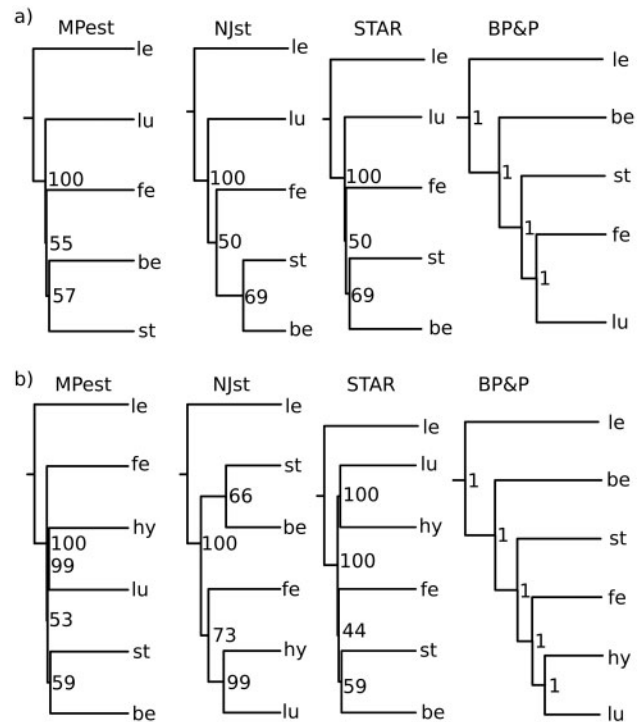

FIGURE 4. Species tree topologies based on 472 loci with >5 informative site and respective bootstrap nodal support obtained with Mpest, Njst, and STAR species tree methods, and posterior probability for the Bayesian species delimitation analyses on BP&P. a) Analyses without intermediate individuals between *Myrmoborus lugubris lugubris* and *M. l. femininus*. b) Analyses assuming intermediate individuals between *M. l. lugubris* and *M. l. femininus* as a distinct taxa. le = *M. leucophrys*; be = *M. l. berlepschi*; st = *M. l. stictopterus*; fe = *M. l. femininus*; lu = *M. l. lugubris*; hy = Intermediate individuals between *M. l. lugubris* and *M. l. femininus*.

(ii) some specimens of *M. l. femininus* and morphological intermediates from the westernmost part of *M. l. lugubris* range (localities 11 and 13; Fig. 1a) also grouped in a well-supported clade. The position of individuals from localities 12 and 14 is congruent with a stepping stone pattern between these two groups, which is in agreement with the presence of intermediate plumage patterns in these localities (see Supplementary Fig. S2 available on dryad).

Assuming that the sNMF clusters ($k=4$) as taxa, the species tree estimation methods (Njst, Star, and MP-EST) recovered topologies with all nodes having bootstrap support <0.75, except for the node clustering all individuals of *M. lugubris* (bootstrap = 100; Fig. 4a). When we considered the intermediate individuals as a separate lineage, in all three methods these intermediate individuals formed a clade with *M. l. lugubris* (bootstrap >99; Fig. 4). Only in Njst the relationship of *M. l. femininus* with the intermediate individuals and *M. l. lugubris* was marginally supported (bootstrap = 73). The full likelihood approach of BP&P produced a totally resolved species tree with maximum posterior probability for all taxa, with and without intermediate individuals (Fig. 4). This topology showed a first split between *M. l. berlepschi* and eastern Amazon taxa, a second split between *M. l.*
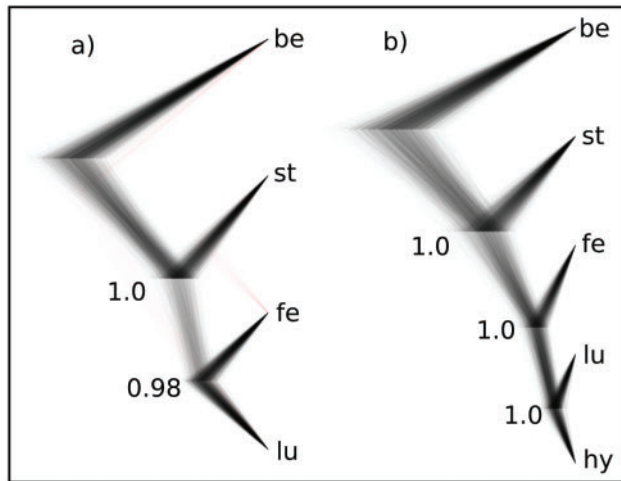
FIGURE 5.   Overlapped species tree topologies obtained with SNAPP based on the complete SNP matrix (1664 SNPs). a) Analysis without intermediate individuals between *Myrmoborus lugubris lugubris* and *M. l.* b) Analysis including intermediate individuals between *M. l. lugubris* and *M. l. femininus* as a distinct taxon. be = *M. l. berlepschi*; st = *M. l. stictopterus*; fe = *M. l. femininus*; lu = *M. l. lugubris*; hy = intermediate individuals between *M. l. lugubris* and *M. l. femininus*.

stictopterus and the remaining individuals, a third split separating *M. l. femninius*, and a fourth split (only in the analyses with five taxa) between the intermediate individuals and *M. l. lugubris* (Fig. 4).

The species trees estimated with SNAPP, assuming the same set of groups (four and five taxa), were identical with the trees obtained with BP&P. The posterior distribution of sampled species tree in DensiTree suggested a well-resolved relationship among the recognized populations with posterior probability of 1.0 for all estimated nodes, except for the relationship between *M. l. lugubris* and *M. l. femininus* (PP = 0.98) in the four taxa analysis (Fig. 5). The BFD results for alternative assignments of individuals to taxa supported with the highest marginal likelihood and a high value of Bayes factor (four taxa analysis, Marginal likelihood = −1770.50, ln(BF) >5; five taxa analysis, Marginal likelihood = −22095.83, ln(BF) >5) indicated the same groups used for the SNAPP analyses as the best arrangement of individuals as independent evolutionary lineages (Tables 1 and 2).

The TREEMIX topology without migration edges was identical to the SNAPP topology assuming five taxa (see Supplementary Fig. S8 available on dryad; likelihood = 131.558). However, the model that best fit the data supported the presence of two migration edges (Fig. 6; likelihood = 135.843). The first took place between *M. l. lugubris* and *M. leucophrys* (the sister taxon to the polytypic *M. lugubris*), which was not significant (Jacknife estimate $P = 0.1606$), and the second between the intermediate individuals and *M. l. femininus*, which was statistically significant ($P = 0.0006$). This latter relationship was the only admixture event supported by the $f3$ statistic, with a highly negative Z-score (*M. l. femininus*; *M. l. stictopterus*, morphologically
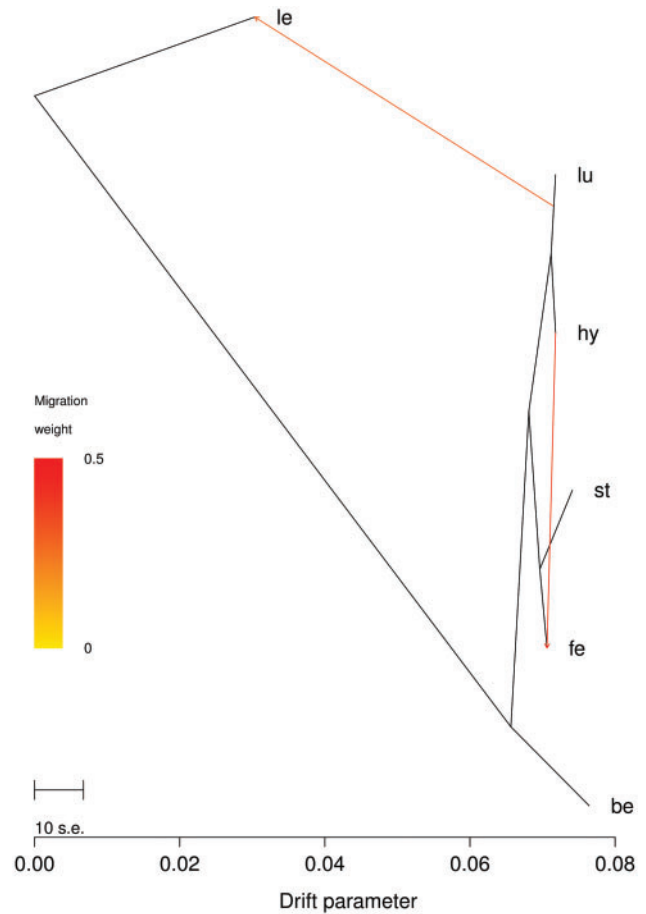

FIGURE 6.   Population relationships and migration edges inferred by TREEMIX. Color-scale indicates the weight of migration edges. Drift parameter is a relative temporal measure and the scale bar indicates 10 times the average standard error of the relatedness among populations based on the variance-covariance matrix of allele frequencies; be = *M. l. berlepschi*; st = *M. l. stictopterus*; fe = *M. l. femininus*; lu = *M. l. lugubris*; hy = intermediate individuals between *M. l. lugubris* and *M. l. femininus*; le = *M. leucophrys* (outgroup).

intermediate individuals; *Zscore* = −6.87209). The presence of this significant migration edge affected the inferred tree topologies, suggesting a sister relationship between *M. l. femininus* and *M. l. stictopterus* (Fig. 6) as observed in the mtDNA results.

The Maximum Pseudo Likelihood algorithm of PhyloNet with introgression edges set to zero recovered similar relationships obtained by BP&P and SNAPP, clustering *M. l. femininus*, *M. l. lugubris* and intermediate individuals as a clade (likelihood (ln(lik)= −739516.4899; Fig. 7). The PhyloNet runs allowing for two introgression edges supported a topology similar to those obtained by TREEMIX and mtDNA, with *M. l. femininus* and *M. l. stictopterus* as sister taxa in the two networks with the highest likelihood (Fig. 7). The inheritance probability of the two best trees were higher for *M. l. lugubris* ($\gamma = 0.70$ and $\gamma = 0.97$ in the first edge and $\gamma = 0.97$ and $\gamma = 0.80$ in the second edge for the first and second best runs, respectively) than for *M. l. stictopterus/M. l. femininus*
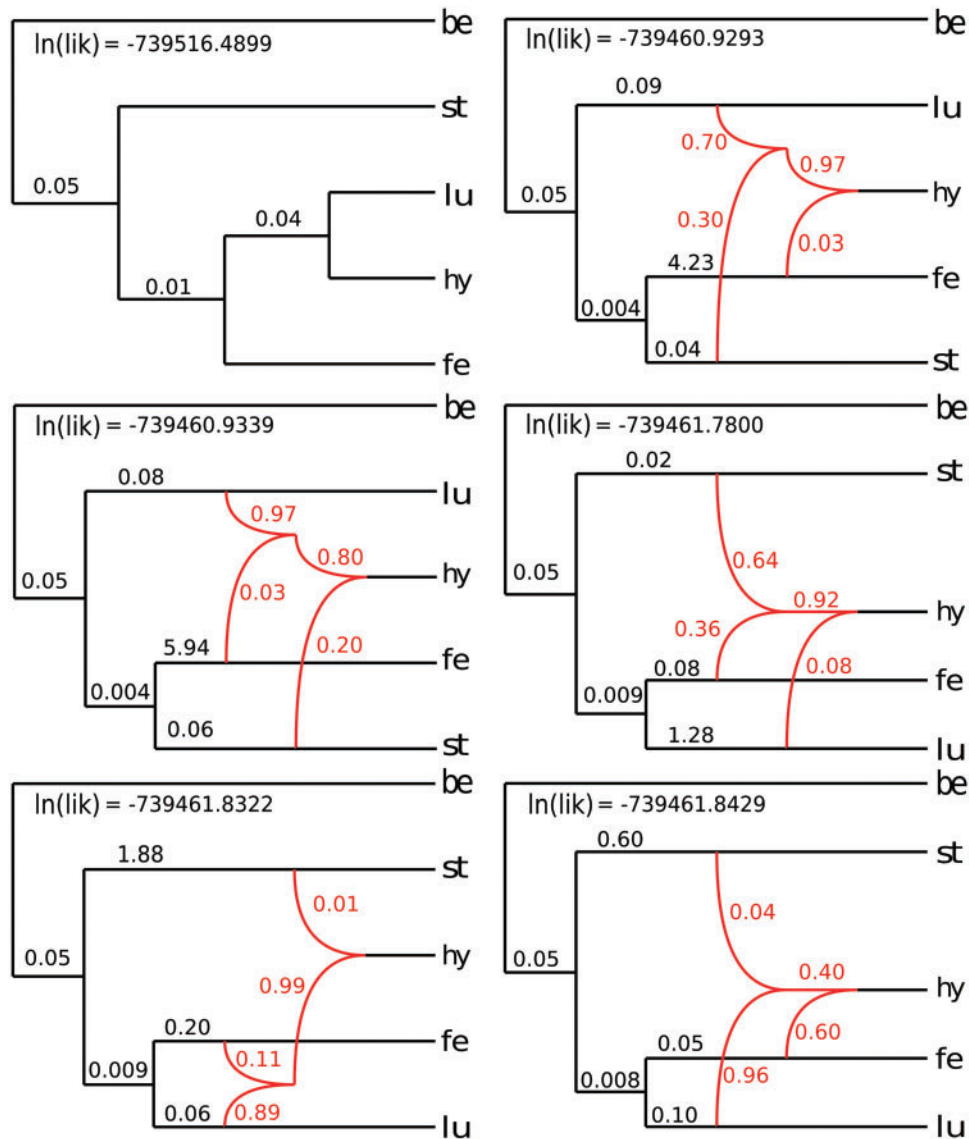
FIGURE 7.   Species tree (top left; no introgression allowed) and the five best phylogenetic networks inferred with the maximum pseudo likelihood algorithm of Phylonet allowing for two introgression events. In red, lines represent the introgression edges and number, the inheritance probabilities of each edge. Numbers in black represent branch length. be = *M. l. berlepschi*; st = *M. l. stictopterus*; fe = *M. l. femininus*; lu = *M. l. lugubris*; hy = Intermediate individuals between *M. l. lugubris* and *M. l. femininus*.

($\gamma = 0.30$ and $\gamma = 0.03$ in the first edge and $\gamma = 0.03$ and $\gamma = 0.20$ in the second edge for the first and second best runs, respectively) with a higher contribution of *M. l. stictopterus* than *M. l. femininus* in both estimations. The three remaining runs supported the same relationship with zero migration edges.

To assess the potential impacts of gene flow in species tree estimation, we used Fastsimcoal2 (Excoffier et al. 2013) to test the fit of demographic models concerning the different topologies with and without post-divergence gene flow. We tested four demographic models (Fig. 8): (i) based on the SNAPP topology, with no gene flow among lineages; (ii) based on the mtDNA/TREEMIX topologies, also with no gene flow among lineages and considering the intermediate

individuals used in SNAPP as a distinct lineage sister to *M. l. lugubris*; (iii) based on the BP&P/SNAPP topologies, but considering the intermediate individuals as a hybrid population between *M. l. femininus* and *M. l. lugubris* with gene flow among *M. l. femininus*, intermediate individuals and *M. l. lugubris*; (iv) based on the mtDNA/TREEMIX topology, with gene flow among *M. l. femininus*, intermediate individuals and *M. l. lugubris*, and assuming that the intermediate individuals form a hybrid population between *M. l. femininus* and *M. l. lugubris*. The results supported a maximum relative contribution, based on AIC values (1.0), for model D which assumed the mtDNA/TREEMIX topology and a hybrid population between *M. l. femininus* and *M. l. lugubris* (Tables 4 and 5; Fig. 8d). Parameter estimation
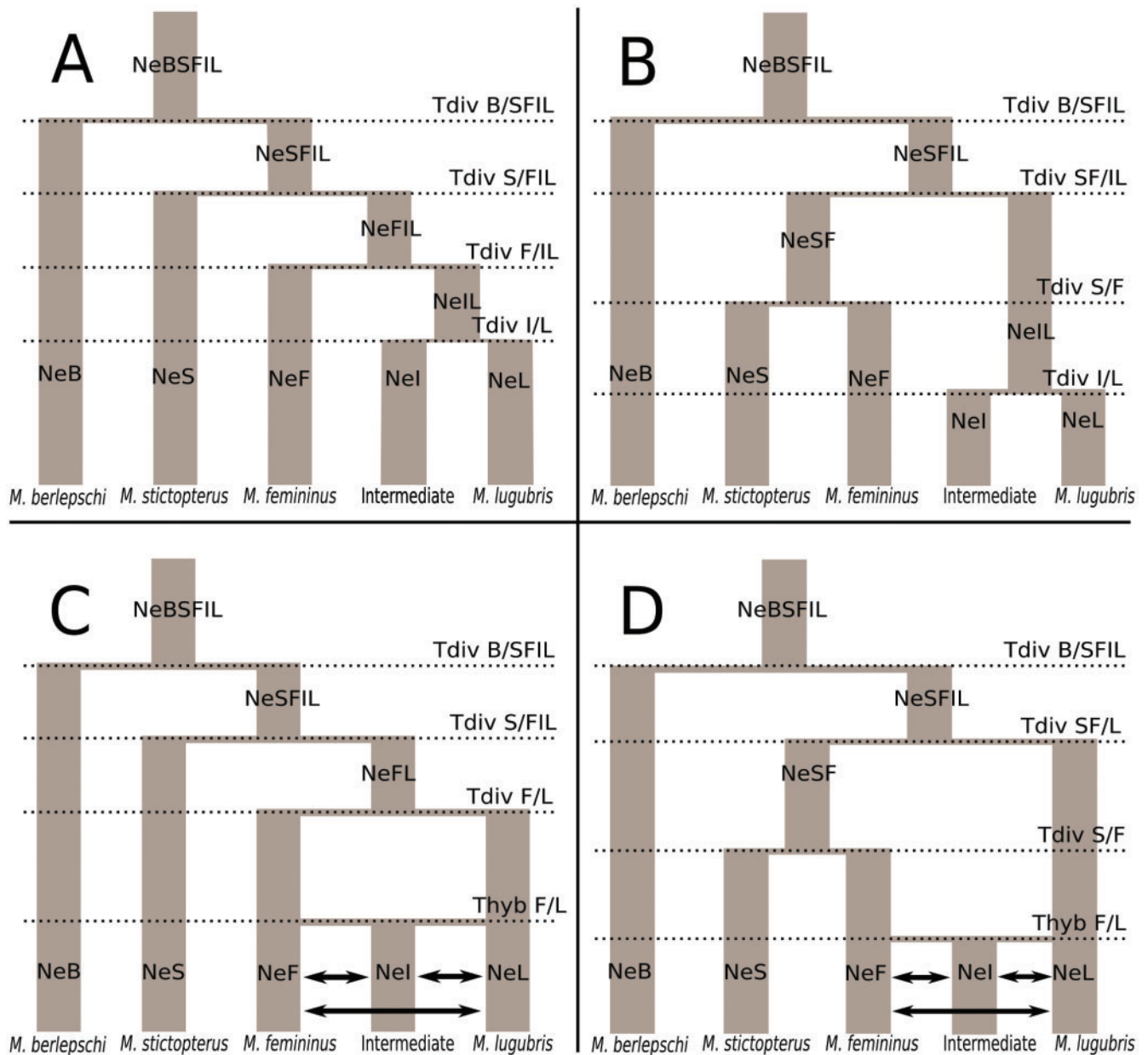
FIGURE 8.   The four demographic models tested in Fastsimcoal2. Model parameters are shown in Table 3, and include divergence times (Tdiv), current and ancestral effective population sizes (Ne) and migration (horizontal arrows).

and CIs obtained through bootstrap replicates supported relatively large current effective population sizes for the four *M. lugubris* taxa but a reduced effective size of hybrids (Table 4). The diversification time estimates supported the split of *M. l. berlepschi* from the remaining taxa at approximately 1.5 Ma, the split between *M. l. lugubris* from *M. l. femininus* and *M. l. stictopterus* at approximately 1.4 Ma, the split between *M. l femininus* and *M. l. stictopterus* at approximately 0.35 Ma and the beginning of the introgression between *M. l. lugubris* and *M. l. femininus* at approximately 0.14 Ma (Table 4). Gene flow estimates supported an asymmetric pattern and relative high number of migrants per generation from

M. l. femininus into the morphologically intermediate population (4.90, 95% CI 4.59–5.01) and from *M. l. lugubris* into *M. l. femininus* (0.91, 95% CI 0.97–1.32; Table 4).

DISCUSSION

Phylogenetic Inference and Species Tree Discordances

Our study indicated that failing to account for gene flow in species tree estimation can affect topology in cases of recent diversification. In addition, our results showed that the use of phenotypic information can help detect the effects of introgression in wild

TABLE 4. Parameters estimated in Fastsimcoal 2 under four demographic models for the diversification of *Myrmoborus lugubris*. Models are described in Fig. 8

Model	N1	N2	N3	N4	N5	Na23	Na34	Na234	Na	NM43
A	1 101 907	389 003	663 662	356 988	145 770	***	1878408	940 731	112 603	***
B	500 690	714 630	1 110 812	654 584	230 709	148 122	***	1 005 070	1 956 792	***
C	1 074 254	645 673	1 085 734	66 633	208 507	***	1570047	1 964 183	123 272	0.00
D	1 075 223	731 159	1 133 674	66 832	590 531	1 556 677	***	2 175 098	2 160 755	0.02
Lo 95%	896 418	661 016	1 230 587	61 972	456 088	1 494 866	***	1 984 074	2 137 099	0.02
Up 95%	962 213	741 371	1 368 173	68 617	457 183	1 574 056	***	2 149 316	2 148 960	0.07

Model	NM34	NM54	NM45	NM53	NM35	TDIV1	TDIV2	TDIV3	TDIV4	Lhood
A	***	***	***	***	***	1 253 010	266216.5	290599.9	1603298.6	−19813.85
B	***	***	***	***	***	443909.3	455759.6	477831.7	494954.9	−19830.184
C	4.65	0.03	0.00	0.02	0.23	169127.7	485341.3	505777.8	1484885.7	−19679.779
D	4.90	0.02	0.01	0.91	0.01	148078.5	350702.2	1486209.1	1508230	−19554.501
Lo 95%	4.59	0.01	0.01	0.97	0.02	125994.7	303123.7	1136881.6	1171892.1	***
Up 95%	5.01	0.03	0.02	1.32	0.07	148453.6	347698.9	1198295.7	1222050	***

Estimated parameters are effective population sizes (Ne) of current and ancestral populations in haploid individuals (N1—Ne *M. l. berlepschi*; N2—Ne *M. l. stictopterus*; N3—Ne *M. l. femininus*; N4—Ne morphologically intermediate group between *M. l. lugubris* and *M. l. femininus*; N5—Ne *M. l. lugubris*; Na23—ancestral of *M. l. femininus* and *M. l. stictopterus*; Na34—ancestral Ne of *M. l. femininus* and *M. l. lugubris*; Na234—ancestral Ne of *M. l. stictopterus, M. l. femininus* and *M. l. lugubris*; Na—ancestral Ne of *M. lugubris*), number of migrants per generation from population *x* to *y* forward in time (NMxy), divergence times backwards in time as in Fig. 8 (TDIV), lower and upper limits of the 95% confidence interval of maximum likelihood parameter estimates based on 100 parametric bootstraps (Lo 95% and Up 95%). *** = not applicable.

TABLE 5. Maximum likelihood (Max ln(L)) obtained for the four demographic models simulated in Fastsimcoal2

Model	Max ln(L)	2*par	AIC	ΔAIC	AIC weight
A	−19 814	18	91264.2	1182.3	2E−257
B	−19 830	18	91339.4	1257.6	8E−274
C	−19 680	30	90658.7	576.9	5E−126
D	−19 555	30	90081.8	0	1

Number of parameters (par); Akaike information criterion (AIC); differences in AIC relative to the best-fit model (ΔAIC); relative Akaike's weight of evidence based on the maximum-likelihood estimates (AIC weight).

populations and in the interpretation of species tree analyses (Edwards and Knowles 2014).

The species tree methods applied here have been widely used to reconstruct shallow diversification scenarios that could be affected by gene flow between populations assumed to be isolated. In these cases the implicit assumption that ILS solely generated the observed genealogical patterns is restrictive and can bias topology, branch length and effective population size estimations if considerable levels of gene flow are present (Leaché et al. 2014a,b; Solís-Lemus et al. 2016). Here, we detected admixture among non-sister lineages, which would have exacerbated this bias and increased the complexity underlying the evolutionary relationship among lineages. Summary-based species tree methods that use individual gene trees as inputs have shown accurate results under low levels of gene flow, although higher levels of gene flow increase gene tree heterogeneity promoting loss of resolution and weak node support in bootstrap replicates (Liu et al. 2009a; Solís-Lemus et al. 2016). Solís-Lemus et al. (2016) reported anomalous results based on concatenated data sets and using two summary-based species tree methods (STAR and Njst) when more than 30% ($\gamma > 0.3$) of loci were inherited through reticulation, even with high numbers of markers such as 10 000 simulated genes. The strong

effects of gene flow reported by Solís-Lemus et al. (2016) is in agreement with our results that showed inconsistencies among results of species tree methods that do not account for gene flow (Figs. 4 and 5) and those of PhyloNet, whose model with the highest likelihood supported values higher than 0.3 (Fig. 7).

On the other hand, little is known about how species tree methods such as SNAPP are affected by high levels of gene flow. Although Rheindt et al. (2014) suggested that SNAPP is robust with limited gene flow between populations of *Zimmerius* flycatchers, here we obtained full statistical support for topologies indicating to a distinct relationship between populations that was not consistent with results derived from our mtDNA phylogeny, phenotypic data, TREEMIX, PhyloNet as well as the best-fit model recovered with fastsimcoal2. The SNAPP algorithm uses SNP allele frequencies assuming that the probability of an allele frequency at a given locus is correlated with the probability of a site in a gene tree multiplied by the gene tree probability given a species tree (Bryant et al. 2012; Leaché et al. 2014a,b). Thus it is possible that a hybrid population, with a unique pattern of allele frequencies (intermediary allele frequency), tends to be weighted as an independent lineage by SNAPP and BFD methods. The results presented here are in agreement with Sukumaran and Knowles (2017) that suggested that MSC methods can produce misleading results when used to directly infer species limits in face of relative high gene flow but are useful tools to estimate phylogenetic relationships and genetic structure if their main assumptions are preserved.

The phylogenetic estimate obtained based on mtDNA and the model-based approach implemented on fastsimcoal2 are partially congruent with the morphological data. The most readily diagnosable taxon is *M. l. berlepschi* given its significantly smaller size and its characteristic female plumage, and it is also the first population to diverge within *M. lugubris*. The

second split within *M. lugubris* separated the taxa of central Amazonia (*M. l. femininus* and *M. l. stictopterus*) from *M. l. lugubris* of the lower Amazon river, which is concordant with the black-faced pattern and the overlap in morphometric variables observed in *M. l. femininus* and *M. l. stictopterus*. However, the pattern of gradual transition in plumage characters, size and the coefficient of ancestry of UCEs observed between *M. l. femininus* and *M. l. lugubris* was not observed in mtDNA, which supported an abrupt transition between these taxa with only Parintins (locality n°14 in Fig. 1a) presenting individuals from both clades (Fig. 3). The absence of mtDNA introgression could be explained by the matrilineal inheritance without recombination of this marker (Carling and Brumfield 2008) in addition to a recent secondary contact, as suggested by fastsimcoal2 (Table 4) or even male biased dispersion. However, given the biology of antbirds, including territorial defense by both sexes (Zimmer and Isler 2003), the latter hypothesis seems to be less likely.

Phenotypes and Hybridization

In scenarios of recent diversification, the use of multiple types of evidence—phenotypic data such as plumage, morphometric, vocal, ecological, and physiological traits—can provide more accurate estimates of and a better understanding of demographic processes than genetic data alone (Edwards and Knowles 2014; Solís-Lemus et al. 2015; Zamudio et al. 2016). The results obtained here revealed high levels of introgression between *M. l. lugubris* and *M. l. femininus* (Table 4; Figs. 1b,c, and 2, Supplementary Figs. S2 and S4 available on dryad), with a wide hybrid zone at least 500 km long and with a clear pattern of transition in female plumage and morphometric measurements (tail length, wing length, and exposed culmem length). The hypothesis of Haffer and Fitzpatrick (1985) that the central Amazonian taxa could be intermediate between the two extreme plumage forms as a result of secondary intergradation of *M. l. berlepschi* and *M. l. lugubris* was not corroborated, since we observed secondary intergradation only between *M. l. femininus* and *M. l. lugubris* and a clear diagnosis of *M. l. berlepschi* that seems restricted to the Solimões river. Size variation over the geographic distribution of *M. lugubris* was first noted by Hellmayr (1910) who, when describing *M. l. femininus* and *M. l. berlepschi*, noticed the smaller size of western populations. The significantly larger size of hybrid individuals from localities 14–17 (Figs. 1a and 2) in most of the morphometric characters measured (see Supplementary Table S3 available on dryad) suggest that the secondary contact and introgression between *M. l. lugubris* and *M. l. femininus* could result on heterosis. However, despite the general larger size of the morphologically intermediate population, more detailed studies must be performed to better understand the potential effects of hybridization on the fitness of these individuals compared with that of the parental forms.

Diversification of the Ash-Breasted Antbird over the Amazonian Floodplains

Large Amazonian rivers have been historically recognized as effective barriers to gene flow in upland forest (*terra-firme*) species, delimiting their geographic distributions (Cracraft 1985; Antonelli et al. 2010; Smith et al. 2014). However, this well documented pattern is not expected to be observed in organisms occurring along riverine environments. Phylogeographic studies of non-aquatic bird species associated to floodplains did not find population structure throughout the Amazon basin, suggesting that the linear distribution and connectivity of floodplains enable high levels of gene flow (Aleixo 2006; Cadena et al. 2011; but see Choueri et al. 2017). Furthermore, species occurring in these environments are expected to be good dispersers due to the constant effect of flooding cycles and sedimentation dynamics, which affect the distribution of river created environments (Aleixo 2006). Against these patterns and assumptions, our study supported the presence of four structured populations within the ash-breasted antbird complex geographically associated with main Amazonian rivers (Solimões, Madeira, Negro and Amazonas). This hitherto uncommon pattern suggests that despite the apparent absence of modern physical barriers, historical processes could have promoted opportunities of isolation along floodplain forests, mainly in organisms such as *M. lugubris*, which is restricted to specific environments of river edge forests and islands (Rosenberg 1990).

The parameter estimates obtained with fastsimcoal2 supported a recent scenario of diversification starting at the mid-Pleistocene with the split between *M. l. berlepschi* and the remaining taxa followed by an almost simultaneous diversification of *M. l. lugubris* and the two taxa of central Amazonia (Table 4). These results indicate a scenario where paleoclimatic fluctuations during the mid and late Pleistocene potentially interrupted the distribution and availability of river-created environments (Latrubesse and Franzinelli 2005; Irion et al. 2009; Rossetti et al. 2015). Similar results were obtained by Choueri et al. (2017) studying landscape genetics of four antbird species—including *M. lugubris*—along the Negro river archipelagos, suggesting that populations restricted to this river diverged from the Amazon and Solimões populations during the late Pleistocene. During this time span, local river channel incisions occurred at the main course of the Solimões river and tributaries, with floodplain environments being replaced by upland forest habitats (Nogueira et al. 2013; Rossetti et al. 2015). This process could have fragmented riverine environments, specially in central Amazon where the sedimentary basin of the Solimões runs through hard rock substrates of the Amazon Craton (Choueri et al. 2017).

River incision cyclically alternated with sediment accumulation periods with the expansion of floodplains environments and potentially connection of previously isolated areas. Although these cycles could be linked

to sea level fluctuations related to the intensification of glacial/interglacial periods during the mid-Pleistocene in the lower Amazon river (Irion et al. 2009; Bertani et al. 2015), areas from the central and western Amazon Basin are more likely to be influenced by paleoclimatic and tectonic processes occurring in source areas of sediments, increasing water discharge and sediment load (Choueri et al. 2017). This intense dynamic cycle of sediment accumulation and erosion could have originated the hybrid zone between *M. l. lugubris* and *M. l. femininus* around 0.14 Ma. Additionally, during glacial cycles, large lakes (the so called Ria lakes) were formed in the lower courses of rivers that carry low amounts of sediments, such as the Negro and Tapajós rivers (Irion et al. 2009; Bertani et al. 2015). The Ria lakes formation produced large areas without islands and with poorly developed floodplains on river banks (Franzinelli and Igreja 2002; Irion et al. 2009), potentially acting as barrier for *M. lugubris* taxa, explaining the possible absence of *M. l. stictopterus* in the lower course of the Negro river. However, despite the clear pattern of genetic structure and the recent diversification scenario reported here, comparative phylogeographic studies including large sample sizes and estimates of population size changes over time, have the potential to reveal a much more refined pattern of diversification of organisms associated with Amazonian floodplain forests.

The first diversification event observed in *M. lugubris* separating *M. l. berlepschi* of the upper Amazon (Solimões) river from the eastern taxa of the Negro, Madeira and lower Amazon rivers is in accordance with studies reporting abrupt transitions in community composition of trees, spiders, fishes, and birds (Hubert and Renno 2006; Albernaz et al. 2012; Cohn-Haft et al. 2007), as well as sedimentation patterns and river dynamics (Mertes et al. 1996), along a strong longitudinal biogeographic gradient encompassing floodplain forests of the entire course of the Amazon river and its main tributaries. However, few phylogeographic studies on populations restricted to the floodplains have been conducted so far to evaluate whether this biogeographic pattern is mirrored by the genetic structure of populations along this region (Choueri et al. 2017). Farias and Hrbek (2008), based on mtDNA data, recovered reciprocally monophyletic groups in discus fishes of the genus *Symphysodon* from the Solimões and Amazon rivers, suggesting that this diversification process was mediated by the breach of Purus Arch during the Pliocene and by the different chemical composition of these rivers. Despite the similarities in the geographic distributions of the taxa studied here and the ones of the *Symphysodon* clades identified by Farias and Hrbek (2008), the more recent divergence times reported in *M. lugubris* suggests that distinct historical processes shaped the current pattern of genetic diversity in these two groups.

Another potentially important process of diversification that we cannot rule out as an explanation for the results we obtained, is the presence of ecological gradients with distinct selective pressures in specific regions of the Amazonian floodplains. For example, the transitions between white water rivers (with high sediment load) and black water rivers (with low sediment load) delimit distinct vegetation types and affect the pattern of genetic diversity in several species of fishes (Cooke et al. 2014).

Species Limits and Taxonomy

Here, we showed significant levels of differentiation among subspecies of *M. lugubris* in phenotypic and genetic characters which, when interpreted together, allow for a redefinition of species limits within this complex. The clear diagnosis of *M. l. berlepschi* by all morphological and genetic analyses, including the absence of gene flow between this taxon and the remaining ones grouped under *M. lugubris* provide the basis for considering it a separate species under distinct species concepts including the biological species concept (Mayr 1942; Queiroz 2007) and the phylogenetic species concept ones (PSC; Cracraft 1983; Queiroz 2007). However the observed pattern does not guarantee total reproductive isolation given their apparent allopatric distribution. The high support obtained for the reciprocal monophyly of *M. l. stictopterus* by most phylogenetic estimates based on UCEs, was consistent with its apparent allopatric distribution and differences in the coloration of the face (pure black in *M. l. stictopterus* without sparse reddish feather as in *M. l. femininus*) thereby supporting a clear diagnosis for these taxa that could be recognized as valid species under distinct species concepts such as the PSC (Queiroz 2007).

The hybridization pattern observed between *M. l. lugubris* and *M. l. femininus* indicates a complex scenario with multiple taxonomic interpretations including: (i) the clearly diagnosable parental forms and the heterosis pattern among hybrids suggest secondary contact and post zygotic incompatibilities between these taxa that could be recognized as valid species under the PSC (Queiroz 2007); (ii) the gene flow pattern with a hybrid zone geographically as large as the distribution of the parental forms could suggest an ephemeral speciation with fusion of non-sister lineages not supporting species status for these taxa. Thus, studies with denser sampling over the hybrid zone to better explore the clinal variation, as well as studies focusing on hybrid fitness, must be performed to better explore the genetic cohesion of these two taxa.

Considerations When Accounting for Gene Flow in Phylogenetic Reconstructions

The importance of MSC methods arose with the assumption that ILS can produce discordance among gene trees, which cannot be modeled with concatenated markers. Moreover, methods that explicitly incorporate gene flow showed that ILS is not the

only source of inconsistencies among gene trees. The increasing amount of evidence for reticulated evolution reinforces this demand (Edwards et al. 2016; Mallet et al. 2016). Phylogenetic inferences are the starting analyses of phylogeographic studies that require a strong phylogenetic hypothesis to explore complex demographic parameters. Commonly used methods to estimate demographic parameters such as Ima2 (Hey and Nielsen 2007) and G-Phocs (Gronau et al. 2011) demand a priori topology for the estimation of demographic parameters, including effective population size, gene flow and divergence times. Under this perspective, network estimation methods that explicitly accept gene flow into species tree estimation can be important tools to understand complex patterns of diversification among relatively large species or population complexes (Bapteste et al. 2013; Nakhleh 2013; Wen et al. 2016). Similarly, model-based approaches such as the composite likelihood based method implemented in fastsimcoal2 or Approximate Bayesian Computation can provide robust parameter estimations to test competing phylogenetic hypotheses (Robinson et al. 2014; Martin et al. 2015; Jackson et al. 2017). However, these methods do not allow for exhaustive tree search, meaning that robust alternative hypotheses need to be constructed based on complementary analyses (Nater et al. 2015).

In recent years, a relatively large number of analytical methods were proposed to accommodate both ILS and gene flow in phylogenetic inferences (Than et al. 2008; Kubatko 2009; Jackson et al. 2017; Solís-Lemus et al. 2016). As examples, there are fast and reliable algorithms such as the Maximum Pseudo Likelihood of PhyloNet (Yu and Nakhleh 2015) and SNaQ (Solís-Lemus et al. 2016) and a model selection framework that allows to test the fit of hundreds of models to varying topologies and gene flow direction as implemented in the software PHRAPL (Jackson et al. 2017). Thereby, given the relative large availability of methods and the results presented here and elsewhere (Meyer et al. 2016; Morales et al. 2017), studies involving phylogenetic estimations of recently diverged taxa in the gray zone of phylogeography must consider the effects of gene flow by applying such alternative methods. Additionally, the application of these alternative methods can help to detect errors in phylogenetic estimations (Knowles and Kubatko 2010) and incongruent topologies can be later on explored by model testing approaches.

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.13b88.

Funding

Acknowledgements

References

Albernaz A.L., Pressey R.L., Costa L.R.F., Moreira M.P., Ramos J.F., Assunção P. A., Franciscon C.H. 2012. Tree species compositional change and conservation implications in the white-water flooded forests of the Brazilian Amazon. J. Biogeogr. 39:869–883.

Aleixo A. 2006. Historical diversification of floodplain forest specialist species in the Amazon: a case study with two species of the avian genus *Xiphorhynchus* (Aves: Dendrocolaptidae). Biol. J. Linn. Soc. 89:383–395.

Alexander D.H., Novembre J., Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655–1664.

Andrews S. 2014. FastQC: a quality control tool for high-throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed 9 March 2017).

Antonelli A., Quijada-Mascareñas A., Crawford A.J., Bates J.M., Velazco P.M., Wüster W. 2010. Molecular studies and phylogeography of Amazonian tetrapods and their relation to geological and climatic models. In: Hoorn, C., Wesselingh, F. editors. Amazonia: landscape and species evolution. 1st ed. Oxford: Wiley-Blackwell. p. 386–404.

Bagley J.C., Mayden R.L., Roe K.J., Holznagel W., Harris P.M. 2011. Congeneric phylogeographical sampling reveals polyphyly and novel biodiversity within black basses (Centrarchidae: Micropterus). Biol. J. Linn. Soc. 104:346–363.

Bapteste E., van Iersel L., Janke A., Kelchner S., Kelk S., McInerney J.O., Morrison D.A., Nakhleh L., Steel M., Stougie L., Whitfield J. 2013. Networks: expanding evolutionary thinking. Trends Genet. 29:439–441.

Barley A.J., Brown J.M., Thomson R.C. 2017. Impact of model violation on the inference of species boundaries under the multispecies coalescent. Syst. Biol. 0:1–17.

Bertani T.C., Rossetti D.F., Hayakawa E.H., Cohen, M.C.L. 2015. Understanding Amazonian fluvial rias based on a Late Pleistocene-Holocene analog. Earth Surf. Process. Landf. 40:285–292.

Bouckaert R.R. 2010. DensiTree: making sense of sets of phylogenetic trees. Bioinformatics 26:1372–1373.

Bouckaert R., Heled J., Khnert D., Vaughan T., Wu C.H., Xie D., Suchard M.A., Rambaut A., Drummond A.J. 2014. BEAST 2: a software platform for Bayesian Evolutionary Analysis. PLoS Comput. Biol. 10:1–6.

Bryant D., Bouckaert R., Felsenstein J., Rosenberg N. A., Roychoudhury A. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29:1917–1932.

Cadena C.D., Gutiérrez-Pinto N., Dávila N., Chesser R.T. 2011. No population genetic structure in a widespread aquatic songbird from the Neotropics. Mol. Phylogenet. Evol. 58:540–545.

Carling M.D., Brumfield R.T. 2008. Haldane's rule in an avian system: using cline theory and divergence population genetics to test for differential introgression of mitochondrial, autosomal, and sex-linked loci across the *Passerina* bunting hybrid zone. Evolution 62:2600–2615.

Choueri E., Gubill C., Borges S., Thom G., Sawakuchi A., Soares E., Ribas C. 2017. Phylogeography and population dynamics of Antbirds (Thamnophilidae from Amazonian fluvial islands. J. Biogeogr. Doi:10.1111/jbi.13042.

Cohn-Haft M., Naka L.N. F.A.M. 2007. Padrões de distribuição da avifauna da várzea dos rios Solimões-Amazonas. In: Albernaz A.L., editor. Conservação da Várzea, Identificação e Caracterização de Regiões Biogeográficas. Manaus: IBAMA/ProVárzea/INPA. p. 287–324.

Cooke G.M., Landguth E.L., Beheregaray L.B. 2014. Riverscape genetics identifies replicated ecological divergence across an Amazonian ecotone. Evolution 68:1947–1960.

Corander J., Marttinen P. 2006. Bayesian identification of admixture events using multilocus molecular markers. Mol. Ecol. 15: 2833–2843.

Cracraft J. 1983. Species concept and speciation analysis. Curr. Ornithol. 1:159–187.

Cracraft J. 1985. Historical biogeography and patterns of differentiation within the South Amercian avifauna: areas of endemism. Ornithol. Monogr. 36:49–84.

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E., Lunter G., Marth G.T., Sherry S.T., McVean G., Durbin R. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

del Hoyo J., Elliott A., Sargatal J., Christie D.A., de Juana E. 2017. Handbook of the birds of the world alive. Barcelona: Lynx Edicions, (retrieved from http://www.hbw.com/ on 01/10/2018).

Edwards D.L., Knowles L.L. 2014. Species detection and individual assignment in species delimitation: can integrative data increase efficacy? Proc. R. Soc. B. 281:20132765.

Edwards S. V., Potter S., Schmitt C.J., Bragg J.G., Moritz C. 2016. Reticulation, divergence, and the phylogeography–phylogenetics continuum. Proc. Natl. Acad. Sci. U.S.A. 113:8025–8032.

Eckert A., Carstens B. C. 2008. Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. Mol. Phylogenet. Evol. 49:832–842.

Excoffier L., Dupanloup I., Huerta-Sanchez E., Sousa V.C., Foll M. 2013. Robust demographic inference from genomic and SNP data. PLoS Genet. 9:e1003905.

Faircloth B.C., McCormack J.E., Crawford N.G., Harvey M.G., Brumfield R.T., Glenn T.C. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61:717–26.

Farias I.P., Hrbek T. 2008. Patterns of diversification in the discus fishes (*Symphysodon* spp. Cichlidae) of the Amazon basin. Mol. Phylogenet. Evol. 49:32–43.

Feder J.L., Egan S.P., Nosil P. 2012. The genomics of speciation-with-gene-flow. Trends Genet. 28:342–350.

Franzinelli E., Igreja H. 2002. Modern sedimentation in the lower Negro River, Amazonas State, Brazil. Geomorphology 44: 259–271.

Frichot E., Mathieu F., Trouillon T., Bouchard G., François O. 2014. Fast and efficient estimation of individual ancestry coefficients. Genetics 196:973–983.

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z.,

Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29:644–652.

Gronau I., Hubisz M. J., Gulko B., Danko C. G., Siepel A. 2011. Bayesian inference of ancient human demography from individual genome sequences. Nat. Genet. 43:1031–1034.

Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59:307–321.

Gutenkunst R.N., Hernandez R.D., Williamson S.H., Bustamante C.D. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5:1–11.

Hackett S.J., Kimball R.T., Reddy S., Bowie R.C.K., Braun E.L., Braun M.J., Chojnowski J.L., Cox W.A., Han K.-L., Harshman J., Huddleston C.J., Marks B.D., Miglia K.J., Moore W.S., Sheldon F.H., Steadman D.W., Witt C.C., Yuri T. 2008. A Phylogenomic Study of Birds Reveals Their Evolutionary History. Science 320:1763–1768.

Haffer J. 1969. Speciation in Amazonian forest birds. Science 165:131–137.

Haffer J., Fitzpatrick J.W. 1985. Geographic variation in some Amazonian forest birds. Ornithol. Monogr. 36:147–168.

Harrison R.G. 1990. Hybrid zones: windows on evolutionary process. Oxf. Surv. Evol. Biol. 7:69–128.

Harris S.E., O'Neill R.J., Munshi-South J. 2015. Transcriptome resources for the white-footed mouse (Peromyscus leucopus): new genomic tools for investigating ecologically divergent urban and rural populations. Mol. Ecol. Resour. 15:382–394.

Harvey M.G., Smith B.T., Glenn T.C., Faircloth B.C., Brumfield R.T. 2016. Sequence capture versus restriction site associated DNA sequencing for shallow systematics. Syst. Biol. 65: 910–924.

Hellmayr C.E. 1910. The birds of the Rio Madeira. Novit. Zool. 17:257–428.

Hellmayr C.E. 1929. A contribution to the ornithology of Northeastern Brazil. Field Mus. Nat. His. Zoo. Ser. 12:235500.

Hewitt G.M. 2004. Genetic consequences of climatic changes in the Quaternary. Phil. Trans. R. Soc. Lond. B 359:183–195.

Hey J., Nielsen R. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proc. Natl. Acad. Sci. U.S.A. 104:2785–2790.

Hickerson M.J., Meyer C.P., Moritz C. 2006. DNA barcoding will often fail to discover new animal species over broad parameter space. Syst. Biol. 55:729–739.

Hosner P., Faircloth B.C., Glenn T.C., Braun E.L., Kimball R.T. 2015. Avoiding missing data biases in phylogenomic inference: An empirical study in the landfowl (Aves: Galliformes). Mol. Biol. Evol. 33:1110–1125.

Hubert N., Renno J.F. 2006. Historical biogeography of South American freshwater fishes. J. Biogeogr. 33:1414–1436.

Irion G., Müller J., Morais J.O., Keim G., de Mello J.N., Junk W.J. 2009. The impact of Quaternary sea level changes on the evolution of the Amazonian lowland. Hydrol. Process. 23:3168–3172.

Isler M.L., Bravo G.A., Brumfield R.T. 2013. Taxonomic revision of *Myrmeciza* (Aves: Passeriformes: Thamnophilidae) into 12 genera based on phylogenetic, morphological, behavioral, and ecological data. Zootaxa 3717:469–497.

Jackson N.D., Carstens B.C., Morales A.E., O'Meara B.C. 2017. Species delimitation with gene flow. Syst. Biol. 66:799–812

Jombart T., Ahmed I. 2011. Adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. Bioinformatics 27:3070–3071.

Jombart T., Devillard S., Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 11:94.

Junk W.J., Piedade M.T.F., Schöngart J., Cohn-Haft M., Adeney J.M., Wittmann F. 2011. A classification of major naturally-occurring amazonian lowland wetlands. Wetlands 31:623–640.

Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

Kimball R.T., Braun E.L., Barker F.K., Bowie R.C., Braun M.J., Chojnowski J.L., Hackett S.J., Han K.-L., Harshman J., Heimer-Torres

V. 2009. A well-tested set of primers to amplify regions spread across the avian genome. Mol. Phyl. Evol. 50:654–660.

Knowles L.L., Kubatko L.S. eds. 2010. Estimating species trees: practical and theoretical aspects. Hoboken: Wiley Blackwell.

Kubatko L.S. 2009. Identifying hybridization events in the presence of coalescence via model selection. Syst. Biol. 58:478–488.

Kubatko L.S., Degnan J.H. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56:17–24.

Lamichhaney S., Berglund J., Almén M.S., Maqbool K., Grabherr M., Martinez-Barrio A., Promerová M., Rubin C.J., Wang C., Zamani N., Grant B.R., Grant P.R., Webster M.T., Andersson L. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature 518:371–375.

Lanier H.C., Huang H., Knowles L.L. 2014. How low can you go? The effects of mutation rate on the accuracy of species-tree estimation. Mol. Phylogenet. Evol. 70:112–119.

Latrubesse E.M., Franzinelli E. 2005. The late Quaternary evolution of the Negro River, Amazon, Brazil: Implications for island and floodplain formation in large anabranching tropical systems. Geomorphology 70:372–397.

Leaché A.D., Fujita M.K. 2010. Bayesian species delimitation in West African forest geckos (Hemidactylus fasciatus). Proc. Biol. Sci. 277:3071–3077.

Leaché A.D., Fujita M.K., Minin V.N., Bouckaert R.R. 2014a. Species delimitation using genome-wide SNP data. Syst. Biol. 63: 534–542.

Leaché A.D., Harris R.B., Rannala B., Yang Z. 2014b. The influence of gene flow on species tree estimation: a simulation study. Syst. Biol. 63:17–30.

Li H., Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Liu L., Yu L., Edwards S. V. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol. Biol. 10:302.

Liu L., Yu L., Kubatko L., Pearl D.K., Edwards S. V. 2009a. Coalescent methods for estimating phylogenetic trees. Mol. Phylogenet. Evol. 53:320–328.

Liu L., Yu L., Pearl D.K., Edwards S. V. 2009b. Estimating species phylogenies using coalescence times among sequences. Syst. Biol. 58:468–77.

Liu L., Yu L. 2011. Estimating species trees from unrooted gene trees. Syst. Biol. 60:661–667.

Maldonado-Coelho M. 2012. Climatic oscillations shape the phylogeographical structure of Atlantic Forest fire-eyes (Aves: Thamnophilidae). Biol. J. Linn. Soc. 105:900–924.

Mallet J. 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. 20:29–37.

Mallet J. 2007. Hybrid speciation. Nature 446:279–283.

Mallet J., Besansky N., Hahn M.W. 2016. How reticulated are species? BioEssays 38:140–149.

Manthey J.D., Campillo L.C., Burns K.J., Moyle R.G. 2016. Comparison of target-capture and restriction-site associated DNA sequencing for phylogenomics: A test in cardinalid tanagers (Aves, Genus: Piranga). Syst. Biol. 65:640–650.

Martin S.H., Eriksson A., Kozak, K.M., Manica A., Jiggins C.D. 2015. Speciation in Heliconius butterflies: minimal contact followed by millions of generations of hybridization. BioRxiv doi: 10.1101/015800.

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

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K., Altshuler D., Gabriel S., Daly M., DePristo M.A. 2010. The genome analysis toolkit: A mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303.

Mertes L. A. K., Dunne T., Martinelli L. A. 1996. Channel-floodplain geomorphology along the Solimões-Amazon River, Brazil. Geol. Soc. Am. Bull. 108:1089–1107.

Meyer B.S., Matschiner M., Salzburger W. 2016. Disentangling incomplete lineage sorting and introgression to refine species-tree estimates for lake Tanganyika cichlid fishes. Syst. Biol. 66:531–550.

Mirarab S., Reaz R., Bayzid M.S., Zimmermann T., S. Swenson M., Warnow T. 2014. ASTRAL: Genome-scale coalescent-based species tree estimation. Bioinformatics. 30:541–548.

Morales A.E., Jackson N., Dewey T.A., O'Meara B., Carstens B.C. 2017. Speciation with gene flow in North American Myotis bats. Syst. Biol. 66:440–452.

Nadachowska-Brzyska K., Li C., Smeds L., Zhang G., Ellegren H. 2015. Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. Curr. Biol. 25:1375–1380.

Nakhleh L. 2013. Computational approaches to species phylogeny inference and gene tree reconciliation. Trends Ecol. Evol. 28:719–728.

Nater A., Burri R., Kawakami T., Smeds L., Ellegren H. 2015. Resolving evolutionary relationships in closely related species with whole-genome sequencing data. Syst. Biol. 46:1–54.

Nogueira A.C.R., Silveira R., Guimarães J.T.F. 2013. Neogene–Quaternary sedimentary and paleovegetation history of the eastern Solimões basin, central Amazon region. J. South Amer. Earth Sciences. 46: 89–99.

Oswald J.A., Overcast I., Mauck, W.M.I., Andersen M.J., Smith B.T. 2017. Isolation with asymmetric gene flow during the nonsynchronous divergence of dry forest birds. Mol. Ecol. 26:1386–1400.

Petermann P. 1997. The birds. In: Junk W.J., editor. The central amazon floodplain: ecology of a pulsing system. Berlin: Springer Verlag. p. 419–452.

Pickrell J.K., Pritchard J.K. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8:e1002967.

Poelstra J.W., Vijay N., Bossu C.M., Lantz H., Ryll B., Müller I., Baglione V., Unneberg P., Wikelski M., Grabherr M.G., Wolf J.B.W. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344:1410–4.

Posada D. 2008. jModelTest: Phylogenetic model averaging. Mol. Biol. Evol. 25:1253–1256.

Pritchard J.K., Stephens M., Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

Queiroz K. De. 2007. Species concepts and species delimitation. Syst. Bot. 56:879–886.

Rambaut A., Drummond A.J. 2007. Tracer v1.4, http://tree.bio.ed.ac.uk/software/tracer/ (accessed 9 March 2017).

Remsen J.V., Parcker III T.A. 1983. Contribution of river-created habitats to bird species richness in Amazonia. Biotropica 15:223.

Rannala B., Yang Z. 2013. Improved reversible jump algorithms for Bayesian species delimitation. Genetics 194:245–253.

Reich D., Thangaraj K., Patterson N., Price A.L., Singh L. 2009. Reconstructing Indian population history. Nature 461:489–94.

Rheindt F.E., Fujita M.K., Wilton P.R., Edwards S.V. 2014. Introgression and phenotypic assimilation in Zimmerius flycatchers (Tyrannidae): population genetic and phylogenetic inferences from genome-wide SNPs. Syst. Biol. 63:134–152.

Robinson J.D., Bunnefeld L., Hearn J., Stone G.N., Hickerson M.J. 2014. ABC inference of multi-population divergence with admixture from unphased population genomic data. Mol. Ecol. 23:4458–4471.

Ronquist F., Teslenko M., Van Der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–542.

Rosenberg G.H. 1990. Habitat specialization and foraging behavior by birds of Amazonian river islands in northeastern Peru. Condor. 92:427.

Rosenblum E.B., Sarver B.A.J., Brown J.W., Des Roches S., Hardwick K.M., Hether T.D., Eastman J.M., Pennell M.W., Harmon L.J. 2012. Goldilocks meets Santa Rosalia: an ephemeral speciation model explains patterns of diversification across time scales. Evol. Biol. 39:255–261.

Rossetti D.F., Cohen M.C.L., Tatumi S.H., Sawakuchi A.O., Cremon E.H., Mittani J.C.R., Bertani T.C., Munita C.J.A.S., Tudela D.R.G., Yee M., Moya G. 2015. Mid-Late Pleistocene OSL chronology in western Amazonia and implications for the transcontinental Amazon pathway. Sediment. Geol. 330: 1–15.

Sambrook J., Fritsch E.F., Maniatis T. 1989. Molecular cloning: a laboratory manual, 2nd ed. New York: Cold Spring Harbor Laboratory Press.

Shaw T.I., Ruan Z., Glenn T.C., Liu L. 2013. STRAW: Species TRee Analysis Web server. Nucleic Acids Res. 41:W238–W241.

Smith B.T., McCormack J.E., Cuervo A.M., Hickerson M.J., Aleixo A., Cadena C.D., Pérez-Emán J., Burney C.W., Xie X., Harvey M.G., Faircloth B.C., Glenn T.C., Derryberry E.P., Prejean J., Fields S., Brumfield R.T. 2014. The drivers of tropical speciation. Nature 515:406–409.

Smithe F.B. 1975. Naturalist's color guide. New York: American Museum of Natural History.

Smithe F.B. 1981. Naturalist's color guide. Part III. New York: American Museum of Natural History.

Solís-Lemus C., Knowles L.L., Ané C. 2015. Bayesian species delimitation combining multiple genes and traits in a unified framework. Evolution 69:492–507.

Solís-Lemus C., Yang M., Ané C. 2016. Inconsistency of species-tree methods under gene flow. Syst. Biol. 65:843–851.

Sousa V., Hey J. 2013. Understanding the origin of species with genome-scale data: modeling gene-flow. Nat. Rev. Genet. 14:404–414.

Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312–1313.

Sukumaran J., Knowles L.L. 2017. Multispecies coalescent delimits structure, not species. Proc. Natl. Acad. Sci. USA. 114:1607–1612.

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

Wen D., Yu Y., Hahn M.W., Nakhleh L. 2016. Reticulate evolutionary history and extensive introgression in mosquito species revealed by phylogenetic network analysis. Mol. Ecol. 25:2361–2372.

Wittmann F., Schongart J., Junk W. J. 2010. Phytogeography, species diversity, community structure and dynamics of central Amazonian floodplais forests. In: Junk W. J., Piedade M. T. F., Wittmann F., Schongart J., Parolin P., editors. Amazonian floodplain forests: Ecophysiology, biodiversity and sustainable management. NewYork, NY: Springer. p. 61–102.

Wittmann F., Householder E., Piedade M.T.F., Assis R.L.D., Schöngart J., Parolin P., Junk W.J. 2012. Habitat specifity, endemism and the neotropical distribution of Amazonian white-water floodplain trees. Ecography 36:690–707.

Yang Z., Rannala B. 2010. Bayesian species delimitation using multilocus sequence data. Proc. Natl. Acad. Sci. USA 107: 9264–9.

Yang Z. Rannala B. 2014. Unguided species delimitation using DNA sequence data from multiple loci. Mol. Biol. Evol. 31: 3125–3135.

Yu Y., Nakhleh L. 2015. A maximum pseudo-likelihood approach for phylogenetic networks. BMC Genomics. 16:S10.

Zamudio K.R., Bell R.C., Mason N.A. 2016. Phenotypes in phylogeography: Species ' traits, environmental variation , and vertebrate diversification. Proc. Natl. Acad. Sci. USA 113: 1–8.

Zhang C., Rannala B., Yang Z. 2014. Bayesian species delimitation can be robust to guide-tree inference errors. Syst. Biol. 63: 993–1004.

Zimmer K.J., Isler M.L. 2003. Family Thamnophilidae (typical antbirds). In: del Hoyo J., Elliot A., Christie D. A., editors. Handbook of the birds of the world, Vol. 8. Broadbills to Tapaculos. Barcelona: Lynx Edicións. p. 448–681.