Genomic phylogeography of the White-crowned Manakin *Pseudopipra pipra* (Aves: Pipridae) illuminates a continental-scale radiation out of the Andes

Author Affiliations:

Jacob S. Berv[1,2*], Leonardo Campagna[1,2], Teresa J. Feo[3], Ivandy Castro-Astor[4], Camila C. Ribas[5], Richard O. Prum[6], Irby J. Lovette[1,2]

1. Fuller Evolutionary Biology Program, Cornell Lab of Ornithology, 159 Sapsucker Woods Road, Ithaca, NY 14850, USA.

2. Department of Ecology and Evolutionary Biology, Cornell University, 215 Tower Road, Ithaca, NY 14853, USA.

3. Department of Vertebrate Zoology, MRC-116, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013, USA.

4. Department of Biology, City College of New York and CUNY Graduate Center, City University of New York, New York, NY 10031, USA

5. Coordenacão de Biodiversidade, Instituto Nacional de Pesquisas da Amazônia, Manaus, AM, Brazil

6. Department of Ecology and Evolutionary Biology, and Peabody Museum of Natural History, Yale University, New Haven, Connecticut, 06520, USA.

*Correspondence to: Jacob S. Berv (jsb439@cornell.edu)

Current address: Department of Ecology and Evolutionary Biology, and University of Michigan Museum of Paleontology, 1105 North University Avenue, Biological Sciences Building, Ann Arbor, MI 48109-1085, USA.

***Abstract*:** The complex landscape history of the Neotropics has generated opportunities for population isolation and subsequent diversification that place this region among the most species-rich in the world. Detailed phylogeographic studies are required to uncover the biogeographic histories of Neotropical taxa, to identify evolutionary correlates of diversity, and to reveal patterns of genetic connectivity, disjunction, and potential differentiation among lineages from different areas of endemism. The White-crowned Manakin (*Pseudopipra pipra*) is a small suboscine passerine bird that is broadly distributed through the subtropical rainforests of Central America, the lower montane cloud forests of the Andes from Colombia to central Peru, the lowlands of Amazonia and the Guianas, and the Atlantic forest of southeast Brazil. *Pseudopipra* is currently recognized as a single, polytypic biological species. We studied the effect of the Neotropical landscape on genetic and phenotypic differentiation within this species using genomic data derived from double digest restriction site associated DNA sequencing (ddRAD), and mitochondrial DNA. Most of the genetic breakpoints we identify among populations coincide with physical barriers to gene flow previously associated with avian areas of endemism. The phylogenetic relationships among these populations imply a novel pattern of Andean origination for this group, with subsequent diversification into the Amazonian lowlands. Our analysis of genomic admixture and gene flow establishes a complex history of introgression between some western Amazonian populations. These reticulate processes confound our application of standard concatenated and coalescent phylogenetic methods and raise the question of whether a lineage in the western Napo area of endemism should be considered a hybrid species. Lastly, analysis of variation in vocal and plumage phenotypes in the context of our phylogeny supports the hypothesis that *Pseudopipra* is a species-complex composed of at least 8, and perhaps up to 17 distinct species which have arisen in the last ~2.5 Ma.

Key words: Neotropics, ddRAD, genomics, phylogeography, manakin, suboscine, speciation, introgression, admixture, reticulation, multi-species coalescent, coancestry, isolation by distance, vocalization

1. *Introduction*

Many kinds of geographic and habitat barriers have been proposed to drive population diversification and speciation in the Neotropics (Cracraft and Prum, 1988; Haffer, 1969, 2008; Smith et al., 2014; Wallace, 1854). Such barriers partition biodiversity into areas of endemism (Cracraft, 1985; Cracraft and Prum, 1988; Crother and Murray, 2011; Da Silva et al., 2005; Linder, 2001; Noguera-urbano, 2016) by acting as impediments to gene flow for dispersal-limited organisms (Brumfield, 2012; Cheviron et al., 2005; Fernandes et al., 2015; Moore et al., 2008; Ribas et al., 2012). Three of the most prominent features implicated in structuring the biodiversity of Neotropical forest birds (Figure 1) include the Andes Mountains and other montane regions (Figure 1, grey relief); the Chaco, Cerrado, and Caatinga biomes, which collectively form a "dry diagonal" of open habitat separating the Amazon forest from the Atlantic Forest; and the large rivers of the complex Amazonian drainage system (Brumfield, 2012; Harvey and Brumfield, 2015; Naka and Brumfield, 2018; Smith et al., 2014).

Montane Andean regions in the Neotropics are known to be exceptionally biodiverse, encompassing at least 15 areas of avian endemism with biotas shaped by a combination of vicariance and dispersal events (Hazzi et al., 2018; Morrone, 2015). Hazzi et al. (2018) concluded that the Marañon, Apurimac, and Yungas Inter-Andean Valleys, as well as the Colombian Massif, Las Cruces Pass, and the Táchira Depression have all likely played significant roles in structuring Andean biodiversity. Elevational gradients in the Andes also contribute substantially to Neotropical diversification metrics and raise fundamental questions about the historical relationships between

lowland rainforest and montane endemics in the Neotropics (Janzen, 1967; Musher et al., 2019; Quintero and Jetz, 2018; Weir, 2006). From a biogeographic perspective, avian evolution has proceeded both "into and out of the Andes" (Brumfield and Edwards, 2007; Nylander et al., 2008). A number of examples illustrate colonization of Andean forest regions by lowland forest ancestors (Bates and Zink, 1994; Fjeldså, 1992; Ribas et al., 2007), whereas examples of diversification into the lowlands are so far restricted to open habitat taxa (da Silva, 1995; van Els et al., 2019; Voelker, 1999). For some tanagers (Aves: Thraupidae), the Northern Andes have been a source of lineages that later dispersed into the Central Andes and Amazonian lowlands (Sedano and Burns, 2010). By contrast, Southern Andean forests may have been connected to the Atlantic Forest via Pleistocene expansions of forests into the region occupied today by the Cerrado biome (Cabanne et al., 2019). Indeed, glacial cycles and climatic fluctuations during the Pleistocene have been implicated as an important factor in montane diversification, disproportionally affecting high elevation forest (Hooghiemstra and Van der Hammen, 2004; Weir, 2006) and transiently connecting highland and lowland habitats (Brumfield and Edwards, 2007; Cabanne et al., 2019; Nylander et al., 2008).

In the lowland Amazon basin, a number of areas of endemism have also been described, delimited primarily by major tributaries of the Amazon River (Borges and Da Silva, 2012; Cracraft, 1985; Da Silva et al., 2005; Haffer, 1974). Wallace (1854) initially suggested that the Amazon basin could be divided into four wide bioregions (which he termed "Guyana," "Ecuador," "Peru," and "Brazil"), based on primate distributions (see also Lynch Alfaro et al., 2015). These areas were subsequently partitioned by later biogeographers into at least eight major areas of endemism for terrestrial vertebrates (Figure 1) (Borges and Da Silva, 2012; Cracraft, 1985; Cracraft and Prum, 1988; Da Silva et al., 2002; Da Silva et al., 2005; Haffer, 1978; Haffer, 1985; Haffer, 1992; Naka, 2011). The aggregate of these geographic partitions has been recognized as the "Amazonian areas

of endemism" and is the basis for the riverine barrier hypothesis (Figure 1, Antonelli et al., 2018; also see Figure 1A in Silva et al., 2019).

Few studies of birds have applied next-generation sequencing datasets to reconstruct the biogeographic history of a Neotropical radiation distributed across lowland and highland areas of endemism. Here, we investigate lineage diversification across the Neotropics in the continentally distributed *Pseudopipra* genus. The monotypic genus *Pseudopipra* (Kirwan et al., 2016), family Pipridae, currently includes a single biological species, the White-crowned Manakin (*Pseudopipra pipra* = *Dixiphia*). *Pseudopipra* is found across a variety of known dispersal barriers, including elevational gradients of the Andes, major Amazonian rivers, the dry diagonal, tepuis, and the Isthmus of Panama. Therefore, this group is particularly appropriate for assessing patterns of phylogenomic differentiation across a nested set of spatial scales in Neotropical forests.

In order to assess the historical relationships among and within Andean and lowland Amazonian forest populations, and the degree to which population structure may be related to landscape features, our study design uses fine-scale sampling across a majority of *Pseudopipra's* range. We use double digest restriction site associated DNA (RAD) sequencing (Peterson et al., 2012)—a reduced representation genomic sequencing technique— to sample thousands of independent markers across the genomes of hundreds of individuals spanning many well-known Amazonian and Andean areas of endemism, from Costa Rica to the Atlantic Forest of Brazil (> 80 localities). We use these data to infer the history of lineages and populations within *Pseudopipra* using a variety of population genomic, landscape genetic, phylogenetic, and comparative approaches. Lastly, we investigated congruence between geographic variation in vocal variation, plumage phenotype, and phylogeographic structure, to evaluate the biological mechanisms that may be contributing to population differentiation and gene flow, and to propose a new hypothesis of species delimitation for the genus.

1.1. *Study organism*

The White-crowned Manakin, *Pseudopipra pipra* (Pipridae, hereafter *Pseudopipra*), is a small (10-12 gram), non-migratory, suboscine passerine bird that is broadly distributed within Central America, the lower montane cloud forests of the Andes (up to ~2000m) from Colombia to central Peru the Amazon basin, and the Atlantic Forest (Figure 2) (Kirwan et al., 2016; Kirwan and Green, 2012). The male's striking white crown and jet-black body makes them among the most easily identified manakins, though the grey-green females are often confused with other manakin species in the field. *Pseudopipra* are typically found in dense humid forest and are predominantly frugivorous. Unlike many manakin species that exhibit concentrated or cooperative lek behavior (e.g. Prum, 1990; Prum, 1994), *Pseudopipra* males display in dispersed leks of 2-5 males (Castro-Astor et al., 2007; Snow, 1961). Castro-Astor et al. (2007) described at least 11 components in the display repertoire of the Atlantic forest *Pseudopipra* population, including rapid turning, jumping, "to-and-fro" flights, and an about-face.

A recent phylogenetic hypothesis of manakins based on mitochondrial and nuclear DNA sequences placed *Pseudopipra* as sister group to the five species of *Ceratopipra* (Ohlson et al., 2013). While the higher level phylogeny and taxonomy of manakins has received significant attention (Ohlson et al., 2013; Tello et al., 2009), work exploring intraspecific genetic variation within manakins has been restricted to relatively few species (e.g. Brumfield et al., 2008; Capurucho et al., 2013; Cheviron et al., 2006; Cheviron et al., 2005; Francisco et al., 2007; Gubili et al., 2016; Luna et al., 2017; McDonald, 2003).

Based on geographic variation in plumage and vocalizations, many previous authors have suggested that the biological species *Pseudopipra pipra* likely includes multiple, distinct phylogenetic species (Freile, 2014; Kirwan and Green, 2012; Ridgely and Greenfield, 2001; Ridgely and Tudor, 2009; Snow, 2004; Spencer, 2012). Thirteen subspecies of *Pseudopipra pipra*

119 have been recognized based primarily on subtle variations in plumage coloration, which are
120 frequently more marked in females than in males (summarized in Dickinson, 2003; Kirwan and
121 Green, 2012; Snow, 2004). Because suboscine passerines generally do not learn their songs (but see
122 Saranathan et al., 2007), the presence of substantial vocal variation across *Pseudopipra* populations
123 further suggests that the genus may contain unrecognized cryptic species (e.g. Campagna et al.,
124 2012). However, all previous authors awaited new information on genetic differentiation within
125 *Pseudopipra* before making taxonomic recommendations.

126 It also remains unclear how the current subspecific classification reflects evolutionary
127 history. The genetic variation within *Pseudopipra* has received prior attention by two studies that
128 used mitochondrial DNA to infer population genetic patterns. Milá et al. (2012) reported
129 intraspecific divergences of up to 3.5% (n=19) across three Amazonian populations, and the highest
130 observed nucleotide diversity ($\pi = 0.266$) among 14 widely distributed Amazonian birds species.
131 Castro-Astor (2014) used a larger sample (n=57), and discovered that at least four *Pseudopipra*
132 subspecies correspond to well supported mitochondrial clades which were generally congruent with
133 Amazonian areas of endemism. Castro-Astor (2014) used a molecular clock model to estimate the
134 age of the *Pseudopipra* complex to coincide with the onset of the Pleistocene ~2.457 Ma (95% HPD
135 1.45-3.97).

136

137 **2. *Material and methods***

138 2.1. *Field and tissue sampling*

139 Muscle tissue samples were obtained from available vouchered avian material from US and
140 Brazilian collections and other institutions (see Acknowledgements and Supplemental Table 1). We
141 sampled from ten of thirteen recognized subspecies (Figure 2) of *Pseudopipra* (Kirwan and Green,
142 2012). Unfortunately, we were unable to obtain material of adequate quality for ddRAD sequencing

143 from the Andes of Colombia (*bolivari*, *unica*, *coracina*, *minima*), Ecuador (*minima*) and the
144 lowlands of Peru (*pygmaea*). However, we were able to obtain mtDNA data from *coracina* and
145 *pygmaea*, which allowed us to make a preliminary assessment of their phylogenetic affinities (see
146 discussion on mtDNA). In sum, after discarding failed samples and samples with high proportions
147 of missing data, we obtained genetic data from 286 individuals (234 from ddRAD, 168 from
148 mtDNA), representing ~80 localities and 10 of 13 subspecies. We also obtained comparable ddRAD
149 and mtDNA data for two specimens of *Ceratopipra rubrocapilla* as outgroups (Supplementary
150 Table 1).
151
152 2.2. *Laboratory methods and data processing*
153 We extracted DNA from avian tissue specimens (Supplementary Table 1) using the DNeasy
154 Blood & Tissue Kit (QIAGEN, CA) and generated ddRADtags following the protocol of Peterson
155 et al. (2012), with modifications described in Thrasher et al. (2018). We used the STACKS 1.44
156 bioinformatics pipeline (Catchen et al., 2013) to generate assemblies of ddRAD loci, using
157 *Manacus vitellinus* as a reference genome (GCA_001715985.1 from www.ncbi.nlm.nih.gov). We
158 obtained mitochondrial ND2 sequences for a set of individuals using standard Sanger Sequencing
159 protocols described in Berv and Prum (2014). Additional sequencing, assembly, and alignment
160 details are reported in the Supplemental Appendix.
161 For ddRAD loci, we used several filters in the "populations" module from STACKS to
162 generate different sets of bi-allelic SNPs: a missing data filter, a minimum depth of coverage filter,
163 a filter that exports only the first SNP per RAD locus, and a minor allele frequency filter (MAF). In
164 order to explore the sensitivity of genetic cluster inference to these important parameters (Linck and
165 Battey, 2017; Paris et al., 2017), we produced eight versions of the SNP dataset by combining these
166 different standard filters (Table 1).

167 Dataset 1 (D1- labeled arbitrarily) included 1,960 SNPs that were generated by setting the
168 MAF to >= 5%, requiring minimum coverage of 20x, requiring the presence of a SNP in at least
169 80% of individuals, and filtering out all but the first SNP in each RAD locus to minimize linkage
170 across the dataset. To evaluate the sensitivity of cluster analyses to rare variants (Linck and Battey,
171 2017; Shafer et al., 2017), Dataset 2 (D2) was exported with identical parameters to dataset D1, but
172 without a minor allele frequency filter; this generated a dataset of 2,581 putatively unlinked SNPs.
173 We generated datasets D3 and D4 to reflect the same filters as D1 and D2 but to also include all
174 SNPs from each locus. Lastly, for each dataset D1-D4, we generated a complimentary dataset which
175 filtered out any sites with missing data. The parameters of these SNP datasets are summarized in
176 Table 1.
177
178 2.3. *Phylogenetic analysis*
179 Because sparse super-matrices may be more prone to systematic error (e.g. Roure et al.,
180 2013), we explored the sensitivity of our phylogenetic inferences to increasing amounts of missing
181 data using three different haplotype datasets (separate from SNP Datasets) with 20%, 50%, and
182 80% thresholds for missing data at the RAD locus level. For example, for an 80% threshold, if the
183 RAD locus is present in at least 20% of individuals, then that locus, containing all variable and
184 invariable sites, was exported. This approach generated datasets comprising 2,584, 4,763, and 7,901
185 RAD loci, respectively. For each dataset, we generated a concatenated supermatrix using Sequence
186 Matrix (Vaidya et al., 2011). These combined datasets had up to ~40% missing sites across an entire
187 matrix, irrespective of columns or rows (i.e., using the 80% threshold above).
188 We estimated phylogenies using each dataset with a concatenated maximum likelihood and
189 a coalescent species-tree approach for a total of six phylogenetic analyses. For analysis of
190 concatenated data, we estimated a maximum likelihood phylogenetic hypothesis using RAxML

8.2.9 (Stamatakis, 2014). We estimated 500 bootstrap replicates followed by a full ML search for each dataset, using a GTRGAMMAX substitution model. We avoid the need for ascertainment bias correction because we use whole short-read haplotype alignments including invariant sites (Leaché et al., 2015). In our case, each locus alignment was 138bp long (supplemental Appendix). We also tested for an effect of data partitioning on phylogenetic inference by generating optimized locus partitioning schemes in PartitionFinder 2 (Lanfear et al., 2017), which were then re-analyzed in RAxML. We found that our topological results were insensitive to data partitioning, and as such we do not present these results. Lastly, we generated a mitochondrial DNA ND2 gene tree using IQ-TREE 1.6.10 (Chernomor et al., 2016; Hoang et al., 2017; Kalyaanamoorthy et al., 2017; Schmidt et al., 2014; Trifinopoulos et al., 2016). We partitioned by codon position and generated a maximum likelihood tree using the MFP+MERGE model search and partitioning option (Supplementary Figure 17).

We evaluated coalescent phylogenetic structure using the SVDquartets method implemented in PAUP* 4.0a159 (Chifman and Kubatko, 2014, 2015; Swofford, 2002). SVDquartets may be a useful tool for testing hypotheses of species delimitation because it can generate a lineage-tree under the multispecies coalescent without assigning individuals to species *a priori*. In this approach, each individual in a dataset is considered to be a separate "species." If a recovered clade in such a lineage-tree has high bootstrap support, this would mean that the descendant taxa have strong support as a group under the multispecies coalescent, which may coincide with species boundaries (*Personal communication*, Laura Kubatko). Our SVDquartets analysis evaluated all possible quartets with 500 bootstrap replicates (other settings left to defaults). SVDQuartets has the advantage of being able to work directly on sequence data (above) and has been shown to perform well in comparison to summary methods which assume fully and correctly resolved gene trees as input (Chou et al., 2015; Schmidt-Lebuhn et al., 2017). SVDquartets has also been shown to

215 produce reliable results under very general conditions, and is theoretically robust to variation in
216 effective population size, molecular clocks, limited gene flow, and incomplete lineage sorting (Long
217 and Kubatko, 2017).

218 While a full exploration of our specific usage of SVDquartets is beyond the scope of the
219 present work, we suggest that this approach may be a useful way to assess support for hypotheses of
220 species delimitation under the multispecies coalescent. We do not take the perspective that all well
221 supported clades from this coalescent assessment *necessarily* represent biological species – such an
222 interpretation often leads to an over-inflation of species estimates (Chambers and Hillis, 2020;
223 Hillis, 2020; Sukumaran and Knowles, 2017). In general, we prefer a species agnostic approach to
224 those which require *a priori* delimitation of species boundaries. Our approach instead focuses on
225 discovering well supported coalescent structure without a prior hypothesis of species delimitation
226 and then assesses the degree to which metrics of support across coalescent and concatenated tree
227 inference methods are concordant. We note that there are many approaches for species delimitation
228 using genetic data (e.g. BP&P, Flouri et al. (2018), BFD, Leaché et al. (2014), PTP, Zhang et al.
229 (2013), also see Leaché et al. (2018)), and that there is lack of consensus on how best to interpret
230 these analyses with respect to various species concepts (Carstens et al., 2013). In the present study,
231 we instead rely on analysis of vocal and plumage phenotypes in combination with genetic patterns,
232 to make taxonomic recommendations.

233

234 2.4. *Descriptive statistics*

235 We estimated global and population pairwise F_{st} (Weir and Cockerham, 1984), regional
236 variation in allelic richness (Goudet, 2005), Hardy-Weinberg equilibrium (Paradis, 2010),
237 inbreeding (Goudet, 2005), linkage disequilibrium (Kamvar et al., 2014), sex-biased dispersal
238 (Goudet et al., 2002), and an analysis of molecular variance (Kamvar et al., 2014).

239 Some of these statistics and analytical details are reported as supplementary material, while others
240 are discussed below when relevant.
241
242 2.5. *Cluster inference and patterns of genomic admixture*
243 We performed cluster analyses intended to quantify the number of similar genetic groups
244 across the range of *Pseudopipra.* First, we evaluated broad-scale variation across our genomic data
245 using phenetic K-means clustering implemented in the find.clusters function in the adegenet R
246 package (Jombart, 2008). This function first transforms the data with a PCA, and then identifies
247 clusters with a K-means algorithm (MacQueen, 1967). Using each SNP dataset, we evaluated K1:20
248 successively with the Bayesian Information Criterion (BIC). K-means clustering is based on the
249 idea of minimizing the variance within clusters and has the potential advantage of making few
250 assumptions about the underlying processes which generated the data. For each analysis, we
251 retained all PC axes, set 200 randomly chosen starting centroids, and ran each search for 10^9
252 iterations.
253 We also estimated the number of differentiated populations using STRUCTURE 2.3.4
254 (Pritchard et al., 2000), a widely-used model-based approach which assigns individuals to
255 populations such that Hardy-Weinberg equilibrium is maximized and linkage among loci is
256 minimized within groups. By comparing the output of STRUCTURE runs fixed to different
257 numbers of populations (K), one can assess the degree of fit of the data to various models of K and
258 assess genomic admixture among individuals. We ran STRUCTURE using SNP dataset D1 (Table
259 1), from K1:20 with 10 iterations per K value. We ran the program for 700,000 generations,
260 discarding the first 200,000 as burn-in, implementing the admixture ancestry model with correlated
261 allele frequencies. We evaluated K by examining $\Pr(X|K)$ or $L(K)$ after summarizing our results
262 using Structure Harvester (Earl and vonHoldt, 2012) and CLUMPP (Jakobsson and Rosenberg,

263 2007). We also used the Evanno method to estimate the rate of change in the log probability of data
264 between values of K, which is a useful metric to detect the uppermost hierarchical level of structure
265 in the data (Evanno et al., 2005). Lastly, we generated 95% Bayesian Credible Intervals around
266 admixture coefficients (presented as averages across individuals for population assignment).

267 To investigate finer scale population genetic differentiation and the genomic composition of
268 populations defined at different hierarchical levels, we used the programs fineRADstructure and
269 RADpainter (Malinsky et al., 2018). RADpainter takes advantage of the information in haplotype
270 data and considers each individual in a dataset as a "recipient" whose genomes are reconstructed
271 using chunks of "donor" DNA from all other available individuals. This approach generates a
272 "coancestry matrix" which combines the information that can be derived from both PCA and
273 model-based clustering approaches and thus can be more sensitive to subtle population structure
274 (Lawson et al., 2012). The fineRADstructure software uses an MCMC approach to explore the
275 space of population assignments based on the coancestry matrix, using an algorithm which merges
276 and splits populations, or moves individuals among populations. We ran fineRADstructure with
277 default priors, but increased the burn in to 200000 iterations, followed by 1000000 iterations,
278 sampling every 1000. We then assessed convergence by 1) considering the assignment of
279 population membership across multiple independent runs, 2) visualizing the MCMC traces of
280 estimated parameters to ensure convergence on the same posterior distributions, and 3) running
281 each chain long enough to achieve effective parameter sample sizes > 100.

282 Both RADpainter and fineRADstructure are based on the recently developed chromoPainter
283 and finestructure software (Lawson et al., 2012), but are optimized to take advantage of the linkage
284 properties of RAD-like datasets. As our data are assembled relative to a *Manacus* reference
285 genome, the order in which loci appear in our data files are related to positions on assembly

286 scaffolds, even though the ddRAD loci in our data are generally unlinked when considered at the
287 population level (see Supplementary Appendix).
288 We compared the inferred population assignments from fineRADstructure and
289 STRUCTURE with K-means clustering (as above) of the coancestry matrix as well as SNP datasets.
290 To determine how much of the variation in our dataset was captured by the RADpainter analysis
291 relative to a standard PCoA of SNP data, we used the normalized PCA approach of (Lawson et al.,
292 2012), using the mypca function provided in the "FinestructureLibrary.R" library in the
293 fineRADstructure package.
294
295 2.6. *Phylogenetic reticulation in the western Napo lineage*
296 Model-based, phenetic, and population genetic analyses revealed that a lineage of
297 individuals presently restricted to the western Amazonian Napo area of endemism has a complex
298 history of introgression between southern and northern Amazonian clades (see Results). To further
299 investigate this subset of individuals and their history, we generated two additional versions of the
300 dataset optimized for different analyses. First- we generated a haplotype dataset of 52 individuals,
301 which included all putatively introgressed western Napo individuals (n=10, clade E), as well as
302 samples from proximate neighbor populations. These populations consist of 1) a geographically
303 adjacent population inferred to have mostly unadmixed ancestry (Southwestern Amazon: Inambari,
304 n=17, clade C1), and 2) the implied source of introgressed genotypes (north-central Amazonian Jaú
305 + eastern Napo, n=25, clade D2).
306 We used a 5% minor allele frequency cutoff, required 20x minimum coverage, and allowed
307 up to 20% missing data per locus—generating a dataset comprising 2,370 loci, including 4,979
308 SNPs (Dataset H1). Dataset H1 is analogous to that generated for coancestry (fineRADstructure)
309 analysis of the entire dataset (above). We also generated a similar dataset (Dataset H2), without a

310 minor allele frequency cutoff—comprising 2,947 loci including variable and invariable sites—as
311 required for inference of demographic parameters (Gronau et al., 2011).

312 Hybrid populations are expected to have higher mean coancestry and lower F_{st} with each of
313 their progenitor lineages than their progenitor lineages will have with each other (Barrera-Guzmán
314 et al., 2018). To test this prediction, we first estimated coancestry values using Dataset H1 and
315 RADpainter (as above). We then compared inferred coancestry of population pairs with a standard
316 ANOVA and the glht function in the multcomp R package (Hothorn et al., 2013). We estimated F_{st}
317 with Weir and Cockerham's estimator (Weir and Cockerham, 1984), and tested the prediction using
318 1000 bootstrapped datasets to estimate 95% confidence intervals.

319 To test whether or not the relationships among these focal populations is best represented by
320 a phylogenetic network, we used PhyloNet 3.6.4 (Wen et al., 2018; Zhu et al., 2018). PhyloNet
321 takes SNP data and uses a reversible-jump MCMC technique to explore the posterior distribution of
322 phylogenetic networks and bi-furcating topologies, while accommodating both reticulation and
323 incomplete lineage sorting. PhyloNet searches the set of all possible reticulation models without
324 requiring *a priori* model specification.

325 For computational tractability, we used SNP dataset D1 (1960 unlinked SNPs), and pruned it
326 to the 52 individuals comprising the groups of interest. We pruned out all sites with missing data,
327 leaving 572 unlinked biallelic SNPs, as required by PhyloNet. For our final runs of the PhyloNet
328 program MCMC_BiMarkers, we randomly subsampled five diploid individuals from each group, as
329 preliminary program runs indicated a severe computational bottleneck with more than five diploid
330 individuals or with larger datasets. We assigned each group of individuals to a lineage, allowed for
331 a maximum of one reticulation, set the chain length for 800000 iterations (sampling every 100), set
332 a burn in of 2.5%, and left all priors as default (i.e. assuming population mutation rates are
333 constant).

Using the final network topology estimated from PhyloNet, we estimated demographic parameters using G-PhoCS version 1.3 (Gronau et al., 2011). We ran the program for 750,000 iterations with a 10% burn-in, estimating a total of 13 parameters: the effective population sizes of the three focal lineages, two ancestral population sizes, two splitting times, and six directional migration parameters. Parameter MCMC traces were inspected in Tracer 1.7.1 (Rambaut et al., 2018). We converted the median and 95% Bayesian credible intervals from mutation scale to generations and individuals as described in Campagna et al. (2015), assuming an approximate mutation rate of 10^{-9} per base pair per generation (Kumar and Subramanian, 2002; Smeds et al., 2016). We calculated the number of migrants per generation as the product of the per generation migration rate multiplied by ¼θ for the receiving population ($m_{a\text{-}b} * (\theta_b/4)$) (Gronau et al., 2011), and considered migration rates to be significant if the 95% lower bound was > 1 (the level of gene flow expected to counteract the effects of drift; Wright, 1931).

2.7. *Spatial distribution of genetic variation*

Dispersal barriers often create sharp genetic discontinuities between adjacent populations that would otherwise be continuous (Petkova et al., 2015). To assess the degree to which spatial variation in genetic diversity within *Pseudopipra* may be attributed to landscape features, we investigated how discrete and continuous patterns of isolation-by-distance (IBD) vary across the landscape. First, we used Mantel and partial Mantel tests to investigate IBD effects within sampling regions delimited by prominent physical barriers, and then separately tested the roles of specific dispersal barriers in structuring genetic variation. Mantel tests evaluate the correlation among two or more matrices, in our case representing pairwise genetic and geographic distance. Significance is assessed by permuting the rows and columns of one of these matrices (Mantel, 1967). Partial Mantel tests incorporate a third "barrier matrix," which contains information about environmental

358 or ecological distance. By evaluating the correlation between genetic and ecological distance while
359 controlling for geographic distance, a partial Mantel test can be used to investigate the effect of a
360 particular geographic barrier on genetic differentiation. Because mantel tests may have an inflated
361 type I error rate (e.g. Bradburd et al., 2018; Guillot and Rousset, 2013; Legendre and Fortin, 2010),
362 we follow the recommendations of Diniz-Filho et al. (2013), and only reject the null hypothesis of
363 no correlation if $p < 0.001$ (~10x more stringent than a standard correction for multiple tests would
364 require).

365 For these tests and others, we generated a pairwise geographic distance matrix by
366 calculating great circle distances using the distm function in the geosphere R package (Hijmans et
367 al., 2015). Individuals sampled at the same GPS coordinates were set to have an arbitrarily small
368 geographic distance of 0.0001 meters. In order to assess spatial variation *within and between*
369 populations, we use estimates of pairwise genetic distance at the individual level, rather than at the
370 population level. Our analyses compare patristic phylogenetic distance between all pairs of
371 individuals (estimated from RAxML) as well as coancestry (from fineRADstructure), as proxies of
372 genetic distance among individuals. These procedures, including the generation of an appropriate
373 barrier matrix, are developed in a set of R functions and scripts that operate on adegenet GENIND
374 objects (see supplementary R code) and internally use the Mantel implementation from the ecodist
375 R package (Goslee and Urban, 2007).

376 We also analyzed continuous patterns of spatial variation in our data with a population
377 genetic model that relates effective migration rates across geographic space to expected genetic
378 dissimilarities among individuals (Petkova et al., 2015). This recently proposed method, termed
379 Estimated Effective Migration Surfaces (EEMS), produces a visualization that emphasizes
380 deviations from isolation by distance to discover migratory corridors and barriers to gene flow. This
381 method is complimentary to more explicit hypothesis-testing approaches (such as Mantel tests),

because it seeks to identify where the assumption of constant IBD is violated in geographic space, and thereby can highlight areas where strong barriers to gene flow may exist. In brief, a region is covered with a dense regular grid connecting subpopulations (demes), among which individuals can migrate with rates varying by location. EEMS uses an MCMC approach to estimate expected genetic dissimilarity between two individuals, integrating over all possible migration histories, and adjusting migration rates among graph edges to match the genetic differences in the data. A migration surface is then interpolated across a region to indicate where genetic similarities decay faster than what a constant isolation by distance model predicts, visually highlighting barriers to gene flow and migration corridors. An effective diversity parameter is also estimated for every deme, reflecting local deviations in heterozygosity. We ran several analyses using default priors, iteratively increasing the grid density and chain length, with up to 2000 estimated demes. For each test, we ran three chains, and after examining chain convergence, we assessed model fit by comparing the observed and fitted dissimilarity between and within demes.

2.8. *Vocal variation*

We obtained sound recordings of vocalizations of *Pseudopipra* from the Macaulay Library at the Cornell Lab of Ornithology (https://www.macaulaylibrary.org) and the Xeno-canto (www.xeno-canto.org) collections (Supplementary Table 4). Vocalizations were analyzed in Raven Pro 1.4 (Bioacoustics_Research_Program, 2011), and converted into spectrograms using a 512-sample Hann window with 50% overlap. Audio recordings ranged from a cut of a single vocalization to longer recordings of a lek with multiple individuals vocalizing. For our initial assessment, we identified qualitatively distinct vocalization types based on our own survey of the available recordings, without regard to the geographic location of the recordings or subspecies identity.

406 *Pseudopipra* vocalizations have 1-3 buzzy and/or tonal notes. We measured: 1) starting
407 frequency, 2) ending frequency, 3) minimum frequency, 4) maximum frequency, 5) number of
408 notes, and 6) duration of the entire vocalization (Supplementary Figure 13). We performed principal
409 components analysis (PCA) and logistic regression on these vocal measurements to test for
410 significant differences between vocal types and to reduce the dimensionality of the data for
411 comparison to results from analysis of genetic data. Because no sound records were directly
412 associated with genetic samples in this study, we used geographic proximity to vocalization
413 recordings and localization to areas of endemism or areas bounded by clear physical barriers to
414 associate vocal types to genetic samples and taxonomic ranks. This approach assumes that
415 genetically and geographically proximate individuals are likely to share the same vocal type, and
416 enabled us to perform a preliminary assessment of how variation in vocalization type maps onto
417 existing genetic variation. Testing the fine-scale association of genetic and vocalization boundaries
418 will require extensive field sampling from individual manakins. Additional analytical details and
419 analyses are reported as supplementary material.
420

421 **3. Results**

422 3.1. Genetic *data collection*

423 Our total genomic data set included an average of 1 ± 0.4 million sequences of 138 bp from
424 232 individuals of *Pseudopipra* and two *Ceratopipra rubrocapilla,* which were used as an outgroup.
425 Our reference-based assembly produced a catalogue with 47,046 RAD loci. Filtered SNP datasets
426 requiring a SNP to be present in at least 80% of individuals ranged from 1,960 unlinked SNPs
427 (dataset D1: 5% MAF with 1st SNP on each locus) to 41,024 SNPs (dataset D4: all SNPs on all loci,
428 no MAF) (Table 1). Datasets requiring a SNP to be present in 100% of individuals were
429 considerably smaller, as expected, and ranged from 175 SNPs in D1mod, to 3,832 SNPs in D4mod

(Table 1). We also obtained 168 mtDNA ND2 sequences using PCR methods which broadly overlapped with our ddRAD dataset. Our mtDNA dataset included two subspecies, *pygmaea* and *coracina*, from which we were not able to obtain enough high-quality material for ddRAD sequencing.

3.2. *Phylogenetic analyses*

RAxML and SVDquartets analyses generated a consistently well-supported phylogenetic hypothesis (Figure 3). The monophyletic lineages of *Pseudopipra* have distributions that are largely congruent with previously recognized avian areas of endemism, and further reveal much finer scale patterns of local biogeographic differentiation. Using the two *C. rubrocapilla* individuals to root the phylogeny, we detected support for at least five major clades across all phylogenetic analyses of nuclear genomic data, with consistently high bootstrap support in both concatenated and coalescent analysis (clades A-E in Figures 3-5). We first describe the structure of these major clades, followed by descriptions of well supported substructure within each of these regional groupings. We indicate six bootstrap values for each clade with at least some high support, from each of the three datasets analyzed in RAxML and SVDquartets (see heatmaps, Figure 3). Phylogenetic analysis of mtDNA (Supplemental Figure 17) generated a congruent result, except where noted below.

3.2.1 *Phylogenetic structure*

The sister group to all the other *Pseudopipra* is a clade including individuals from subtropical Central America and the northern Andes. The northern Andes are represented in our ddRAD dataset by a single individual from subtropical forests (~1,000 m) of the eastern slopes of the Andes in San Martín, Peru. This subtropical San Martín specimen (identified as *P. pipra occulta*) was the only viable sample in our genomic dataset from the entire northern Andes region

(A1 in Figure 3). Mitochondrial data from three additional specimens from the Andes of southern Ecuador (subspecies *coracina*) form a monophyletic group with the San Martín specimen; thus, this northern Andean sister lineage to the Central American lineage likely includes all montane Andean populations of *Pseudopipra* in Ecuador, Colombia, and Venezuela. A well supported Central American clade (A2 in Figure 3) was identified in all analyses, with Costa Rican (A3 in Figure 3) and Panamanian subclades (A4 in Figure 3) supported in RAxML.

The second successively nested clade in *Pseudopipra* includes individuals from subtropical forests of the Andes of Southern Peru, which are sister to all remaining lineages (Clade B in Figure 3). The southern Andean Peru clade is further subdivided into two very well resolved clades: a northern clade (B2 in Figure 3), sampled from the southern Cordillera Azul in southwestern Loreto, and a southern clade from Pasco, Junín, and Cusco (labeled B1 in Figure 3). This southern Andean Peru clade is the sister group to a monophyletic, lowland clade (F in Figure 3) that is further subdivided into a southern Amazon clade + Atlantic Forest clade (Clade C in Figure 3), and a Northern Amazonian + Guiana Shield clade (Clade D in Figure 3). Our analysis of mtDNA with sampling of subspecies *pygmaea*, from the tropical forest of lower Rio Huallaga Valley, Peru, places this population as sister group to the entire lowland radiation (Clade F; Supplemental Appendix).

Within the broad lowland Northern Amazonian + Guiana Shield clade (Clade D), virtually no phylogenetic substructure is associated with strong bootstrap support. The maximum likelihood tree however does have substantial biogeographic coherence, which we use as additional justification for population subdivision when monophyletic groups are identified to be coincident with geographic features. For example, all but one individual from coastal Suriname and Amapá form a monophyletic assemblage (D1 in Figure 3), east of the Essequibo river. Likewise, all but one individual from the far eastern Napo area of endemism, east of the Rio Putumayo, form a clade (D2

in Figure 3). This coherence with landscape features is unlikely to be coincidental, though support for these groups is weak. We treat clade D1 and D2 as well as two other Guianan regions, as separate populations for descriptive population genetic statistics.

A widely distributed set of individuals encompassing the western portion of the Napo area of endemism (from the northern bank of the confluence of the Rio Solimões and the Rio Napo north of Iquitos, then west and south) was recovered as the sister group to the Guiana Shield + Northern Amazon clade (Clade E in Figure 3). Analysis of mtDNA (Supplementary Figure 17) further resolved (BS > 99) two subgroups delimited by the Rio Napo. Importantly, and in contrast to nuDNA, we infer that these mtDNA haplotypes are derived from Inambari lineages (BS = 98, Clades C1 + C3 in Figure 3), and are not sister to Clade E. These conflicting patterns are explained by subsequent population genetic analyses, which indicate that this Napo population is a distinct lineage of the southern Amazonian clade (Clade C in Figure 3) that has experienced substantial hybrid introgression from the northern Amazonian clade D (see Population Genetic analyses and Discussion).

Within the large lowland Southern Amazonia clade (Clade C in Figure 3), we recovered eight moderately supported and hierarchically nested clades, which are successive sister groups to each other, and are congruent in distribution with recognized areas of endemism. First is a moderately supported clade (C1 in Figure 3) of individuals from the western Inambari area of endemism, from west of the Rio Ucayali to the Rio Purus. Although the first lineages to diverge within clade C1 have low bootstrap support, a monophyletic subgroup of three montane (>1000m) individuals from the highlands between the Rio Huallaga and Rio Ucayali cluster together with high bootstrap support (C2 in Figure 3).

The next successively nested clade includes individuals from eastern Inambari, between the Rio Purus and the Rio Madeira (C3 in Figure 3). The next successive clade includes two individuals

sampled from the Rondônia area of endemism, between the Rio Madeira and Rio Tapajós (C4 in Figure 3). The next successive clade includes a group of individuals from the Tapajós area of endemism between the Rio Tapajós and the Rio Xingu (C5 in Figure 3). Next, a clade comprised of individuals from the northern Xingu area of endemism, east of the Rio Xingu (C6 in Figure 3), is sister to a well-supported Brazilian Atlantic Forest clade (C7 in Figure 3), which is further subdivided into three successive clades (north to south) in Bahia, Espírito Santo, and Rio de Janeiro, respectively (C8, C9, and C10 in Figure 3). In summary, the subdivisions of the Southern Amazonian clade reflect a stepwise west-east progression, with a hierarchically nested structure: (West Inambari (East Inambari (Rondônia (Tapajós (Xingu (Bahia (Espírito Santo (Rio de Janeiro))))))).

3.3. *Cluster inference and patterns of genomic admixture*

3.3.1. *Broad-scale population structure and admixture*

Weir and Cockerham's F_{st} was moderate when comparing populations within *Pseudopipra*; for Dataset 1, overall F_{st} was 0.195 [95% CI: 0.187-0.203], and never dipped below 0.171 [0.168-0.175] (Dataset 4). Notably, in SNP datasets filtered to remove sites with missing data, overall F_{st} estimates were on average ~20% higher and ranged from 0.188 [0.175-0.201] (D4) to 0.261 [0.191-0.336] (D2). Pairwise population F_{st} estimates from Dataset 1 (other datasets showed similar patterns, see Supplementary Figure 10a-h), ranged from essentially undifferentiated, to almost entirely distinct. At the extremes: comparing the geographically proximate Jaú and Guiana Shield populations indicated a particularly low F_{st}: 0.005 [0.003 - 0.007]. By contrast, comparing the Atlantic Forest Espírito Santo population to the Panama population recovered an F_{st} of 0.81 [0.79 - 0.83], or almost entirely differentiated. For reference, Hartl and Clark (1997) indicate $F_{st} > 0.15$ may be considered "great" genetic differentiation.

K-means clustering of SNP datasets on the basis of BIC score identified groups of individuals that differed somewhat with respect to dataset filtering strategies (Summarized in Figure 5, right). For D1, K-means clustering identified five major groups: Central America (Clade A2), South Andean Peru (Clade B), Atlantic Forest (Clade C7), Southern Amazon including the western Napo population (the rest of Clade C + Clade E), all Guiana Shield + Northern Amazon (Clade D) (Supplementary Figure 2a-h). For dataset D1, the montane sample from San Martín (North Andean Peru, 5444.PE.MAR) clustered with Central American populations (Clade A), as it does in all phylogenetic analyses. These clusters are essentially similar to those detected by STRUCTURE at K=5 (Figure 4, 5, below). All other SNP datasets grouped the San Martín specimen with the Southern Amazon populations. D1m further identified the Tapajós + Xingu individuals as a distinct cluster, while D2, D3, D3m, D4, and D4m generated sets of clusters identical to D1, with the exception of the clustering of the San Martín specimen, which clustered with the Southern Amazon in those datasets. Additional field sampling of the North Andean population is required to understand why this lineage may share signals of ancestry with northern and southern groups. K-means clustering of D2m failed to discriminate any lowland Amazonian populations (Supplementary Figure 2d). Thus, only 75 low frequency SNPs which are removed in D1m but are present in D2m obfuscate the signal of population differentiation detected by K-means clustering of D1m. Lastly, across datasets, standard PCoA explained 13-26% of the variance in the SNP data on the first two axes (Supplementary Figure 2).

Visual inspection of $\Pr(X|K)$ from summarized STRUCTURE runs on dataset D1 indicated that the likelihood of each successive K from 1:20 plateaued at K = 5, with the standard deviation across runs increasing rapidly after this point (Supplementary Figure 3). We describe patterns of inferred admixture as they appear in Figure 5, reading from top to bottom. For K5, individuals from Central America are unambiguously assigned to their own cluster (Clade A2 in Figure 3, 5). The

north Andean montane specimen from San Martín, (North Andean Peru, 5444.PE.MAR) is reconstructed as being highly admixed, with a genomic composition including 4 of the 5 inferred clusters. However, this inference is tenuous because we have only one individual to represent this likely distinct and highly diverse north Andean lineage (though see RADpainter below, which is consistent with these results). Samples from the southern Amazon and Atlantic Forest (Clade C), have no detected admixture with northern populations, except for some individuals from the western (Clade C1), and to a more limited degree, the eastern (Clade C3) Inambari populations, which are inferred to have limited (~<5%) admixture with individuals assigned to the Northern Amazon + Guiana Shield (Clade D, see Figure 5, 6, and discussion below).

Going from West to East across the Southern Amazon, very limited admixture with Atlantic Forest alleles (green, Clade C7) is first inferred in the Tapajos (Clade C5) area of Endemism. This admixture component increases in the Xingu area of endemism (Clade C8) and increases further in a geographic and genetic cline to 100% assignment probability to a distinct cluster in the Rio de Janeiro clade (Clade C10). Given the geographic context of this cline in admixture, we suggest this pattern may a product of isolation by distance, hierarchical structure, and geographic barriers (Bradburd et al., 2018), and not isolated introgression events. We note that the conStruct method (Bradburd et al., 2018) may be useful for explicitly testing this hypothesis (Supplemental Appendix), however our evaluation of this method suggested that it is presently intractable for a dataset of this size.

Next, all individuals from the western portion of the Napo area of endemism are inferred to be heavily admixed (Clade E; Figure 5). On average, at K=5, STRUCTURE assigns 40.2% of the these individuals' (n=10) genetic composition to the Guiana Shield population (SD 4.7%, average Bayesian credible interval (aBCI): [0.33-0.47]), 50.4% to the Southern Amazon population (SD 3.4%, aBCI: [0.43-0.58]), as well as limited 4.15% probability to Central Peru (SD

3.1%, aBCI: [0.01-0.08]), and 5.2% to Central America (SD 4.8%, aBCI: [0.02-0.1]). The three localities included in this clade– (1) north of Iquitos, on the east bank of the confluence of the Rio Solimões and the Rio Napo, (2) southwest of Iquitos in Allpahuayo Mishana Reserve (represented by three mtDNA samples), and (3) near San Jacinto, Ecuador– indicate that this heavily admixed population (corresponding to subspecies *discolor*) may occur across a large area of the western Napo region of endemism, and therefore may not represent a narrow hybrid zone (Figures 2-7).

Lastly, at K=5, all members of the Guiana Shield + Northern Amazon clade (Clade D) are inferred to be virtually unadmixed. Individuals in the far eastern Napo area of endemism on the east bank of the Rio Putumayo (brown dots in Figure 3, 4) cluster with other Guiana Shield groups in phylogenetic analyses but are detected to have limited (~5%) admixture with pure individuals of Southern Amazon provenance. At higher values of K, the broad-scale population assignments inferred at K=5 are mostly unchanged, however additional admixture components are inferred for most groups. The introgressed western Napo clade is eventually placed into its own cluster at K=9-10 (see Supplementary Figure 4, Supplementary Figure 8).

3.3.2. *Fine-scale population structure and admixture*

Admixture analysis using the RAD locus optimized chromosome painting approach in RADpainter and then population assignment with fineRADstructure (Lawson et al., 2012; Malinsky et al., 2018), was congruent with other analyses, and also detected finer partitions of population structure. The fineRADstructure program detected 15 groups of individuals which were each supported with a posterior probability of 100% and which also overlapped with sampling regions which we delimited for population genetic analyses based on geographic barriers (Black dots in Figure 3, Supplementary Figure 5). The only sampling regions which fineRADstructure failed to confidently partition into separate groups consisted of a group of individuals in the Jaú area of

598 endemism, and a group of individuals in the eastern Napo area of endemism (both groups are poorly
599 resolved in phylogenetic analyses, but unambiguously members of the broader Northern Amazon +
600 Guiana Shield clade D). The eastern Napo individuals (D2 in figure 3) are identified as a low
601 support group (*pp* = 0.49), while the Jaú individuals are lumped in with other western Guiana Shield
602 individuals (D in figure 3). See Supplementary Figure 5 (fineRADstructure dendrogram), and the
603 full dataset coancestry matrix plots Supplementary Figures 6-7.

604 Patterns of coancestry show a complex mosaic of shared ancestry among populations
605 (Supplementary Figures 6-7). In general, coancestry patterns appear consistent with admixture
606 patterns detected by STRUCTURE, including the case of the highly introgressed western Napo
607 lineage. Therefore, we interpret patterns indicated by STRUCTURE to be reliable. Notably, the
608 high level of admixture initially inferred for our San Martín sample is recapitulated in RADpainter
609 coancestry analysis – this sample has relatively high coancestry across most populations (as does
610 the pattern of coancestry we report for the western Napo lineage, below).

611 Lastly, the Lawson et al. (2012) "normalized PCA" approach provided with
612 fineRADstructure captured 89% of the variance in the genetic data on the first four component axes
613 (51.4%, 24.9%, 7.77%, 4.94%) indicating that the coancestry matrix captures much more
614 information about genetic variance than standard PCoA/PCA of SNP data (Supplementary Figure
615 8a-c). It is therefore not surprising that K-means clustering of coancestry principle component axes
616 better identified groups of individuals which match previously described subspecies designations
617 and phylogenetic structure (Figure 5, CoA).

618

619 3.4. *Phylogenetic reticulation in the western Napo lineage*

620 Separate RADpainter and PhyloNet analysis of the populations in the western Napo area of
621 endemism and their putative progenitor lineages support the hypothesis that this region has a

complex history of introgression. One-tailed tests of mean coancestry indicated the null hypothesis that the difference in means among groups is less than or equal to zero can be strongly rejected (ghlt $p < 2e\text{-}16$), and that individuals from the western Napo population have greater median/mean coancestry with each progenitor lineage than their progenitor lineages share with each other (PHvGS/progenitors = ~1.078, PHvInambari/progenitors = ~1.16). These analyses establish the asymmetry in admixture patterns detected by STRUCTURE. On average, western Napo individuals share ~ 7.5% greater coancestry with Inambari individuals (southern Amazon) than they do with other Northern Amazon individuals (glht $p < 2e\text{-}16$), indicating these sampled populations do not include F1 hybrids. We did not observe any patterns of clinal variation within our samples of this lineage (spanning ~20,500 km^2).

fineRADstructure analysis of the coancestry matrix, as well as visualization of the coancestry matrix indicated that the western Napo population may be differentiated, as well as intermediate when compared to its progenitors (Figure 6a-d). Principle components analysis (Figure 6d) of the coancestry matrix indicate that the western Napo lineage is intermediate on PC1 (compared to progenitor lineages), but differentiated on PC2, which would be expected if the Napo lineage has had sufficient time in reproductive isolation for sorting of ancestral alleles or for the evolution of new alleles (Barrera-Guzmán et al., 2018). This hypothesis is supported by the observation that this lineage has distinct mitochondrial haplotypes derived from Inambari populations (~separated from Inambari haplotypes by ~6.5 substitutions or 0.63%), as well as analyses of IBD (below).

Bootstrapped estimates of F_{st} indicated the same general patterns: 95%CI; western Napo vs Eastern Napo (Northern progenitor, 0.07 - 0.090), western Napo v Inambari (Southern progenitor, 0.058 - 0.077): Eastern Napo vs Inambari: (between progenitors, 0.122 - 0.144). These patterns of F_{st} estimates were robust to alternative assignments of geographically proximate individuals to

646 progenitor linages (for example, including an additional adjacent population for both of the putative
647 progenitor lineages). Thus, the introgressed western Napo lineage is less differentiated from *each* of
648 its progenitor lineages than its progenitor lineages are from each other and it is *also* biased in its
649 genetic composition toward Inambari populations.

650 PhyloNet detected that an evolutionary reticulation is the best hypothesis to explain the
651 origin of the western Napo lineage (Figure 6a), which provides a plausible evolutionary scenario
652 upon which we can interpret other statistics. All parameters in the RJMCMC search achieved an
653 effective sample size > 150 (most >1000), and PhyloNet recovered two network topologies
654 representing 98.6% of the posterior distribution. The top ranked network (*pp* = 57%, Figure 6a)
655 implies the MRCA of the northern Amazonian (Eastern Napo + Jaú) and southern (Inambari +
656 Western Napo) lineages diverged first (clade F). This initial split was followed by a split within the
657 southern lineage that led to the Inambari (S1) and progenitor western Napo (S2a) populations
658 (MRCA of S1 and S2a: Figure 6a). At some point, genetically distinct nuclear genomes from the
659 Northern lineage (clade D/D2) introgressed into S2a (ancestral western Napo), which gave rise to
660 the contemporary western Napo population (S2).

661 The second-best network (*pp* = 41%, Supplemental Data) depicts a scenario of classic
662 hybrid speciation. First, northern and southern lineages of Clade F first diverged, as seen for the top
663 ranked network. Then, members of S1 and the NW Amazonian lineages merged to form a hybrid
664 lineage. We base our subsequent discussion on the top ranked network, as this hypothesis is more
665 consistent with patterns we detect in other analyses. Support for these two reticulation scenarios is
666 similar, however, and they present interesting alternatives which both require a complex history of
667 reticulation.

668 Taken in the context of STRUCTURE (Figure 5) and fineRADstructure (Figure 6) results,
669 the gene flow implied by the top ranked network seems to have been asymmetrical, leaving very

little evidence of southern admixture among northern lineages (see Figure 5, "weakly resolved eastern Napo," which are detected to have very limited (< ~5%) southern admixture). This observation, taken alone, could have been produced by sampling biases. However, demographic parameters estimated with G-PhoCS also support this hypothesis: the only significant migration rate (lower 95% CI > 1 migrant per generation) was inferred from the eastern Napo (member of Guiana Shield lineage) lineage into S2/S2a (Figure 6, Supplementary Table 5).

G-PhoCS also allows us to place a preliminary maximum age constraint on the timing of introgression into S2a. Assuming 2.47 years per generation in *Pseudopipra* (estimated in Bird et al., 2020) and a constant mutation rate, the absolute splitting time between S2a and S1 may have been ~1.2 Ma (Figure 6, Supplementary Table 5). Though we emphasize that this is a rough approximation, this result is loosely consistent with prior divergence time estimates for several groups in the genus based on molecular clock analysis of mtDNA (Castro-Astor, 2014); that study estimated ssp *pipra* (Clade D; Northern Amazon + Guiana Shield lineage) to have branched off from its closest relatives ~ 0.67 Ma [95% HPD: 0.3506-1.1287]. Therefore, introgression into S2a must have occurred more recently.

Lastly, patterns of admixture (Figure 5, Supplemental Figure 6, 7) suggest some evidence of limited introgression of Northern Amazon alleles into the southwestern Amazonian Inambari population (Figure 5). This could be a consequence of limited gene flow into the Inambari population concurrently with or after the primary introgression event(s) which generated the introgressed western Napo lineage. While our interpretations are conditional on whether or not our genetic sample is indicative of a broader pattern, the product of this process appears to be a distinct, geographically isolated and introgressed lineage within the western Napo area of endemism.

3.5. *Spatial distribution of genetic variation*

694 Under isolation by distance (IBD), the empirical relationship between geographic distance
695 and genetic distance is related to offspring dispersal distance and population size (Rousset, 1997).
696 Isolation by distance effects are thus predicted when individual dispersal distances are smaller than
697 a species' range (Teske et al., 2018). Fifteen out of eighteen considered areas (see Population
698 Genetics, Figure 4, and Supplemental Appendix for a description of how these areas were defined)
699 contained > 2 individuals, and thus could be evaluated for isolation by distance effects within them.
700 Five of these regions were identified to have a significant signal of IBD within them at the 0.001
701 alpha level (Table 2). This signal, however, was not consistent across different proxies of genetic
702 and geographic distance, suggesting that different processes may be contributing to spatial
703 correlations within these areas.

704 Partial Mantel tests of barriers to gene flow indicated that all but one evaluated physical
705 barrier likely play significant roles in structuring genetic variation between areas (Table 2). Taken
706 together, Mantel and partial Mantel tests suggest individuals are generally able to disperse within
707 localized regions (with some signal of spatial autocorrelation within regions) but are unable to cross
708 prominent geographic barriers delimiting these areas (Table 2). Notably, the Rio Negro was the
709 only evaluated barrier which is not detected to exhibit a significant effect— (Table 2 and
710 Supplementary Table 2).

711 Model based analysis of migration rates with EEMS largely corroborated the specific
712 hypotheses implied by Mantel tests. For dissimilarities between pairs of demes, the correlation
713 between observed and fitted values was high – $R^2 = 0.871$ – within demes somewhat less – $R^2 =$
714 0.579 (Supplementary Figures 10, 11). Thus, the EEMS model appears to do a good job of
715 describing spatially structured variation in this dataset, and a non-stationary migration surface is a
716 better fit to the data than simple isolation by distance. The EEMS visualization detects many strong
717 signals (posterior probability > 0.9) of barriers to gene flow (Figure 7). As expected, the Andes and

the Amazon river were unambiguously detected as the most significant barriers to gene flow and are clearly recovered as landscape features, even though the model does not know these features exist. Remarkably, EEMS almost perfectly infers the spatial extents of the lowland Amazonian range (range map outlined in black), despite very limited genetic sampling toward the edge of the range.

EEMS detects signals of deviation from constant IBD that likely correspond to several Amazonian tributaries in the Amazon basin which delimit areas of endemism (Figure 7). An additional strong barrier to gene flow was inferred in the south-central amazon (south-central brown patch in Figure 7), corresponding to the Caatinga - Cerrado – Chaco "dry-diagonal" biomes. Strong signals of genetic connectivity are inferred in many localities (blue), notably west of the Andes, connecting north Andean and Central American lineages (despite a lack of sampling in that region). This link perhaps reflects the common ancestry among montane lineages in these regions. Consistent with other analyses, EEMS detects the large Guianan + Northern Amazonian region (E. Napo + Jaú + Guiana) to have relatively weak population structure across a variety of barriers (including the Rio Negro). EEMS detects several discrete areas of continuous gene flow within the Atlantic Forest, likely corresponding to the Coastal Bahia and Serra do Mar areas identified by da Silva et al. (2004) (Figure 7). The introgressed hybrid lineage from the western Napo area is inferred to be genetically isolated from neighboring populations, with apparent reductions in gene flow on all sides, perhaps corresponding to the effects of the Rio Marañón in the south and Rio Putumayo/Iça to the north.

Lastly, EEMS detected two broad clusters of relatively high genetic diversity (relative heterozygosity: Supplementary Figure 12). One is centered along the Amazon river and was relatively uniform within the northern Amazonian basin, reflecting relatively high genetic diversity in Guiana Shield, but also some southern Amazonian populations. Notably, the other prominent high diversity area overlaps with our sampling of the introgressed western Napo lineage. Areas of

742 relatively low genetic diversity included the Atlantic Forest, the Peruvian Andes, and Central
743 American lineages. These patterns are consistent with direct estimates of regional variation in allelic
744 richness (Supplementary Appendix).
745
746 3.6. *Vocal variation*
747 We examined 198 recordings of *Pseudopipra* and identified 14 distinct vocalizations that
748 can be diagnosed by ear or visual inspection of sonograms (arbitrarily labeled numerically in Figure
749 8). This set of vocalization types includes both advertisement vocalizations from males at leks and
750 calls by males and females in other contexts. After examining recordist notes, we generated a
751 preliminary hypothesis which distinguishes eleven highly variable and distinct, lekking
752 advertisements (Figure 8a) from three broad ranging call types (Figure 8b). Our delimitation of
753 vocalizations into advertisement and call categories partly reflects expert opinion, as the behavioral
754 context of vocal records was not always clear (see Supplemental Appendix). Work by Castro-Astor
755 et al. (2007) provides important insights of the vocal behavior of Atlantic forest populations, but
756 most *Pseudopipra* populations lack similarly detailed studies. Vocalization types within each
757 category can be statistically discriminated based on a combination of the frequency, duration, and
758 number of notes. PCA and logistic regression on lekking vocalizations (n=114, Supplementary
759 Table 3), and calls (n=33) found significant differences among types ($p < 0.001$). The first three
760 axes of a PCA explained ~90% of the variation in lekking vocalization characters, with PC1 (~64%)
761 primarily explaining variation in note number and frequency (Supplementary Figure 15).
762 Our assignments of vocal types to well supported genetic groups are preliminary in the
763 absence of recordings from the same individuals from which we have genetic samples.
764 Nevertheless, we found that most of the 11 lek advertisement vocalizations are phylogenetically
765 diagnostic of the populations in which they were recorded (except for two cases of apparent

plesiomorphy, Supplementary Figure 16b). In contrast, the three call vocalization types we identified were diagnostic of one lineage (type 5, clade D), or shared plesiomorphically across two ancient clades (type 10 and 13, Supplementary Figure 16a). See Taxonomic Summary, below.

4. Discussion

In comparison to the great diversity of Neotropical organisms available for study, there have been few studies focused on understanding both fine and broad scale diversity patterns within broadly distributed Neotropical taxa, and this has slowed the biogeographic interpretation of the effects that many landscape features have on the evolution of Neotropical organisms (Fernandes et al., 2013; Harvey et al., 2017; Harvey and Brumfield, 2015; Marks et al., 2002; Milá et al., 2012; Nyári, 2007). We investigated the phylogenetic and population genetic history of a continentally distributed taxon currently recognized as a single biological species– *Pseudopipra pipra*– using a genomic dataset which samples thousands of loci across the genomes of 232 individuals representing eight of thirteen *Pseudopipra* subspecies. Two additional subspecies, *pygmaea* and *coracina*, were represented by mtDNA only. Our genetic sampling encompasses a majority of the broad range of *P. pipra*, spanning Neotropical forests from Central America to the Atlantic Forest of Brazil, including highland and lowland areas of endemism. Our sampling of vocal variation fills in many important gaps in our genetic sampling—therefore, our intersecting datasets allow us to assess at least some data from all thirteen previously identified subspecies.

The predominant biogeographic signal we detect in our dataset is of two successively nested montane clades as the sister groups to several major lowland Amazonian clades (Figure 5, see *Historical Biogeography*, below). This pattern implies that lowland Amazonian and montane Central American populations of *Pseudopipra* expanded historically "out of the Andes." Ancient lineages endemic to the Andes may have originated around the onset of the Pleistocene ~2.5 Ma

790 (molecular clock estimates from Castro-Astor, 2014). The lowland clade within *Pseudopipra*
791 originated (~1.085 [0.57-1.79] Ma) in southwestern Amazonia (Figure 4) and expanded, occupying
792 Amazonia on both margins of the Amazon river, and eventually reached the Atlantic forest (see
793 below). Therefore, our study provides an unusual example of a widespread suboscine passerine
794 which originated in the Andean highlands.
795
796 4.1 *Population genetic variation*
797 We examined our dataset using a variety of population genetic approaches intended to assess
798 the number of natural genetic groupings that may be present in *Pseudopipra*. Phenetic K-means
799 clustering of SNP data and analysis with STRUCTURE estimated a minimum of 5 genetic clusters
800 with broad geographic ranges (Figure 5, Supplementary Figure 2, Supplementary Figure 4). K-
801 means clustering of the coancestry matrix (which captured substantially more variation than
802 standard PCoA) more finely partitioned the dataset into 8 genetic clusters (Figure 5, Supplementary
803 Figure 8), which closely recapitulates previously identified subspecies boundaries (Figure 2).
804 Phylogenetic (RAxML, SVDquartets) and model-based population genetic (fineRADstructure)
805 analyses of nuclear genomic data detected fifteen identical groups of individuals which coincided
806 with sampling regions delimited by clear geographic barriers (results, Figure 3, 4, Supplementary
807 Appendix). Two additional subspecies represented by mtDNA (results) bring the total number of
808 distinct genetic populations delimited by geographic barriers to at least seventeen. The close
809 concordance observed between population structure and landscape features implies that the
810 evolutionary history of *Pseudopipra* within the Neotropics is deeply connected to the South
811 American landscape, adding support to a rich body of literature endorsing this hypothesis
812 (Brumfield, 2012; Cracraft and Prum, 1988; Silva et al., 2019).

Our sampling reveals a "horseshoe" pattern of genetic differentiation across the Amazon basin (Figure 4), with no evidence of gene flow across the Amazon river, except in the western headwaters where the Amazon is narrowest and floodplains are dynamic and young (Haffer, 1992; Pupim et al., 2019; Räsänen et al., 1987). Given that the genus originated in the west, we hypothesize that northern and southern clades diverged across the Amazon river in this region. For virtually all evaluated cases, we find significant effects of geographic barriers on subdividing genetic variation within this species complex, including across the Amazon River and most of its larger tributaries, with the notable exception of the Rio Negro (Figure 7 and Supplementary Table 2a, 2b). As expected, the "dry-diagonal" Caatinga, Cerrado, and Chaco belt is a strong barrier to gene flow, isolating Atlantic Forest lineages from their southeastern Amazonian Xingu relatives—as are the Andes, which exhibit a disproportionate effect on divergence between Peruvian foothills populations and Central American lineages (see Supplementary Appendix for additional discussion).

Multiple analyses indicate that the western Napo population has a complex history of initial and early differentiation from the broader Southern Amazonian clade, followed by subsequent, asymmetric genetic introgression from Jaú and eastern Napo populations of the Guianan Shield + Northern Amazon clade (Clade D; see discussion below). When the genetic data are analyzed under either concatenated or coalescent phylogenetic frameworks which do not allow for phylogenetic reticulation, the western Napo population is reconstructed as most closely related to Guiana Shield + Northern Amazon populations (Figure 3, Figure 5, Clade D). In contrast, phylogenetic network analysis indicates that this population is derived from individuals of Inambari (southern Amazon) provenance (Figure 6, also summarized in Figure 4). This incongruence between methods is likely due to the complexity of the constituent gene trees in this population-level differentiation induced by asymmetric introgression from the Jaú and eastern Napo populations (Thom et al., 2018). We

hypothesize that introgression has introduced more recently derived variation from the Guiana Shield + Northern Amazon clade D into an older, differentiated, population of the Southern Amazonian clade, pulling the phylogenetic affinity toward the northern clade.

4.2 *History of an introgressed hybrid lineage*

The substantial level of observed introgression (>40%) in the western Napo individuals, as well as a suite of genomic and phenotypic evidence, suggests that this genetic population may be best recognized as a hybrid lineage, though it is uncertain if it can be considered a product of hybrid speciation. To date, five prior examples of hybrid species formation in birds have been evaluated with genomic data: (1) the Golden-crowned manakin *Lepidothrix vilasboasi* (Barrera-Guzmán et al., 2018), (2), the Audubon's warbler *Setophaga auduboni* (Brelsford et al., 2011), (3) the Italian sparrow *Passer italiae* (Hermansen et al., 2011), (4) the Hawaiian duck *Anas wyvilliana* (Lavretsky et al., 2015), and (5) a putative species of Darwin's finches *Geospiza* (Lamichhaney et al., 2018). These examples of mosaic genome hybrid speciation, in which the fusion of two or more progenitor lineages leads to a reproductively isolated hybrid species (Counterman, 2016; Jiggins et al., 2008), repeatedly demonstrate that modern patterns of avian diversity may have been significantly influenced by historical phylogenetic reticulation (Suh, 2016). Our analyses of genomic data across *Pseudopipra* imply that the western Napo lineage is likely a differentiated, allopatric phylogenetic species which has a genomic composition derived from heavy historical admixture with a non-sister lineage. This historical scenario is not the same as those examples cited above, but it has produced a population genetic signature which is extremely similar (compare to Barrera-Guzmán et al., 2018). These observations force us to ask whether historical process, or contemporary genetic pattern should be paramount in delineating a hybrid species.

Some authors have proposed strict population genetic criteria for identifying homoploid hybrid speciation, including demonstrating (1) evidence of reproductive isolation from parental species, (2) evidence of past hybridization, and (3) evidence that intrinsic reproductive isolation was derived from the hybridization event (Schumer et al., 2014). The first two conditions are satisfied by the western Napo *Pseudopipra* lineage. First, phylogenetic and population genetic analyses are suggestive of contemporary reproductive isolation. Phylogenetic analyses of nuclear genome data, including concatenated and coalescent analyses (Figure 3, 4, 6), recover this lineage as a well-supported reciprocally monophyletic group. Differentiated mtDNA haplotypes from a broader geographic sample (Figure 6) also form subgroups derived from southern Inambari haplotypes (results, Supplementary Figure 17), implying reproductive isolation from female northern progenitors at a minimum.

PCA-based and model-based population genetic cluster analyses (Figure 6b, 6c, Supplementary Figures 4-9), help establish this lineage as intermediate *and* differentiated when compared to the living descendants of its inferred progenitors. The level of differentiation on PC2 corresponds to about 10% of the total variation explained by PC1 and PC2 (~85%). At the same time, Mantel tests, as well as model-based analysis of isolation by distance, support the hypothesis that contemporary gene flow into and out of this population is at least restricted and probably not occurring (Figure 7, Supplementary Figure 12). Thus, we interpret our G-PhoCS results of a significant migration parameter (results) as indicative of a window of historical introgression, which could be further explored with simulation studies. Estimates of inbreeding coefficients support this hypothesis and suggest similar levels of inbreeding among *Pseudopipra* populations in the western Amazon ($F_{IS} > 0.1$, Supplementary Figure 9, and Supplementary Appendix). The western Napo population is thus not detected to be more outbred than most populations (i.e. recently introgressed).

883 In sum, these patterns observed in our genetic data are consistent with the hypothesis of
884 contemporary reproductive isolation for the introgressed Western Napo populations.

885 The second criterion from Schumer et al. (2014), evidence of past hybridization, is directly
886 supported by multiple model-based analysis of genomic admixture and phylogenetic reticulation
887 (Figure 5, 6), and indirectly supported by model based and phenetic cluster analysis (Supplementary
888 Figures 4-9). We hypothesize that hybridization initially occurred between progenitor lineage S2a
889 and western lineages of the Guiana Shield + Northern Amazon clade (Figure 6a). Lineage S2a was
890 then replaced by the novel introgressed lineage, which exchanged genes with neighboring
891 populations within a finite time window.

892 The third criterion from Schumer et al. (2014), evidence that intrinsic reproductive isolation
893 was derived from the hybridization event, is recognized as the most difficult to demonstrate, and
894 probably the most tenuous. As discussed by Nieto Feliner et al. (2017) in critique of Schumer et al.
895 (2014), this criterion may be useful for framing the strength of evidence for or against the existence
896 of a homoploid hybrid species, but only if one adheres strictly to the biological species concept
897 (Mayr, 1942). If one applies a more evolutionary definition of species, like the metapopulation
898 lineage concept, in which the only necessary requirement is demonstration of separately evolving
899 lineages (de Queiroz, 2005), the importance of demonstrating intrinsic reproductive isolation to
900 claim species status for a putative hybrid lineage (or any lineage at all) is less clear to us. Speciation
901 in birds in particular is thought to proceed relatively fast, and may be more affected by external
902 factors than internal ones (Grant and Grant, 1997). Further, requiring hybridization to be the *source*
903 of intrinsic isolation excludes the possibility that extrinsic or pre-zygotic isolation may evolve from
904 changes that enable hybrids access to new niches, which may also be geographically isolated
905 relative to parental forms (Nieto Feliner et al., 2017).

Intriguingly, the evolutionary scenario implied by phylogenetic reticulation analysis for the introgressed western Napo hybrids provides a mechanism to explain a confusing distribution of vocalization types (Figure 8, Supplementary Appendix). The males from the Napo area of endemism share vocalization type 2 with birds in the southeastern Amazon (Figure 8, Supplementary Figure 16), and with the restricted Huallaga valley *pygmaea* subspecies (the sister group to all lowland Amazonian forms, according to mtDNA). Manakins are suboscine passerines and have innate vocalizations (though see Saranathan et al., 2007). Therefore, the introgressed Napo lineage appears to have inherited its vocalizations from the MRCA of groups S1 and S2 (Figure 6a) and retained it through subsequent introgression with northern lineages that have a different vocal type (Type 3 and/or 5). In this scenario, vocal type 7 in the western Inambari represents more recently derived variation in the southern Amazonian clade (Figure 8, Supplementary Figure 16).

As noted earlier, the western Napo lineage also has mtDNA haplotypes derived from Inambari populations (Supplementary Figure 17). In addition to implying reproductive isolation from the Guiana Shield clade, this mitochondrial pattern implies that the introgressed Napo lineage may be descended from the union of progenitor female S2a and progenitor western males of the Guiana Shield + Northern Amazon clade, perhaps through male biased dispersal or mate choice (a preliminary test of male biased dispersal did not detect a signal, Supplemental Appendix). Regardless, this pattern raises fascinating questions about why and how the western Napo lineage may have retained its southern song type.

Taking all of the available genetic and phenotypic evidence into account, we suggest that the western Napo lineage may be an additional example of the formation of a hybrid lineage in birds, and one that was produced via an underappreciated historical mechanism. Assuming that our limited sampling is representative of the *discolor* subspecies, the patterns we observe appear to be most similar to “hybrid trait speciation,” in which a hybrid species is formed after introgression

930 from one species into a genomic background of a close relative (Counterman, 2016; Jiggins et al.,
931 2008; Marques et al., 2019). This type of combinatorial hybrid speciation was originally proposed
932 from studies of *Heliconius* butterflies (Jiggins et al., 2008; Salazar et al., 2010).
933
934 4.3 *Comparative and Historical Biogeography*
935 The spatial pattern of population structure and evolutionary relationships we identify here
936 are broadly congruent with other recent studies of Neotropical biogeography in upland forest birds,
937 contributing another example of the substantial phylogeographic pseudocongruence across taxa at
938 this wide scale (e.g. Cunningham and Collins, 1994; Harvey et al., 2017). As expected, the two
939 primary barriers identified for *Pseudopipra* in the South American lowlands are the Amazon River
940 and the dry diagonal. Amazonian tributaries seem to delimit recently isolated populations in
941 southern Amazonia, while northern *Pseudopipra* populations show weaker evidence of being
942 isolated by the Negro and Japurá rivers. By contrast, southern populations seem to be strongly
943 isolated by rivers with smaller discharge, such as the Xingu, Tapajós and Purus.
944 A survey of the suboscine ovenbirds in the *Xenops* complex (Harvey and Brumfield, 2015)
945 identified similar patterns of genetic structuring coincident with major Amazonian tributaries, as did
946 Ribas et al. (2012) in the *Psophia* trumpeters, though the inferred relationships among areas of
947 endemism are not the same. A recent comparative analysis shows that large rivers delimit genetic
948 clusters in as many as 23 groups of upland Amazonian forest birds, strongly arguing in favor of
949 these rivers acting as barriers to gene flow (Silva et al., 2019). The somewhat distinct pattern found
950 for *Pseudopipra* may be related to its recent expansion throughout the eastern Andean lowlands and
951 its habitat tolerance.
952 Within the manakins, Ohlson et al. (2013) placed *Pseudopipra* as sister to the genus
953 *Ceratopipra*, which includes five well-recognized species that are extensively codistribtuted with

Pseudopipra. The montane origin of *Pseudopipra* is congruent with the observation that *Pseudopipra* is the sister group to *Ceratopipra,* which is also broadly distributed in the Neotropical lowlands from Central America to the Atlantic Forest, with two secondary expansions into the subtropical forests in the tepuis (*C. cornuta*) and the southern Andes (*C. chloromeros*) (Ohlson et al., 2013). The breakpoints among these *Ceratopipra* species are highly concordant with the breakpoints among the genetic clusters within the *Pseudopipra* complex that we have presented here, implying that these closely related taxa have components of their phylogeographic history in common. See the Supplemental Appendix for additional notes, as well as Castro-Astor (2014).

A key feature of the phylogeny of *Pseudopipra* is that montane Andean lineages are sister groups to both the montane Central American and the lowland Amazonian-southeast Brazilian lineages (Figure 5). Ancestral character reconstruction under Bayesian stochastic mapping and maximum parsimony therefore unambiguously reconstructs the most recent common ancestor of *Pseudopipra* as Andean (Figure 3-5, Supplemental Appendix). Thus, it appears that the most recent common ancestor of *Pseudopipra* was an Andean lineage restricted to subtropical, lower montane forest. The earliest diversification event in the genus was likely the differentiation of the ancestral Andean populations north and south of the headwaters of the Rio Huallaga, Peru. The northern lineage gave rise to the subtropical montane lineages of the northern Andes and Central America (Clade A). The southern lineage gave rise to a subtropical montane Southern Peruvian lineage (Clade B) and a southwestern lowland Amazonian lineage (*pygmaea*, represented in our dataset by mtDNA). Subsequently, the Amazonian lineage expanded to the eastern lowlands and differentiated into the northern lowland Amazonian + Guiana Shield (Clade D in Figure 3) and southern lowland Amazonian/Atlantic Forest (Clade C in Figure 3) clades. Within the Southern Amazonian clade, differentiation proceeded from west to east: (Figure 3, 4):

977 (West Inambari (East Inambari (Rondônia (Tapajós (Xingu (Bahia (Espírito Santo (Rio de Janerio))))))))).
978 Contrastingly, the northern Amazonian lineage also expanded all the way to the Atlantic Ocean, but
979 without producing strong phylogenetic structure (Figures 3, 4). In the Napo area of endemism,
980 substantial introgression occurred from east Napo and Jaú populations into the western Napo
981 population, with limited backcrossing into Inambari and east Napo populations.
982 Intriguingly, a well-supported distinct montane (>1000m) (BS: 100) (Clade C2 in Figure 3;
983 Cushabatay area, northern Cordillera Azul, Loreto Prov., Peru) is recovered as nested within the
984 well supported lowland western Inambari subclade (Clade C1 in Figure 3). Thus, there may have
985 been at least one secondary invasion of montane Andean habitats from a lowland ancestor. Patterns
986 of admixture support this hypothesis – montane individuals from the northern Cordillera Azul and
987 their proximate lowland relatives all show limited genomic admixture from derived lineages of the
988 Guiana Shield + Northern Amazon clade (yellow in Figure 5). This result is consistent with a
989 scenario of secondary invasion of the highlands from lowland Inambari lineages after weak
990 introgression from the western Napo (see below).
991
992 4.4 *North Andean diversity*
993 Our final genomic data set does not include any samples from Andean Ecuador or
994 Colombia, a region that includes four previously recognized subspecies and at least four distinct
995 vocalization types (see Taxonomic Summary). Unfortunately, the three Andean Ecuadorian tissue
996 (representing subspecies *coracina*) specimens in our original sample were not of high enough
997 preservation quality to be viable for ddRAD sequencing (Supplementary Table 1). We were able to
998 obtain mtDNA from these samples however, which unambiguously formed a monophyletic cluster
999 with our single San Martín (North Andean Peru) specimen in mtDNA gene tree analyses
1000 (Supplemental Appendix, Supplementary Figure 17). A prior study by Castro-Astor (2014) based

on mtDNA also included one Andean Ecuadorian sample and one Central American sample, which unambiguously clustered together (*pp*=100). Thus, it is likely that our unsampled Andean lineages of Ecuador and Colombia (including the subspecies *minimus, bolivari, unica,* and *coracina*) are likely members of the northern Andean clade represented in our genomic data set by our single San Martín specimen. As noted earlier, our San Martín specimen had approximately equal probability of assignment to multiple populations, which could indicate heavy admixture, or that this is a single sample from a highly distinct population which has its own distinct history. These patterns may also be consistent with the hypothesis that the North Andean clade is the source of montane and lowland *Pseudopipra* diversity.

4.5 *Phenotypic and Vocal Evolution*

Thirteen subspecies have been previously described based on variations in plumage color and, to a lesser extent, size (See Taxonomic Summary in Supplementary Appendix). Male plumage coloration varies among the subspecies of *Pseudopipra* in the glossiness of the black body feathers, the length of the white crown feathers, and the color– white, gray, or black– of the bases of these crown feathers. Female plumage coloration often shows more striking differentiation among subspecies than does male plumage, including variation in the shade of olive green on the body, the shade and extent of gray on the crown and face, and olive, yellow, or gray coloration on the belly (Chapman et al., 1914; de Schauensee, 1945, 1950; Ridgway, 1906; Wetmore, 1972; Zimmer, 1936).

Although we did not conduct a detailed analysis of plumage variation among populations of *Pseudopipra*, we did observe study skin specimens of all but one (*bolivari*) of the previously described subspecies in museum collections. All of the ten subspecies that were included in our genetic samples were identified as distinct, diagnosable, monophyletic groups. In other words, in all

cases that we were able to test, traditional taxonomic practices conducted between 1758 through 1936 successfully identified and named distinct evolutionary lineages within *Pseudopipra*. We did not have genetic samples of the northern Andean taxa *minima, unica,* and *bolivari*, but we note that they all share entirely white crown feather bases (forecrown feathers only in *bolivari*)(de Schauensee, 1945, 1950), implying that this unique plumage character state is derived within *Pseudopipra*, and that *minima*, *unica*, and *bolivari* form a clade. One lineage containing two subspecies– *separabilis* and *cephaleucos*- has evolved another unique, shared, derived plumage character– a distinctive, second, predefinitive male plumage which has also further differentiated between the two subspecies. In *cephaleucos,* predefinitive males have an olive green back, a white crown, and slate gray face and belly. In *separabilis,* predefinitive males are similar with lighter gray belly, and a medium gray, instead of white, crown. In conclusion, plumage coloration appears to provide highly informative evidence of evolutionary lineage status in this genus.

Our analysis of vocal behavior indicates that vocalizations are also highly informative of lineage identity within *Pseudopipra*. The complex phylogenetic history we have documented with genetic data has clearly had a strong impact on the evolution of vocalizations in *Pseudopipra.* Because lek vocalizations are a focus of female choice in manakins and because suboscine passerines have largely innate vocalizations, the extensive vocal and genetic differentiation among populations of *Pseudopipra* strongly indicates the existence of many species within *Pseudopipra*.

We identified fourteen distinct vocalization types (Figure 8), and nearly all were restricted to and diagnostic of a single previously recognized subspecies or a broader monophyletic group (See Expanded Taxonomic Summary in Supplementary Appendix for details). One of the two exceptions was vocalization type 2, which may be shared plesiomorphically across the Amazon basin between populations of *pygmaea* in the Huallaga Valley, *discolor* in the western Napo region of Ecuador and Peru, and *separabilis* in Para, Brazil, but not in the intervening southwestern Amazonian

populations of *microlopha* (Supplementary Figure 16). Unfortunately, available vocal sampling is very limited in the intervening regions along the south side of the Amazon Basin, which are populated by a nested series of genetically distinct lineages with successively closer relationships to southern Amazonian *separabilis* and *cephaleucos* from the Atlantic forest of Brazil.

The other exception is the sharing of vocal type 1 ('*trill-dink*') between populations of *occulta* from the eastern slopes of the Andes from San Martin, Peru to southeastern Ecuador (A1) and populations of *comata* from the southern Cordillera Azul in Loreto (B1). The song of the southern clade (B2) from southern Huánuco (AMNH 820866, 820952), Pasco, Junín, and Cusco is currently unknown. Surprisingly, these geographically adjacent lineages sit on either side of the most basal phylogenetic separation within *Pseudopipra*. The *comata* populations are phylogenetically closer to northern Andean and Central American lineages that exhibit substantial vocal diversity. Likewise, the southern Cordillera Azul populations are more closely related to other south Peruvian and Amazonian populations which also exhibit great vocal diversity. This result implies that the song type 1 was shared by the ancestral lineage and has been retained in these lineages (Supplemental Figure 16). We identified the same pattern for call vocalization type 10.

Our phylogenetic analysis also identified a distinct montane clade from the northern Cordillera Azul (Clade C2; Cushabatay area; Alverson et al. 2001) which is closely related to populations from lowland forests of eastern Peru south of the Rio Marañón and east to Rio Purus, Brazil, currently recognized as *P. p. microlopha*. The hypothesis that this lineage is a secondary expansion into montane forests from the nearby Amazonian populations is strongly supported by the evidence that this montane population shares same song type 7 ('*jeer'*) with the lowland populations of *microlopha* (ML-190390, ML-190388, and ML-190262).

4.6 *Taxonomy and Revised Classification*

The present set of analyses demonstrate a number of genetically well-differentiated and phenotypically diagnosable lineages and provides compelling new evidence for evaluating the species status of *Pseudopipra*. Although there are gaps in our sampling, we find that there are at least eight genetically well-differentiated, and phenotypically diagnosable lineages of *Pseudopipra*, and this is reflected in our proposed taxonomic revision below. Nonetheless, it should be noted that there is more genetic variation than observed vocal variation implies, and more vocal variation beyond our geographic coverage of genetic samples (especially in northern Andes). Thus, our taxonomic recommendations for *Pseudopipra* almost certainly represent an underestimate of extant diversity, underscoring the need for continued reassessment of species limits and of diversity patterns in the Neotropics.

Our analysis provides the first comprehensive opportunity to reevaluate species limits within *Pseudopipra* based on phylogenetic, population genetic, vocal, and plumage differentiation among populations and named subspecies (See the Expanded Taxonomic Summary in the Supplementary Appendix). This proposed classification is a conservative treatment that recognizes the limitations on our genetic sampling of populations in the northern Andean clade from Ecuador and Colombia. Four of these unsampled (and under-sampled) northern Andean subspecies – *coracina*, *minima*, *bolivari*, and *unica* – have unique and highly differentiated vocalization types, and diagnosable plumage differences. Our phylogenetic results strongly suggest that most of the *Pseudopipra* subspecies are distinct evolutionary lineages deserving species status under either the phylogenetic or biological species concepts. In sum, *Pseudopipra* may include 15-17 distinct species that have rapidly arisen in the last ~2.5 Ma, but further reclassification will require more vocal data and detailed analyses of plumage variation.

5. Taxonomic Summary

Pseudopipra coracina (Sclater 1856) **Andean White-crowned Manakin**

Distribution: Subtropical Andes from Venezuela south to Esmeraldas, Ecuador and San Martín, Peru.

Phylogenetic Position: Clade A1

P. c. coracina (Sclater 1856)

Type Locality: Villavicencio, Meta, Colombia.

Distribution: Subtropical forests of the eastern slope of the Andes from western Venezuela to Morona-Santiago, Ecuador.

Phylogenetic Position: (mtdna) Member of Clade A1

Lek Vocal Type: 8 (*errrwer*).

Call Vocal Type: Unknown

P. c. minima (Chapman 1914)

Type Locality: West of Popayan, Cauca, Colombia

Distribution: Subtropical forests of western Cauca, Colombia south to Esmeraldas, Ecuador

Phylogenetic Position: (unsampled) Member of Clade A/A1

Lek Vocal Type: 9 (*reeee*)

Call Vocal Type: unknown

P. c. bolivari (de Schauensee 1950)

Type Locality: Murucucu, Bolivar, Colombia

Distribution: Subtropical forests of southern Córdoba, Colombia. (Not Sampled)

Phylogenetic Position: (unsampled) Member of Clade A1

Lek Vocal Type: Unknown

Call Vocal Type: Unknown

P. c. unica (de Schauensee 1945)

Type Locality: Lomas de Isnos, Huila, Colombia

Distribution: Subtropical forests of Magdalena Valley, Antioquia to Huila, Colombia.

Phylogenetic Position: (unsampled) Member of Clade A1

Lek Vocal Type: 11a (*weer-dink*) and 11b (*shureeep*)

Call Vocal Type: Unknown

P. c. occulta (Zimmer 1936)

Type Locality: Ucho, east of Chachapoyas, Amazonas/San Martin, Peru

Distribution: Eastern slope of the Andes from Zamora-Chinchipe, Ecuador (Freile, 2014) south to San Martín, and Huánuco, Peru, west of the Rio Huallaga

Phylogenetic Position: Member of Clade A/A1

Lek Vocal Type: 1 (*trill-dink*)

Call Vocal Type: 10 (*bree*)

Pseudopipra anthracina (Ridgway 1906) **Western White-crowned Manakin**

Type Locality: Moravia, Costa Rica

Distribution: Subtropical Costa Rica to Western Panama

Phylogenetic Position: Clade A2

Lek Vocal Type: 4 (*jureeee*)

Call Vocal Type: Unknown

Pseudopipra comata (Berlepsch and Stolzmann 1894) **Junín White-crowned Manakin**

Type Locality: Garita del Sol, Junin, Peru.

Distribution: Subtropical Andes of Peru from Cordillera Azul, Loreto (east and south of the Rio Huallaga) to southern Huánuco Pasco, Junín, and northern Cusco.

Phylogenetic Position: Clade B

Lek Vocal Type: 1 (*trill-dink*)

Call Vocal Type: 10 (*bree*)

Pseudopipra pygmaea (Zimmer 1936) **Huallaga White-crowned Manakin**

Type Locality: Chamicuros, Loreto, Peru

Distribution: Tropical forest of lower Rio Huallaga Valley, Peru

Phylogeneic Position: (mtDNA) Sister to Clade F

Lek Vocal Type: 2 (*deeeer*)

Call Vocal Type: 13 (*weeu*)

Pseudopipra discolor (Zimmer 1936) **Napo White-crowned Manakin**

Type Locality: Puerto Indiana, Loreto, Peru

Distribution: Tropical forest in Napo, Ecuador and Loreto, Peru south to the Rio Marañón.

Phylogenetic Position: Clade E

Lek Vocal Type: 2 (*deeeer*)

Call Vocal Type: 13 (*weeu*)

Pseudopipra pipra (Linneaus 1758) **Northern White-crowned Manakin**

Type Locality: Suriname

Distribution: Tropical forest of eastern Colombia, southern Venezuela, the Guianas, and Brazil north of the Amazon and west to the right (west) bank of the Rio Putumayo, Colombia.

Phylogenetic Position: Clade D

Lek Vocal Type: 3 (*buzzzz*)

Call Vocal Type: 5 (*zeee*)

Pseudopipra microlopha (Zimmer 1929) **Southern White-crowned Manakin**

Distribution: Tropical forest of eastern Peru south of the Rio Huallaga, Rio Marañón, and the Amazon River east to Pará, Brazil, and subtropical forests between the Rio Huallaga and Rio Ucayali.

Phylogenetic Position: Paraphyletic, including Clade C without Clade C7

***P. m.* undescribed subspecies**

Distribution: Subtropical forest from the Cushabatay area, northern Cordillera Azul, between Rio Huallaga and Rio Ucayali, Loreto Province, Peru.

Phylogenetic Position: Clade C2

Lek Vocal Type: 7 (*jeer*)

Call Vocal Type: 13 (*weeu*)

P. m. microlopha (Zimmer 1929)

Distribution: Eastern Peru south of the Rio Marañón and Rio Huallaga, west to Rio Juruá and Rio Purus, Brazil.

Phylogenetic Position: Clade C1 excluding C2

Lek Vocal Type: 7 (*jeer*)

Call Vocal Type: 13 (*weeu*)

***P. m.* undescribed subspecies**

Distribution: Right (east) bank of the Rio Purus to the left (west) bank Rio Madeira

Phylogenetic Position: Clade C3

Plumage: Not examined

Lek Vocal Type: Unknown

Call Vocal Type: 13 (*weeu*)

***P. m.* undescribed subspecies**

Distribution: Right (east) bank of the Rio Madeira to the left (west) bank the Rio Tapajos.

Phylogenetic Position: Clade C4

Plumage: Not examined

Lek Vocal Type: Unknown

Call Vocal Type: 13 (*weeu*)

***P. m.* undescribed subspecies**

Distribution: Right (east) bank of the Rio Tapajos to the left (west) bank of the Rio Xingu

Phylogenetic Position: Clade C5

Plumage: Not examined

Lek Vocal Type: 6b (*zeeeee-tonk*)

Call Vocal Type: 13 (*weeu*)

P. m. separabilis (Zimmer 1936)

Type Locality: Tapara, Rio Xingu, Brazil

Distribution: Tropical forest of Rio Xingu east to central and southern Pará.

Phylogenetic Position: Clade C6

Lek Vocal Type: 2 (*deeeeer*)

Call Vocal Type: 13 (*weeu*)

Pseudopipra cephaleucos (Thunberg 1822) **Atlantic White-crowned Manakin**

Type Locality: Bahia, Brazil

Distribution: Tropical forest from Bahia south to northern Rio de Janeiro, Brazil. **Phylogenetic Position:** Clade C7

Lek Vocal Type: 6a (*zeeeee-tonk*)

Call Vocal Type: 13 (*weeu*)

Supplementary Material: Supplementary Data, including data files and R code can be found in the Mendeley Data Repository [temporary link for review only]:
https://data.mendeley.com/datasets/...
and GenBank XXX-XXX.

Funding: J. S. B. was supported by a National Science Foundation Graduate Research Fellowship, NSF Doctoral Dissertation Improvement Grant [DGE-1650441, DEB-1700786], and University of Michigan Life Sciences Fellowship. L. C. was supported by NSF grant DEB-1555754 to I.J.L. T. J. F. was supported by an NSF Post-doctoral Fellowship in Biology [DBI-1523857]. C. C. R. was funded by CNPq (308927/2016-8), FAPEAM, FAPESP/NSF (NSF DEB-1241066, FAPESP, grant #2012/50260-6) and USAID (PEER AID-OAA-A-11-00012, cycle 5). J. S. B. and R. O. P. were

funded by the W. R. Coe Fund of Yale University. Data collection was paid for by Athena Funds from the Cornell Laboratory of Ornithology, and W. R. Coe Fund from Yale University. This work was also supported in part by the facilities and staff of the Yale University Faculty of Arts and Sciences High Performance Computing Center, and with the resources of the Cornell University BRC Bioinformatics Facility, which is partially funded by Microsoft Corporation.

Acknowledgements:

We thank the numerous researchers whose field collection efforts made this research possible including: Angelo P. Capparella, Camila Ribas, Ivandy Castro-Astor, Cecilia Fox, David E. Willard, Diego Ocampo, Donna Dittmann, Jacob S. Berv, John P, O'Neill, Karla Conejo, Katya Balta, Kenneth V Rosenberg, Louise M. Augustine, Peter E. Scott, Richard O. Prum, Steven W. Cardiff, Susan Allen-Stotz, Thomas Valqui, Tristan J. Davis, Ivan Prates and others. Tissue loans or skins were kindly received from the Divisão de Aves do Museu de Zoologia da Universidade Estadual de Feira de Santana (MZFS), the Field Museum of Natural History (FMNH), Instituto Nacional de Pesquisas da Amazônia (INPA), Kansas University Natural History Museum (KUNHM), Louisiana Museum of Natural History (LSUMZ), Museu de Zoologia da Universidade de São Paulo (MZUSP), Museu Paraense Emilio Goeldi (MPEG), the Academy of Natural Sciences of Drexel University (ANSP), Universidade do Estado do Rio de Janeiro (UERJ), the University of Alaska Museum (UAM), the University of Nevada, Las Vegas (UNLV), the Yale Peabody Museum of Natural History (YPM), the Cornell Museum of Vertebrates (CUMV), the American Museum of Natural History (AMNH), the Museum of Comparative Zoology (MCZ), and the United States Museum of Natural History (NMNH). We thank the curators and collections managers of these institutions for facilitating these loans: Camila Ribas, John Bates, Mark Robbins, Donna Dittman, Alexandre Aleixo, Caio Graco Machado, Joel Cracraft, Luís Fábio Silveira, Kevin Winker, Maria

Alice Santos Alves, Kristof Zyskowski, Nate Rice, Charles Dardia, John Klicka, and others. We thank Eliot Miller, Emma Greig, Matthew Medler, Greg Budney, Eduardo Iñigo and Andrew Spencer at the Macaulay Library for assistance with audio recordings. We thank members of the Lovette Lab, Daniel Lane and **XXX** anonymous reviewers, for comments on the research and the manuscript. The illustration depicted in Fig. 1 is reproduced with permission from the Handbook of Birds of the World, Lynx Editions, Barcelona. J. S. B. and R. O. P. conceived of the research. J. S. B. and L. C. performed laboratory and bioinformatic processing, with input and support from I. J. L. We thank Bronwyn Butcher for help with laboratory protocols. J. S. B. and L. C. analyzed the genomic data and interpretation of the results was supported by input from all authors. T. J. F. downloaded, processed, and analyzed lekking vocalization data. C. C. R. and I. C. R. facilitated critical sample acquisitions from Amazonia and the Atlantic Forest. I. J. L. and R. O. P. provided detailed guidance and feedback on the manuscript. R. O. P. examined plumage phenotypes and proposed a revised taxonomy. The manuscript was written by J. S. B. with input from all authors.

Literature Cited

Antonelli, A., Ariza, M., Albert, J., Andermann, T., Azevedo, J., Bacon, C., Faurby, S., Guedes, T., Hoorn, C., Lohmann, L.G., Matos-Maraví, P., Ritter, C.D., Sanmartín, I., Silvestro, D., Tejedor, M., ter Steege, H., Tuomisto, H., Werneck, F.P., Zizka, A., Edwards, S.V., 2018. Conceptual and empirical advances in Neotropical biodiversity research. PeerJ 6, e5644.

Barrera-Guzmán, A.O., Aleixo, A., Shawkey, M.D., Weir, J.T., 2018. Hybrid speciation leads to novel male secondary sexual ornamentation of an Amazonian bird. Proceedings of the National Academy of Sciences 115, E218-E225.

Bates, J.M., Zink, R.M., 1994. Evolution into the Andes: Molecular Evidence for Species Relationships in the Genus Leptopogon. The Auk 111, 507-515.

Berv, J.S., Prum, R.O., 2014. A comprehensive multilocus phylogeny of the Neotropical cotingas (Cotingidae, Aves) with a comparative evolutionary analysis of breeding system and plumage dimorphism and a revised phylogenetic classification. Molecular Phylogenetics and Evolution 81, 120-136.

Bioacoustics_Research_Program, 2011. Raven Pro: Interactive Sound Analysis Software (Version 1.4) [Computer software]. The Cornell Laboratory of Ornithology. Available from http://www.birds.cornell.edu/raven., Ithaca, NY.

Bird, J.P., Martin, R., Akçakaya, H.R., Gilroy, J., Burfield, I.J., Garnett, S., Symes, A., Taylor, J., Şekercioğlu, Ç.H., Butchart, S.H.M., 2020. Generation lengths of the world's birds and their implications for extinction risk. Conservation Biology n/a.

1310 Borges, S.H., Da Silva, J.M.C., 2012. A New Area of Endemism for Amazonian Birds in the Rio
1311 Negro Basin. The Wilson Journal of Ornithology 124, 15-23.

1312 Bradburd, G.S., Coop, G.M., Ralph, P.L., 2018. Inferring Continuous and Discrete Population
1313 Genetic Structure Across Space. Genetics 210, 33-52.

1314 Brelsford, A., Milá, B., Irwin, D.E., 2011. Hybrid origin of Audubon's warbler. Molecular Ecology
1315 20, 2380-2389.

1316 Brumfield, R.T., 2012. Inferring the origins of lowland Neotropical birds. The Auk 129, 367-376.

1317 Brumfield, R.T., Edwards, S.V., 2007. EVOLUTION INTO AND OUT OF THE ANDES: A
1318 BAYESIAN ANALYSIS OF HISTORICAL DIVERSIFICATION IN THAMNOPHILUS
1319 ANTSHRIKES. Evolution 61, 346-367.

1320 Brumfield, R.T., Liu, L., Lum, D.E., Edwards, S.V., 2008. Comparison of Species Tree Methods for
1321 Reconstructing the Phylogeny of Bearded Manakins (Aves: Pipridae, Manacus) from Multilocus
1322 Sequence Data. Systematic Biology 57, 719-731.

1323 Cabanne, G.S., Campagna, L., Trujillo-Arias, N., Naoki, K., Gómez, I., Miyaki, C.Y., Santos, F.R.,
1324 Dantas, G.P.M., Aleixo, A., Claramunt, S., Rocha, A., Caparroz, R., Lovette, I.J., Tubaro, P.L.,
1325 2019. Phylogeographic variation within the Buff-browed Foliage-gleaner (Aves: Furnariidae:
1326 Syndactyla rufosuperciliata) supports an Andean-Atlantic forests connection via the Cerrado.
1327 Molecular Phylogenetics and Evolution 133, 198-213.

1328 Campagna, L., Gronau, I., Silveira, L.F., Siepel, A., Lovette, I.J., 2015. Distinguishing noise from
1329 signal in patterns of genomic divergence in a highly polymorphic avian radiation. Molecular
1330 Ecology 24, 4238-4251.

Campagna, L., St. Clair, J.J.H., Lougheed, S.C., Woods, R.W., Imberti, S., Tubaro, P.L., 2012. Divergence between passerine populations from the Malvinas – Falkland Islands and their continental counterparts: a comparative phylogeographical study. Biological Journal of the Linnean Society 106, 865-879.

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

Carstens, B.C., Pelletier, T.A., Reid, N.M., Satler, J.D., 2013. How to fail at species delimitation. Molecular Ecology 22, 4369-4383.

Castro-Astor, I.N., 2014. A Phylogeographic study of two neotropical bird species, the Red-Headed Manakin (Pipna rubrocapilla) and the White-crowned Manakin (Dixiphia pipra). Biology. CUNY, CUNY Academic Works.

Castro-Astor, I.N., Alves, M.A.S., Cavalcanti, R.B., 2007. DISPLAY BEHAVIOR AND SPATIAL DISTRIBUTION OF THE WHITE-CROWNED MANAKIN IN THE ATLANTIC FOREST OF BRAZIL. The Condor 109, 155-166.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology 22, 3124-3140.

Chambers, E.A., Hillis, D.M., 2020. The Multispecies Coalescent Over-Splits Species in the Case of Geographically Widespread Taxa. Systematic Biology 69, 184-193.

1351 Chapman, F.M., Miller, L.E., Allen, A.A., Richardson, W.B., Cherrie, G.K., 1914. Diagnoses of
1352 apparently new Colombian birds. 3. Bulletin of the AMNH; v. 33, article 40.

1353 Chernomor, O., Minh, B.Q., von Haeseler, A., 2016. Terrace Aware Data Structure for
1354 Phylogenomic Inference from Supermatrices. Systematic Biology 65, 997-1008.

1355 Cheviron, Z.A., Hackett, S.J., Brumfield, R.T., 2006. Sequence variation in the coding region of the
1356 melanocortin-1 receptor gene (MC1R) is not associated with plumage variation in the blue-crowned
1357 manakin (Lepidothrix coronata). Proceedings of the Royal Society B: Biological Sciences 273,
1358 1613-1618.

1359 Cheviron, Z.A., Hackett, S.J., Capparella, A.P., 2005. Complex evolutionary history of a
1360 Neotropical lowland forest bird (Lepidothrix coronata) and its implications for historical hypotheses
1361 of the origin of Neotropical avian diversity. Molecular Phylogenetics and Evolution 36, 338-357.

1362 Chifman, J., Kubatko, L., 2014. Quartet Inference from SNP Data Under the Coalescent Model.
1363 Bioinformatics 30, 3317-3324.

1364 Chifman, J., Kubatko, L., 2015. Identifiability of the unrooted species tree topology under the
1365 coalescent model with time-reversible substitution processes, site-specific rate variation, and
1366 invariable sites. Journal of Theoretical Biology 374, 35-47.

1367 Chou, J., Gupta, A., Yaduvanshi, S., Davidson, R., Nute, M., Mirarab, S., Warnow, T., 2015. A
1368 comparative study of SVDquartets and other coalescent-based species tree estimation methods.
1369 BMC Genomics 16, S2.

1370 Counterman, B.A., 2016. Hybrid Speciation. In: Kliman, R.M. (Ed.), Encyclopedia of Evolutionary
1371 Biology. Academic Press, Oxford, pp. 242-248.

Cracraft, J., 1985. Historical Biogeography and Patterns of Differentiation within the South American Avifauna: Areas of Endemism. Ornithological Monographs, 49-84.

Cracraft, J., Prum, R.O., 1988. PATTERNS AND PROCESSES OF DIVERSIFICATION: SPECIATION AND HISTORICAL CONGRUENCE IN SOME NEOTROPICAL BIRDS. Evolution 42, 603-620.

Crother, B.I., Murray, C.M., 2011. Ontology of areas of endemism. Journal of Biogeography 38, 1009-1015.

Cunningham, C.W., Collins, T.M., 1994. Developing model systems for molecular biogeography: Vicariance and interchange in marine invertebrates. In: Schierwater, B., Streit, B., Wagner, G.P., DeSalle, R. (Eds.), Molecular Ecology and Evolution: Approaches and Applications. Birkhäuser Basel, Basel, pp. 405-433.

Da Silva, J., Novaes, F., Oren, D., 2002. Differentiation of Xiphocolaptes (Dendrocolaptidae) across the river Xingu, Brazilian Amazonia: recognition of a new phylogenetic species and biogeographic implications. BULLETIN-BRITISH ORNITHOLOGISTS CLUB 122, 185-193.

da Silva, J.M.C., 1995. Biogeographic analysis of the South American Cerrado avifauna. Steenstrupia 21, 49-67.

da Silva, J.M.C., de Sousa, M.C., H. M. Castelletti, C., 2004. Areas of Endemism for Passerine Birds in the Atlantic Forest, South America. Global Ecology and Biogeography 13, 85-92.

Da Silva, J.M.C., Rylands , A.B., Da Fonseca, G.A.B., 2005. The Fate of the Amazonian Areas of Endemism. Conservation Biology 19, 689-694.

1392 de Queiroz, K., 2005. Ernst Mayr and the modern concept of species. Proceedings of the National
1393 Academy of Sciences 102, 6600-6607.

1394 de Schauensee, R.M., 1945. Notes on Colombian flycatchers, manakins and cotingas. Proceedings
1395 of the Academy of Natural Sciences of Philadelphia 97, 41-57.

1396 de Schauensee, R.M., 1950. Colombian zoological survey. Part V: New birds from Colombia.
1397 Notulae Naturae (221), 1-13.

1398 Dickinson, E., 2003. The Howard and Moore Complete Checklist of the Birds of the World
1399 Christopher Helm. London.

1400 Diniz-Filho, J.A.F., Soares, T.N., Lima, J.S., Dobrovolski, R., Landeiro, V.L., de Campos Telles,
1401 M.P., Rangel, T.F., Bini, L.M., 2013. Mantel test in population genetics. Genetics and Molecular
1402 Biology 36, 475-485.

1403 Earl, D.A., vonHoldt, B.M., 2012. STRUCTURE HARVESTER: a website and program for
1404 visualizing STRUCTURE output and implementing the Evanno method. Conservation Genet
1405 Resour 4, 359-361.

1406 Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the
1407 software structure: a simulation study. Molecular Ecology 14, 2611-2620.

1408 Fernandes, A., Cohn-Haft, M., Hrbek, T., Farias, I., 2015. Rivers acting as barriers for bird dispersal
1409 in the Amazon. Revista Brasileira de Ornitologia-Brazilian Journal of Ornithology 22, 363-373.

1410 Fernandes, A.M., Gonzalez, J., Wink, M., Aleixo, A., 2013. Multilocus phylogeography of the
1411 Wedge-billed Woodcreeper Glyphorynchus spirurus (Aves, Furnariidae) in lowland Amazonia:

Widespread cryptic diversity and paraphyly reveal a complex diversification pattern. Molecular Phylogenetics and Evolution 66, 270-282.

Fjeldså, J., 1992. Biogeographic patterns and evolution of the avifauna of relict high-attitude woodlands of the Andes. Steenstrupia, 9-62.

Flouri, T., Jiao, X., Rannala, B., Yang, Z., 2018. Species Tree Inference with BPP Using Genomic Sequences and the Multispecies Coalescent. Molecular Biology and Evolution 35, 2585-2593.

Francisco, M.R., Gibbs, H.L., Galetti, M., Lunardi, V.O., Junior, P.M.G., 2007. Genetic structure in a tropical lek-breeding bird, the blue manakin (Chiroxiphia caudata) in the Brazilian Atlantic Forest. Molecular Ecology 16, 4908-4918.

Freile, J.F., 2014. First record of White-crowned Manakin Dixiphia pipra in western Ecuador. Bull. Brit. Orn. Club 134, 238-240.

Goslee, S.C., Urban, D.L., 2007. The ecodist package for dissimilarity-based analysis of ecological data. Journal of Statistical Software 22, 1-19.

Goudet, J., 2005. hierfstat, a package for r to compute and test hierarchical F-statistics. Molecular Ecology Notes 5, 184-186.

Goudet, J., Perrin, N., Waser, P., 2002. Tests for sex-biased dispersal using bi-parentally inherited genetic markers. Molecular Ecology 11, 1103-1114.

Grant, P.R., Grant, B.R., 1997. Genetics and the origin of bird species. Proceedings of the National Academy of Sciences 94, 7768-7775.

Gronau, I., Hubisz, M.J., Gulko, B., Danko, C.G., Siepel, A., 2011. Bayesian inference of ancient human demography from individual genome sequences. Nature Genetics 43, 1031.

Gubili, C., Ritter, C.D., Motta, J., Farias, I.P., Bates, J., Canton, R., Capurucho, J.M.G., Cornelius, C., Feldheim, K.A., Ribas, C.C., 2016. Isolation and characterization of polymorphic microsatellite DNA Markers from an Amazonian white-sand vegetation specialist bird, Xenopipo atronitens (Aves: Pipridae). The Wilson Journal of Ornithology 128, 668-672.

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

Haffer, J., 1969. Speciation in Amazonian Forest Birds. Science 165, 131-137.

Haffer, J., 1974. Avian speciation in tropical South America, with a systematic survey of the toucans (Ramphastidae) and jacamars (Galbulidae) (Publications of the Nuttall Ornithological Club No. 14, Cambridge, MA).

Haffer, J., 1978. Distribution of Amazon forest birds. Bonner Zoologische Beiträge 29, 38-78.

Haffer, J., 1985. Avian zoogeography of the Neotropical lowlands. Ornithological Monographs, 113-146.

Haffer, J., 1992. On the river effect in some forest birds of southern Amazonia. Boletim do Museu Paraense Emílio Goeldi. Nova série. Zoologia 8, 217-245.

Haffer, J., 2008. Hypotheses to explain the origin of species in Amazonia. Brazilian Journal of Biology 68, 917-947.

Hartl, D.L., Clark, A.G., 1997. Principles of population genetics. Sinauer associates Sunderland, MA.

Harvey, M.G., Aleixo, A., Ribas, C.C., Brumfield, R.T., 2017. Habitat Association Predicts Genetic Diversity and Population Divergence in Amazonian Birds. The American Naturalist 190, 631-648.

Harvey, M.G., Brumfield, R.T., 2015. Genomic variation in a widespread Neotropical bird (Xenops minutus) reveals divergence, population expansion, and gene flow. Molecular Phylogenetics and Evolution 83, 305-316.

Hazzi, N.A., Moreno, J.S., Ortiz-Movliav, C., Palacio, R.D., 2018. Biogeographic regions and events of isolation and diversification of the endemic biota of the tropical Andes. Proceedings of the National Academy of Sciences.

Hermansen, J.S., Sæther, S.A., Elgvin, T.O., Borge, T., Hjelle, E., Sætre, G.-P., 2011. Hybrid speciation in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene flow. Molecular Ecology 20, 3812-3822.

Hijmans, R.J., Williams, E., Vennes, C., 2015. Geosphere: spherical trigonometry. R package version 1.3-11. URL: https://CRAN. R-project. org/package= geosphere.

Hillis, D.M., 2020. The Detection and Naming of Geographic Variation Within Species. Herpetological Review, 51:52-56.

Hoang, D.T., Vinh, L.S., Chernomor, O., Minh, B.Q., von Haeseler, A., 2017. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution 35, 518-522.

Hooghiemstra, H., Van der Hammen, T., 2004. Quaternary Ice-Age dynamics in the Colombian Andes: developing an understanding of our legacy. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 359, 173-181.

Hothorn, T., Bretz, F., Westfall, P., Heiberger, R.M., Schuetzenmeister, A., Scheibe, S., 2013. multcomp: simultaneous inference in general parametric models. R package version 1.2-18. Foundation for Statistical Computing, Vienna, Austria.

Jakobsson, M., Rosenberg, N.A., 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23, 1801-1806.

Janzen, D.H., 1967. Why Mountain Passes are Higher in the Tropics. The American Naturalist 101, 233-249.

Jiggins, C.D., Salazar, C., Linares, M., Mavarez, J., 2008. Hybrid trait speciation and Heliconius butterflies. Philosophical Transactions of the Royal Society B: Biological Sciences 363, 3047-3054.

Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403-1405.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 14, 587.

Kamvar, Z.N., Tabima, J.F., Grünwald, N.J., 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2, e281.

1488 Kirwan, G.M., David, N., Gregory, S.M., Jobling, J.A., Steinheimer, F.D., Brito, G.R.R., 2016. The
1489 mistaken manakin: a new genus-group name for Parus pipra Linnaeus, 1758 (Aves: Passeriformes:
1490 Pipridae). Zootaxa 4121, 89-94.

1491 Kirwan, G.M., Green, G., 2012. Cotingas and Manakins. Princeton University Press, Princeton, NJ.

1492 Kumar, S., Subramanian, S., 2002. Mutation rates in mammalian genomes. Proceedings of the
1493 National Academy of Sciences 99, 803-808.

1494 Lamichhaney, S., Han, F., Webster, M.T., Andersson, L., Grant, B.R., Grant, P.R., 2018. Rapid
1495 hybrid speciation in Darwin's finches. Science 359, 224-228.

1496 Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. PartitionFinder 2: New
1497 Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological
1498 Phylogenetic Analyses. Molecular Biology and Evolution 34, 772-773.

1499 Lavretsky, P., Engilis, A., Eadie, J.M., Peters, J.L., 2015. Genetic admixture supports an ancient
1500 hybrid origin of the endangered Hawaiian duck. Journal of Evolutionary Biology 28, 1005-1015.

1501 Lawson, D.J., Hellenthal, G., Myers, S., Falush, D., 2012. Inference of Population Structure using
1502 Dense Haplotype Data. PLOS Genetics 8, e1002453.

1503 Leaché, A.D., Banbury, B.L., Felsenstein, J., de Oca, A.n.-M., Stamatakis, A., 2015. Short Tree,
1504 Long Tree, Right Tree, Wrong Tree: New Acquisition Bias Corrections for Inferring SNP
1505 Phylogenies. Systematic Biology 64, 1032-1047.

1506 Leaché, A.D., Fujita, M.K., Minin, V.N., Bouckaert, R.R., 2014. Species Delimitation using
1507 Genome-Wide SNP Data. Systematic Biology 63, 534-542.

1508 Leaché, A.D., Zhu, T., Rannala, B., Yang, Z., 2018. The Spectre of Too Many Species. Systematic
1509 Biology, syy051-syy051.

1510 Legendre, P., Fortin, M.-J., 2010. Comparison of the Mantel test and alternative approaches for
1511 detecting complex multivariate relationships in the spatial analysis of genetic data. Molecular
1512 Ecology Resources 10, 831-844.

1513 Linck, E.B., Battey, C.J., 2017. Minor allele frequency thresholds strongly affect population
1514 structure inference with genomic datasets. bioRxiv.

1515 Linder, H.P., 2001. On Areas of Endemism, with an Example from the African Restionaceae.
1516 Systematic Biology 50, 892-912.

1517 Long, C., Kubatko, L., 2017. The effect of gene flow on coalescent-based species-tree inference.
1518 arXiv:1710.03806.

1519 Luna, L.W., Souza, T.O., e Silva, W.A.d.G., Schneider, H., Sampaio, I., 2017. Genetic variation of
1520 the endangered Araripe Manakin (Antilophia bokermanni) indicates a history of demographic
1521 decline. Revista Brasileira de Ornitologia 25, 60-66.

1522 Lynch Alfaro, J.W., Boubli, J.P., Paim, F.P., Ribas, C.C., Silva, M.N.F.d., Messias, M.R., Röhe, F.,
1523 Mercês, M.P., Silva Júnior, J.S., Silva, C.R., Pinho, G.M., Koshkarian, G., Nguyen, M.T.T.,
1524 Harada, M.L., Rabelo, R.M., Queiroz, H.L., Alfaro, M.E., Farias, I.P., 2015. Biogeography of
1525 squirrel monkeys (genus Saimiri): South-central Amazon origin and rapid pan-Amazonian
1526 diversification of a lowland primate. Molecular Phylogenetics and Evolution 82, 436-454.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics. University of California Press, Berkeley, Calif., pp. 281-297.

Malinsky, M., Trucchi, E., Lawson, D.J., Falush, D., 2018. RADpainter and fineRADstructure: Population Inference from RADseq Data. Molecular Biology and Evolution 35, 1284-1290.

Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. Cancer research 27, 209-220.

Marks, B.D., Hackett, S.J., Capparella, A.P., 2002. Historical relationships among Neotropical lowland forest areas of endemism as determined by mitochondrial DNA sequence variation within the Wedge-billed Woodcreeper (Aves: Dendrocolaptidae: Glyphorynchus spirurus). Molecular Phylogenetics and Evolution 24, 153-167.

Marques, D.A., Meier, J.I., Seehausen, O., 2019. A Combinatorial View on Speciation and Adaptive Radiation. Trends in Ecology & Evolution.

Mayr, E., 1942. Systematics and the origin of species from the viewpoint of a zoologist. Columbia University Press.

McDonald, D.B., 2003. MICROSATELLITE DNA EVIDENCE FOR GENE FLOW IN NEOTROPICAL LEK-MATING LONG-TAILED MANAKINS. The Condor 105, 580-586.

Milá, B., Tavares, E.S., Muñoz Saldaña, A., Karubian, J., Smith, T.B., Baker, A.J., 2012. A Trans-Amazonian Screening of mtDNA Reveals Deep Intraspecific Divergence in Forest Birds and Suggests a Vast Underestimation of Species Diversity. PLOS ONE 7, e40541.

Moore, R.P., Robinson, W.D., Lovette, I.J., Robinson, T.R., 2008. Experimental evidence for extreme dispersal limitation in tropical forest birds. Ecology Letters 11, 960-968.

Morrone, J.J., 2015. Biogeographical regionalisation of the Andean region. Zootaxa 3936, 207-236.

Musher, L.J., Ferreira, M., Auerbach, A.L., McKay, J., Cracraft, J., 2019. Why is Amazonia a source of biodiversity? Climate-mediated dispersal and synchronous speciation across the Andes in an avian group (Tityrinae). Proceedings of the Royal Society B: Biological Sciences 286, 20182343.

Naka, L.N., 2011. Avian distribution patterns in the Guiana Shield: implications for the delimitation of Amazonian areas of endemism. Journal of Biogeography 38, 681-696.

Naka, L.N., Brumfield, R.T., 2018. The dual role of Amazonian rivers in the generation and maintenance of avian diversity. Science Advances 4.

Nieto Feliner, G., Álvarez, I., Fuertes-Aguilar, J., Heuertz, M., Marques, I., Moharrek, F., Piñeiro, R., Riina, R., Rosselló, J.A., Soltis, P.S., Villa-Machío, I., 2017. Is homoploid hybrid speciation that rare? An empiricist's view. Heredity 118, 513.

Noguera-urbano, E.A., 2016. Areas of endemism: travelling through space and the unexplored dimension. Systematics and Biodiversity 14, 131-139.

Nyári, Á.S., 2007. Phylogeographic patterns, molecular and vocal differentiation, and species limits in Schiffornis turdina (Aves). Molecular Phylogenetics and Evolution 44, 154-164.

Nylander, J.A.A., Olsson, U., Alström, P., Sanmartín, I., 2008. Accounting for Phylogenetic Uncertainty in Biogeography: A Bayesian Approach to Dispersal-Vicariance Analysis of the Thrushes (Aves: Turdus). Systematic Biology 57, 257-268.

Ohlson, J.I., Fjeldså, J., Ericson, P.G.P., 2013. Molecular phylogeny of the manakins (Aves: Passeriformes: Pipridae), with a new classification and the description of a new genus. Molecular Phylogenetics and Evolution 69, 796-804.

Paradis, E., 2010. pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics 26, 419-420.

Paris, J.R., Stevens, J.R., Catchen, J.M., Johnston, S., 2017. Lost in parameter space: a road map for stacks. Methods in Ecology and Evolution 8, 1360-1373.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PLOS ONE 7, e37135.

Petkova, D., Novembre, J., Stephens, M., 2015. Visualizing spatial population structure with estimated effective migration surfaces. Nature Genetics 48, 94.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of Population Structure Using Multilocus Genotype Data. Genetics 155, 945-959.

Prum, R.O., 1990. Phylogenetic Analysis of the Evolution of Display Behavior in the Neotropical Manakins (Aves: Pipridae). Ethology 84, 202-231.

Prum, R.O., 1994. PHYLOGENETIC ANALYSIS OF THE EVOLUTION OF ALTERNATIVE SOCIAL BEHAVIOR IN THE MANAKINS (AVES: PIPRIDAE). Evolution 48, 1657-1675.

Pupim, F.N., Sawakuchi, A.O., Almeida, R.P., Ribas, C.C., Kern, A.K., Hartmann, G.A., Chiessi, C.M., Tamura, L.N., Mineli, T.D., Savian, J.F., Grohmann, C.H., Bertassoli, D.J., Stern, A.G., Cruz,

F.W., Cracraft, J., 2019. Chronology of Terra Firme formation in Amazonian lowlands reveals a dynamic Quaternary landscape. Quaternary Science Reviews 210, 154-163.

Quintero, I., Jetz, W., 2018. Global elevational diversity and diversification of birds. Nature 555, 246.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Systematic Biology 67, 901-904.

Räsänen, M.E., Salo, J.S., Kalliola, R.J., 1987. Fluvial perturbance in the western Amazon basin: regulation by long-term sub-Andean tectonics. Science 238, 1398-1401.

Ribas, C.C., Aleixo, A., Nogueira, A.C.R., Miyaki, C.Y., Cracraft, J., 2012. A palaeobiogeographic model for biotic diversification within Amazonia over the past three million years. Proceedings of the Royal Society B: Biological Sciences 279, 681-689.

Ribas, C.C., Moyle, R.G., Miyaki, C.Y., Cracraft, J., 2007. The assembly of montane biotas: linking Andean tectonics and climatic oscillations to independent regimes of diversification in Pionus parrots. Proceedings of the Royal Society B: Biological Sciences 274, 2399-2408.

Ridgely, R.S., Greenfield, P.J., 2001. The birds of Ecuador: status, distribution, and taxonomy. Cornell University Press.

Ridgely, R.S., Tudor, G., 2009. Field guide to the songbirds of South America: the passerines. University of Texas Press.

Ridgway, R., 1906. Descriptions of some new forms of oligomyodian birds. Proceedings of the Biological Society of Washington, 115-120.

1608 Roure, B., Baurain, D., Philippe, H., 2013. Impact of Missing Data on Phylogenies Inferred from
1609 Empirical Phylogenomic Data Sets. Molecular Biology and Evolution 30, 197-214.

1610 Rousset, F., 1997. Genetic Differentiation and Estimation of Gene Flow from F Statistics Under
1611 Isolation by Distance. Genetics 145, 1219-1228.

1612 Salazar, C., Baxter, S.W., Pardo-Diaz, C., Wu, G., Surridge, A., Linares, M., Bermingham, E.,
1613 Jiggins, C.D., 2010. Genetic Evidence for Hybrid Trait Speciation in Heliconius Butterflies. PLOS
1614 Genetics 6, e1000930.

1615 Saranathan, V., Hamilton, D., Powell, G.V.N., Kroodsma, D.E., Prum, R.O., 2007. Genetic
1616 evidence supports song learning in the three-wattled bellbird Procnias tricarunculata (Cotingidae).
1617 Molecular Ecology 16, 3689-3702.

1618 Schmidt, H.A., Minh, B.Q., von Haeseler, A., Nguyen, L.-T., 2014. IQ-TREE: A Fast and Effective
1619 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and
1620 Evolution 32, 268-274.

1621 Schmidt-Lebuhn, A.N., Aitken, N.C., Chuah, A., 2017. Species trees from consensus single
1622 nucleotide polymorphism (SNP) data: Testing phylogenetic approaches with simulated and
1623 empirical data. Molecular Phylogenetics and Evolution 116, 192-201.

1624 Schumer, M., Rosenthal, G.G., Andolfatto, P., 2014. HOW COMMON IS HOMOPLOID HYBRID
1625 SPECIATION? Evolution 68, 1553-1560.

1626 Sedano, R.E., Burns, K.J., 2010. Are the Northern Andes a species pump for Neotropical birds?
1627 Phylogenetics and biogeography of a clade of Neotropical tanagers (Aves: Thraupini). Journal of
1628 Biogeography 37, 325-343.

1629 Shafer, A.B.A., Peart, C.R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C.W., Wolf, J.B.W., 2017.
1630 Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic
1631 inference. Methods in Ecology and Evolution 8, 907-917.

1632 Silva, S.M., Peterson, A.T., Carneiro, L., Burlamaqui, T.C.T., Ribas, C.C., Sousa-Neves, T.,
1633 Miranda, L.S., Fernandes, A.M., d'Horta, F.M., Araújo-Silva, L.E., Batista, R., Bandeira,
1634 C.H.M.M., Dantas, S.M., Ferreira, M., Martins, D.M., Oliveira, J., Rocha, T.C., Sardelli, C.H.,
1635 Thom, G., Rêgo, P.S., Santos, M.P., Sequeira, F., Vallinoto, M., Aleixo, A., 2019. A dynamic
1636 continental moisture gradient drove Amazonian bird diversification. Science Advances 5, eaat5752.

1637 Smeds, L., Qvarnström, A., Ellegren, H., 2016. Direct estimate of the rate of germline mutation in a
1638 bird. Genome Research 26, 1211-1218.

1639 Smith, B.T., McCormack, J.E., Cuervo, A.M., Hickerson, M.J., Aleixo, A., Cadena, C.D., Perez-
1640 Eman, J., Burney, C.W., Xie, X., Harvey, M.G., Faircloth, B.C., Glenn, T.C., Derryberry, E.P.,
1641 Prejean, J., Fields, S., Brumfield, R.T., 2014. The drivers of tropical speciation. Nature 515, 406-
1642 409.

1643 Snow, D., 1961. THE DISPLAYS OF THE MANAKINS PIPRA PIPRA AND TYRANNEUTES
1644 VIRESCENS*. Ibis 103A, 110-113.

1645 Snow, D.W., 2004. Family Pipridae (Manakins). In: del Hoyo, J., Elliot, A., Christie, D.A. (Eds.),
1646 Handbook of the Birds of the World. Lynx Edicions, Barcelona, pp. 110-169.

1647 Spencer, A., 2012. White-crowned Manakin vocal variation. Xeno-Cango.org.

1648 Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
1649 phylogenies. Bioinformatics 30, 1312-1313.

1650 Suh, A., 2016. The phylogenomic forest of bird trees contains a hard polytomy at the root of
1651 Neoaves. Zoologica Scripta 45, 50-62.

1652 Sukumaran, J., Knowles, L.L., 2017. Multispecies coalescent delimits structure, not species.
1653 Proceedings of the National Academy of Sciences 114, 1607-1612.

1654 Swofford, D.L., 2002. Phylogenetic analysis using parsimony (*and Other Methods). Version 4.
1655 Sinauer Associates, Sunderland, Massachusetts.

1656 Tello, J.G., Moyle, R.G., Marchese, D.J., Cracraft, J., 2009. Phylogeny and phylogenetic
1657 classification of the tyrant flycatchers, cotingas, manakins, and their allies (Aves: Tyrannides).
1658 Cladistics 25, 429-467.

1659 Teske, P.R., Golla, T.R., Sandoval-Castillo, J., Emami-Khoyi, A., van der Lingen, C.D., von der
1660 Heyden, S., Chiazzari, B., Jansen van Vuuren, B., Beheregaray, L.B., 2018. Mitochondrial DNA is
1661 unsuitable to test for isolation by distance. Scientific Reports 8, 8448.

1662 Thom, G., Amaral, F.R.D., Hickerson, M.J., Aleixo, A., Araujo-Silva, L.E., Ribas, C.C., Choueri,
1663 E., Miyaki, C.Y., 2018. Phenotypic and Genetic Structure Support Gene Flow Generating Gene
1664 Tree Discordances in an Amazonian Floodplain Endemic Species. Systematic Biology 67, 700-718.

1665 Thrasher, D.J., Butcher, B.G., Campagna, L., Webster, M.S., Lovette, I.J., 2018. Double-digest
1666 RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness:
1667 A proof of concept in a highly promiscuous bird. Molecular Ecology Resources 0.

1668 Trifinopoulos, J., Nguyen, L.-T., Minh, B.Q., von Haeseler, A., 2016. W-IQ-TREE: a fast online
1669 phylogenetic tool for maximum likelihood analysis. Nucleic Acids Research 44, W232-W235.

1670 Vaidya, G., Lohman, D.J., Meier, R., 2011. SequenceMatrix: concatenation software for the fast
1671 assembly of multi-gene datasets with character set and codon information. Cladistics 27, 171-180.

1672 van Els, P., Norambuena, H.V., Etienne, R.S., 2019. From pampa to puna: Biogeography and
1673 diversification of a group of Neotropical obligate grassland birds (Anthus: Motacillidae). Journal of
1674 Zoological Systematics and Evolutionary Research 00: 1– 12.

1675 Voelker, G., 1999. DISPERSAL, VICARIANCE, AND CLOCKS: HISTORICAL
1676 BIOGEOGRAPHY AND SPECIATION IN A COSMOPOLITAN PASSERINE GENUS
1677 (ANTHUS: MOTACILLIDAE). Evolution 53, 1536-1552.

1678 Wallace, A.R., 1854. On the monkeys of the Amazon. Annals and Magazine of Natural History 14,
1679 451-454.

1680 Weir, B.S., Cockerham, C.C., 1984. ESTIMATING F-STATISTICS FOR THE ANALYSIS OF
1681 POPULATION STRUCTURE. Evolution 38, 1358-1370.

1682 Weir, J.T., 2006. DIVERGENT TIMING AND PATTERNS OF SPECIES ACCUMULATION IN
1683 LOWLAND AND HIGHLAND NEOTROPICAL BIRDS. Evolution 60, 842-855.

1684 Wen, D., Yu, Y., Zhu, J., Nakhleh, L., 2018. Inferring Phylogenetic Networks Using PhyloNet.
1685 Systematic Biology, syy015-syy015.

1686 Wetmore, A., 1972. The birds of the Republic of Panama. Part 3—Passeriformes: Dendrocolaptidae
1687 (woodcreepers) to Oxyruncidae (sharpbills). Smithsonian Miscellaneous Collection 150.
1688 Smithsonian Institution Press, Washington, DC.

1689 Wright, S., 1931. Evolution in Mendelian populations. Genetics 16, 97.

Zhang, J., Kapli, P., Pavlidis, P., Stamatakis, A., 2013. A general species delimitation method with applications to phylogenetic placements. Bioinformatics 29, 2869-2876.

Zhu, J., Wen, D., Yu, Y., Meudt, H.M., Nakhleh, L., 2018. Bayesian inference of phylogenetic networks from bi-allelic genetic markers. PLOS Computational Biology 14, e1005932.

Zimmer, J.T., 1936. Studies of Peruvian birds. XXII. Notes on the Pipridae. American Museum novitates; no. 889. 1-29.

Figure 1. Neotropical areas of endemism and sampling localities. Amazonian lowland areas of endemism as portrayed in Da Silva et al. (2005) are highlighted in shades of green, as well as the Jaú area of endemism (Borges and Da Silva, 2012). Montane Andean, Central American, Guianan, and dry diagonal regions (>1000m) are emphasized in grey. As summarized in Da Silva et al. (2005), a sequence of authors identified seven areas of endemism for lowland birds that retained Wallace's "Guyana" (Wallace, 1854), split "Ecuador" into "Imeri" and "Napo," renamed "Peru" to "Inambari," and split "Brazil" into "Rondônia," "Pará," and "Belém" (Cracraft, 1985; Cracraft and Prum, 1988; Haffer, 1978; Haffer, 1985; Haffer, 1992). "Pará" was subsequently further partitioned into two regions, "Tapajós" and "Xingu," (Da Silva et al., 2002). Recently, additional sub-partitions have been proposed for the "Napo" (Jaú - Borges and Da Silva, 2012), and Guianan areas of endemism (Naka, 2011). Plotted points indicate field sampling localities for 286 *Pseudopipra* individuals sequenced for genetic analysis in the present study. Inset top right is an illustration of the *pipra* subspecies, reproduced from the Handbook of the Birds of the World, Lynx Edicions.

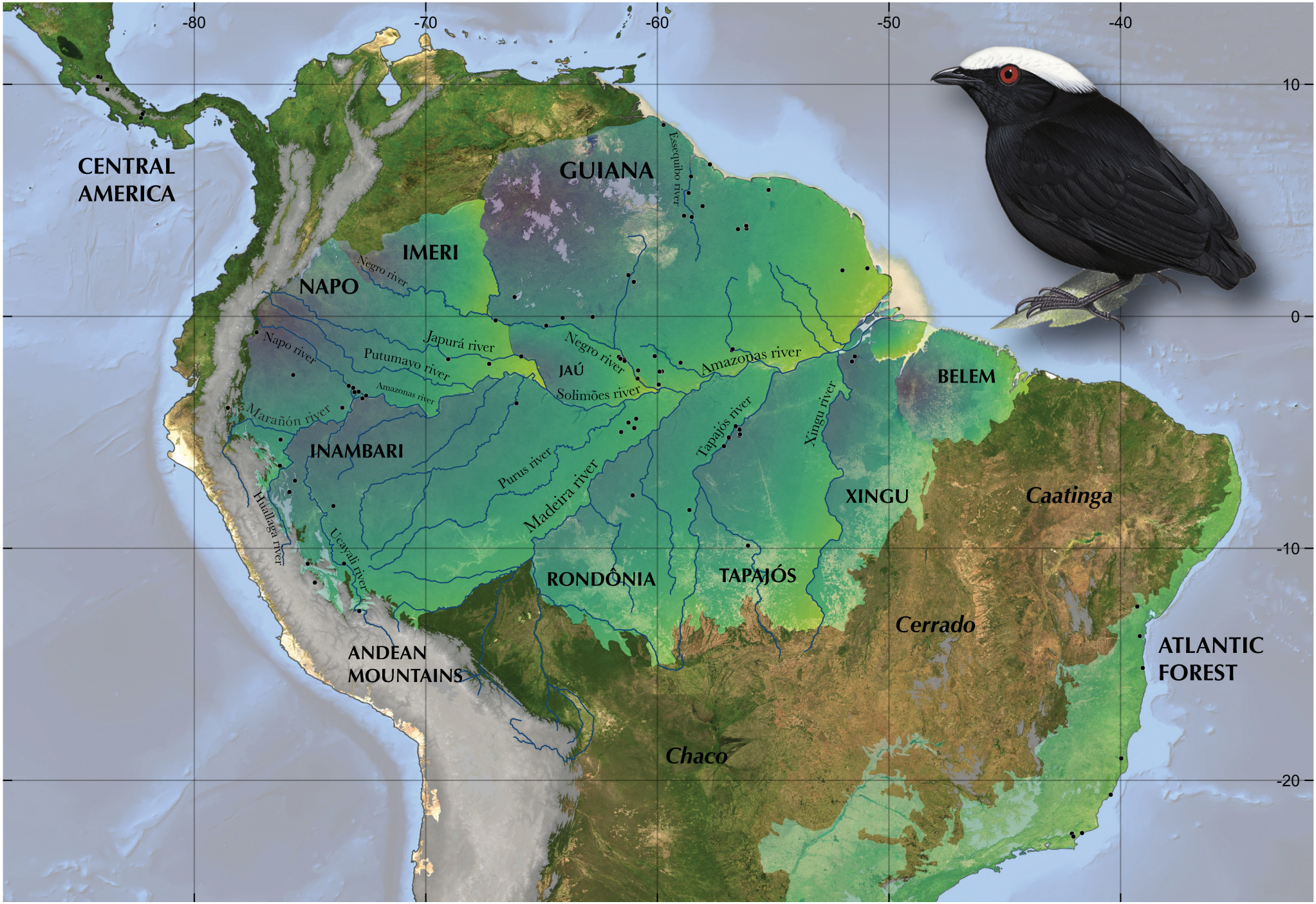
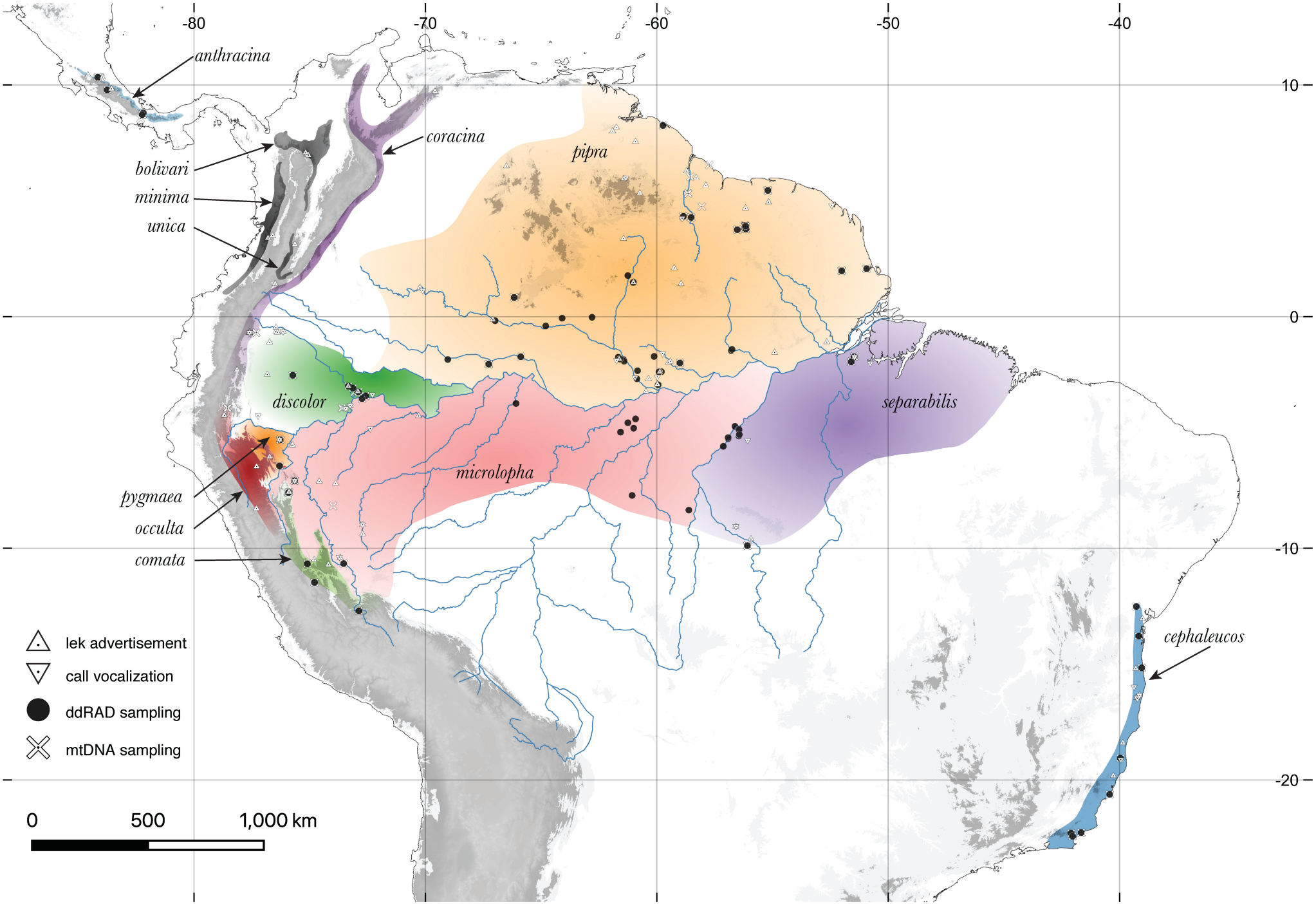

-80
-70
-60
-50
-40
10
0
-10
-20
CENTRAL AMERICA
GUIANA
Essequibo river
IMERI
Negro river
NAPO
Napo river
Japurá river
Putumayo river
Amazonas river
Negro river
JAÚ
Solimões river
Amazonas river
BELEM
Marañón river
INAMBARI
Purus river
Madeira river
Tapajós river
Xingu river
XINGU
Caatinga
Huallaga river
Ucayali river
RONDÔNIA
TAPAJÓS
Cerrado
ANDEAN MOUNTAINS
ATLANTIC FOREST
Chaco

Figure 2. Subspecies sampling. Our genetic sampling of *Pseudopipra pipra* intersects with the ranges of ten named subspecies. In this figure, we indicate sampling localities as they overlap with previously identified subspecies ranges. Areas greater than 1000m in elevation are emphasized in light grey; major Amazonian rivers are indicated in light blue. Overall distribution reflects the BirdLife International approximation, updated to reflect current knowledge (BirdLife International Species factsheet: *Pseudopipra pipra*, 2018, also see Supplemental Figure 1). The ranges of Andean taxa were adjusted based on unpublished data provided Andrés Cuervo (*personal communication*). Hypothetical subspecies ranges are based on descriptions in Kirwan and Green (2012) and imply maximal ranges highlighting key geographic barriers delimiting subspecies. Black dots indicate sampling for ddRADseq, while crosses indicate sampling for mtDNA ND2. Triangles indicate sampling localities for male advertisement vocalizations (upward) and call vocalizations (downward). Our sampling encompasses the majority of the lowland extents of the recognized range of *Pseudopipra*, as well as key montane populations in Peru, Ecuador, and Central America. Ssp *anthracina, bolivari, minima, unica, coracina, occulta, and comata* are exclusively montane.


-80
-70
-60
-50
-40
10
0
-10
-20
anthracina
coracina
bolivari
minima
unica
pipra
discolor
separabilis
microlopha
pygmaea
occulta
comata
cephaleucos
lek advertisement
call vocalization
ddRAD sampling
mtDNA sampling
0
500
1,000 km

1758 **Figure 3.** Phylogenetic hypothesis for *Pseudopipra*. Phylogenetic analysis of concatenated ddRAD
1759 sequences generated a well-resolved phylogenetic hypothesis which was largely congruent across
1760 datasets and analytical approaches. Montane clades, including a Northern Andean (A1) + Central
1761 American (A2) clade (A), and a central Peruvian clade (B), are recovered as nested sister groups to
1762 two wide ranging lowland clades (C and D), indicating that the lowland clades are descended from
1763 Andean lineages (Figure 5). Shown is the RAxML topology generated using the 50% haplotype
1764 dataset. Branch lengths are set to equal for display (see Supplementary Data for original tree files).
1765 Colored circles at tips indicate group membership to one of eighteen population-areas (matching
1766 Figure 4 and 5). Heatmaps at key nodes indicate bootstrap scores from each of six phylogenetic
1767 analyses for 20/50/80% datasets, as indicated in the legend. Black dots on branches indicate
1768 populations which are identified with 100% posterior probability in fineRADstructure analysis.
1769 Deep phylogenetic structure is recovered by both RAxML and SVDquartets with high support,
1770 though RAxML recovers additional low-support substructure in the Northern Amazonian + Guiana
1771 Shield clade (D) which is coincident with several geographic features (clades D1 and D2). A clade
1772 of introgressed western Napo individuals (E) is recovered by all standard phylogenetic analyses as
1773 sister to the northern Amazonian + Guiana shield clade (D). As the history of this clade is
1774 characterized by a complex introgression scenario between northern and southern lineages, a purely
1775 bifurcating tree model is inadequate to describe its relations with other groups (see text and Figure
1776 6). Also shown are the well supported phylogenetic positions of two distinct subspecies lineages
1777 (*pygmaea* and *coracaina*) from which we were able to obtain mtDNA (Supplemental Appendix).
1778
1779
1780

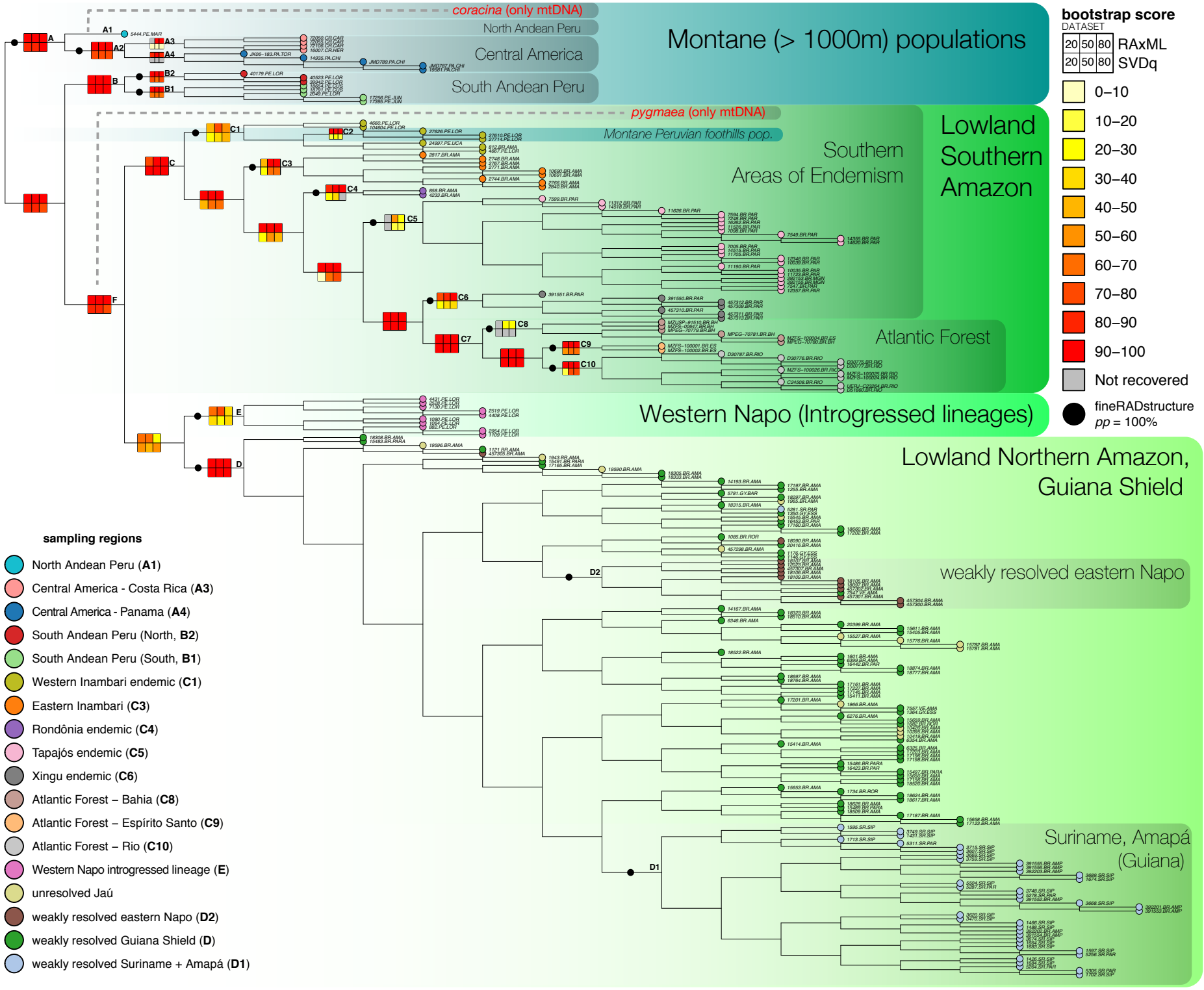
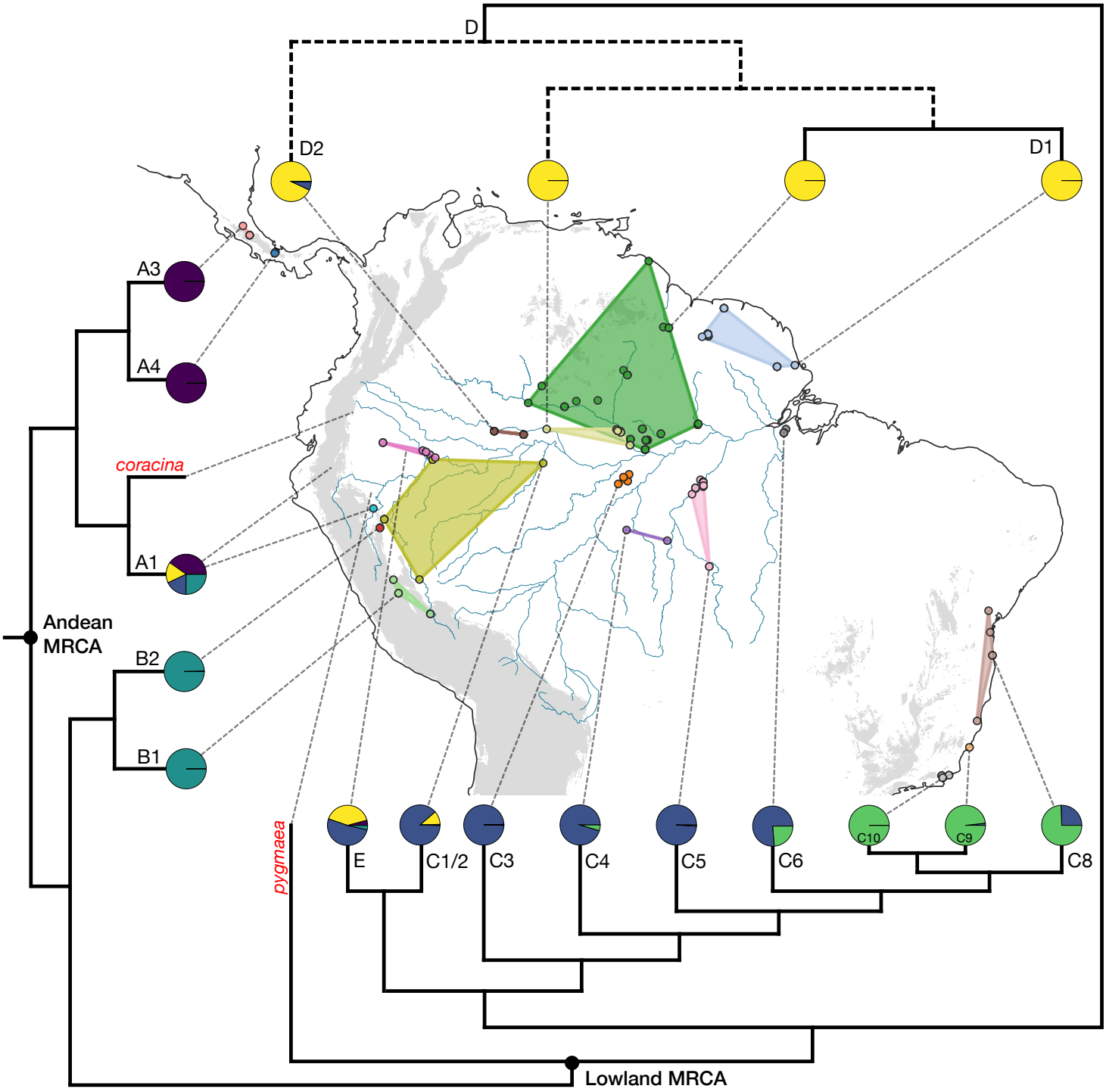

coracina (only mtDNA)
North Andean Peru
Central America
South Andean Peru
Montane (> 1000m) populations
pygmaea (only mtDNA)
Montane Peruvian foothills pop.
Southern
Areas of Endemism
Lowland
Southern
Amazon
Atlantic Forest
Western Napo (Introgressed lineages)
Lowland Northern Amazon,
Guiana Shield
weakly resolved eastern Napo
Suriname, Amapá
(Guiana)
bootstrap score
DATASET
20 50 80 RAxML
20 50 80 SVDq
0–10
10–20
20–30
30–40
40–50
50–60
60–70
70–80
80–90
90–100
Not recovered
fineRADstructure
pp = 100%
sampling regions
North Andean Peru (A1)
Central America - Costa Rica (A3)
Central America - Panama (A4)
South Andean Peru (North, B2)
South Andean Peru (South, B1)
Western Inambari endemic (C1)
Eastern Inambari (C3)
Rondônia endemic (C4)
Tapajós endemic (C5)
Xingu endemic (C6)
Atlantic Forest – Bahia (C8)
Atlantic Forest – Espírito Santo (C9)
Atlantic Forest – Rio (C10)
Western Napo introgressed lineage (E)
unresolved Jaú
weakly resolved eastern Napo (D2)
weakly resolved Guiana Shield (D)
weakly resolved Suriname + Amapá (D1)

Figure 4. Mapping of phylogenetic relationships into geographic space. Here, a simplified topology is plotted with tips anchored to sampling coordinates. The colors of convex hulls (i.e., hypotheses of minimum clade ranges) match tip colors in Figure 3 and 5. These colored hulls represent distinct non-overlapping geographic areas delimited by physical barriers (though this is somewhat visually confounded in the Northern Amazon + Guiana Shield clade (D), with two weakly resolved lineages (e.g. pale yellow Jaú and green Guiana Shield population groups appear to have some overlap in this figure, but they are delimited by the Rio Negro). Clade D bipartitions are shown with dashed branches (top) to indicate low support for phylogenetic structure (Figure 3, results). The ancestral habit of the genus is inferred to be Andean (Figure 5, Supplementary Appendix), with a single origin of lowland lineages. Also shown at the tips of the simplified tree are median admixture proportions across sampling localities as inferred with STRUCTURE (Figure 5).


D
D2
D1
A3
A4
coracina
A1
Andean
MRCA
B2
B1
pygmaea
E
C1/2
C3
C4
C5
C6
C10
C9
C8
Lowland MRCA

1804 **Figure 5.** Cluster and admixture analysis of SNP data. Left: RAxML topology using the 50%
1805 haplotype dataset, with tip labels and colors indicating group membership to one of eighteen
1806 population-areas (matching those in Figures 3, 4). Branch colors indicate ancestral area, according
1807 to the results from SIMMAP and maximum parsimony analysis (Supplemental Appendix), and
1808 unambiguously reconstruct the *Pseudopipra* MRCA to have been Andean. Center: STRUCTURE
1809 output for Dataset 1, depicting population assignment and admixture estimates for K=5. The
1810 likelihood of each evaluated number of K clusters from 1:20 plateaued at K = 5, with the standard
1811 deviation across runs increasing rapidly after this point (Supplemental Figure 2). The five optimized
1812 clusters broadly correspond to wide biogeographic Neotropical regions which coincide with
1813 lowland areas of endemism, Central American, and Peruvian montane regions (Figure 4). Given
1814 these five clusters, each labeled focal region is inferred to have a unique combination of admixture
1815 proportions. The admixture cline inferred from the Xingu to the southern Atlantic Forest appears to
1816 be a product of isolation by distance (Supplementary Appendix), while the signature of
1817 introgression inferred for western Napo lineages appears to be a product of substantial historical
1818 introgression from a northern lineage, into a distinct Southern Amazonian lineage (see Figure 6 and
1819 discussion). See Supplementary Figure 3 for K2-10 results. Right: clustering solutions for SNP
1820 dataset D1-D4m (described in Table 1, also see Supplementary Figure 2), as well as for the
1821 coancestry matrix (CoA). Black bars indicate distinct clusters which are also detected as
1822 monophyletic groups in the phylogenetic analysis. Red bars indicate one distinct cluster in a given
1823 dataset, but which is paraphyletic with respect to the phylogenetic analysis. The grey bars for D2m
1824 indicate several groups of individuals that were not resolvable in this dataset (Supplementary Figure
1825 2).
1826

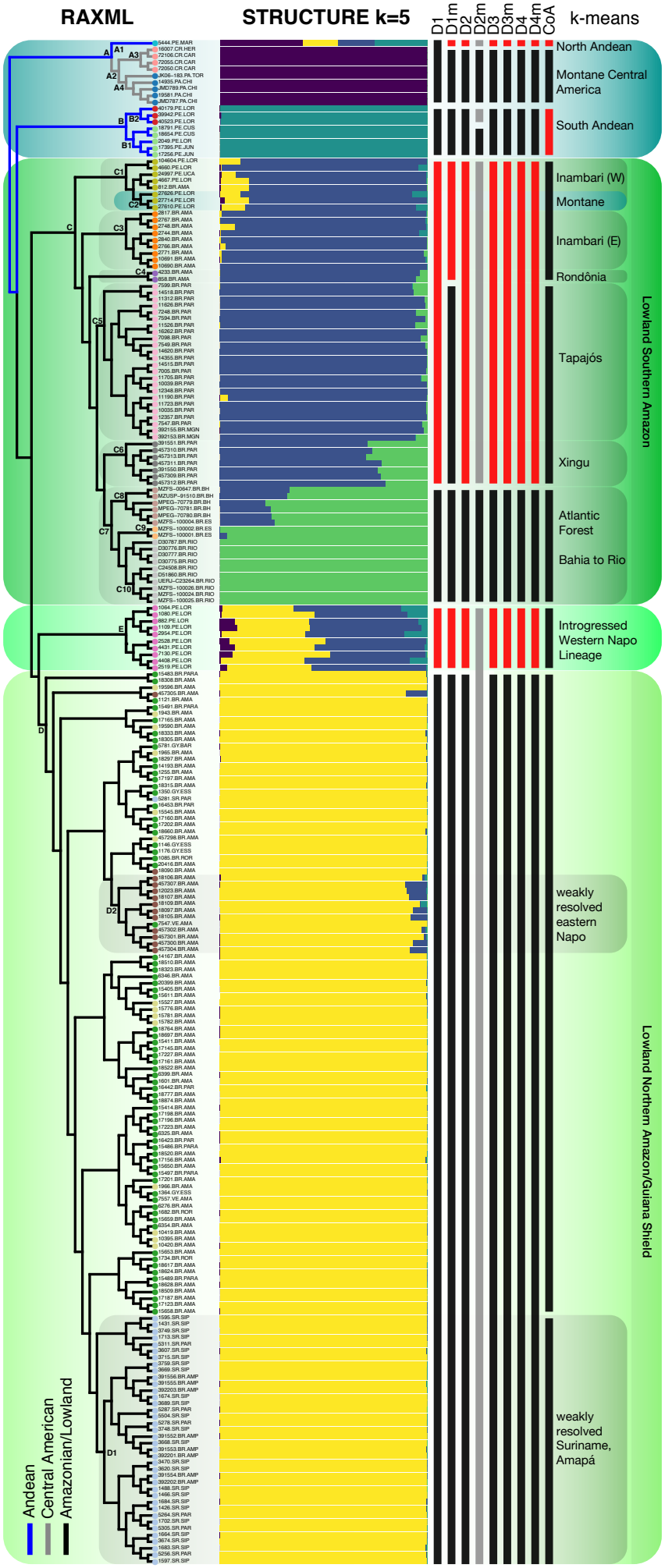

RAXML
STRUCTURE k=5
D1
D1m
D2
D2m
D3
D3m
D4
D4m
CoA
k-means
North Andean
Montane Central America
South Andean
Inambari (W)
Montane
Inambari (E)
Rondônia
Tapajós
Xingu
Lowland Southern Amazon
Atlantic Forest
Bahia to Rio
Introgressed Western Napo Lineage
weakly resolved eastern Napo
Lowland Northern Amazon/Guiana Shield
weakly resolved Suriname, Amapá
Andean
Central American
Amazonian/Lowland

Figure 6. Examination of the introgressed lineage in the western Napo area of endemism. All of the available genomic evidence is consistent with the hypothesis that individuals in the western Napo area of endemism are derived from a distinct southern Amazonian lineage (S2a in Panel A) which is not represented in contemporary samples. This lineage (S2a) experienced substantial introgression from the ancestral populations of what are today eastern Napo and Jaú restricted linages. Panel A shows the phylogenetic network (in black) inferred by PhyloNet, with arrows indicating the direction of introgression estimated in G-PhoCS. The red arrows along the PhyloNet reticulation indicate inferred introgression from the northern Amazonian lineages (clade C1 in Figure 3) into this southern Amazonian lineage (S2a). The dashed grey line connecting S2a/S2 to S1 indicates possible introgression into Inambari lineages, which may have occurred after or concurrently with the primary introgression event and brought northern (Clade D) alleles into SW Amazonian populations (see text and Figure 5). Panel B displays the individual level coancestry matrix for introgressed Napo individuals and their source/progenitor lineages; darker colors indicate greater coancestry. Introgressed (PH: putative hybrid) hybrids have greater coancestry with each of their progenitor lineages (S2a and Northern Amazon+Guiana shield (abbreviated GS)) than their progenitor lineages share with each other (PHvGS/progenitors = 1.078, PHvInambari/progenitors = 1.17, see text). In the proposed scenario, the southern progenitor lineage (S2a) is either unsampled or extinct and has not persisted sympatrically with S2. Thus, the western Inambari southern lineage (S1) comprises the closest living relatives to S2a (see discussion). Panel C summarizes coancestry variance as projected into component space after normalizing. Panel D summarizes the minimum implied geographic extents of the groups involved (as sampled in this study). For the convex hull representing the range of the western Napo lineage, dark shading indicates localities for ddRAD sampling, while light shading indicates indicates additional localities of the Napo lineage as implied by mtDNA samples. G-PhoCS demographic parameter estimates are also depicted in panel D (m:

1851 migrants per generation, when lower 95% CI > 1, τ: splitting time in generations (x10^6), θ: effective
1852 population size (10^6), Supplemental Table 5). In A-D, two subpopulations (East and West) are
1853 indicated for each progenitor group, as outlined with outlined minimum convex hulls in D.
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873

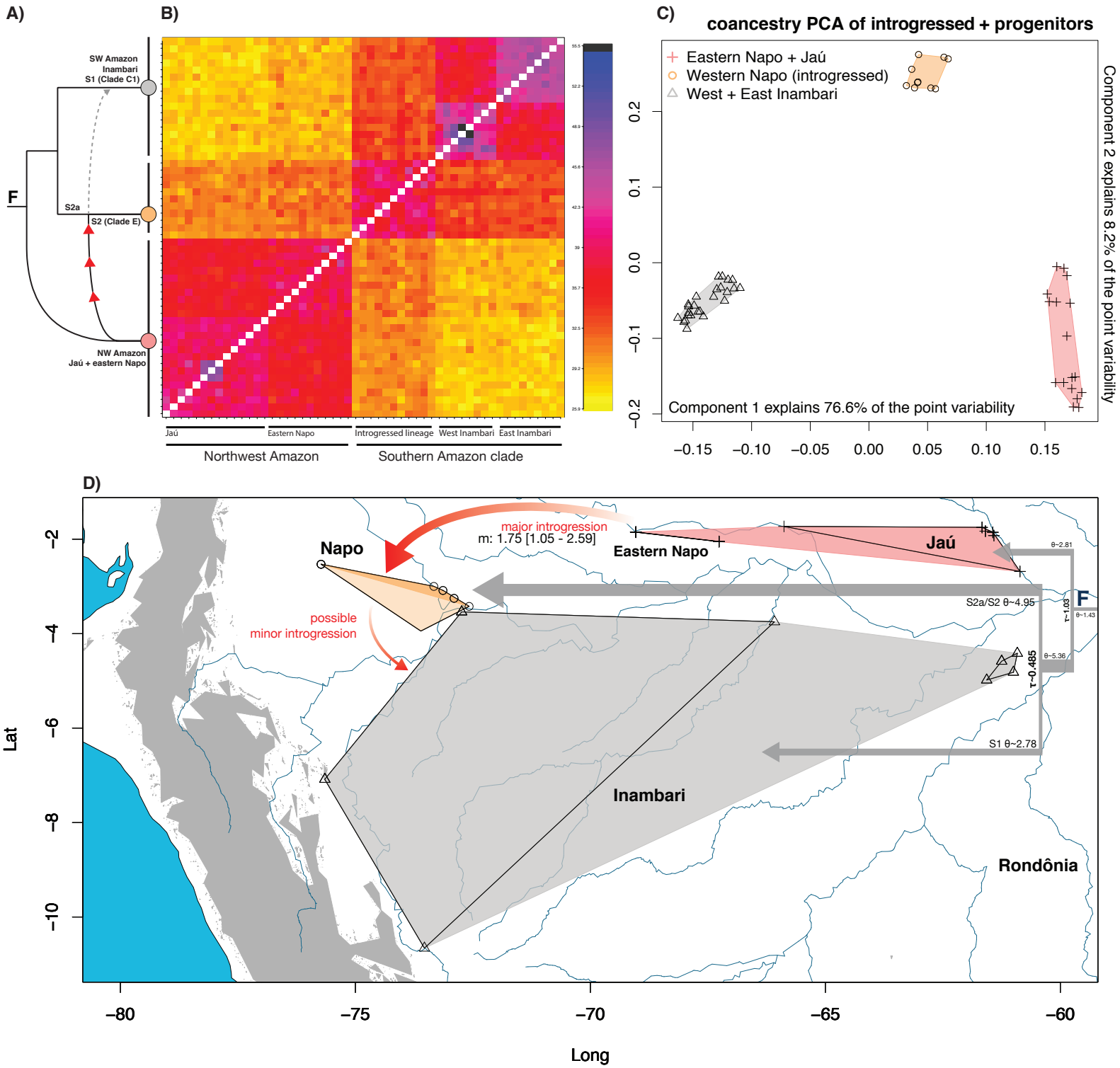

A)
B)
C)
D)
SW Amazon
Inambari
S1 (Clade C1)
F
S2a
S2 (Clade E)
NW Amazon
Jaú + eastern Napo
Jaú
Eastern Napo
Introgressed lineage
West Inambari
East Inambari
Northwest Amazon
Southern Amazon clade
coancestry PCA of introgressed + progenitors
Eastern Napo + Jaú
Western Napo (introgressed)
West + East Inambari
Component 1 explains 76.6% of the point variability
Component 2 explains 8.2% of the point variability
major introgression
m: 1.75 [1.05 - 2.59]
Napo
Eastern Napo
Jaú
possible
minor introgression
S2a/S2 θ~4.95
θ~2.81
θ~1.43
τ~1.03
θ~5.36
τ~0.485
S1 θ~2.78
Inambari
Rondônia
Lat
Long

Figure 7. Estimated Effective migration surface (EEMS). The estimated migration surface links the topography and drainage system of the Neotropics to spatial patterns of genetic variation. Shown are results estimated from a model with 2000 demes (vertices). Blue colors indicate areas where gene flow is more likely to be occurring (i.e., more connectivity than isolation by distance), while brown colors indicate areas where barriers to gene flow are highly probable (i.e. less connectivity than isolation by distance). In this analysis, the Andes and the Amazon River are the most clearly inferred barriers to gene flow. Major rivers, as well as a >1000m contour (light grey) are overlaid to aid interpretation. See text for additional discussion.

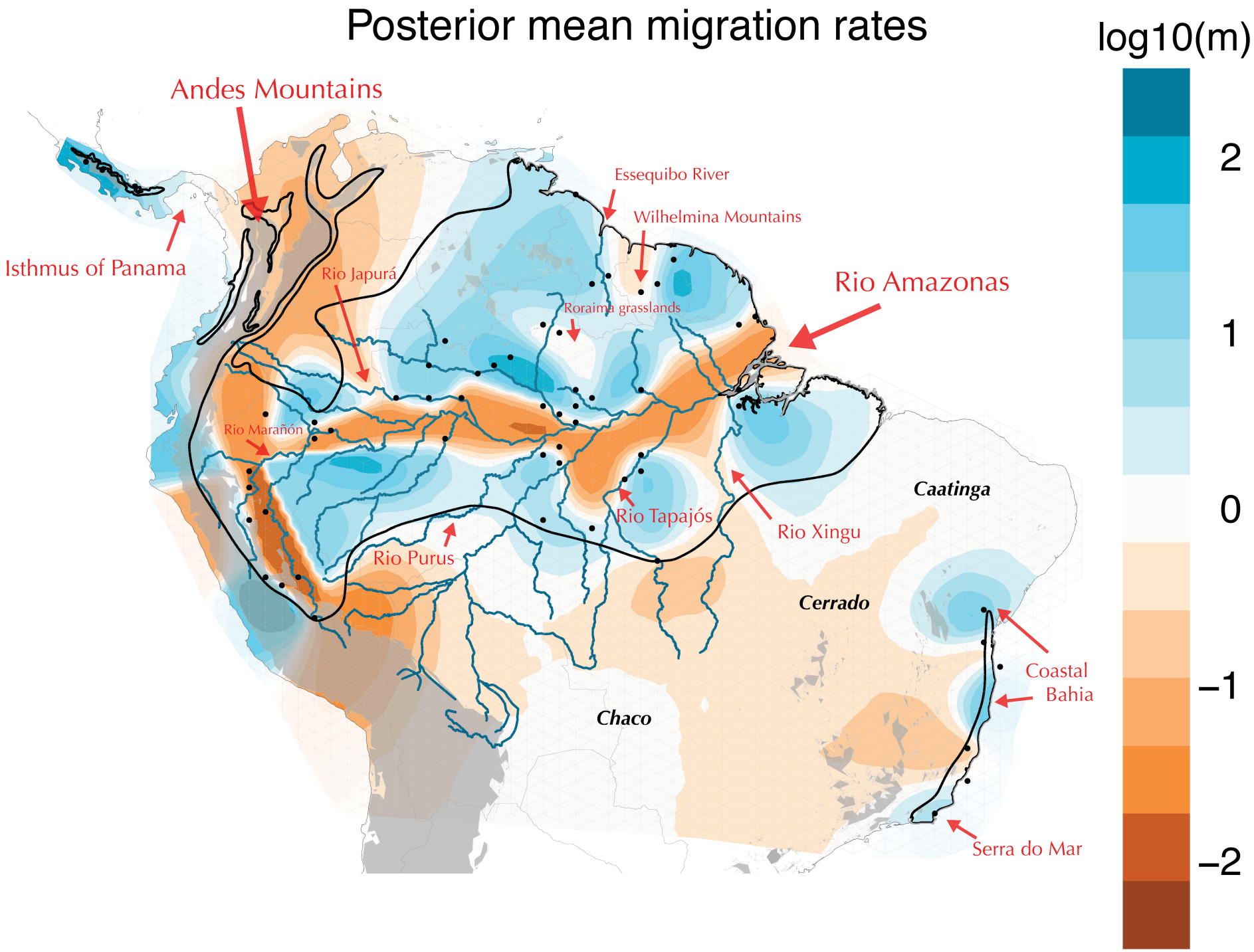
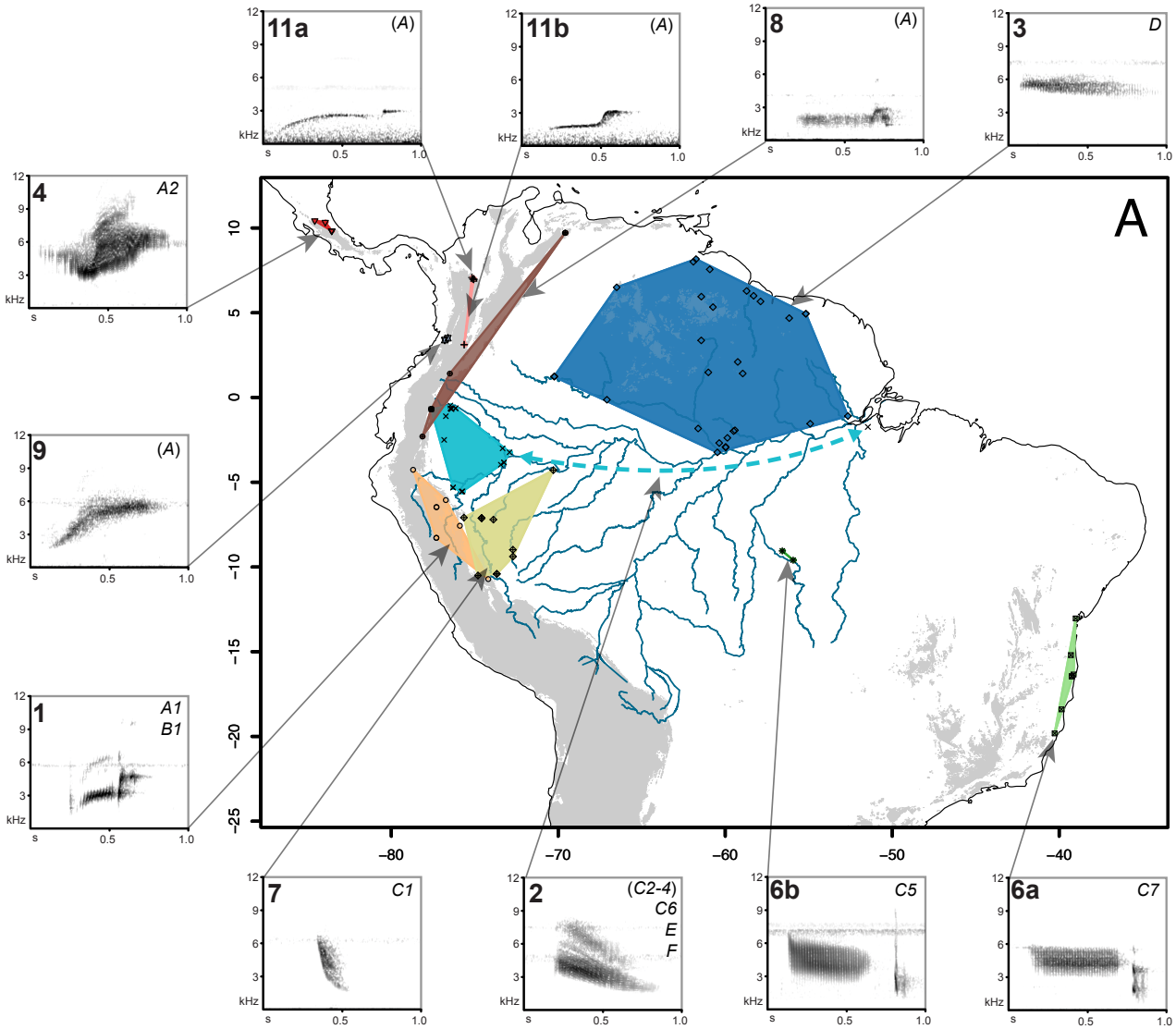

Posterior mean migration rates
log10(m)
2
1
0
−1
−2
Andes Mountains
Isthmus of Panama
Rio Japurá
Rio Marañón
Essequibo River
Wilhelmina Mountains
Roraima grasslands
Rio Amazonas
Rio Tapajós
Rio Xingu
Rio Purus
Caatinga
Cerrado
Chaco
Coastal Bahia
Serra do Mar

Figure 8. Panel A: Summary of lek vocalization phenotypes. Our analysis identified 11 qualitatively distinct lek vocalization types that can be easily diagnosed by ear, or by visual inspection of sonograms (arbitrarily labeled numerically). See Supplementary Figure 13 for vocalization measures. The inset map indicates the minimum implied ranges for each identified vocalization type, with colors and plotting symbols matching those in PCA plots (Supplemental Figures 14, 15). The thicker pale blue dashed line indicates where the 'southern Amazon' vocalization type 2 may be shared, apparently plesiomorphically (Supplementary Figure 16). Vocalization type 2 may therefore occur in other unsampled areas of the southern Amazon. Type 1 also appears to be shared pleisiomorphically between populations on the eastern slopes of the Andes (Clade A1) and in populations from the southern Cordillera Azul in Loreto (Clade B1). Panel B: Summary of call vocalization phenotypes. Our analysis identified three distinct but similar call vocalization types that can be diagnosed by ear, or by visual inspection of sonograms. Vocalization type 5 was primarily recorded in the northern Amazon basin, coinciding with the broad ranging Northern Amazon + Guiana Shield lineage (Clade D). From our initial classification, one recording from the Atlantic Forest (XC427315) was assigned to type 5, suggesting this type may be shared between the Guiana Shield clade and the Atlantic Forest clade (not shown, supplemental discussion). As with lek vocalization type 1, call vocalization type 10 was shared pleisiomorphically across Andean clades (A1 + B1). Call vocalization type 13 was entirely restricted to but shared across diverse lowland southern Amazonian forms (Clade C). Letters to the top right of sonograms indicate explicit [i.e., *A*] or implicit clade associations [i.e., (*A*)]. See Supplemental Appendix for additional details and discussion.


A
11a (A)
11b (A)
8 (A)
3 D
4 A2
9 (A)
1 A1 B1
7 C1
2 (C2-4) C6 E F
6b C5
6a C7
kHz
s
0.5
1.0
10
5
0
−5
−10
−15
−20
−25
−80
−70
−60
−50
−40

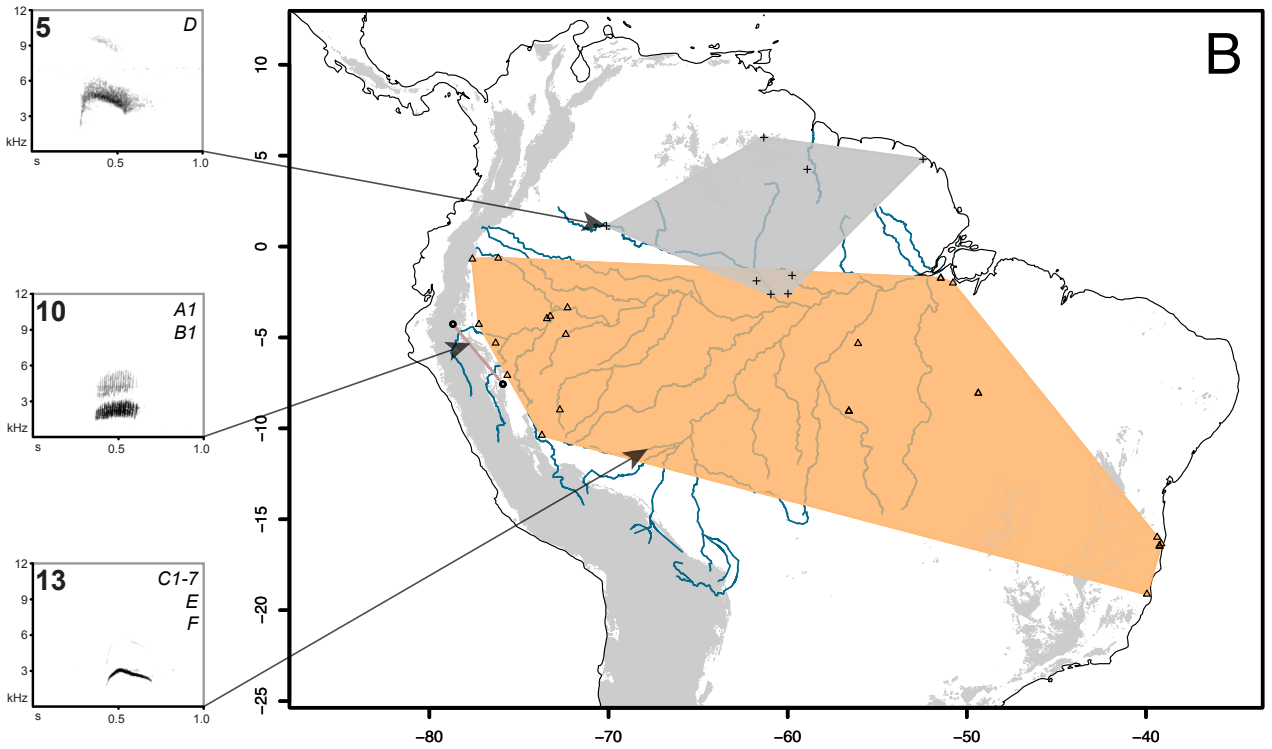

B
5 D
10 A1 B1
13 C1-7 E F
kHz
s
0.5
1.0
10
5
0
−5
−10
−15
−20
−25
−80
−70
−60
−50
−40

Table 1. Summary of dataset filtering parameters. We generated eight SNP datasets under different filtering schemes to evaluate the sensitivity of cluster analyses. For eight SNP datasets, dataset D1 – D4mod, we note the parameters for each filter: the required sequencing coverage, the missing data threshold, whether or not a single SNP (single nucleotide polymorphism) or all SNPs per locus were retained, the minor allele frequency (MAF), the number of loci, the total number of SNPS represented in a given dataset, and the average number of SNPs per locus.

Table 1: Summary of dataset filtering parameters

Dataset	coverage	SNP present in	linkage	MAF	loci	SNPs	SNPs/locus
D1	>20x	>80%	1st SNP	>= 5%	1960	1960	1
D1mod	>20x	100%	1st SNP	>= 5%	175	175	1
D2	>20x	>80%	1st SNP	NA	2581	2581	1
D2mod	>20x	100%	1st SNP	NA	250	250	1
D3	>20x	>80%	all SNPs	>= 5%	1960	5099	2.6
D3mod	>20x	100%	all SNPs	>= 5%	177	391	2.2
D4	>20x	>80%	all SNPs	NA	2581	41024	15.9
D4mod	>20x	100%	all SNPs	NA	251	3832	15.3

1943 **Table 2.** Summary of Mantel Tests. For each geographic barrier and sampling region we were able
1944 to evaluate, we report results for analyses using phylogenetic patristic distance and coancestry,
1945 under both linear and log scales. Asterisks indicate *p* values less than 0.001, which we interpret as
1946 statistical significance (see discussion in text).

Table 2: Summary of Mantel Tests

	Patristic distance		Coancestry	
Approximate Barrier	p < 0.001	p < 0.001	p < 0.001	p < 0.001
Cordillera de Talamanca	-	*	-	-
Andes (1)	-	*	-	*
Andes (2)	-	*	-	*
Andes (3)	*	*	-	*
Rio Ucayali	*	*	*	*
Eastern Marañón + Hauallaga Rivers	*	*	*	*
Rio Putumayo	*	*	*	*
Rio Purus	*	*	-	*
Rio Madeira	*	*	*	-
Rio Tapajós	-	*	-	-
Rio Xingu	*	*	-	-
Cerrado (1)	*	*	-	*
Cerrado (2)	*	*	*	*
Rio Japurá	-	*	*	*
Rio Negro	-	-	-	-
Rio Essequibo	*	*	*	*
Rio Amazonas	*	*	*	*
Regions				
Full dataset	*	*	*	*
Central America - Costa Rica (Clade A3)	-	-	-	-
Central America - Panama (Clade A4)	-	-	-	-
North Andean – San Martín (Clade A1)	?	?	?	?
South Andean Peru (North) (Clade B2)	-	-	-	-
South Andean Peru (South) (Clade B1)	-	-	-	-
weakly resolved Guiana Shield (Clade D)	-	-	-	*
unresolved Jaú	-	*	-	*
weakly resolved eastern Napo (Clade D2)	-	-	-	-
weakly resolved Suriname + Amapá (D1)	-	-	-	*
Western Napo introgressed lineage (Clade E)	-	-	-	-
Western Inambari endemic (Clade C1)	*	*	-	*
Eastern Inambari endemic (Clade C3)	-	-	-	-
Rondônia endemic (Clade C4)	?	?	?	?
Tapajós endemic (Clade C5)	-	-	-	*
Xingu endemic (Clade C6)	-	-	-	-
Atlantic Forest – Bahia (Clade C8)	-	-	-	-
Atlantic Forest – Espírito Santo (Clade C9)	?	?	?	?
Atlantic Forest – Rio (Clade C10)	*	-	*	*
	linear	**log**	**linear**	**log**