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.

A CONTINENTAL-SCALE AVIAN RADIATION OUT OF THE ANDES

***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

23 context of our phylogeny supports the hypothesis that *Pseudopipra* is a species-complex
24 composed of 15-17 distinct lineages which have arisen in the last ~2.5 Ma.
25
26 **Key words**: Neotropics, ddRAD, genomics, phylogeography, manakin, suboscine,
27 speciation, introgression, admixture, reticulation, multi-species coalescent, coancestry,
28 isolation by distance, vocalization
29

30 **1. *Introduction***

31 Many kinds of geographic and habitat barriers have been proposed to drive
32 population diversification and speciation in the Neotropics (Cracraft and Prum, 1988;
33 Haffer, 1969, 2008; Smith et al., 2014; Wallace, 1854). Such barriers partition biodiversity
34 into areas of endemism (Cracraft, 1985; Cracraft and Prum, 1988; Crother and Murray,
35 2011; Da Silva et al., 2005; Linder, 2001; Noguera-urbano, 2016) by acting as impediments
36 to gene flow for dispersal-limited organisms (Brumfield, 2012; Cheviron et al., 2005;
37 Fernandes et al., 2015; Moore et al., 2008; Ribas et al., 2012). Three of the most prominent
38 features implicated in structuring the biodiversity of Neotropical forest birds (Figure 1)
39 include the Andes Mountains and other montane regions (Figure 1, grey relief); the Chaco,
40 Cerrado, and Caatinga biomes, which collectively form a 'dry diagonal' of open habitat
41 separating the Amazon forest from the Atlantic Forest; and the large rivers of the complex
42 Amazonian drainage system (Brumfield, 2012; Harvey and Brumfield, 2015; Naka and
43 Brumfield, 2018; Smith et al., 2014).

44 Montane Andean regions in the Neotropics are known to be exceptionally
45 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', '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 to propose a new hypothesis of species delimitation for the genus.

1.1. *Study organism and Taxonomic history*

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 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, 1990a; 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.

For most of its history, the species *pipra* was placed in the genus *Pipra* (Snow, 1979). Prum (1990a, 1990b); and Prum (1992) recognized that the traditional *Pipra* (*sensu* Snow 1979) was polyphyletic and placed the species *pipra* in *Dixiphia* (Reichenbach, 1850) which was recognized as a junior synonym of *Pipra*. Recently, Kirwan et al. (2016) demonstrated that the genus *Dixiphia* was unavailable for *pipra* because it is actually a junior synonym of the tyrannid genus *Arundinicola*. They named the new genus *Pseudopipra* for *pipra*.

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

136 Greenfield, 2001; Ridgely and Tudor, 2009; Snow, 2004; Spencer, 2012). Thirteen
137 subspecies of *Pseudopipra pipra* have been recognized based primarily on subtle variations
138 in plumage coloration, which are frequently more marked in females than in males
139 (summarized in Dickinson, 2003; Kirwan and Green, 2012; Snow, 2004). Because
140 suboscine passerines generally do not learn their songs (but see Saranathan et al., 2007),
141 the presence of substantial vocal variation across *Pseudopipra* populations further suggests
142 that the genus may contain unrecognized cryptic species (e.g. Campagna et al., 2012).
143 However, all previous authors awaited new information on genetic differentiation within
144 *Pseudopipra* before making taxonomic recommendations.
145 It also remains unclear how the current subspecific classification reflects
146 evolutionary history. The genetic variation within *Pseudopipra* has received prior attention
147 by two studies that used mitochondrial DNA to infer population genetic patterns. Milá et al.
148 (2012) reported intraspecific divergences of up to 3.5% (n=19) across three Amazonian
149 populations, and the highest observed nucleotide diversity ($\pi = 0.266$) among 14 widely
150 distributed Amazonian birds species. Castro-Astor (2014) used a larger sample (n=57), and
151 discovered that at least four *Pseudopipra* subspecies correspond to well supported
152 mitochondrial clades which were generally congruent with Amazonian areas of endemism.
153 Castro-Astor (2014) used a molecular clock model to estimate the age of the *Pseudopipra*
154 complex to coincide with the onset of the Pleistocene ~2.457 Ma (95% HPD 1.45-3.97).
155
156 **2. *Material and methods***
157
158 2.1. *Field and tissue sampling*

159 Muscle tissue samples were obtained from available vouchered avian material from
160 US and Brazilian collections and other institutions (see Acknowledgements and
161 Supplemental Table 1). We sampled from ten of thirteen recognized subspecies (Figure 2)
162 of *Pseudopipra* (Kirwan and Green, 2012). Unfortunately, we were unable to obtain
163 material of adequate quality for ddRAD sequencing from the Andes of Colombia (*bolivari*,
164 *unica*, *coracina*, *minima*), Ecuador (*minima*) and the lowlands of Peru (*pygmaea*). However,
165 we were able to obtain mtDNA data from *coracina* and *pygmaea*, which allowed us to make
166 a preliminary assessment of their phylogenetic affinities (see discussion on mtDNA). In
167 sum, after discarding failed samples and samples with high proportions of missing data, we
168 obtained genetic data from 286 individuals (234 from ddRAD, 168 from mtDNA),
169 representing ~80 localities and 10 of 13 subspecies. We also obtained comparable ddRAD
170 and mtDNA data for two specimens of *Ceratopipra rubrocapilla* as outgroups
171 (Supplementary Table 1).
172
173 2.2. *Laboratory methods*
174 We extracted DNA from avian tissue specimens (Supplementary Table 1) using the
175 DNeasy Blood & Tissue Kit (QIAGEN, CA) and generated ddRADtags following the protocol
176 of Peterson et al. (2012), with modifications described in Thrasher et al. (2018). Briefly, we
177 digested equal quantities of genomic DNA from each sample in individual reactions with
178 two restriction enzymes, SbfI and MspI (New England Biolabs, MA), and ligated adapters on
179 both ends. The 5' adapters contained one of 20 unique 5-7 bp barcodes, whereas the 3'
180 adapter was common to all samples. We pooled groups of samples with unique 5' barcodes
181 and subsequently size selected DNA fragments that were between 400-700 bp using a Blue

182 Pippin (Sage Science, MA). For each group of size-selected samples, we incorporated
183 unique Illumina TruSeq adapters by performing 11 cycles of PCR. The combination of 5'
184 barcodes and TruSeq adapters was unique to each sample. A total of 13 groups of pooled
185 samples with different Illumina TruSeq adapters were combined in equimolar proportions
186 into two libraries. We sequenced both libraries on four lanes of Illumina HiSeq 2500 at the
187 Cornell University Institute for Biotechnology, obtaining single-end 151 bp sequences. We
188 obtained mitochondrial ND2 sequences for a set of individuals using standard Sanger
189 Sequencing protocols described in Berv and Prum (2014).
190
191 2.3. *Assembly of sequencing reads into RAD loci*
192 We obtained 497 million raw 151 bp reads (~75 Gb) for 241 individuals. We first
193 assessed the overall read quality with FastQC (Andrews, 2010) and trimmed lower quality
194 bases at the 3' end with FASTX-Toolkit (Gordon and Hannon, 2010). The trimmed
195 sequences were 145 bp in length. Using FASTX-Toolkit we filtered-out lower quality reads
196 if they had a single base below a Phred quality score of 10 and/or more than 5% of bases
197 with quality between Phred 10 and 20.
198 The quality filtered reads were demultiplexed with the 'process_radtags' program
199 from the STACKS 1.44 bioinformatics pipeline (Catchen et al., 2013). During the
200 demultiplexing step we also discarded reads that had not passed the Illumina filter, had
201 adapter contamination, lacked barcodes used for multiplexing, or did not contain an SbfI
202 cut site. This step removed the inline barcodes and trimmed all reads to an equal length,
203 the length of the reads that contained a 7 bp barcode. After demultiplexing and filtering, we
204 retained an average of 1.1 ± 0.4 million 138 bp sequences per individual.

205 We downloaded the *Manacus vitellinus* (GCA_001715985.1) reference genome from
206 www.ncbi.nlm.nih.gov, and aligned the reads from each individual using bowtie2 2.3
207 (Langmead et al., 2009), as recommended by Paris et al. (2017); Shafer et al. (2017). We
208 assembled the mapped reads into RAD loci using the reference-based pipeline in STACKS,
209 executed with the ref_map script which runs the modules "pstacks/cstacks/sstacks". We
210 subsequently ran the error correction module "rxstacks" and a final iteration of
211 "cstacks/sstacks". We set the parameters to a minimum coverage of 20 (*m*) and up to two
212 differences among aligned loci of different individuals (*n*). The reference-based assembly
213 produced a catalogue with 47,046 RAD loci. Seven samples were discarded due to high
214 proportions of missing data in the final assemblies; our final RAD dataset thus comprised
215 234 individuals.

216 We used several filters in the "populations" module from STACKS to export different
217 sets of bi-allelic SNPs: a missing data filter, a minimum depth of coverage filter, a filter that
218 exports only the first SNP per RAD locus, and a minor allele frequency filter (MAF). In order
219 to explore the sensitivity of genetic cluster inference to these important parameters (Linck
220 and Battey, 2017; Paris et al., 2017), we produced eight versions of the SNP dataset by
221 combining these different standard filters (Table 1).

222 Dataset 1 (D1- labeled arbitrarily) included 1,960 SNPs that were generated by
223 setting the MAF to >= 5%, requiring minimum coverage of 20x, requiring the presence of a
224 SNP in at least 80% of individuals, and filtering out all but the first SNP in each RAD locus to
225 minimize linkage across the dataset. To evaluate the sensitivity of cluster analyses to rare
226 variants (Linck and Battey, 2017; Shafer et al., 2017), Dataset 2 (D2) was exported with
227 identical parameters to dataset D1, but without a minor allele frequency filter; this

228 generated a dataset of 2,581 putatively unlinked SNPs. We generated datasets D3 and D4 to
229 reflect the same filters as D1 and D2 but to also include all SNPs from each locus. Lastly, for
230 each dataset D1-D4, we generated a complimentary dataset which filtered out any sites
231 with missing data. The parameters of these SNP datasets are summarized in Table 1.
232
233 2.4. *Phylogenetic analysis*
234 Because sparse super-matrices may be more prone to systematic error (e.g. Roure et
235 al., 2013), we also explored the sensitivity of our phylogenetic inferences to increasing
236 amounts of missing data using three different haplotype datasets (separate from SNP
237 Datasets) with 20%, 50%, and 80% thresholds for missing data at the RAD locus level. For
238 example, for an 80% threshold, if the RAD locus is present in at least 20% of individuals,
239 then that locus, containing all variable and invariable sites, was exported. This approach
240 generated datasets comprising 2,584, 4,763, and 7,901 RAD loci, respectively. For each
241 dataset, we generated a concatenated supermatrix using Sequence Matrix (Vaidya et al.,
242 2011). These combined datasets had up to ~40% missing sites across an entire matrix,
243 irrespective of columns or rows (i.e., using the 80% threshold above).
244 We estimated phylogenies using each dataset with a concatenated maximum
245 likelihood and a coalescent species-tree approach for a total of six phylogenetic analyses.
246 For analysis of concatenated data, we estimated a maximum likelihood phylogenetic
247 hypothesis using RAxML 8.2.9 (Stamatakis, 2014). We estimated 500 bootstrap replicates
248 followed by a full ML search for each dataset, using a GTRGAMMAX substitution model. We
249 avoid the need for ascertainment bias correction because we use whole short-read
250 haplotype alignments including invariant sites (Leaché et al., 2015). In our case, each locus

251 alignment was 138bp long. We also tested for an effect of data partitioning on phylogenetic
252 inference by generating optimized locus partitioning schemes in PartitionFinder 2 (Lanfear
253 et al., 2017), which were then re-analyzed in RAxML. We found that our topological results
254 were insensitive to data partitioning, and as such we do not present these results. Lastly,
255 we generated a mitochondrial DNA ND2 gene tree using IQ-TREE 1.6.10 (Chernomor et al.,
256 2016; Hoang et al., 2017; Kalyaanamoorthy et al., 2017; Schmidt et al., 2014; Trifinopoulos
257 et al., 2016). We partitioned by codon position and generated a maximum likelihood tree
258 using the MFP+MERGE model search and partitioning option (Supplementary Figure 17).
259 We evaluated coalescent phylogenetic structure using the SVDquartets method
260 implemented in PAUP* 4.0a159 (Chifman and Kubatko, 2014, 2015; Swofford, 2002).
261 SVDquartets may be a useful tool for testing hypotheses of species delimitation because it
262 can generate a lineage-tree under the multispecies coalescent without assigning individuals
263 to species *a priori*. In this approach, each individual in a dataset is considered to be a
264 separate "species." If a recovered clade in such a lineage-tree has high bootstrap support,
265 this would mean that the descendant taxa have strong support as a group under the
266 multispecies coalescent, which may coincide with species boundaries (*Personal*
267 *communication*, Laura Kubatko). Our SVDquartets analysis evaluated all possible quartets
268 with 500 bootstrap replicates (other settings left to defaults). SVDQuartets has the
269 advantage of being able to work directly on sequence data (above) and has been shown to
270 perform well in comparison to summary methods which assume fully and correctly
271 resolved gene trees as input (Chou et al., 2015; Schmidt-Lebuhn et al., 2017). SVDquartets
272 has also been shown to produce reliable results under very general conditions, and is

theoretically robust to variation in effective population size, molecular clocks, limited gene flow, and incomplete lineage sorting (Long and Kubatko, 2017).

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

2.5. *Descriptive statistics*

We estimated global and population pairwise F_{st} (Weir and Cockerham, 1984), regional variation in allelic richness (Goudet, 2005), Hardy-Weinberg equilibrium (Paradis,

295 2010), inbreeding (Goudet, 2005), linkage disequilibrium (Kamvar et al., 2014), sex-biased
296 dispersal (Goudet et al., 2002), and an analysis of molecular variance (Kamvar et al., 2014).
297 Some of these statistics and analytical details are reported as supplementary material,
298 while others are discussed below when relevant.
299
300 2.6. *Cluster inference and patterns of genomic admixture*
301 We performed cluster analyses intended to quantify the number of similar genetic
302 groups across the range of *Pseudopipra.* First, we evaluated broad-scale variation across
303 our genomic data using phenetic K-means clustering implemented in the find.clusters
304 function in the adegenet R package (Jombart, 2008). This function first transforms the data
305 with a PCA, and then identifies clusters with a K-means algorithm (MacQueen, 1967). Using
306 each SNP dataset, we evaluated K1:20 successively with the Bayesian Information Criterion
307 (BIC). K-means clustering is based on the idea of minimizing the variance within clusters
308 and has the potential advantage of making few assumptions about the underlying
309 processes which generated the data. For each analysis, we retained all PC axes, set 200
310 randomly chosen starting centroids, and ran each search for 10^9 iterations.
311 We also estimated the number of differentiated populations using STRUCTURE 2.3.4
312 (Pritchard et al., 2000), a widely-used model-based approach which assigns individuals to
313 populations such that Hardy-Weinberg equilibrium is maximized and linkage among loci is
314 minimized within groups. By comparing the output of STRUCTURE runs fixed to different
315 numbers of populations (K), one can assess the degree of fit of the data to various models of
316 K and assess genomic admixture among individuals. We ran STRUCTURE using SNP dataset
317 D1 (Table 1), from K1:20 with 10 iterations per K value. We ran the program for 700,000

318 generations, discarding the first 200,000 as burn-in, implementing the admixture ancestry
319 model with correlated allele frequencies. We evaluated K by examining $\Pr(X|K)$ or $L(K)$
320 after summarizing our results using Structure Harvester (Earl and vonHoldt, 2012) and
321 CLUMPP (Jakobsson and Rosenberg, 2007). We also used the Evanno method to estimate
322 the rate of change in the log probability of data between values of K, which is a useful
323 metric to detect the uppermost hierarchical level of structure in the data (Evanno et al.,
324 2005). Lastly, we generated 95% Bayesian Credible Intervals around admixture
325 coefficients (presented as averages across individuals for population assignment).
326 To investigate finer scale population genetic differentiation and the genomic
327 composition of populations defined at different hierarchical levels, we used the programs
328 fineRADstructure and RADpainter (Malinsky et al., 2018). RADpainter takes advantage of
329 the information in haplotype data and considers each individual in a dataset as a
330 "recipient" whose genomes are reconstructed using chunks of "donor" DNA from all other
331 available individuals. This approach generates a "co-ancestry matrix" which combines the
332 information that can be derived from both PCA and model-based clustering approaches and
333 thus can be more sensitive to subtle population structure (Lawson et al., 2012). The
334 fineRADstructure software uses an MCMC approach to explore the space of population
335 assignments based on the co-ancestry matrix, using an algorithm which merges and splits
336 populations, or moves individuals among populations. We ran fineRADstructure with
337 default priors, but increased the burn in to 200000 iterations, followed by 1000000
338 iterations, sampling every 1000. We then assessed convergence by 1) considering the
339 assignment of population membership across multiple independent runs, 2) visualizing the
340 MCMC traces of estimated parameters to ensure convergence on the same posterior

341 distributions, and 3) running each chain long enough to achieve effective parameter sample
342 sizes > 100.

343 Both RADpainter and fineRADstructure are based on the recently developed
344 chromoPainter and finestructure software (Lawson et al., 2012), but are optimized to take
345 advantage of the linkage properties of RAD-like datasets. As our data are assembled
346 relative to a *Manacus* reference genome, the order in which loci appear in our data files are
347 related to positions on assembly scaffolds, even though the ddRAD loci in our data are
348 generally unlinked when considered at the population level (see Supplementary Appendix).

349 We compared the inferred population assignments from fineRADstructure and
350 STRUCTURE with K-means clustering (as above) of the co-ancestry matrix as well as SNP
351 datasets. To determine how much of the variation in our dataset was captured by the
352 RADpainter analysis relative to a standard PCoA of SNP data, we used the normalized PCA
353 approach of (Lawson et al., 2012), using the mypca function provided in the
354 "FinestructureLibrary.R" library in the fineRADstructure package.

355

356 2.7. *Phylogenetic reticulation in the western Napo lineage*

357 Model-based, phenetic, and population genetic analyses revealed that a lineage of
358 individuals presently restricted to the western Amazonian Napo area of endemism has a
359 complex history of introgression between southern and northern Amazonian clades (see
360 Results). In order to further investigate this subset of individuals and their history, we
361 generated two additional versions of the dataset optimized for different analyses. First- we
362 generated a haplotype dataset of 52 individuals, which included all introgressed western
363 Napo individuals (n=10, clade E), as well as samples from proximate neighbor populations.

364 These populations consist of 1) a geographically adjacent population inferred to have
365 mostly unadmixed ancestry (Southwestern Amazon: Inambari, n=17, clade C1), and 2) the
366 inferred source of introgressed genotypes (north-central Amazonian Jaú + eastern Napo,
367 n=25, clade D2).

368 We used a 5% minor allele frequency cutoff, required 20x minimum coverage, and
369 allowed up to 20% missing data per locus—generating a dataset comprising 2,370 loci,
370 including 4,979 SNPs (Dataset H1). Dataset H1 is analogous to that generated for co-
371 ancestry (fineRADstructure) analysis of the entire dataset (above). We also generated a
372 similar dataset (Dataset H2), without a minor allele frequency cutoff—comprising 2,947
373 loci including variable and invariable sites—as required for inference of demographic
374 parameters (Gronau et al., 2011).

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

383 To test whether or not the relationships among these focal populations may be best
384 represented by a phylogenetic network, we used PhyloNet 3.6.4 (Wen et al., 2018; Zhu et
385 al., 2018). PhyloNet takes SNP data and uses a reversible-jump MCMC technique to explore
386 the posterior distribution of phylogenetic networks and bi-furcating topologies, while

387 accommodating both reticulation and incomplete lineage sorting. PhyloNet searches the set
388 of all possible reticulation models without requiring *a priori* model specification.
389 For computational tractability, we used SNP dataset D1 (1960 unlinked SNPs), and
390 pruned it to the 52 individuals comprising the groups of interest. We pruned out all sites
391 with missing data, leaving 572 unlinked biallelic SNPs, as required by PhyloNet. For our
392 final runs of the PhyloNet program MCMC_BiMarkers, we randomly subsampled five
393 diploid individuals from each group, as preliminary program runs indicated a severe
394 computational bottleneck with more than five diploid individuals or with larger datasets.
395 We assigned each group of individuals to a lineage, allowed for a maximum of one
396 reticulation, set the chain length for 800000 iterations (sampling every 100), set a burn in
397 of 2.5%, and left all priors as default (i.e. assuming population mutation rates are constant).
398 Using the final network topology estimated from PhyloNet, we estimated
399 demographic parameters using G-PhoCS version 1.3 (Gronau et al., 2011). We ran the
400 program for 750,000 iterations with a 10% burn-in, estimating a total of 13 parameters:
401 the effective population sizes of the three focal lineages, two ancestral population sizes, two
402 splitting times, and six directional migration parameters. Parameter MCMC traces were
403 inspected in Tracer 1.7.1 (Rambaut et al., 2018). We converted the median and 95%
404 Bayesian credible intervals from mutation scale to generations and individuals as
405 described in Campagna et al. (2015), assuming an approximate mutation rate of 10^{-9} per
406 base pair per generation (Kumar and Subramanian, 2002; Smeds et al., 2016). We
407 calculated the number of migrants per generation as the product of the per generation
408 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).

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

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

443 We also analyzed continuous patterns of spatial variation in our data with a
444 population genetic model that relates effective migration rates across geographic space to
445 expected genetic dissimilarities among individuals (Petkova et al., 2015). This recently
446 proposed method, termed Estimated Effective Migration Surfaces (EEMS), produces a
447 visualization that emphasizes deviations from isolation by distance to discover migratory
448 corridors and barriers to gene flow. This method is complimentary to more explicit
449 hypothesis-testing approaches (such as Mantel tests), because it seeks to identify where
450 the assumption of constant IBD is violated in geographic space, and thereby can highlight
451 areas where strong barriers to gene flow may exist. In brief, a region is covered with a
452 dense regular grid connecting subpopulations (demes), among which individuals can
453 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.

2.9. *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.

Pseudopipra vocalizations have 1-3 buzzy and/or tonal notes. We measured: 1) starting frequency, 2) ending frequency, 3) minimum frequency, 4) maximum frequency, 5)

477 number of notes, and 6) duration of the entire vocalization (Supplementary Figure 13). We
478 performed principal components analysis (PCA) and logistic regression on these vocal
479 measurements to test for significant differences between vocal types and to reduce the
480 dimensionality of the data for comparison to results from analysis of genetic data. Because
481 no sound records were directly associated with genetic samples in this study, we used
482 geographic proximity to vocalization recordings and localization to areas of endemism or
483 areas bounded by clear physical barriers to associate vocal types to genetic samples and
484 taxonomic ranks. This approach assumes that genetically and geographically proximate
485 individuals are likely to share the same vocal type and enabled us to perform a preliminary
486 assessment of how variation in vocalization type maps onto existing genetic variation.
487 Testing the fine-scale association of genetic and vocalization boundaries will require
488 extensive field sampling from individual manakins. Additional analytical details and
489 analyses are reported as supplementary material.
490
491
492 **3. Results**
493
494 3.1. Genetic *data collection*
495 Our total genomic data set included an average of 1 ± 0.4 million sequences of 138
496 bp from 232 individuals of *Pseudopipra* and two *Ceratopipra rubrocapilla,* which were used
497 as an outgroup. Our reference-based assembly produced a catalogue with 47,046 RAD loci.
498 Filtered datasets requiring a SNP to be present in at least 80% of individuals ranged from
499 1,960 unlinked SNPs (dataset D1: 5% MAF with 1st SNP on each locus) to 41,024 SNPs

500 (dataset D4: all SNPs on all loci, no MAF) (Table 1). Datasets requiring a SNP to be present
501 in 100% of individuals were considerably smaller, as expected, and ranged from 175 SNPs
502 in D1mod, to 3,832 SNPs in D4mod (Table 1). We also obtained 168 mtDNA ND2 sequences
503 using PCR methods which broadly overlapped with our ddRAD dataset. Our mtDNA dataset
504 included two subspecies, *pygmaea* and *coracina*, from which we were not able to obtain
505 enough high-quality material for ddRAD sequencing.
506
507 3.2. *Phylogenetic analyses*
508 RAxML and SVDquartets analyses generated a consistently well-supported
509 phylogenetic hypothesis (Figure 3). The monophyletic lineages of *Pseudopipra* have
510 distributions that are largely congruent with previously recognized avian areas of
511 endemism, and further reveal much finer scale patterns of local biogeographic
512 differentiation. Using the two *C. rubrocapilla* individuals to root the phylogeny, we detected
513 support for at least five major clades across all phylogenetic analyses of nuclear genomic
514 data, with consistently high bootstrap support in both concatenated and coalescent
515 analysis (clades A-E in Figures 3-5). We first describe the structure of these major clades,
516 followed by descriptions of well supported substructure within each of these regional
517 groupings. We indicate six bootstrap values for each clade with at least some high support,
518 from each of the three datasets analyzed in RAxML and SVDquartets (see heatmaps, Figure
519 3). Phylogenetic analysis of mtDNA (Supplemental Figure 17) generated a congruent result,
520 except where noted below.
521
522 3.2.1 *Phylogenetic structure*

523 The sister group to all the other *Pseudopipra* is a clade including individuals from
524 subtropical Central America and the northern Andes. The northern Andes are represented
525 in our ddRAD dataset by a single individual from subtropical forests (~1,000 m) of the
526 eastern slopes of the Andes in San Martín, Peru. This subtropical San Martín specimen
527 (identified as *P. pipra occulta*) was the only viable sample in our genomic dataset from the
528 entire northern Andes region (A1 in Figure 3). Mitochondrial data from three additional
529 specimens from the Andes of southern Ecuador (subspecies *coracina*) form a monophyletic
530 group with the San Martín specimen; thus, this northern Andean sister lineage to the
531 Central American lineage likely includes all montane Andean populations of *Pseudopipra* in
532 Ecuador, Colombia, and Venezuela. A well supported Central American clade (A2 in Figure
533 3) was identified in all analyses, with Costa Rican (A3 in Figure 3) and Panamanian
534 subclades (A4 in Figure 3) supported in RAxML.

535 The second successively nested clade in *Pseudopipra* includes individuals from
536 subtropical forests of the Andes of Southern Peru, which are sister to all remaining lineages
537 (Clade B in Figure 3). The southern Andean Peru clade is further subdivided into two very
538 well resolved clades: a northern clade (B2 in Figure 3), sampled from the southern
539 Cordillera Azul in southwestern Loreto, and a southern clade from Pasco, Junín, and Cusco
540 (labeled B1 in Figure 3). This southern Andean Peru clade is the sister group to a
541 monophyletic, lowland clade (F in Figure 3) that is further subdivided into a southern
542 Amazon clade + Atlantic Forest clade (Clade C in Figure 3), and a Northern Amazonian +
543 Guiana Shield clade (Clade D in Figure 3). Our analysis of mtDNA with sampling of
544 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 population genetic analyses, which indicate that this Napo population is a distinct lineage of the southern Amazonian clade

567 (Clade C in Figure 3) that has experienced substantial hybrid introgression from the
568 northern Amazonian clade D (see Population Genetic analyses and Discussion).
569 Within the large lowland Southern Amazonia clade (Clade C in Figure 3), we
570 recovered eight moderately supported and hierarchically nested clades, which are
571 successive sister groups to each other, and are congruent in distribution with recognized
572 areas of endemism. First is a moderately supported clade (C1 in Figure 3) of individuals
573 from the western Inambari area of endemism, from west of the Rio Ucayali to the Rio
574 Purus. Although the first lineages to diverge within clade C1 have low bootstrap support, a
575 monophyletic subgroup of three montane (>1000m) individuals from the highlands
576 between the Rio Huallaga and Rio Ucayali cluster together with high bootstrap support (C2
577 in Figure 3).
578 The next successively nested clade includes individuals from eastern Inambari,
579 between the Rio Purus and the Rio Madeira (C3 in Figure 3). The next successive clade
580 includes two individuals sampled from the Rondônia area of endemism, between the Rio
581 Madeira and Rio Tapajós (C4 in Figure 3). The next successive clade includes a group of
582 individuals from the Tapajós area of endemism between the Rio Tapajós and the Rio Xingu
583 (C5 in Figure 3). Next, a clade comprised of individuals from the northern Xingu area of
584 endemism, east of the Rio Xingu (C6 in Figure 3), is sister to a well-supported Brazilian
585 Atlantic Forest clade (C7 in Figure 3), which is further subdivided into three successive
586 clades (north to south) in Bahia, Espírito Santo, and Rio de Janeiro, respectively (C8, C9,
587 and C10 in Figure 3). In summary, the subdivisions of the Southern Amazonian clade reflect
588 a stepwise west-east progression, with a hierarchically nested structure:
589 (West Inambari (East Inambari (Rondônia (Tapajós (Xingu (Bahia (Espírito Santo (Rio de Janeiro))))))).

590

591 3.3. *Cluster inference and patterns of genomic admixture*

592 3.3.1. *Broad-scale population structure and admixture*

593

594 Weir and Cockerham's F_{st} was moderate when comparing populations within
595 *Pseudopipra*; for Dataset 1, overall F_{st} was 0.195 [95% CI: 0.187-0.203], and never dipped
596 below 0.171 [0.168-0.175] (Dataset 4). Notably, in SNP datasets filtered to prevent missing
597 data, overall F_{st} estimates were on average ~20% higher and ranged from 0.188 [0.175-
598 0.201] (D4) to 0.261 [0.191-0.336] (D2). Pairwise population F_{st} estimates from Dataset 1
599 (other datasets showed similar patterns, see Supplementary Figure 10a-h), ranged from
600 essentially undifferentiated, to almost entirely distinct. At the extremes: comparing the
601 geographically proximate Jaú and Guiana Shield populations indicated a particularly low
602 F_{st}: 0.005 [0.003 - 0.007]. By contrast, comparing the Atlantic Forest Espírito Santo
603 population to the Panama population recovered an F_{st} of 0.81 [0.79 - 0.83], or almost
604 entirely differentiated. For reference, Hartl and Clark (1997) indicated $F_{st} > 0.15$ may be
605 considered "great" genetic differentiation.

606 K-means clustering of SNP datasets on the basis of BIC score identified groups of
607 individuals that differed somewhat with respect to dataset filtering strategies (Summarized
608 in Figure 5, right). For D1, K-means clustering identified five major groups: Central America
609 (Clade A2), South Andean Peru (Clade B), Atlantic Forest (Clade C7), Southern Amazon
610 including the western Napo population (the rest of Clade C + Clade E), all Guiana Shield +
611 Northern Amazon (Clade D) (Supplementary Figure 2a-h). For dataset D1, the montane
612 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

659 of the confluence of the Rio Solimões and the Rio Napo, (2) southwest of Iquitos in
660 Allpahuayo Mishana Reserve (represented by three mtDNA samples), and (3) near San
661 Jacinto, Ecuador– indicate that this heavily admixed population (corresponding to
662 subspecies *discolor*) may occur across a large area of the western Napo region of
663 endemism, and therefore may not represent a narrow hybrid zone (Figures 2-7).
664 Lastly, at K=5, all members of the Guiana Shield + Northern Amazon clade (Clade D)
665 are inferred to be virtually unadmixed. Individuals in the far eastern Napo area of
666 endemism on the east bank of the Rio Putumayo (brown dots in Figure 3, 4) cluster with
667 other Guiana Shield groups in phylogenetic analyses but are detected to have limited
668 (~5%) admixture with pure individuals of Southern Amazon provenance. At higher values
669 of K, the broad-scale population assignments inferred at K=5 are mostly unchanged,
670 however additional admixture components are inferred for most groups. The introgressed
671 western Napo clade is eventually placed into its own cluster at K=9-10 (see Supplementary
672 Figure 4, Supplementary Figure 8).
673
674 3.3.2. *Fine-scale population structure and admixture*
675 Admixture analysis using the RAD locus optimized chromosome painting approach
676 in RADpainter and then population assignment with fineRADstructure (Lawson et al., 2012;
677 Malinsky et al., 2018), was congruent with other analyses, and also detected finer partitions
678 of population structure. The fineRADstructure program detected 15 groups of individuals
679 which were each supported with a posterior probability of 100% and which also
680 overlapped with sampling regions which we delimited for population genetic analyses
681 based on geographic barriers (Black dots in Figure 3, Supplementary Figure 5). The only

682 sampling regions which fineRADstructure failed to confidently partition into separate
683 groups consisted of a group of individuals in the Jaú area of endemism, and a group of
684 individuals in the eastern Napo area of endemism (both groups are poorly resolved in
685 phylogenetic analyses, but unambiguously members of the broader Northern Amazon +
686 Guiana Shield clade D). The eastern Napo individuals (D2 in figure 3) are identified as a low
687 support group (*pp* = 0.49), while the Jaú individuals are lumped in with other western
688 Guiana Shield individuals (D in figure 3). See Supplementary Figure 5 (fineRADstructure
689 dendrogram), and the full dataset co-ancestry matrix plots Supplementary Figures 6-7.

690 Patterns of co-ancestry show a complex mosaic of shared ancestry among
691 populations (Supplementary Figures 6-7). In general, admixture patterns appear consistent
692 with those detected by STRUCTURE, including the case of the highly introgressed western
693 Napo lineage; therefore, we interpret patterns indicated by STRUCTURE to be reliable.
694 Notably, the high level of admixture initially inferred for our San Martín sample is
695 recapitulated in RADpainter coancestry analysis – this sample has relatively high co-
696 ancestry across most populations (as does the pattern of co-ancestry we report for the
697 western Napo lineage, below).

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

705

706 3.4. *Phylogenetic reticulation in the western Napo lineage*

707 Separate RADpainter and PhyloNet analysis of the populations in the western Napo
708 area of endemism and their putative progenitor lineages support the hypothesis that this
709 region has a complex history of introgression. One-tailed tests of mean co-ancestry
710 indicated the null hypothesis that the difference in means among groups is less than or
711 equal to zero can be strongly rejected (ghlt $p < 2e-16$), and that individuals from the
712 western Napo population have greater median/mean co-ancestry with each progenitor
713 lineage than their progenitor lineages share with each other (PHvGS/progenitors = ~1.078,
714 PHvInambari/progenitors = ~1.16). These analyses establish the asymmetry in admixture
715 patterns detected by STRUCTURE. On average, western Napo individuals share ~ 7.5%
716 greater co-ancestry with Inambari individuals (southern Amazon) than they do with other
717 Northern Amazon individuals (glht $p < 2e-16$), indicating these sampled populations do not
718 include F1 hybrids. We did not observe any patterns of clinal variation within our samples
719 of this lineage (spanning ~20,500 km^2).

720 fineRADstructure analysis of the co-ancestry matrix, as well as visualization of the
721 co-ancestry matrix indicated that the western Napo population may be differentiated, as
722 well as intermediate when compared to its progenitors (Figure 6a-d). Principle
723 components analysis (Figure 6d) of the co-ancestry matrix indicate that the western Napo
724 lineage is intermediate on PC1 (compared to progenitor lineages), but differentiated on
725 PC2, which would be expected if the Napo lineage has had sufficient time in reproductive
726 isolation for sorting of ancestral alleles or for the evolution of new alleles (Barrera-Guzmán
727 et al., 2018). This hypothesis is supported by the observation that this lineage has distinct

728 mitochondrial haplotypes derived from Inambari populations (~separated from Inambari
729 haplotypes by ~6.5 substitutions or 0.63%), as well as analyses of IBD (below).

730 Bootstrapped estimates of F_{st} indicated the same general patterns: 95%CI; western
731 Napo vs Eastern Napo (Northern progenitor, 0.07 - 0.090), western Napo v Inambari
732 (Southern progenitor, 0.058 - 0.077): Eastern Napo vs Inambari: (between progenitors,
733 0.122 - 0.144). These patterns of F_{st} estimates were robust to alternative assignments of
734 geographically proximate individuals to progenitor linages (for example, including an
735 additional adjacent population for both of the putative progenitor lineages). Thus, the
736 introgressed western Napo lineage is less differentiated from *each* of its progenitor
737 lineages than its progenitor lineages are from each other and it is *also* biased in its genetic
738 composition toward Inambari populations.

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

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

758 Taken in the context of STRUCTURE (Figure 5) and fineRADstructure (Figure 6)
759 results, the gene flow implied by the top ranked network seems to have been asymmetrical,
760 leaving very little evidence of southern admixture among northern lineages (see Figure 5,
761 'weakly resolved eastern Napo', which are detected to have very limited (< ~5%) southern
762 admixture). This observation, taken alone, could have been produced by sampling biases.
763 However, demographic parameters estimated with G-PhoCS also support this hypothesis:
764 the only significant migration rate (lower 95% CI > 1 migrant per generation) was inferred
765 from the eastern Napo (member of Guiana Shield lineage) lineage into S2/S2a (Figure 6,
766 Supplementary Table 5).

767 G-PhoCS also allows us to place a preliminary maximum age constraint on the
768 timing of introgression into S2a. Assuming 2.47 years per generation in *Pseudopipra*
769 (estimated in Bird et al., 2020) and a constant mutation rate, the absolute splitting time
770 between S2a and S1 may have been ~1.2 Ma (Figure 6, Supplementary Table 5). Though we
771 emphasize that this is a rough approximation, this result is loosely consistent with prior
772 divergence time estimates for several groups in the genus based on molecular clock
773 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, if these estimates are correct, 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 limited sampling 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*

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

796 distance, suggesting that different processes may be contributing to spatial correlations
797 within these areas.

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

805 Model based analysis of migration rates with EEMS largely corroborated the specific
806 hypotheses implied by Mantel tests. For dissimilarities between pairs of demes, the
807 correlation between observed and fitted values was high – $R^2 = 0.871$ – within demes
808 somewhat less – $R^2 = 0.579$ (Supplementary Figure 11). Thus, the EEMS model appears to
809 do a good job of describing spatially structured variation in this dataset, and a non-
810 stationary migration surface is a better fit to the data than simple isolation by distance. The
811 EEMS visualization detects many strong signals (posterior probability > 0.9) of barriers to
812 gene flow (Figure 7). As expected, the Andes and the Amazon river were unambiguously
813 detected as the most significant barriers to gene flow and are clearly recovered as
814 landscape features, even though the model does not know these features exist.

815 EEMS also detected signals of deviation from constant IBD that likely correspond to
816 several Amazonian tributaries in the Amazon basin which delimit areas of endemism
817 (Figure 7). An additional strong barrier to gene flow was inferred in the south-central
818 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 relatively low genetic diversity included the Atlantic Forest, the Peruvian Andes, and Central American lineages. These patterns are consistent with direct estimates of regional variation in allelic richness (Supplementary Appendix).

3.6. *Vocal variation*

We examined 198 recordings of *Pseudopipra* and identified 14 distinct vocalizations that can be diagnosed by ear or visual inspection of sonograms (arbitrarily labeled numerically in Figure 8). This set of vocalization types includes both advertisement vocalizations from males at leks and calls by males and females in other contexts. After examining recordist notes, we generated a preliminary hypothesis which distinguishes eleven highly variable and distinct, lekking advertisements (Figure 8a) from three broad ranging call types (Figure 8b). Our delimitation of vocalizations into advertisement and call categories partly reflects expert opinion, as the behavioral context of vocal records was not always clear (see Supplemental Appendix). Work by Castro-Astor et al. (2007) provides important insights of the vocal behavior of Atlantic forest populations, but most *Pseudopipra* populations lack similarly detailed studies. Vocalization types within each category of male advertisement vocalizations or calls can be statistically discriminated based on a combination of the frequency, duration, and number of notes. PCA and logistic regression on lekking vocalizations (n=114, Supplementary Table 3), and calls (n=33) found significant differences among types ($p < 0.001$). The first three axes of a PCA explained ~90% of the variation in lekking vocalization characters, with PC1 (~64%) primarily explaining variation in note number and frequency (Supplementary Figure 15).

Our assignments of vocal types to well supported genetic groups are preliminary in the absence of matching data (i.e., recordings from the same individuals from which we have genetic samples). Nevertheless, we found that most of the 11 lek advertisement vocalizations are phylogenetically 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. 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*.

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

888 datasets allow us to assess at least some data from all thirteen previously identified
889 subspecies.

890 The predominant biogeographic signal we detect in our dataset is of two
891 successively nested montane clades as the sister groups to several major lowland
892 Amazonian clades (Figure 5, see *Historical Biogeography*, below). This pattern implies that
893 lowland Amazonian and montane Central American populations of *Pseudopipra* expanded
894 historically "out of the Andes." Ancient lineages endemic to the Andes may have originated
895 around the onset of the Pleistocene ~2.5 Ma (molecular clock estimates from Castro-Astor,
896 2014). The lowland clade within *Pseudopipra* originated (~1.085 [0.57-1.79] Ma) in
897 southwestern Amazonia (Figure 4) and expanded, occupying Amazonia on both margins of
898 the Amazon river, and eventually reached the Atlantic forest (see below). Therefore, our
899 study provides an unusual example of a widespread suboscine passerine which originated
900 in the Andean highlands.

901

902 4.1 *Population genetic variation*

903 We examined our dataset using a variety of population genetic approaches intended
904 to assess the number of natural genetic groupings that may be present in *Pseudopipra*.
905 Phenetic K-means clustering of SNP data and analysis with STRUCTURE estimated a
906 minimum of 5 genetic clusters with broad geographic ranges (Figure 5, Supplementary
907 Figure 2, Supplementary Figure 4). K-means clustering of the co-ancestry matrix (which
908 captured substantially more variation than standard PCoA) more finely partitioned the
909 dataset into 8 genetic clusters (Figure 5, Supplementary Figure 8), which closely
910 recapitulates previously identified subspecies boundaries (Figure 2). Phylogenetic (RAxML,

SVDquartets) and model-based population genetic (fineRADstructure) analyses of nuclear genomic data detected fifteen identical groups of individuals which coincided with sampling regions delimited by clear geographic barriers (results, Figure 3, 4, Supplementary Appendix). Two additional subspecies represented by mtDNA (results) bring the total number of distinct genetic populations delimited by geographic barriers to at least seventeen. The close concordance observed between population structure and landscape features implies that the evolutionary history of *Pseudopipra* within the Neotropics is deeply connected to the South American landscape, adding support to a rich body of literature endorsing this hypothesis (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

933 populations and Central American lineages (see Supplementary Appendix for additional
934 discussion).

935 Our analyses also lend themselves to some general conclusions about the sensitivity
936 of various methods to detecting population genetic structure in ddRAD data. Overall,
937 methods based on clustering of SNP data were the least sensitive (prone to lumping) to
938 population structure, whereas phylogenetic analyses were the most sensitive to population
939 structure (prone to splitting). Thus, different phylogenetic and population genetic cluster
940 analyses have detected different levels of hierarchical structure in the data.

941 Multiple analyses indicate that the western Napo population has a complex history
942 of initial and early differentiation from the broader Southern Amazonian clade, followed by
943 subsequent, asymmetric genetic introgression from Jaú and eastern Napo populations of
944 the Guianan Shield + Northern Amazon clade (Clade D; see discussion below). When the
945 genetic data are analyzed under either concatenated or coalescent phylogenetic
946 frameworks which do not allow for phylogenetic reticulation, the western Napo population
947 is reconstructed as most closely related to Guiana Shield + Northern Amazon populations
948 (Figure 3, Figure 5, Clade D). In contrast, phylogenetic network analysis indicates that this
949 population is derived from individuals of Inambari (southern Amazon) provenance (Figure
950 6, also summarized in Figure 4). This incongruence between methods is likely due to the
951 complexity of the constituent gene trees in this population-level differentiation induced by
952 asymmetric introgression from the Jaú and eastern Napo populations (Thom et al., 2018).
953 We hypothesize that introgression has introduced more recently derived variation from
954 the Guiana Shield + Northern Amazon clade D into an older, differentiated, population of
955 the Southern Amazonian clade, pulling the phylogenetic affinity toward the northern clade.

4.2 *History of an introgressed hybrid lineage*

Analysis of genomic admixture and phylogenetic reticulation revealed a population of individuals restricted to the western Napo area of endemism which have highly introgressed genomes and a complex evolutionary history (Figure 6). Our investigation revealed this northwestern lowland Amazonian group is most likely derived from a southwestern Amazonian lineage which experienced introgression from the Guiana Shield + Northern Amazon clade D. The substantial level of observed introgression (>40%), as well as a suite of genomic and phenotypic evidence, suggests that this lineage may be best recognized as a hybrid lineage, though it remains to be determined 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), highlight the emerging paradigm 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

979 historical admixture with a non-sister lineage. This historical scenario is not the same as
980 those examples cited above, but it has produced a population genetic signature which is
981 extremely similar (compare to Barrera-Guzmán et al., 2018). Notably, PhyloNet found
982 similar, but nonetheless asymmetric support for an introgression model over a hybrid
983 speciation model (see results, 57% vs 41%, respectively).

984 Some authors have proposed strict population genetic criteria for identifying
985 homoploid hybrid speciation, including demonstrating (1) evidence of reproductive
986 isolation from parental species, (2) evidence of past hybridization, and (3) evidence that
987 intrinsic reproductive isolation was derived from the hybridization event (Schumer et al.,
988 2014). The first two conditions are satisfied by the western Napo *Pseudopipra* lineage.
989 First, phylogenetic and population genetic analyses are suggestive of contemporary
990 reproductive isolation. Phylogenetic analyses of nuclear genome data, including
991 concatenated and coalescent analyses (Figure 3, 4, 6), recover this lineage as a well-
992 supported reciprocally monophyletic group. Differentiated mtDNA haplotypes from a
993 broader geographic sample (Figure 6) of this lineage also form subgroups derived from
994 southern Inambari haplotypes (results, Supplementary Figure 17).

995 PCA-based and model-based population genetic cluster analyses (Figure 6b, 6c,
996 Supplementary Figures 4-9), also support the hypothesis that this lineage is intermediate
997 *and* differentiated when compared to the living descendants of its inferred progenitors.
998 Frequentist Mantel tests, as well as Bayesian model-based analysis of isolation by distance,
999 support the hypothesis that contemporary gene flow into and out of this population is at
1000 least restricted and probably not occurring (Figure 7, Supplementary Figure 12). Thus, we
1001 interpret our G-PhoCS results of a significant migration parameter (results) as indicative of

1002 a window of historical introgression, which could be further explored with simulation
1003 studies.

1004 Estimates of inbreeding coefficients are also consistent with the hypothesis that
1005 introgression into the western Napo lineage is not ongoing (Supplementary Appendix). The
1006 similar levels of inbreeding among *Pseudopipra* populations in the western Amazon (F_{IS},
1007 Supplementary Figure 9, and Supplementary Appendix) imply that the western Napo
1008 population is not more outbred (i.e. recently introgressed) than other inbred populations.
1009 Notably, the consistently positive level of inbreeding (median ~0.16, 95% CI [0.11-0.21])
1010 among most *Pseudopipra* populations may be attributable to a highly polygynous lek
1011 mating system (Smith, 1979; Stopher et al., 2012; Waser et al., 1986). In sum, these
1012 patterns observed in our nuclear genetic data (and mtDNA, discussed below and earlier)
1013 are consistent with the hypothesis of contemporary reproductive isolation for the
1014 introgressed Western Napo populations.

1015 The second criterion from Schumer et al. (2014), evidence of past hybridization, is
1016 directly supported by multiple model-based analysis of genomic admixture and
1017 phylogenetic reticulation (Figure 5, 6), and indirectly supported by model based and
1018 phenetic cluster analysis (Supplementary Figures 4-9). We hypothesize that hybridization
1019 initially occurred between progenitor lineage S2a and western lineages of the Guiana
1020 Shield + Northern Amazon clade (Figure 6a). Unadmixed lineage S2a was then replaced by
1021 the novel introgressed lineage, which exchanged genes with neighboring populations
1022 within a finite time window. An intrinsic assumption of our analyses is that S1 (western
1023 Inambari) is an appropriate substitute for S2a (which we infer to have existed). We use S1

1024 as proxy for S2a in population genetic analyses because it is the closest and only available
1025 relative.

1026 The third criterion from Schumer et al. (2014), evidence that intrinsic reproductive
1027 isolation was derived from the hybridization event, is recognized as the most difficult to
1028 demonstrate, and probably the most tenuous. As discussed by Nieto Feliner et al. (2017) in
1029 critique of Schumer et al. (2014), this criterion may be useful for framing the strength of
1030 evidence for or against the existence of a homoploid hybrid species, but only if one adheres
1031 strictly to the biological species concept (Mayr, 1942). If one applies a more evolutionary
1032 definition of species, like the metapopulation lineage concept, in which the only necessary
1033 requirement is demonstration of separately evolving lineages (de Queiroz, 2005), the
1034 importance of demonstrating intrinsic reproductive isolation to claim species status for a
1035 putative hybrid lineage (or any lineage at all) is less clear to us. Speciation in birds in
1036 particular is thought to proceed relatively fast, and may be more affected by external
1037 factors than internal ones (Grant and Grant, 1997). Further, requiring hybridization to be
1038 the *source* of intrinsic isolation excludes the possibility that extrinsic or pre-zygotic
1039 isolation may evolve from changes that enable hybrids access to new niches, which may
1040 also be geographically isolated relative to parental forms (Nieto Feliner et al., 2017).

1041 Intriguingly, the evolutionary scenario implied by phylogenetic reticulation analysis
1042 for the introgressed western Napo hybrids provides a mechanism to explain a confusing
1043 distribution of vocalization types (Figure 8, Supplementary Appendix). The males from the
1044 Napo area of endemism share vocalization type 2 with birds in the southeastern Amazon
1045 (Figure 8, Supplementary Figure 16), and with the restricted Huallaga valley *pygmaea*
1046 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 of this possibility (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 from one species into a genomic background of a close relative (Counterman, 2016; Jiggins et al., 2008; Marques et al., 2019). This type of combinatorial hybrid speciation was originally proposed from studies of *Heliconius*

butterflies (Jiggins et al., 2008; Salazar et al., 2010), and suggests introgression of a "magic trait" can instantaneously lead to hybrid speciation without intrinsic barriers to reproduction (Counterman, 2016). While we have not examined specimens for evidence of "magic traits" representative of the Northern Amazon + Guiana Shield clade (D) that have influenced reproductive isolation in the western Napo lineage, our study predicts that such traits may exist in these populations and are worthy of additional study.

4.3 *Comparative and Historical Biogeography*

The spatial pattern of population structure and evolutionary relationships we identify here are broadly congruent with other recent studies of Neotropical biogeography in upland forest birds, contributing another example of the substantial phylogeographic pseudocongruence across taxa at this wide scale (e.g. Cunningham and Collins, 1994; Harvey et al., 2017). As expected, the two main barriers identified for *Pseudopipra* in the South American lowlands are the Amazon River and the dry diagonal. Amazonian tributaries seem to delimit recently isolated populations in southern Amazonia, while northern *Pseudopipra* populations show weaker evidence of being isolated by the Negro and Japurá rivers. By contrast, southern populations seem to be strongly isolated by rivers with smaller discharge, such as the Xingu, Tapajós and Purus.

A comparable survey of the suboscine ovenbirds in the *Xenops* complex (Harvey and Brumfield, 2015) identified similar patterns of genetic structuring coincident with major Amazonian tributaries, as did Ribas et al. (2012) in the *Psophia* trumpeters, though the inferred relationships among areas of endemism are not the same. A recent comparative analysis shows that large rivers delimit genetic clusters in as many as 23 groups of upland

Amazonian forest birds, strongly arguing in favor of these rivers acting as barriers to gene flow (Silva et al., 2019). The somewhat distinct pattern found for *Pseudopipra* may be related to its recent expansion throughout the eastern Andean lowlands and its habitat tolerance.

Within the manakins, a recent molecular phylogeny (Ohlson et al., 2013) placed *Pseudopipra* as sister to the genus *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): (West Inambari (East Inambari (Rondônia (Tapajós (Xingu (Bahia (Espírito Santo (Rio de Janerio))))))))). Contrastingly, the northern Amazonian lineage also expanded all the way to the Atlantic Ocean, but without producing strong phylogenetic structure (Figures 3, 4). In the Napo area of endemism, substantial introgression occurred from east Napo and Jaú populations into the western Napo population, with limited backcrossing into Inambari and east Napo populations.

Intriguingly, a well-supported distinct montane (>1000m) (BS: 100) (Clade C2 in Figure 3; Cushabatay area, northern Cordillera Azul, Loreto Prov., Peru) is recovered as nested within the well supported lowland western Inambari subclade (Clade C1 in Figure 3). Thus, there may have been at least one secondary invasion of montane Andean habitats from a lowland ancestor. Patterns of admixture support this hypothesis – montane individuals from the northern Cordillera Azul and their proximate lowland relatives all show limited genomic admixture from derived lineages of the Guiana Shield + Northern Amazon clade (yellow in Figure 5). This result is consistent with a scenario of secondary

1139 invasion of the highlands from lowland Inambari lineages after weak introgression from
1140 the western Napo (see below).
1141
1142 4.4 *North Andean diversity*
1143 Our final genomic data set does not include any samples from Andean Ecuador or
1144 Colombia, a region that includes four previously recognized subspecies and at least four
1145 distinct vocalization types (see Taxonomic Summary). Unfortunately, the three Andean
1146 Ecuadorian tissue (representing subspecies *coracina*) specimens in our original sample
1147 were not of high enough preservation quality to be viable for ddRAD sequencing
1148 (Supplementary Table 1). We were able to obtain mtDNA from these samples however,
1149 which unambiguously formed a monophyletic cluster with our single San Martín (North
1150 Andean Peru) specimen in mtDNA gene tree analyses (Supplemental Appendix,
1151 Supplementary Figure 17). A prior study by Castro-Astor (2014) based on mtDNA also
1152 included one Andean Ecuadorian sample and one Central American sample, which
1153 unambiguously clustered together (*pp*=100). Thus, it is likely that our unsampled Andean
1154 lineages of Ecuador and Colombia (including the subspecies *minimus, bolivari, unica,* and
1155 *coracina*) are likely members of the northern Andean clade represented in our genomic
1156 data set by our single San Martín specimen.
1157 As noted earlier, our San Martín specimen had approximately equal probability of
1158 assignment to multiple populations, which could indicate heavy admixture, or that this is a
1159 single sample from a highly distinct population which has its own distinct history. These
1160 patterns may also be consistent with the hypothesis that the North Andean clade is the
1161 source of montane and lowland *Pseudopipra* diversity. In summary, the genus *Pseudopipra*

appears to have radiated 'out of the Andes,' expanding into the Central America highlands, and into the eastern lowland Amazonian and Atlantic forests.

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*. 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*

Our analysis identifies 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.* 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). Three of these species are polytypic (i.e. contain multiple subspecies). One of the species we recognize– *P. microlopha* – is paraphyletic with respect to another species, *P. cephaleucos,* because we recognize that speciation may not threaten the lineage identity of a paraphyletic group.

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 for other lineages in the genus strongly suggest that each of these subspecies is a distinct evolutionary lineage deserving species status under either the phylogenetic or biological species concepts. Plumage and vocal data from other newly identified clades may support the recognition of additional species including a lowland-derived montane population from the northern Cordillera Azul (Clade C2; Fig. 3), and the currently unnamed lineages from

the southern Amazon Basin (Clades C3, C4, and C5). In conclusion, *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]:
LINK REMOVED FOR BIORXIV
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

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)

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.

1480 Bird, J.P., Martin, R., Akçakaya, H.R., Gilroy, J., Burfield, I.J., Garnett, S., Symes, A., Taylor, J.,
1481 Şekercioğlu, Ç.H., Butchart, S.H.M., 2020. Generation lengths of the world's birds and their
1482 implications for extinction risk. Conservation Biology n/a.

1483 Borges, S.H., Da Silva, J.M.C., 2012. A New Area of Endemism for Amazonian Birds in the Rio
1484 Negro Basin. The Wilson Journal of Ornithology 124, 15-23.

1485 Bradburd, G.S., Coop, G.M., Ralph, P.L., 2018. Inferring Continuous and Discrete Population
1486 Genetic Structure Across Space. Genetics 210, 33-52.

1487 Brelsford, A., Milá, B., Irwin, D.E., 2011. Hybrid origin of Audubon's warbler. Molecular
1488 Ecology 20, 2380-2389.

1489 Brumfield, R.T., 2012. Inferring the origins of lowland Neotropical birds. The Auk 129, 367-
1490 376.

1491 Brumfield, R.T., Edwards, S.V., 2007. EVOLUTION INTO AND OUT OF THE ANDES: A
1492 BAYESIAN ANALYSIS OF HISTORICAL DIVERSIFICATION IN THAMNOPHILUS
1493 ANTSHRIKES. Evolution 61, 346-367.

1494 Brumfield, R.T., Liu, L., Lum, D.E., Edwards, S.V., 2008. Comparison of Species Tree Methods
1495 for Reconstructing the Phylogeny of Bearded Manakins (Aves: Pipridae, Manacus) from
1496 Multilocus Sequence Data. Systematic Biology 57, 719-731.

1497 Cabanne, G.S., Campagna, L., Trujillo-Arias, N., Naoki, K., Gómez, I., Miyaki, C.Y., Santos, F.R.,
1498 Dantas, G.P.M., Aleixo, A., Claramunt, S., Rocha, A., Caparroz, R., Lovette, I.J., Tubaro, P.L.,
1499 2019. Phylogeographic variation within the Buff-browed Foliage-gleaner (Aves:

Furnariidae: Syndactyla rufosuperciliata) supports an Andean-Atlantic forests connection via the Cerrado. Molecular Phylogenetics and Evolution 133, 198-213.

Campagna, L., Gronau, I., Silveira, L.F., Siepel, A., Lovette, I.J., 2015. Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation. Molecular 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.

1518 Castro-Astor, I.N., Alves, M.A.S., Cavalcanti, R.B., 2007. DISPLAY BEHAVIOR AND SPATIAL
1519 DISTRIBUTION OF THE WHITE-CROWNED MANAKIN IN THE ATLANTIC FOREST OF
1520 BRAZIL. The Condor 109, 155-166.

1521 Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis
1522 tool set for population genomics. Molecular Ecology 22, 3124-3140.

1523 Chambers, E.A., Hillis, D.M., 2020. The Multispecies Coalescent Over-Splits Species in the
1524 Case of Geographically Widespread Taxa. Systematic Biology 69, 184-193.

1525 Chapman, F.M., Miller, L.E., Allen, A.A., Richardson, W.B., Cherrie, G.K., 1914. Diagnoses of
1526 apparently new Colombian birds. 3. Bulletin of the AMNH; v. 33, article 40.

1527 Chernomor, O., Minh, B.Q., von Haeseler, A., 2016. Terrace Aware Data Structure for
1528 Phylogenomic Inference from Supermatrices. Systematic Biology 65, 997-1008.

1529 Cheviron, Z.A., Hackett, S.J., Brumfield, R.T., 2006. Sequence variation in the coding region
1530 of the melanocortin-1 receptor gene (MC1R) is not associated with plumage variation in
1531 the blue-crowned manakin (Lepidothrix coronata). Proceedings of the Royal Society B:
1532 Biological Sciences 273, 1613-1618.

1533 Cheviron, Z.A., Hackett, S.J., Capparella, A.P., 2005. Complex evolutionary history of a
1534 Neotropical lowland forest bird (Lepidothrix coronata) and its implications for historical
1535 hypotheses of the origin of Neotropical avian diversity. Molecular Phylogenetics and
1536 Evolution 36, 338-357.

1537 Chifman, J., Kubatko, L., 2014. Quartet Inference from SNP Data Under the Coalescent
1538 Model. Bioinformatics 30, 3317-3324.

1539 Chifman, J., Kubatko, L., 2015. Identifiability of the unrooted species tree topology under
1540 the coalescent model with time-reversible substitution processes, site-specific rate
1541 variation, and invariable sites. Journal of Theoretical Biology 374, 35-47.

1542 Chou, J., Gupta, A., Yaduvanshi, S., Davidson, R., Nute, M., Mirarab, S., Warnow, T., 2015. A
1543 comparative study of SVDquartets and other coalescent-based species tree estimation
1544 methods. BMC Genomics 16, S2.

1545 Counterman, B.A., 2016. Hybrid Speciation. In: Kliman, R.M. (Ed.), Encyclopedia of
1546 Evolutionary Biology. Academic Press, Oxford, pp. 242-248.

1547 Cracraft, J., 1985. Historical Biogeography and Patterns of Differentiation within the South
1548 American Avifauna: Areas of Endemism. Ornithological Monographs, 49-84.

1549 Cracraft, J., Prum, R.O., 1988. PATTERNS AND PROCESSES OF DIVERSIFICATION:
1550 SPECIATION AND HISTORICAL CONGRUENCE IN SOME NEOTROPICAL BIRDS. Evolution
1551 42, 603-620.

1552 Crother, B.I., Murray, C.M., 2011. Ontology of areas of endemism. Journal of Biogeography
1553 38, 1009-1015.

1554 Cunningham, C.W., Collins, T.M., 1994. Developing model systems for molecular
1555 biogeography: Vicariance and interchange in marine invertebrates. In: Schierwater, B.,

1556 Streit, B., Wagner, G.P., DeSalle, R. (Eds.), Molecular Ecology and Evolution: Approaches and
1557 Applications. Birkhäuser Basel, Basel, pp. 405-433.

1558 Da Silva, J., Novaes, F., Oren, D., 2002. Differentiation of Xiphocolaptes (Dendrocolaptidae)
1559 across the river Xingu, Brazilian Amazonia: recognition of a new phylogenetic species and
1560 biogeographic implications. BULLETIN-BRITISH ORNITHOLOGISTS CLUB 122, 185-193.

1561 da Silva, J.M.C., 1995. Biogeographic analysis of the South American Cerrado avifauna.
1562 Steenstrupia 21, 49-67.

1563 da Silva, J.M.C., de Sousa, M.C., H. M. Castelletti, C., 2004. Areas of Endemism for Passerine
1564 Birds in the Atlantic Forest, South America. Global Ecology and Biogeography 13, 85-92.

1565 Da Silva, J.M.C., Rylands , A.B., Da Fonseca, G.A.B., 2005. The Fate of the Amazonian Areas of
1566 Endemism. Conservation Biology 19, 689-694.

1567 de Queiroz, K., 2005. Ernst Mayr and the modern concept of species. Proceedings of the
1568 National Academy of Sciences 102, 6600-6607.

1569 de Schauensee, R.M., 1945. Notes on Colombian flycatchers, manakins and cotingas.
1570 Proceedings of the Academy of Natural Sciences of Philadelphia 97, 41-57.

1571 de Schauensee, R.M., 1950. Colombian zoological survey. Part V: New birds from Colombia.
1572 Notulae Naturae (221), 1-13.

1573 Dickinson, E., 2003. The Howard and Moore Complete Checklist of the Birds of the World
1574 Christopher Helm. London.

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

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

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

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

Fernandes, A.M., Gonzalez, J., Wink, M., Aleixo, A., 2013. Multilocus phylogeography of the 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.

1595 Francisco, M.R., Gibbs, H.L., Galetti, M., Lunardi, V.O., Junior, P.M.G., 2007. Genetic structure
1596 in a tropical lek-breeding bird, the blue manakin (Chiroxiphia caudata) in the Brazilian
1597 Atlantic Forest. Molecular Ecology 16, 4908-4918.

1598 Freile, J.F., 2014. First record of White-crowned Manakin Dixiphia pipra in western
1599 Ecuador. Bull. Brit. Orn. Club 134, 238-240.

1600 Gordon, A., Hannon, G., 2010. Fastx-toolkit. FASTQ/A short-reads preprocessing tools
1601 (unpublished) http://hannonlab. cshl. edu/fastx_toolkit 5.

1602 Goslee, S.C., Urban, D.L., 2007. The ecodist package for dissimilarity-based analysis of
1603 ecological data. Journal of Statistical Software 22, 1-19.

1604 Goudet, J., 2005. hierfstat, a package for r to compute and test hierarchical F-statistics.
1605 Molecular Ecology Notes 5, 184-186.

1606 Goudet, J., Perrin, N., Waser, P., 2002. Tests for sex-biased dispersal using bi-parentally
1607 inherited genetic markers. Molecular Ecology 11, 1103-1114.

1608 Grant, P.R., Grant, B.R., 1997. Genetics and the origin of bird species. Proceedings of the
1609 National Academy of Sciences 94, 7768-7775.

1610 Gronau, I., Hubisz, M.J., Gulko, B., Danko, C.G., Siepel, A., 2011. Bayesian inference of ancient
1611 human demography from individual genome sequences. Nature Genetics 43, 1031.

1612 Gubili, C., Ritter, C.D., Motta, J., Farias, I.P., Bates, J., Canton, R., Capurucho, J.M.G., Cornelius,
1613 C., Feldheim, K.A., Ribas, C.C., 2016. Isolation and characterization of polymorphic

1614 microsatellite DNA Markers from an Amazonian white-sand vegetation specialist bird,
1615 Xenopipo atronitens (Aves: Pipridae). The Wilson Journal of Ornithology 128, 668-672.

1616 Guillot, G., Rousset, F., 2013. Dismantling the Mantel tests. Methods in Ecology and
1617 Evolution 4, 336-344.

1618 Haffer, J., 1969. Speciation in Amazonian Forest Birds. Science 165, 131-137.

1619 Haffer, J., 1974. Avian speciation in tropical South America, with a systematic survey of the
1620 toucans (Ramphastidae) and jacamars (Galbulidae) (Publications of the Nuttall
1621 Ornithological Club No. 14, Cambridge, MA).

1622 Haffer, J., 1978. Distribution of Amazon forest birds. Bonner Zoologische Beiträge 29, 38-
1623 78.

1624 Haffer, J., 1985. Avian zoogeography of the Neotropical lowlands. Ornithological
1625 Monographs, 113-146.

1626 Haffer, J., 1992. On the river effect in some forest birds of southern Amazonia. Boletim do
1627 Museu Paraense Emílio Goeldi. Nova série. Zoologia 8, 217-245.

1628 Haffer, J., 2008. Hypotheses to explain the origin of species in Amazonia. Brazilian Journal
1629 of Biology 68, 917-947.

1630 Hartl, D.L., Clark, A.G., 1997. Principles of population genetics. Sinauer associates
1631 Sunderland, MA.

1632 Harvey, M.G., Aleixo, A., Ribas, C.C., Brumfield, R.T., 2017. Habitat Association Predicts
1633 Genetic Diversity and Population Divergence in Amazonian Birds. The American Naturalist
1634 190, 631-648.

1635 Harvey, M.G., Brumfield, R.T., 2015. Genomic variation in a widespread Neotropical bird
1636 (Xenops minutus) reveals divergence, population expansion, and gene flow. Molecular
1637 Phylogenetics and Evolution 83, 305-316.

1638 Hazzi, N.A., Moreno, J.S., Ortiz-Movliav, C., Palacio, R.D., 2018. Biogeographic regions and
1639 events of isolation and diversification of the endemic biota of the tropical Andes.
1640 Proceedings of the National Academy of Sciences.

1641 Hermansen, J.S., Sæther, S.A., Elgvin, T.O., Borge, T., Hjelle, E., Sætre, G.-P., 2011. Hybrid
1642 speciation in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene
1643 flow. Molecular Ecology 20, 3812-3822.

1644 Hijmans, R.J., Williams, E., Vennes, C., 2015. Geosphere: spherical trigonometry. R package
1645 version 1.3-11. URL: https://CRAN. R-project. org/package= geosphere.

1646 Hillis, D.M., 2020. The Detection and Naming of Geographic Variation Within Species.
1647 Herpetological Review, 51:52-56.

1648 Hoang, D.T., Vinh, L.S., Chernomor, O., Minh, B.Q., von Haeseler, A., 2017. UFBoot2:
1649 Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution 35,
1650 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.

1670 Kamvar, Z.N., Tabima, J.F., Grünwald, N.J., 2014. Poppr: an R package for genetic analysis of
1671 populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2, e281.

1672 Kirwan, G.M., David, N., Gregory, S.M., Jobling, J.A., Steinheimer, F.D., Brito, G.R.R., 2016. The
1673 mistaken manakin: a new genus-group name for Parus pipra Linnaeus, 1758 (Aves:
1674 Passeriformes: Pipridae). Zootaxa 4121, 89-94.

1675 Kirwan, G.M., Green, G., 2012. Cotingas and Manakins. Princeton University Press,
1676 Princeton, NJ.

1677 Kumar, S., Subramanian, S., 2002. Mutation rates in mammalian genomes. Proceedings of
1678 the National Academy of Sciences 99, 803-808.

1679 Lamichhaney, S., Han, F., Webster, M.T., Andersson, L., Grant, B.R., Grant, P.R., 2018. Rapid
1680 hybrid speciation in Darwin's finches. Science 359, 224-228.

1681 Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. PartitionFinder 2:
1682 New Methods for Selecting Partitioned Models of Evolution for Molecular and
1683 Morphological Phylogenetic Analyses. Molecular Biology and Evolution 34, 772-773.

1684 Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient
1685 alignment of short DNA sequences to the human genome. Genome Biology 10, R25.

1686 Lavretsky, P., Engilis, A., Eadie, J.M., Peters, J.L., 2015. Genetic admixture supports an
1687 ancient hybrid origin of the endangered Hawaiian duck. Journal of Evolutionary Biology 28,
1688 1005-1015.

1689 Lawson, D.J., Hellenthal, G., Myers, S., Falush, D., 2012. Inference of Population Structure
1690 using Dense Haplotype Data. PLOS Genetics 8, e1002453.

1691 Leaché, A.D., Banbury, B.L., Felsenstein, J., de Oca, A.n.-M., Stamatakis, A., 2015. Short Tree,
1692 Long Tree, Right Tree, Wrong Tree: New Acquisition Bias Corrections for Inferring SNP
1693 Phylogenies. Systematic Biology 64, 1032-1047.

1694 Leaché, A.D., Fujita, M.K., Minin, V.N., Bouckaert, R.R., 2014. Species Delimitation using
1695 Genome-Wide SNP Data. Systematic Biology 63, 534-542.

1696 Leaché, A.D., Zhu, T., Rannala, B., Yang, Z., 2018. The Spectre of Too Many Species.
1697 Systematic Biology, syy051-syy051.

1698 Legendre, P., Fortin, M.-J., 2010. Comparison of the Mantel test and alternative approaches
1699 for detecting complex multivariate relationships in the spatial analysis of genetic data.
1700 Molecular Ecology Resources 10, 831-844.

1701 Linck, E.B., Battey, C.J., 2017. Minor allele frequency thresholds strongly affect population
1702 structure inference with genomic datasets. bioRxiv.

1703 Linder, H.P., 2001. On Areas of Endemism, with an Example from the African Restionaceae.
1704 Systematic Biology 50, 892-912.

1705 Long, C., Kubatko, L., 2017. The effect of gene flow on coalescent-based species-tree
1706 inference. arXiv:1710.03806.

1707 Luna, L.W., Souza, T.O., e Silva, W.A.d.G., Schneider, H., Sampaio, I., 2017. Genetic variation
1708 of the endangered Araripe Manakin (Antilophia bokermanni) indicates a history of
1709 demographic decline. Revista Brasileira de Ornitologia 25, 60-66.

1710 Lynch Alfaro, J.W., Boubli, J.P., Paim, F.P., Ribas, C.C., Silva, M.N.F.d., Messias, M.R., Röhe, F.,
1711 Mercês, M.P., Silva Júnior, J.S., Silva, C.R., Pinho, G.M., Koshkarian, G., Nguyen, M.T.T., Harada,
1712 M.L., Rabelo, R.M., Queiroz, H.L., Alfaro, M.E., Farias, I.P., 2015. Biogeography of squirrel
1713 monkeys (genus Saimiri): South-central Amazon origin and rapid pan-Amazonian
1714 diversification of a lowland primate. Molecular Phylogenetics and Evolution 82, 436-454.

1715 MacQueen, J., 1967. Some methods for classification and analysis of multivariate
1716 observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and
1717 Probability, Volume 1: Statistics. University of California Press, Berkeley, Calif., pp. 281-
1718 297.

1719 Malinsky, M., Trucchi, E., Lawson, D.J., Falush, D., 2018. RADpainter and fineRADstructure:
1720 Population Inference from RADseq Data. Molecular Biology and Evolution 35, 1284-1290.

1721 Mantel, N., 1967. The detection of disease clustering and a generalized regression
1722 approach. Cancer research 27, 209-220.

1723 Marks, B.D., Hackett, S.J., Capparella, A.P., 2002. Historical relationships among Neotropical
1724 lowland forest areas of endemism as determined by mitochondrial DNA sequence variation
1725 within the Wedge-billed Woodcreeper (Aves: Dendrocolaptidae: Glyphorynchus spirurus).
1726 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.

1746 Naka, L.N., Brumfield, R.T., 2018. The dual role of Amazonian rivers in the generation and
1747 maintenance of avian diversity. Science Advances 4.

1748 Nieto Feliner, G., Álvarez, I., Fuertes-Aguilar, J., Heuertz, M., Marques, I., Moharrek, F.,
1749 Piñeiro, R., Riina, R., Rosselló, J.A., Soltis, P.S., Villa-Machío, I., 2017. Is homoploid hybrid
1750 speciation that rare? An empiricist's view. Heredity 118, 513.

1751 Noguera-urbano, E.A., 2016. Areas of endemism: travelling through space and the
1752 unexplored dimension. Systematics and Biodiversity 14, 131-139.

1753 Nyári, Á.S., 2007. Phylogeographic patterns, molecular and vocal differentiation, and
1754 species limits in Schiffornis turdina (Aves). Molecular Phylogenetics and Evolution 44, 154-
1755 164.

1756 Nylander, J.A.A., Olsson, U., Alström, P., Sanmartín, I., 2008. Accounting for Phylogenetic
1757 Uncertainty in Biogeography: A Bayesian Approach to Dispersal-Vicariance Analysis of the
1758 Thrushes (Aves: Turdus). Systematic Biology 57, 257-268.

1759 Ohlson, J.I., Fjeldså, J., Ericson, P.G.P., 2013. Molecular phylogeny of the manakins (Aves:
1760 Passeriformes: Pipridae), with a new classification and the description of a new genus.
1761 Molecular Phylogenetics and Evolution 69, 796-804.

1762 Paradis, E., 2010. pegas: an R package for population genetics with an integrated–modular
1763 approach. Bioinformatics 26, 419-420.

1764 Paris, J.R., Stevens, J.R., Catchen, J.M., Johnston, S., 2017. Lost in parameter space: a road
1765 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., 1990a. Phylogenetic Analysis of the Evolution of Display Behavior in the Neotropical Manakins (Aves: Pipridae). Ethology 84, 202-231.

Prum, R.O., 1990b. A test of the monophyly of the manakins (Pipridae) and of the cotingas (Cotingidae) based on morphology. Occasional Papers of the Museum of Zoology of the University of Michigan 723, 1-44.

Prum, R.O., 1992. Syringeal morphology, phylogeny, and evolution of the neotropical manakins (Aves, Pipridae). American Museum novitates; no. 3043.

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.

Reichenbach, H.G., 1850. Avium Systema Naturale: Vorläufer einer Iconographie der Arten der Vögel aller Welttheile, welche, nachdem bereits fast dreitausend Abbildungen erschienen sind, ununterbrochen fortgesetzt wird. Hofmeister.

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.

1806 Ridgway, R., 1906. Descriptions of some new forms of oligomyodian birds. Proceedings of
1807 the Biological Society of Washington, 115-120.

1808 Roure, B., Baurain, D., Philippe, H., 2013. Impact of Missing Data on Phylogenies Inferred
1809 from Empirical Phylogenomic Data Sets. Molecular Biology and Evolution 30, 197-214.

1810 Rousset, F., 1997. Genetic Differentiation and Estimation of Gene Flow from F Statistics
1811 Under Isolation by Distance. Genetics 145, 1219-1228.

1812 Salazar, C., Baxter, S.W., Pardo-Diaz, C., Wu, G., Surridge, A., Linares, M., Bermingham, E.,
1813 Jiggins, C.D., 2010. Genetic Evidence for Hybrid Trait Speciation in Heliconius Butterflies.
1814 PLOS Genetics 6, e1000930.

1815 Saranathan, V., Hamilton, D., Powell, G.V.N., Kroodsma, D.E., Prum, R.O., 2007. Genetic
1816 evidence supports song learning in the three-wattled bellbird Procnias tricarunculata
1817 (Cotingidae). Molecular Ecology 16, 3689-3702.

1818 Schmidt, H.A., Minh, B.Q., von Haeseler, A., Nguyen, L.-T., 2014. IQ-TREE: A Fast and
1819 Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular
1820 Biology and Evolution 32, 268-274.

1821 Schmidt-Lebuhn, A.N., Aitken, N.C., Chuah, A., 2017. Species trees from consensus single
1822 nucleotide polymorphism (SNP) data: Testing phylogenetic approaches with simulated and
1823 empirical data. Molecular Phylogenetics and Evolution 116, 192-201.

1824 Schumer, M., Rosenthal, G.G., Andolfatto, P., 2014. HOW COMMON IS HOMOPLOID HYBRID
1825 SPECIATION? Evolution 68, 1553-1560.

1826 Sedano, R.E., Burns, K.J., 2010. Are the Northern Andes a species pump for Neotropical
1827 birds? Phylogenetics and biogeography of a clade of Neotropical tanagers (Aves:
1828 Thraupini). Journal of Biogeography 37, 325-343.

1829 Shafer, A.B.A., Peart, C.R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C.W., Wolf, J.B.W., 2017.
1830 Bioinformatic processing of RAD-seq data dramatically impacts downstream population
1831 genetic inference. Methods in Ecology and Evolution 8, 907-917.

1832 Silva, S.M., Peterson, A.T., Carneiro, L., Burlamaqui, T.C.T., Ribas, C.C., Sousa-Neves, T.,
1833 Miranda, L.S., Fernandes, A.M., d'Horta, F.M., Araújo-Silva, L.E., Batista, R., Bandeira,
1834 C.H.M.M., Dantas, S.M., Ferreira, M., Martins, D.M., Oliveira, J., Rocha, T.C., Sardelli, C.H.,
1835 Thom, G., Rêgo, P.S., Santos, M.P., Sequeira, F., Vallinoto, M., Aleixo, A., 2019. A dynamic
1836 continental moisture gradient drove Amazonian bird diversification. Science Advances 5,
1837 eaat5752.

1838 Smeds, L., Qvarnström, A., Ellegren, H., 2016. Direct estimate of the rate of germline
1839 mutation in a bird. Genome Research 26, 1211-1218.

1840 Smith, B.T., McCormack, J.E., Cuervo, A.M., Hickerson, M.J., Aleixo, A., Cadena, C.D., Perez-
1841 Eman, J., Burney, C.W., Xie, X., Harvey, M.G., Faircloth, B.C., Glenn, T.C., Derryberry, E.P.,
1842 Prejean, J., Fields, S., Brumfield, R.T., 2014. The drivers of tropical speciation. Nature 515,
1843 406-409.

1844 Smith, R.H., 1979. On selection for inbreeding in polygynous animals. Heredity 43, 205.

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

Snow, D.W., 1979. Tityrinae, Pipridae, Cotingidae. In: M. A. Traylor, J. (Ed.), Check-list of birds of the world. Museum of Comparative Zoology, Harvard Univ., Cambridge, Massachusetts, pp. 229-308.

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

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

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

Stopher, K.V., Nussey, D.H., Clutton-Brock, T.H., Guinness, F., Morris, A., Pemberton, J.M., 2012. Re-mating across years and intralineage polygyny are associated with greater than expected levels of inbreeding in wild red deer. Journal of Evolutionary Biology 25, 2457-2469.

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

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

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

1865 Tello, J.G., Moyle, R.G., Marchese, D.J., Cracraft, J., 2009. Phylogeny and phylogenetic
1866 classification of the tyrant flycatchers, cotingas, manakins, and their allies (Aves:
1867 Tyrannides). Cladistics 25, 429-467.

1868 Teske, P.R., Golla, T.R., Sandoval-Castillo, J., Emami-Khoyi, A., van der Lingen, C.D., von der
1869 Heyden, S., Chiazzari, B., Jansen van Vuuren, B., Beheregaray, L.B., 2018. Mitochondrial DNA
1870 is unsuitable to test for isolation by distance. Scientific Reports 8, 8448.

1871 Thom, G., Amaral, F.R.D., Hickerson, M.J., Aleixo, A., Araujo-Silva, L.E., Ribas, C.C., Choueri, E.,
1872 Miyaki, C.Y., 2018. Phenotypic and Genetic Structure Support Gene Flow Generating Gene
1873 Tree Discordances in an Amazonian Floodplain Endemic Species. Systematic Biology 67,
1874 700-718.

1875 Thrasher, D.J., Butcher, B.G., Campagna, L., Webster, M.S., Lovette, I.J., 2018. Double-digest
1876 RAD sequencing outperforms microsatellite loci at assigning paternity and estimating
1877 relatedness: A proof of concept in a highly promiscuous bird. Molecular Ecology Resources
1878 0.

1879 Trifinopoulos, J., Nguyen, L.-T., Minh, B.Q., von Haeseler, A., 2016. W-IQ-TREE: a fast online
1880 phylogenetic tool for maximum likelihood analysis. Nucleic Acids Research 44, W232-
1881 W235.

1882 Vaidya, G., Lohman, D.J., Meier, R., 2011. SequenceMatrix: concatenation software for the
1883 fast assembly of multi-gene datasets with character set and codon information. Cladistics
1884 27, 171-180.

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

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

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

Waser, P.M., Austad, S.N., Keane, B., 1986. When Should Animals Tolerate Inbreeding? The American Naturalist 128, 529-537.

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

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

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

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

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

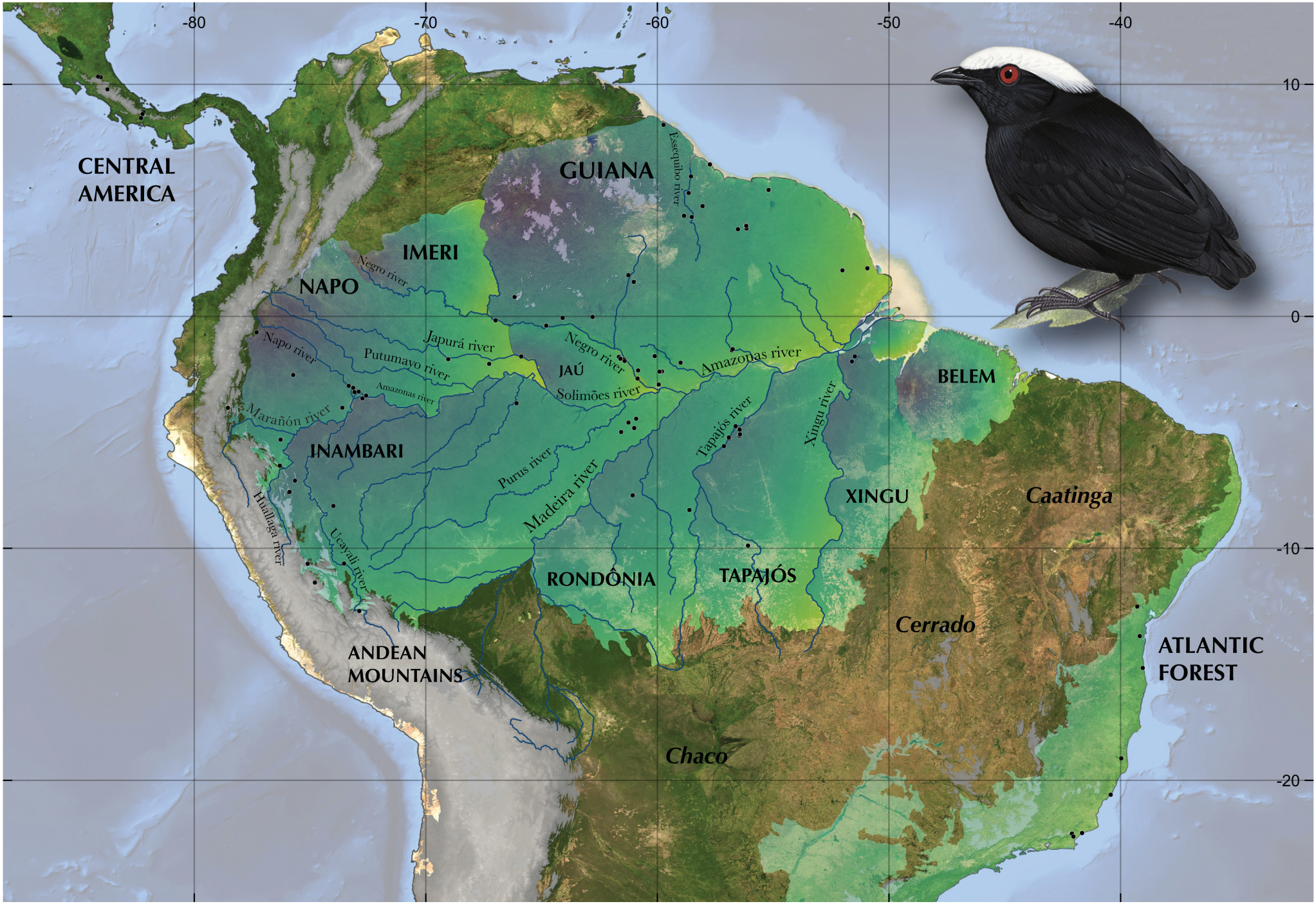
1905 Zhang, J., Kapli, P., Pavlidis, P., Stamatakis, A., 2013. A general species delimitation method
1906 with applications to phylogenetic placements. Bioinformatics 29, 2869-2876.

1907 Zhu, J., Wen, D., Yu, Y., Meudt, H.M., Nakhleh, L., 2018. Bayesian inference of phylogenetic
1908 networks from bi-allelic genetic markers. PLOS Computational Biology 14, e1005932.

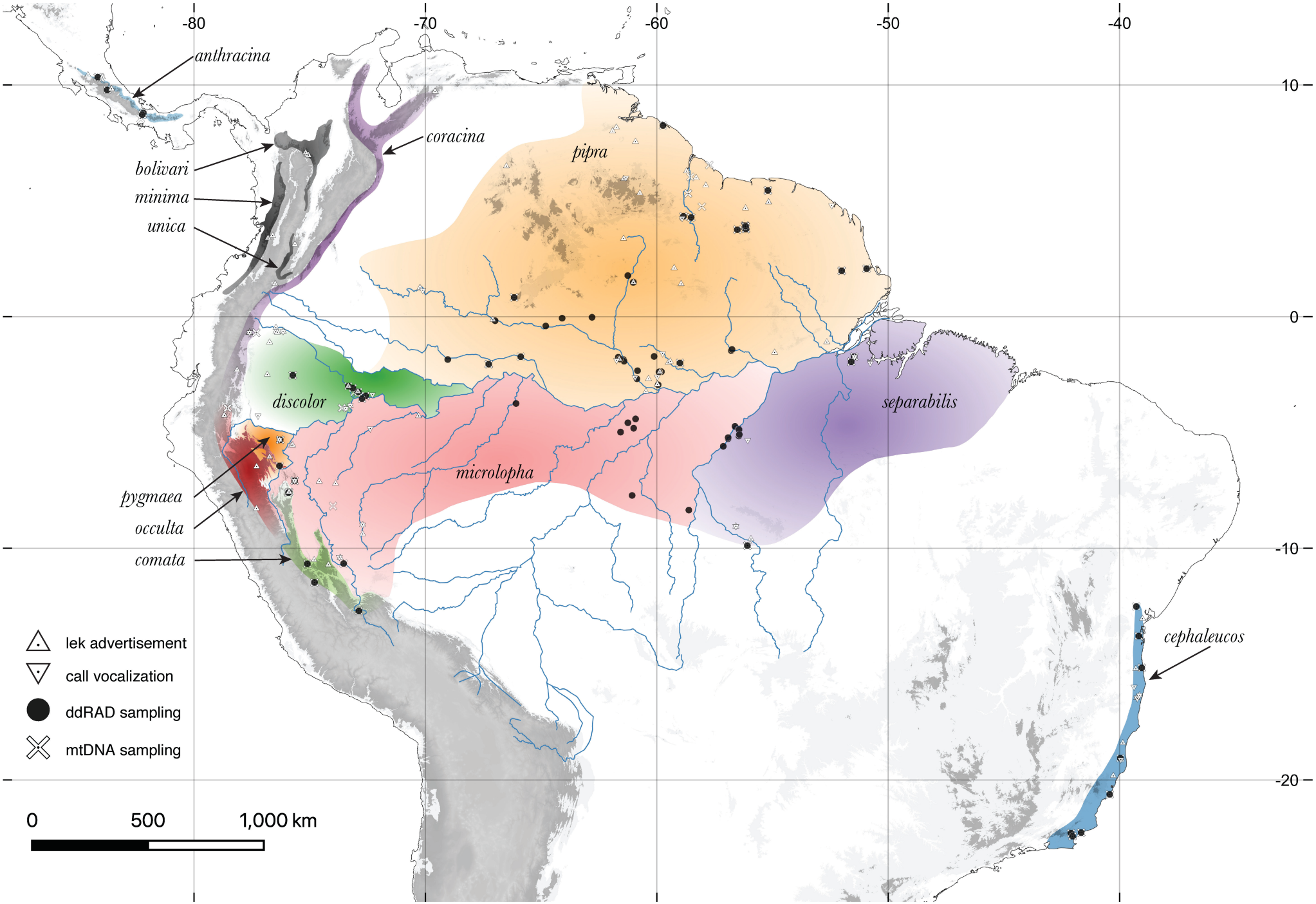
1909 Zimmer, J.T., 1936. Studies of Peruvian birds. XXII. Notes on the Pipridae. American
1910 Museum novitates; no. 889. 1-29.

1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926

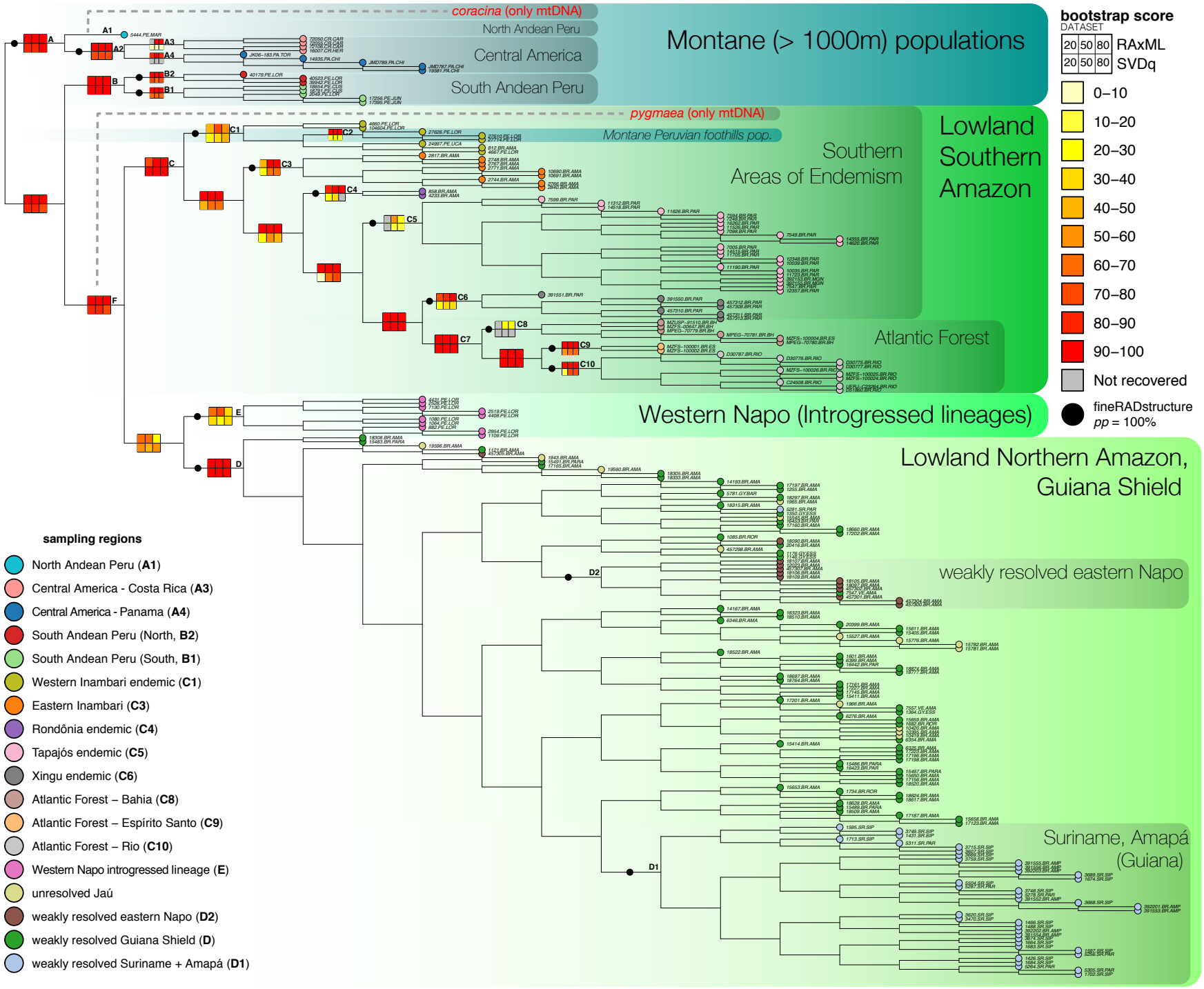
Figure 1. Neotropical areas of endemism and sampling localities. Amazonian lowland areas of endemism as portrayed in Da Silva et al. (2005) are highlighted in shades of green, as well as the Jaú area of endemism (Borges and Da Silva, 2012). Montane Andean, Central American, Guianan, and dry diagonal 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

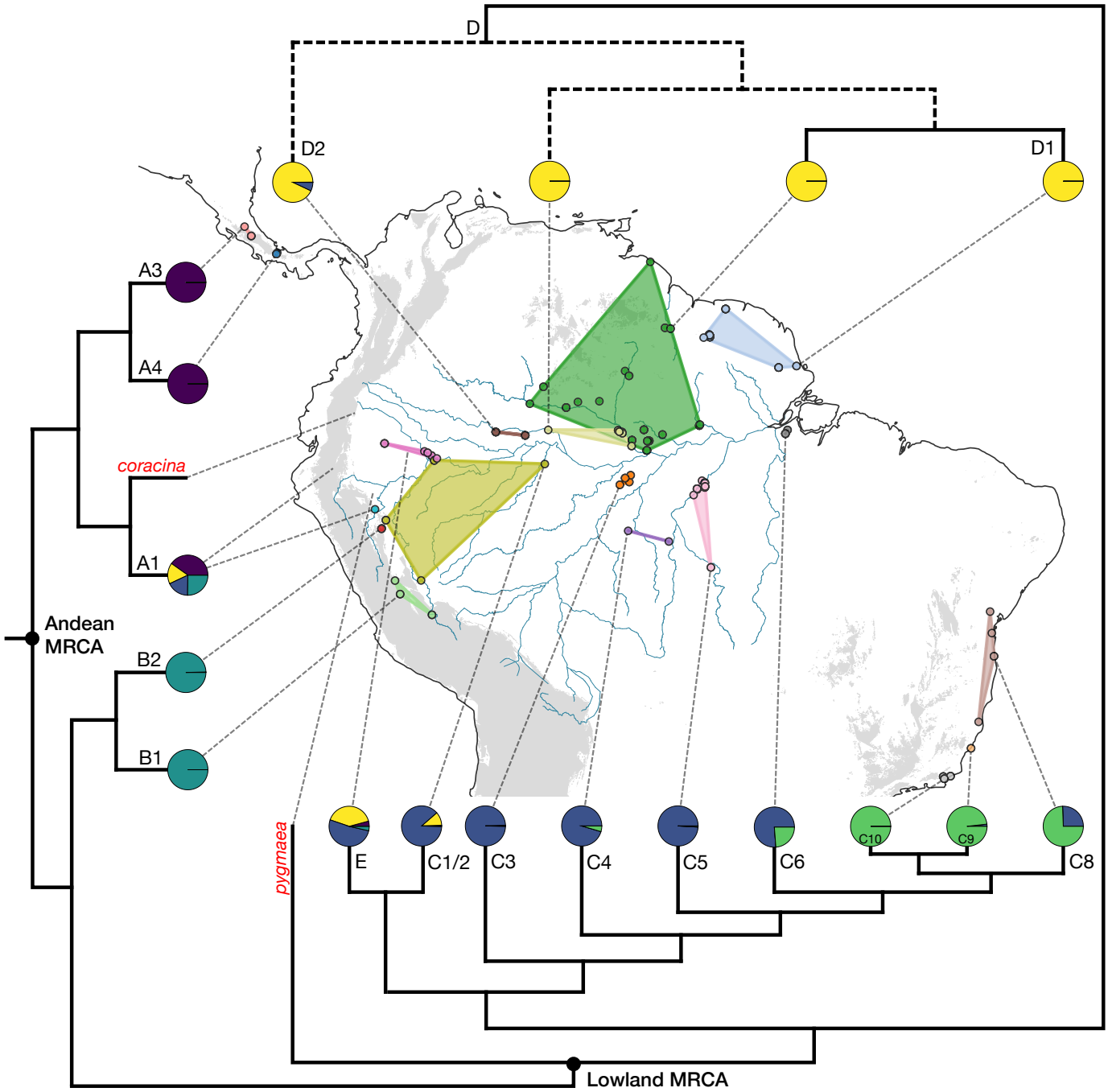
1950 **Figure 2.** Subspecies sampling. Our genetic sampling of *Pseudopipra pipra* intersects with
1951 the ranges of ten named subspecies. In this figure, we indicate sampling localities as they
1952 overlap with previously identified subspecies ranges. Areas greater than 1000m in
1953 elevation are emphasized in light grey; major Amazonian rivers are indicated in light blue.
1954 Overall distribution reflects the BirdLife International approximation, updated to reflect
1955 current knowledge (BirdLife International Species factsheet: *Pseudopipra pipra*, 2018, also
1956 see Supplemental Figure 1). The ranges of Andean taxa were adjusted based on
1957 unpublished data provided Andrés Cuervo (*personal communication*). Hypothetical
1958 subspecies ranges are based on descriptions in Kirwan and Green (2012) and imply
1959 maximal ranges highlighting key geographic barriers delimiting subspecies. Black dots
1960 indicate sampling for ddRADseq, while crosses indicate sampling for mtDNA ND2.
1961 Triangles indicate sampling localities for male advertisement vocalizations (upward) and
1962 territorial vocalizations (downward). Our sampling encompasses the majority of the
1963 lowland extents of the recognized range of *Pseudopipra*, as well as key montane
1964 populations in Peru, Ecuador, and Central America. Ssp *anthracina, bolivari, minima, unica,*
1965 *coracina, occulta, and comata* are exclusively montane.
1966
1967
1968
1969
1970
1971
1972


-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

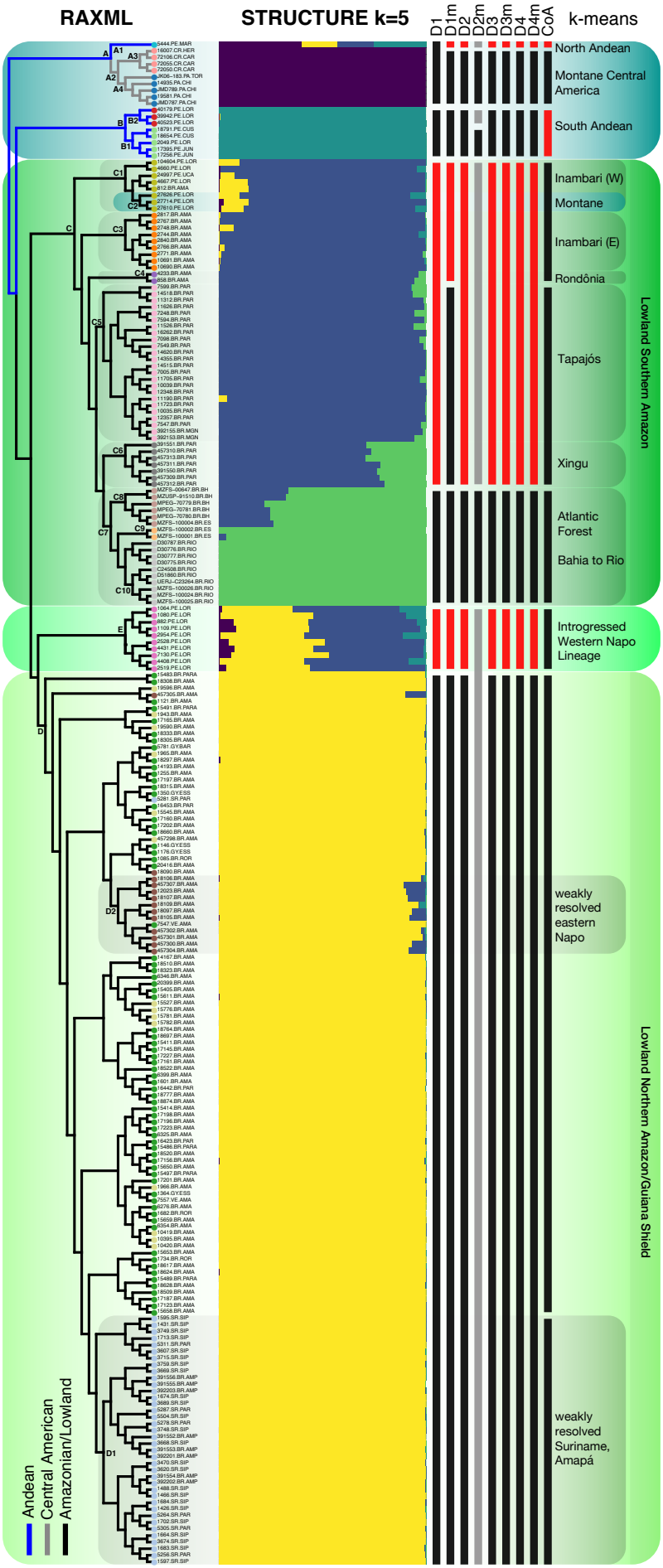
1973 **Figure 3.** Phylogenetic hypothesis for *Pseudopipra*. Phylogenetic analysis of concatenated
1974 ddRAD sequences generated a well-resolved phylogenetic hypothesis which was largely
1975 congruent across datasets and analytical approaches. Montane clades, including a Northern
1976 Andean (A1) + Central American (A2) clade (A), and a central Peruvian clade (B), are
1977 recovered as nested sister groups to two wide ranging lowland clades (C and D), indicating
1978 that the lowland clades are descended from Andean lineages (Figure 5). Shown is the
1979 RAxML topology generated using the 50% haplotype dataset. Branch lengths are set to
1980 equal for graphical interpretability (see Supplementary Data for original tree files). Colored
1981 circles at tips indicate group membership to one of eighteen population-areas (matching
1982 Figure 4 and 5). Heatmaps at key nodes indicate bootstrap scores from each of six
1983 phylogenetic analyses for 20/50/80% datasets, as indicated in the legend. Black dots on
1984 branches indicate populations which are identified with 100% posterior probability in
1985 fineRADstructure analysis. Most of the major structure is recovered by both RAxML and
1986 SVDquartets with high support, though RAxML recovers additional low-support
1987 substructure in the Northern Amazonian + Guiana Shield clade (D) which is coincident with
1988 several geographic features (clades D1 and D2). A clade of introgressed western Napo
1989 individuals (E) is recovered by all standard phylogenetic analyses as sister to the northern
1990 Amazonian + Guiana shield clade (D). As the history of this clade is characterized by a
1991 complex introgression scenario between northern and southern lineages, a purely
1992 bifurcating tree model is inadequate to describe its relations with other groups (see text
1993 and Figure 6). Also shown are the well supported phylogenetic positions of two distinct
1994 subspecies lineages (*pygmaea* and *coracaina*) from which we were able to obtain mtDNA
1995 (Supplemental Appendix).


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)

1996 **Figure 4.** Mapping of phylogenetic relationships into geographic space. Here, a simplified
1997 topology is plotted with tips anchored to sampling coordinates. The colors of convex hulls
1998 (i.e., hypotheses of minimum clade ranges) match tip colors in Figure 3 and 5. These
1999 colored hulls represent distinct non-overlapping geographic areas delimited by physical
2000 barriers (though this is somewhat visually confounded in the Northern Amazon + Guiana
2001 Shield clade (D), with two weakly resolved lineages (e.g. pale yellow Jaú and green Guiana
2002 Shield population groups appear to have some overlap in this figure, but they are delimited
2003 by the Rio Negro). Clade D bipartitions are shown with dashed branches (top) to indicate
2004 low support for phylogenetic structure (Figure 3, results). The ancestral habit of the genus
2005 is inferred to be Andean (Figure 5, Supplementary Appendix), with a single origin of
2006 lowland lineages. Also shown at the tips of the simplified tree are median admixture
2007 proportions across sampling localities as inferred with STRUCTURE (Figure 5).
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017

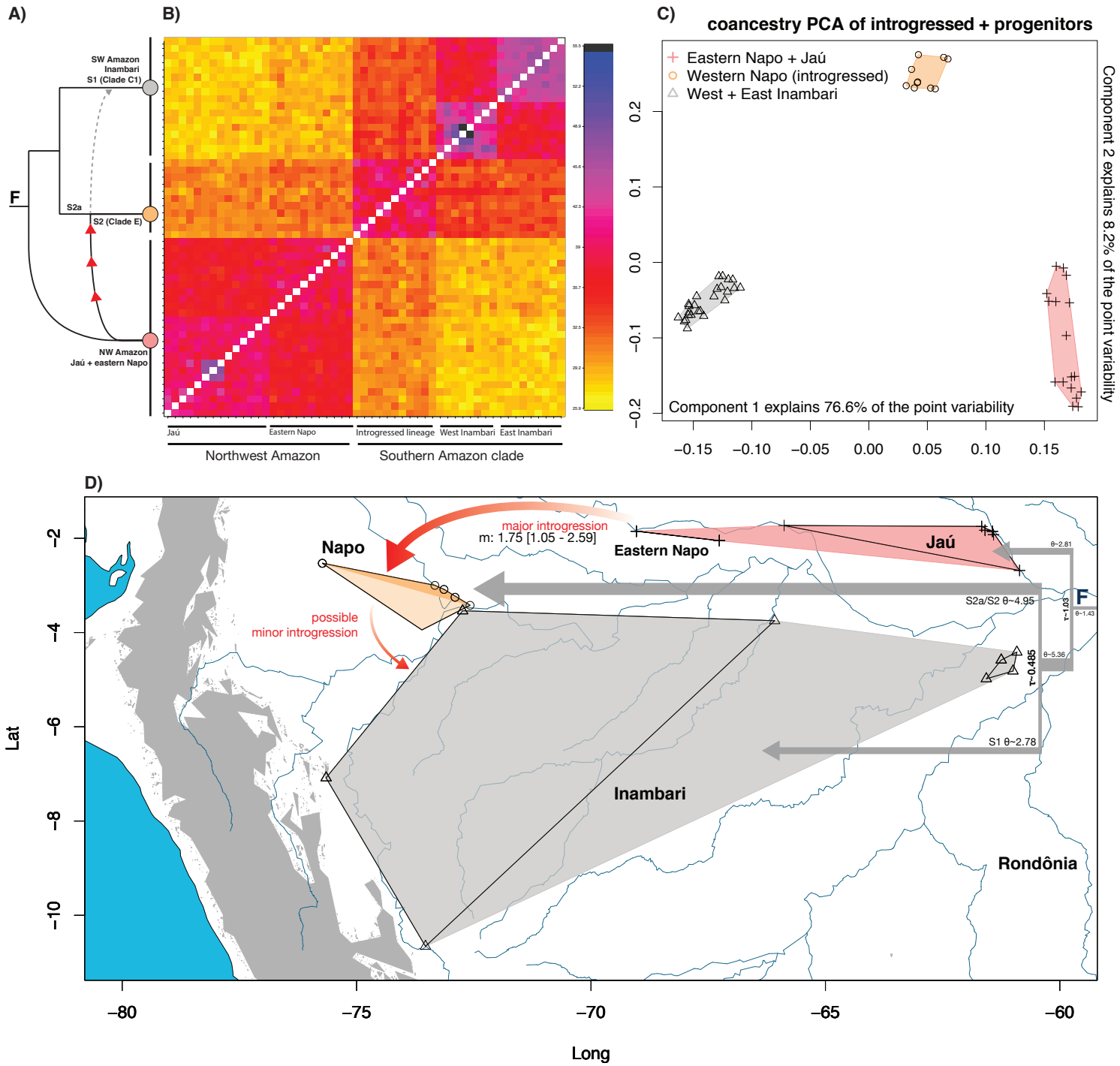

D
D2
D1
A3
A4
coracina
A1
Andean
MRCA
B2
B1
pygmaea
E
C1/2
C3
C4
C5
C6
C10
C9
C8
Lowland MRCA

2018 **Figure 5.** Cluster and admixture analysis of SNP data. Left: RAxML topology using the 50%
2019 haplotype dataset, with tip labels and colors indicating group membership to one of
2020 eighteen population-areas (matching those in Figures 3, 4). Branch colors indicate
2021 ancestral area, according to the results from SIMMAP and maximum parsimony analysis
2022 (Supplemental Appendix), and unambiguously reconstruct the *Pseudopipra* MRCA to have
2023 been Andean. Center: STRUCTURE output for Dataset 1, depicting population assignment
2024 and admixture estimates for K=5. The likelihood of each evaluated number of K clusters
2025 from 1:20 plateaued at K = 5, with the standard deviation across runs increasing rapidly
2026 after this point (Supplemental Figure 2). The five optimized clusters broadly correspond to
2027 wide biogeographic Neotropical regions which coincide with lowland areas of endemism,
2028 Central American, and Peruvian montane regions (Figure 4). Given these five clusters, each
2029 labeled focal region is inferred to have a unique combination of admixture proportions. The
2030 admixture cline inferred from the Xingu to the southern Atlantic Forest appears to be a
2031 product of isolation by distance (Supplementary Appendix), while the signature of
2032 introgression inferred for western Napo lineages appears to be a product of substantial
2033 historical introgression from a northern lineage, into a distinct Southern Amazonian
2034 lineage (see Figure 6 and discussion). See Supplementary Figure 3 for K2-10 results. Right:
2035 clustering solutions for SNP dataset D1-D4m (described in Table 1, also see Supplementary
2036 Figure 2). Black bars indicate distinct clusters which are also detected as monophyletic
2037 groups in the phylogenetic analysis. Red bars indicate one distinct cluster in a given
2038 dataset, but which is paraphyletic with respect to the phylogenetic analysis. The grey bars
2039 for D2m indicate several groups of individuals that were not resolvable in this dataset
2040 (Supplementary Figure 2).


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
weakly resolved Suriname, Amapá
Lowland Northern Amazon/Guiana Shield
Andean
Central American
Amazonian/Lowland

2041 **Figure 6.** Examination of the introgressed lineage in the western Napo area of endemism.
2042 All of the available genomic evidence is consistent with the hypothesis that individuals in
2043 the western Napo area of endemism are derived from a distinct southern Amazonian
2044 lineage (S2a in Panel A) which is not represented in contemporary samples. This lineage
2045 (S2a) experienced substantial introgression from the ancestral populations of what are
2046 today eastern Napo and Jaú restricted linages. Panel A shows the phylogenetic network (in
2047 black) inferred by PhyloNet, with arrows indicating the direction of introgression
2048 estimated in G-PhoCS. The red arrows along the PhyloNet reticulation indicate inferred
2049 introgression from the northern Amazonian lineages (clade C1 in Figure 3) into this
2050 southern Amazonian lineage (S2a). The dashed grey line connecting S2a/S2 to S1 indicates
2051 possible introgression into Inambari lineages, which may have occurred after or
2052 concurrently with the primary introgression event and brought northern (Clade D) alleles
2053 into SW Amazonian populations (see text and Figure 5). Panel B displays the individual
2054 level co-ancestry matrix for introgressed Napo individuals and their source/progenitor
2055 lineages; darker colors indicate greater co-ancestry. Introgressed (PH: putative hybrid)
2056 hybrids have greater co-ancestry with each of their progenitor lineages (S2a and Northern
2057 Amazon+Guiana shield (abbreviated GS)) than their progenitor lineages share with each
2058 other (PHvGS/progenitors = 1.078, PHvInambari/progenitors = 1.17, see text). In the
2059 proposed scenario, the southern progenitor lineage (S2a) is either unsampled or extinct
2060 and has not persisted sympatrically with S2. Thus, the western Inambari southern lineage
2061 (S1) comprises the closest living relatives to S2a (see discussion). Panel C summarizes co-
2062 ancestry variance as projected into component space after normalizing; the introgressed
2063 lineage is intermediate along PC1 and differentiated along PC2. The level of differentiation

on PC2 corresponds to about 10% of the total variation explained by PC1 and PC2 (~85%). 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 depicited in panel D (m: migrants per generation, when lower 95% CI > 1, τ: splitting time in generations ($x10^6$), θ: effective population size (10^6), Supplemental Table 5). In A-D, two subpopulations (East and West) are indicated for each progenitor group, as outlined with outlined minimum convex hulls in D.


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
Napo
Eastern Napo
Jaú
major introgression
m: 1.75 [1.05 - 2.59]
possible
minor introgression
S2a/S2 θ~4.95
S1 θ~2.78
θ~2.81
θ~1.43
θ~5.36
τ~1.03
τ~0.485
Inambari
Rondônia
Lat
Long

2087 **Figure 7.** Estimated Effective migration surface (EEMS). The estimated migration surface
2088 links the topography and drainage system of the Neotropics to spatial patterns of genetic
2089 variation. Shown are results estimated from a model with 2000 demes (vertices). Blue
2090 colors indicate areas where gene flow is more likely to be occurring (i.e., more connectivity
2091 than isolation by distance), while brown colors indicate areas where barriers to gene flow
2092 are highly probable (i.e. less connectivity than isolation by distance). In this analysis, the
2093 Andes and the Amazon River are the most clearly inferred barriers to gene flow. Major
2094 rivers, as well as a >1000m contour (light grey) are overlaid to aid interpretation. EEMS
2095 does a very good job of explaining spatial variance in this dataset (R^2 = 0.871,
2096 Supplementary Figures 10, 11). The model performs poorly in the far western Amazon,
2097 where the inference of a major barrier to gene flow (the Andes) is inferred a few degrees
2098 too far east, pushing other western patterns eastwards. This is likely a consequence of
2099 uneven genetic sampling in this area. Remarkably, EEMS almost perfectly infers the spatial
2100 extents of the lowland Amazonian range (range map outlined in black), despite very limited
2101 genetic sampling toward the edge of the range. EEMS also detects an isolated region of high
2102 connectivity, which corresponds to the introgressed western Napo hybrid lineage.
2103
2104
2105
2106
2107
2108

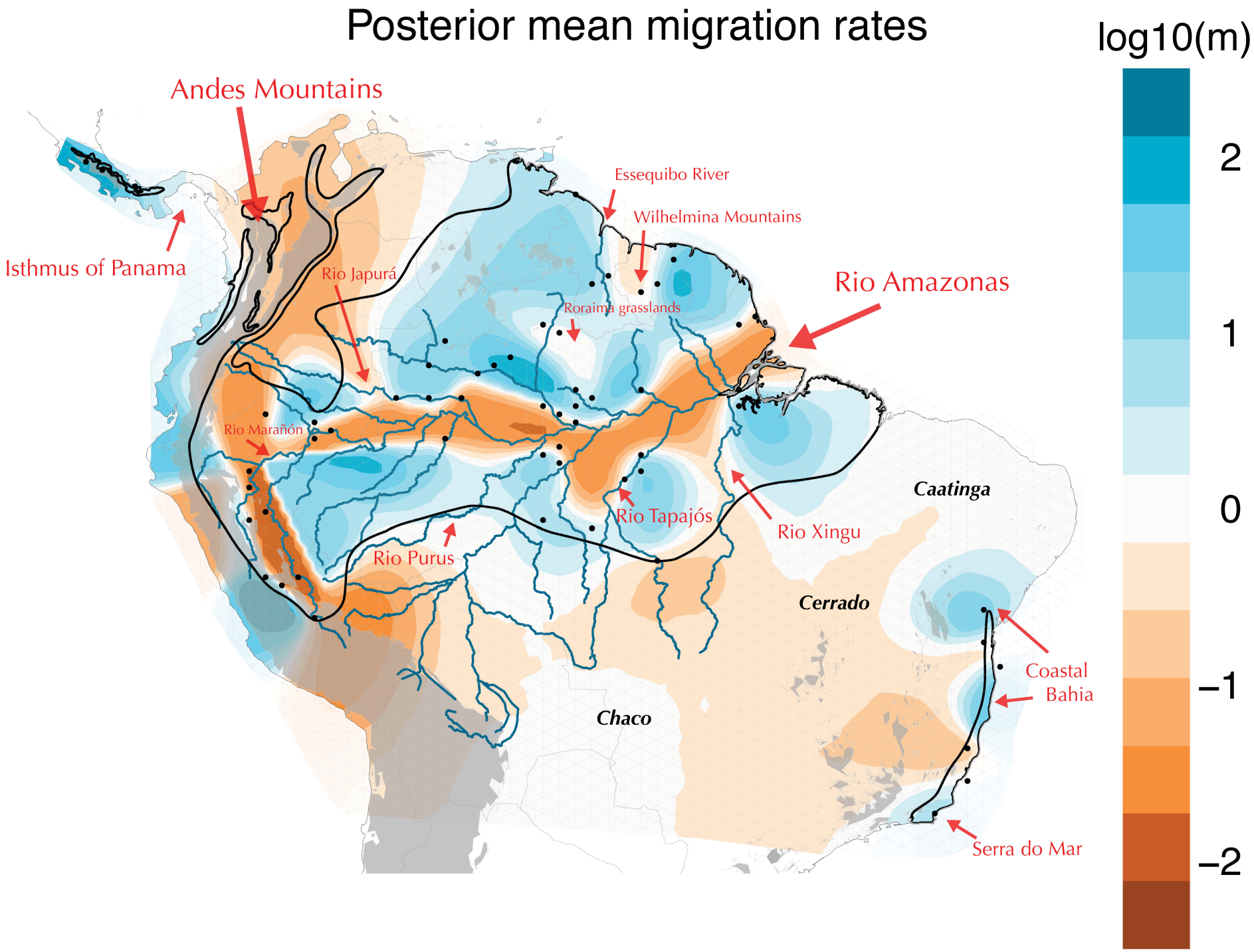
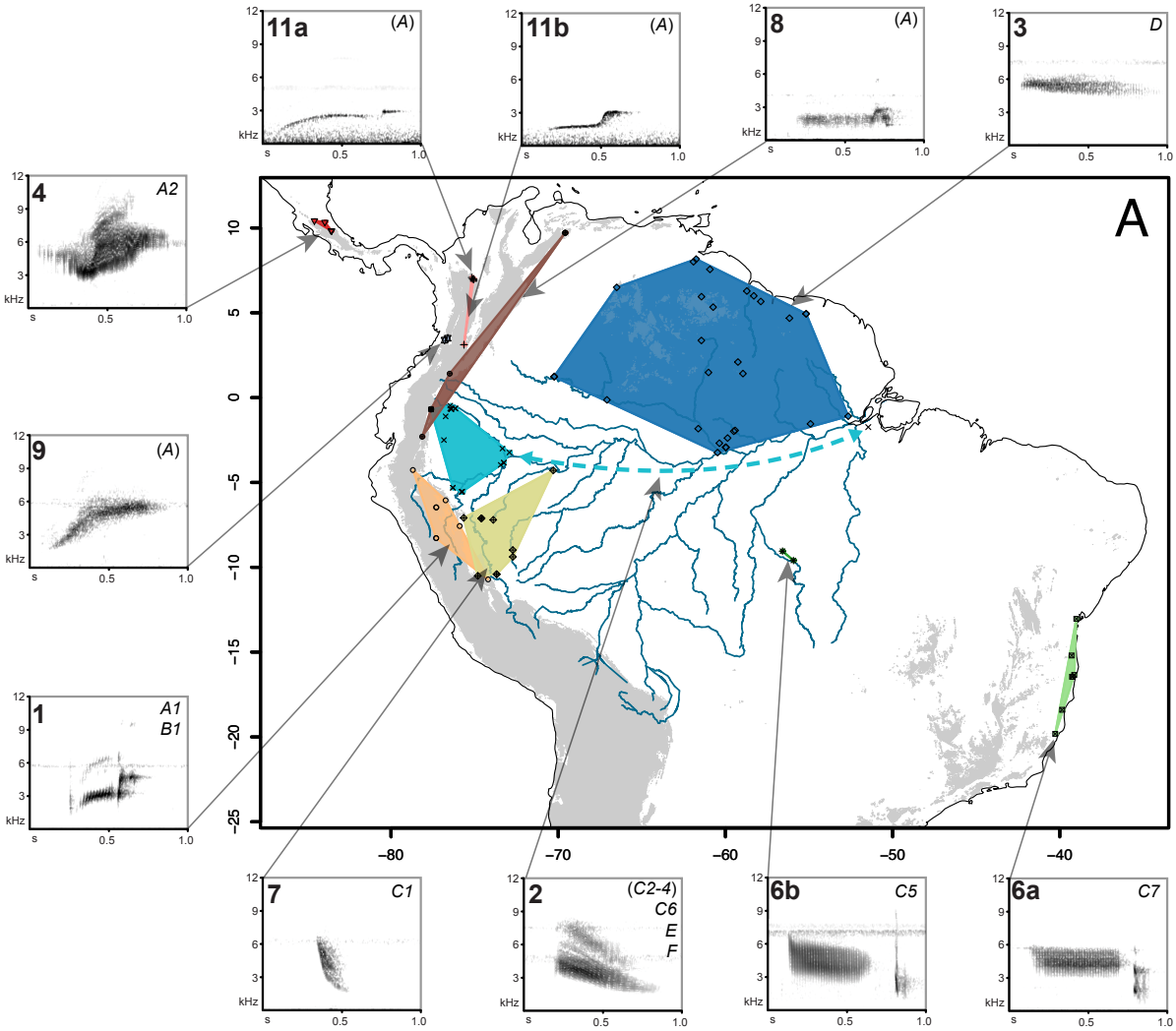
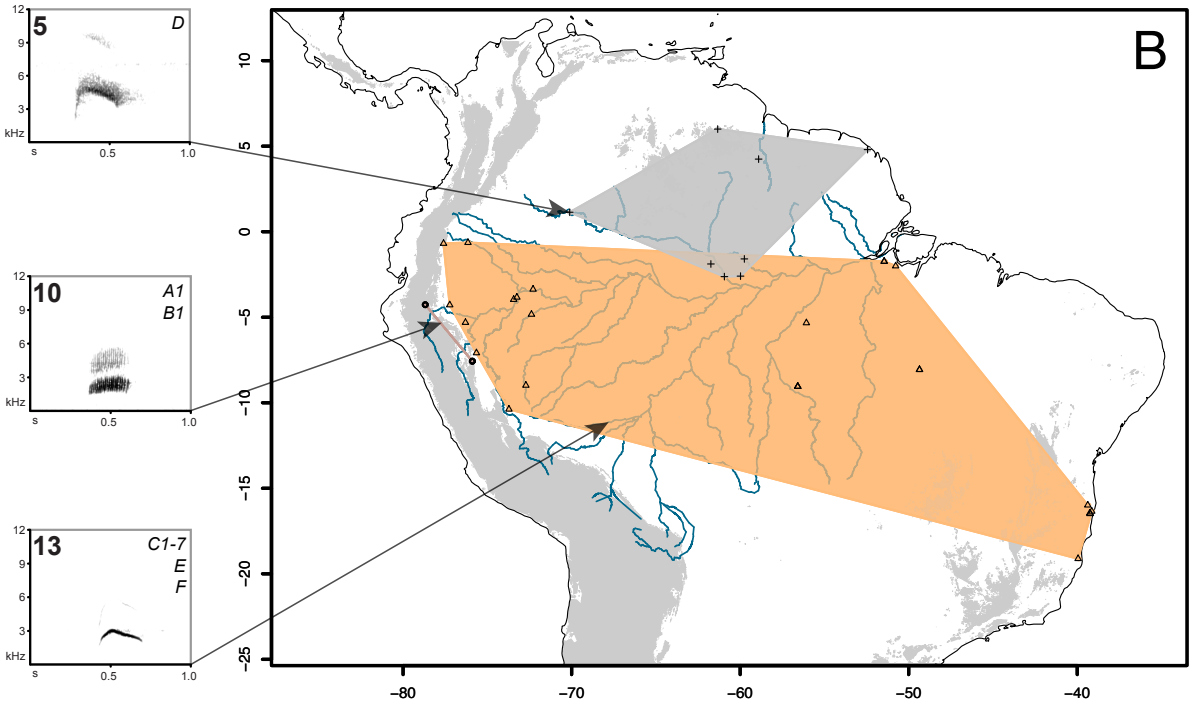

Posterior mean migration rates
log10(m)
2
1
0
−1
−2
Andes Mountains
Isthmus of Panama
Rio Japurá
Rio Marañón
Essequibo River
Wilhelmina Mountains
Roraima grasslands
Rio Amazonas
Rio Tapajós
Rio Xingu
Rio Purus
Caatinga
Cerrado
Chaco
Coastal Bahia
Serra do Mar

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

2132 **Table 1.** Summary of dataset filtering parameters. We generated eight SNP datasets under
2133 different filtering schemes to evaluate the sensitivity of cluster analyses. For eight SNP
2134 datasets, dataset D1 – D4mod, we note the parameters for each filter: the required
2135 sequencing coverage, the missing data threshold, whether or not a single SNP (single
2136 nucleotide polymorphism) or all SNPs per locus were retained, the minor allele frequency
2137 (MAF), the number of loci, the total number of SNPS represented in a given dataset, and the
2138 average number of SNPs per locus.
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153

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

Table 2. Summary of Mantel Tests. For each geographic barrier and sampling region we were able to evaluate, we report results for analyses using phylogenetic patristic distance and coancestry, under both linear and log scales. Asterisks indicate *p* values less than 0.001, which we interpret as 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**