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

1. Fuller Evolutionary Biology Program, Cornell Lab of Ornithology, 159 Sapsucker Woods Road, Ithaca, NY 14850, USA. Department of Ecology and Evolutionary Biology, Cornell University, 215 Tower Road, Ithaca, NY 14853, USA.

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

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

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

5. 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)

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 historical and current 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. Our analyses identify five ancient clades, which encompass seventeen well-differentiated populations. Most of the breakpoints among populations coincide with physical barriers to gene flow previously associated with avian areas of endemism, and generally coincide with subspecies boundaries. The phylogenetic relationships among these populations imply a unique pattern of a montane Andean origin for the genus, with a subsequent expansion and radiation into the Amazonian lowlands. Analyses of genomic admixture demonstrate a complex history of introgression between some western Amazonian populations, which confound standard concatenated and coalescent phylogenetic analyses, and raise the possibility that a lineage in the western Napo area of endemism is of hybrid origin. Lastly, we analyze variation in vocal phenotypes in the

context of our phylogeny and propose that *Pseudopipra* is a species-complex composed of 15-17 distinct species which have arisen in the last ~2.5 Ma.

Key Words: Neotropics, Amazon, Andes, Areas of Endemism, Suboscine, Speciation Genomics, Phylogeography, Lekking, hybrid speciation

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

Montane Andean regions in the Neotropics are known to be exceptionally biodiverse, and encompass at least 15 areas of endemism with biotas shaped by a combination of vicariance and dispersal events (Hazzi et al. 2018). Elevational gradients in the Andes contribute substantially to Neotropical diversification metrics and raise

fundamental questions about the historical relationships between lowland rainforest and montane endemics in the Neotropics (Weir 2006, Quintero and Jetz 2018, Musher et al. 2019). From a biogeographic perspective, 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 regions by lowland ancestors (Fjeldså 1992, Bates and Zink 1994, Ribas et al. 2007), whereas fewer examples suggest colonization of lowlands (grasslands) by Andean ancestors (da Silva 1995, Voelker 1999, van Els et al. 2019). 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). 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, Nylander et al. 2008).

By contrast, in the lowland Amazon basin, a number of areas of endemism have been described, delimited primarily by major tributaries of the Amazon River (Haffer 1974, Cracraft 1985, Da Silva et al. 2005, Borges and Da Silva 2012). 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) (Haffer 1978, Cracraft 1985, Haffer 1985, Cracraft and Prum 1988, Haffer 1992, Da Silva et al. 2002, Da Silva et al. 2005, Naka 2011, Borges and Da Silva 2012). The aggregate of these geographic

69 partitions has been recognized as the "Amazonian areas of endemism" and is the basis for
70 the riverine barrier hypothesis (Figure 1, Antonelli et al. 2018, Silva et al. 2019 Figure 1A).
71 Few studies have applied modern phylogeographic techniques leveraging next-
72 generation sequencing datasets to reconstruct the phylogenetic and population genetic
73 history of a Neotropical radiation distributed across many major Neotropical forest areas
74 of endemism. To our knowledge, there has only been one other study to apply a genomic
75 sequencing technique to study continental-scale phylogeography of an Amazonian
76 suboscine passerine bird (*Xenops minutus*; Harvey and Brumfield 2015, also see Harvey et
77 al. 2017). Here, we investigate lineage diversification across the Neotropics using the
78 continentally distributed *Pseudopipra* genus. The monotypic genus *Pseudopipra* (family
79 Pipridae) currently includes a single biological species, the White-crowned Manakin
80 (*Pseudopipra pipra* = *Dixiphia*). *Pseudopipra* is found on both sides of a variety of known
81 dispersal barriers, including elevational gradients of the Andes, major Amazonian rivers,
82 the dry diagonal, tepuis, and the Isthmus of Panama. Thus, this radiation is particularly
83 appropriate for assessing patterns of phylogenomic differentiation across a nested set of
84 spatial scales in Neotropical forests distributed in both highland and lowland areas of
85 endemism.
86 In order to assess the historical relationships among Andean and lowland
87 Amazonian populations, and the degree to which population structure corresponds to
88 landscape features, our study design uses fine-scale sampling across a majority of
89 *Pseudopipra's* range. We use double digest restriction site associated DNA (RAD)
90 sequencing (Peterson et al. 2012)—a reduced representation genomic sequencing
91 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. We use these data to infer biogeographic history using a variety of methods, including concatenated data and coalescent-based phylogenetic inference.

Our investigation led us to examine the spatial distribution of genetic diversity across the range of *Pseudopipra* and to investigate the roles of many specific barriers to gene flow on population genomic differentiation. We tested for signals of isolation by distance (IBD) within areas delimited by prominent geographic barriers and assessed the degree to which such barriers generate discontinuities from IBD predictions. Our analyses inferred a spatially explicit model that explains a majority of the observed genetic dissimilarity among sampling localities and of the overall genetic variance, implying a strong connection between landscape features and diversification at multiple scales. 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 (e.g. Zamudio et al. 2016), and to reevaluate species delimitation. Overall, these investigations highlight the significant impact of the topography and drainage systems of the Neotropics in structuring genetic diversity across *Pseudopipra*, as well as the complex influence of introgression in generating biogeographic patterns (e.g. Burbrink and Gehara 2018).

Study organism and Taxonomic summary

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 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, exhibit dispersed lek breeding behavior, 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 (Snow 1961, Castro-Astor et al. 2007). 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 (1990b), Prum (1990a), 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 (Tello et al. 2009, Ohlson et al. 2013), work exploring intraspecific genetic variation within manakins has been restricted to a relatively few

species (e.g. McDonald 2003, Cheviron et al. 2005, Cheviron et al. 2006, Francisco et al. 2007, Brumfield et al. 2008, Capurucho et al. 2013, Gubili et al. 2016, Luna et al. 2017).

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 (Ridgely and Greenfield 2001, Snow 2004, Ridgely and Tudor 2009, Kirwan and Green 2012, Spencer 2012, Freile 2014). Thirteen subspecies of *Pseudopipra pipra* have been recognized based primarily on subtle variations in plumage coloration, which are frequently more marked in females than in males (summarized in Dickinson 2003, Snow 2004, Kirwan and Green 2012). Because suboscine passerines generally do not learn their songs (but see Saranathan et al. 2007), the presence of substantial vocal variation across *Pseudopipra* populations further suggests that the genus may contain unrecognized cryptic species (e.g. Campagna et al. 2012). However, all previous authors awaited new information on genetic differentiation within *Pseudopipra* before making taxonomic recommendations.

It also remains unclear how subspecific designations reflect evolutionary history. The genetic variation within *Pseudopipra* has received prior attention by two studies that used mitochondrial DNA to infer population genetic patterns. Milá et al. (2012) reported intraspecific divergences of up to 3.5% (n=19) across three Amazonian populations, and the highest observed nucleotide diversity ($\pi = 0.266$) among 14 widely distributed Amazonian birds species. Castro-Astor (2014) used a larger sample (n=57), and discovered that at least four *Pseudopipra* subspecies correspond to well supported mitochondrial clades which were generally congruent with Amazonian areas of endemism. Castro-Astor

160 (2014) also estimated the age of the *Pseudopipra* complex to coincide with the onset of the
161 Pleistocene ~2.457 Ma (95% HPD 1.45-3.97).
162
163 **Materials & Methods**
164 *Field and tissue sampling*
165 Muscle tissue samples were obtained from available vouchered avian material from
166 US and Brazilian collections and other institutions (see Acknowledgements and
167 Supplemental Table 1). We sampled from ten of thirteen recognized subspecies (Figure 2)
168 of *Pseudopipra* (Kirwan and Green 2012). Unfortunately, we were unable to obtain material
169 of adequate quality for ddRAD sequencing from the Andes of Ecuador, Colombia, and for
170 lowland Peruvian populations representing the *coracina*, *bolivari*, *minima*, *unica*, and
171 *pygmaea* subspecies. However, we were able to obtain mtDNA data from *coracina* and
172 *pygmaea*, which allowed us to make a preliminary assessment of their phylogenetic
173 affinities (see discussion on mtDNA). In total, after discarding failed samples and samples
174 with high proportions of missing data, we obtained new genetic data from 277 individuals
175 (232 from ddRAD, 168 from mtDNA), representing ~80 localities and 10 subspecies. We
176 also obtained comparable ddRAD data for two specimens of *Ceratopipra rubrocapilla* as
177 outgroups. See Supplementary Table 1 for details.
178
179 *Laboratory methods*
180 We extracted DNA from avian tissue specimens (Table S1) using the DNeasy Blood
181 & Tissue Kit (QIAGEN, CA) and generated ddRADtags following the protocol of Peterson et
182 al. (2012), with modifications described in Thrasher et al. (2018). Briefly, we digested

183 equal quantities of genomic DNA from each sample in individual reactions with two
184 restriction enzymes, SbfI and MspI (New England Biolabs, MA), and ligated adapters on
185 both ends. The 5' adapters contained one of 20 unique 5-7 bp barcodes, whereas the 3'
186 adapter was common to all samples. We pooled groups of samples with unique 5' barcodes
187 and subsequently size selected DNA fragments that were between 400-700 bp using a Blue
188 Pippin (Sage Science, MA). For each group of size-selected samples, we incorporated
189 unique Illumina TruSeq adapters by performing 11 cycles of PCR. The combination of 5'
190 barcodes and TruSeq adapters was unique to each sample. A total of 13 groups of pooled
191 samples with different Illumina TruSeq adapters were combined in equimolar proportions
192 into two libraries. We sequenced both libraries on four lanes of Illumina HiSeq 2500 at the
193 Cornell University Institute for Biotechnology, obtaining single-end 151 bp sequences. We
194 obtained a representative sample of mitochondrial ND2 sequences using standard Sanger
195 Sequencing protocols described in Berv and Prum (2014).
196
197 *Assembly of sequencing reads into RAD loci*
198 We obtained 497 million raw 151 bp reads (~75 Gb) for 241 individuals. We first
199 assessed the overall read quality with FastQC (Andrews 2010) and trimmed lower quality
200 bases at the 3' end with FASTX-Toolkit (Gordon and Hannon 2010). The trimmed
201 sequences were 145 bp in length. Using FASTX-Toolkit we filtered-out lower quality reads
202 if they had a single base below a Phred quality score of 10 and/or more than 5% of bases
203 with quality between Phred 10 and 20.
204 The quality filtered reads were demultiplexed with the 'process_radtags' program
205 from the STACKS 1.44 bioinformatics pipeline (Catchen et al. 2013). During the

206 demultiplexing step we also discarded reads that had not passed the Illumina filter, had
207 adapter contamination, lacked barcodes used for multiplexing, or did not contain an SbfI
208 cut site. This step removed the inline barcodes and trimmed all reads to an equal length,
209 the length of the reads that contained a 7 bp barcode. After demultiplexing and filtering, we
210 retained an average of 1.1 ± 0.4 million 138 bp sequences per individual.

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

222 We used several filters in the "populations" module from STACKS to export different
223 sets of bi-allelic SNPs: a missing data filter, a minimum depth of coverage filter, a filter that
224 exports only one SNP per RAD locus, and a minor allele frequency filter (MAF). By
225 combining these different filters, we produced three versions of the dataset which varied in
226 important parameters that could potentially bias downstream population genetic inference
227 (Linck and Battey 2017, Paris et al. 2017). Dataset 1 (labeled arbitrarily) included 1,960
228 SNPs that were generated by setting the MAF to >= 5%, requiring minimum coverage of

229 20x, requiring the presence of a SNP in at least 80% of individuals, and filtering out all but
230 the first SNP in each RAD locus to minimize linkage across the dataset (Paris et al. 2017).
231 To evaluate the sensitivity of our analyses to rare variants (Linck and Battey 2017, Shafer
232 et al. 2017), dataset 2 was exported with identical parameters to dataset 1, but without a
233 MAF; this generated a dataset of 2,581 SNPs. Lastly, for non-model-based cluster analyses
234 which should be insensitive to linkage, a third dataset was exported which included all
235 SNPs from each locus (dataset 3, 5,099 SNPs). We generated several additional datasets for
236 specific analytical cases when required and describe them later as necessary.
237
238 *Phylogenetic analysis*
239 Because sparse super-matrices may be more prone to systematic error (e.g. Roure et
240 al. 2013), we explored the sensitivity of our phylogenetic inferences to increasing amounts
241 of missing data using three different datasets with 20%, 50%, and 80% thresholds for
242 missing data at the RAD locus level (i.e., for the 80% threshold, if the RAD locus is present
243 in at least 20% of individuals, then that locus, containing all variable and invariable sites,
244 was exported). This approach generated datasets comprising 2,584, 4,763, and 7,901 RAD
245 loci, respectively. For each dataset, we obtained complete haplotype data for every
246 locus/individual and generated a concatenated supermatrix using Sequence Matrix (Vaidya
247 et al. 2011). These matrices had up to ~40% missing sites across the entire matrix,
248 irrespective of columns or rows (i.e., using the 80% threshold above).
249 We estimated phylogenies using each dataset with a concatenated maximum
250 likelihood and a coalescent species-tree approach for a total of six phylogenetic analyses.
251 For analysis of concatenated data, we estimated a maximum likelihood phylogenetic

252 hypothesis using RAxML 8.2.9 (Stamatakis 2014). We estimated 500 bootstrap replicates
253 followed by a full ML search for each dataset, using a GTRGAMMAX substitution model. We
254 avoid the need for ascertainment bias correction by using short-read haplotype alignments
255 including invariant sites (Leaché et al. 2015). We also tested for an effect of data
256 partitioning on phylogenetic inference by generating optimized locus partitioning schemes
257 in PartitionFinder 2 (Lanfear et al. 2017), which were then re-analyzed in RAxML. We
258 found that our topological results were insensitive to data partitioning, and as such we do
259 not present these results.

260 We evaluated coalescent phylogenetic structure using the SVDquartets method
261 implemented in PAUP* 4.0a159 (Swofford 2002, Chifman and Kubatko 2014, Chifman and
262 Kubatko 2015). SVDquartets may be a useful tool for testing hypotheses of species
263 delimitation because it can be used to generate a lineage-tree under the multispecies
264 coalescent without assigning individuals to species *a priori*. In this approach, each
265 individual in a dataset is considered to be a separate 'species.' If a clade in such a lineage-
266 tree has high bootstrap support, this would mean that the descendant taxa have strong
267 support as a group under the multispecies coalescent, and may coincide with species
268 boundaries (*Personal communication*, Laura Kubatko). Our SVDquartets analysis evaluated
269 all possible quartets with 500 bootstrap replicates (other settings left to defaults).
270 SVDQuartets has the advantage of working directly on sequence data (above) and has been
271 shown to perform well in comparison to other coalescent models which assume fully and
272 correctly resolved gene trees as input (Chou et al. 2015, Schmidt-Lebuhn et al. 2017).
273 SVDquartets has also been shown to produce reliable results under very general
274 conditions, and is theoretically robust to variation in effective population size, molecular

clocks, limited gene flow, and incomplete lineage sorting (Long and Kubatko 2017). Finally, SVDquartets is very computationally efficient compared to other multispecies coalescent models.

While a full statistical exploration of our specific usage of SVDquartets is beyond the scope of the present work, we suggest that this approach may be a useful and efficient 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 our coalescent analysis necessarily represent biological species – such an interpretation would likely lead to an over-inflation of species estimates (Sukumaran and Knowles 2017, Leaché et al. 2018, Chambers and Hillis 2019). In general however, we prefer such a species agnostic approach to those which require *a priori* delimitation of species boundaries (e.g SNAPP, Bouckaert et al. 2014), as it enables us to discover well supported coalescent structure without an *a priori* hypothesis. Lastly, we generated a mitochondrial DNA ND2 gene tree using IQ-TREE 1.6.10 (Schmidt et al. 2014, Chernomor et al. 2016, Trifinopoulos et al. 2016, Hoang et al. 2017, Kalyaanamoorthy et al. 2017). We partitioned by codon position and generated a maximum likelihood tree using the MFP+MERGE model search and partitioning option (Supplemental Appendix).

Descriptive statistics

Estimates of regional variation in allelic richness (Goudet 2005), Hardy-Weinberg equilibrium (Paradis 2010), inbreeding (Goudet 2005), linkage disequilibrium (Kamvar et al. 2014), and an analysis of molecular variance across multiple strata (Kamvar et al. 2014), are reported as supplementary material.

298

299 *Cluster inference and patterns of genomic admixture*

300 We performed exploratory population genetic analyses intended to quantify the
301 number of similar genetic clusters across the range of *Pseudopipra.* First, we evaluated
302 broad-scale variation across our genomic data using phenetic K-means clustering
303 implemented in the find.clusters function in the adegenet R package (Jombart 2008). This
304 function first transforms the data with a PCA, and then identifies clusters with a phenetic K-
305 means algorithm. Using datasets 1-3, we evaluated K1:20 successively by the Bayesian
306 Information Criterion (BIC). K-means clustering is based on minimizing the variance within
307 clusters, and has the potential advantage of making few prior assumptions about the
308 underlying processes which generated the data (MacQueen 1967). For each analysis, we
309 retained all PC axes, set 200 randomly chosen starting centroids, and ran each search for
310 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 dataset 1, from
317 K1:20 with 10 iterations per K value. We ran the program for 700,000 generations,
318 discarding the first 200,000 as burn-in, implementing the admixture ancestry model with
319 correlated allele frequencies. We evaluated K by examining $\Pr(X|K)$ or $L(K)$ after
320 summarizing our results using Structure Harvester (Earl and vonHoldt 2012) and CLUMPP

321 (Jakobsson and Rosenberg 2007). We also used the Evanno method to estimate the rate of
322 change in the log probability of data between values of K, which has been suggested to be a
323 useful metric to detect the uppermost hierarchical level of structure in the data (Evanno et
324 al. 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 contained 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 (Lawson et al. 2012), but are optimized to take advantage
345 of the linkage properties of RAD-like datasets. As our data were assembled relative to a
346 *Manacus* reference genome, the order in which loci appear in our data files are related to
347 positions on assembly scaffolds, even though the ddRAD loci in our data are generally
348 unlinked when considered at the population level (see Supplementary Appendix). We
349 compared the inferred population assignments from fineRADstructure and STRUCTURE
350 with K-means clustering (as above) of the co-ancestry matrix. To determine how much of
351 the variation in our dataset was captured by the RADpainter analysis relative to a standard
352 PCoA on SNP data, we used the normalized PCA approach of (Lawson et al. 2012), using the
353 mypca function provided in 'FinestructureLibrary.R' Library in the fineRADstructure
354 package.
355
356 *Phylogenetic reticulation in the western Napo lineage*
357 Model-based, cluster, and population genetic analyses revealed that a distinct
358 lineage of individuals 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). To investigate this pattern, we followed the approach of Barrera-Guzmán et al.
361 (2018) to quantify the degree to which the inferred co-ancestry of this Napo population
362 reflected the genomic composition of each putative progenitor lineage which contributed
363 genotypes to contemporary individuals.
364 For these analyses, we exported a separate haplotype dataset of 52 individuals
365 which included all potentially introgressed western Napo individuals (n=10) and their

366 most geographically proximate populations. We used a 5% minor allele frequency cutoff,
367 required 20x minimum coverage, and allowed up to 20% missing data per locus (2,370 loci,
368 including 4,979 SNPs). Potential progenitor populations consisted of 1) the most
369 geographically proximate population inferred to have mostly unadmixed ancestry
370 (Southwestern Amazon: Inambari, n=17), and 2) the primary inferred source of
371 introgressed genotypes (as initially implied by STRUCTURE, north-central Amazonian Jaú +
372 eastern Napo, n=25). This dataset was also analyzed separately with RADpainter as
373 described above.

374 Heavily introgressed or hybrid populations are expected to have higher mean co-
375 ancestry and lower F_{st} with each of their progenitor lineages than their progenitor lineages
376 will have with each other (Barrera-Guzmán et al. 2018). For co-ancestry comparisons, we
377 used a standard ANOVA and the glht function in the multcomp R package (Hothorn et al.
378 2013) to test these predictions. We estimated population differentiation due to genetic
379 structure (F_{st}) with Weir and Cockerham's (Weir and Cockerham 1984) F_{st}, and tested the
380 above prediction using 1000 bootstrapped datasets to estimate 95% confidence intervals.

381 To test whether or not the relationships among these three focal populations may
382 be best represented by a phylogenetic reticulation, we estimated a phylogenetic network
383 using PhyloNet 3.6.4 (Wen et al. 2018, Zhu et al. 2018). PhyloNet takes SNP data and uses a
384 reversible-jump MCMC technique to explore the posterior distribution of phylogenetic
385 networks and bi-furcating topologies, while accommodating both reticulation and
386 incomplete lineage sorting. This approach searches the set of all possible reticulation
387 models without requiring *a priori* model specification. Starting with dataset 1 (1960
388 unlinked SNPs), we pruned the dataset to the 52 individuals that comprise the three groups

389 of primary interest (above). We then pruned out all sites with missing data, leaving 572
390 biallelic SNPs (PhyloNet cannot presently consider sites with missing data). For our final
391 runs of the PhyloNet program MCMC_BiMarkers, we randomly subsampled five diploid
392 individuals from each group, as preliminary program runs indicated a computational
393 bottleneck with more than five individuals. We assigned each group of five individuals to
394 represent a 'species,' allowed for a maximum of one reticulation, set the chain length for
395 100000 iterations with sampling every 100, set a burn in of 20%, and as our interest is
396 primarily concerned with topology, left all priors as default (i.e. assuming population
397 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 exported
400 haplotype data for the focal samples without a minor allele frequency filter, generating an
401 alignment of 2,947 variant and invariant loci. We ran the program for 750,000 iterations
402 with a 10% burn-in, estimating a total of 13 parameters: the effective population sizes of
403 the three focal lineages, two ancestral population sizes, two splitting times, and six
404 directional migration parameters. Parameter MCMC traces were inspected in Tracer 1.7.1
405 (Rambaut et al. 2018). We converted the median and 95% Bayesian credible intervals from
406 mutation scale to generations and individuals as described in Campagna et al. (2015),
407 assuming an approximate mutation rate of 10^{-9} per base pair per generation (Kumar and
408 Subramanian 2002, Smeds et al. 2016). We calculated the number of migrants per
409 generation as the product of the per generation migration rate multiplied by ¼θ for the
410 receiving population ($m_{a\text{-}b} * (\theta_b/4)$) (Gronau et al. 2011).
411

412 *Spatial distribution of genetic variation*

413 In practice, dispersal barriers often create sharp genetic discontinuities between
414 adjacent populations that would otherwise be continuous (Petkova et al. 2015). To assess
415 the degree to which spatial variation in genetic diversity within *Pseudopipra* can be
416 attributed to landscape features, we investigated how patterns of isolation-by-distance
417 (IBD) vary across the landscape. First, we used Mantel and partial Mantel tests to
418 investigate IBD effects within sampling regions delimited by prominent physical barriers,
419 and then separately tested the roles of specific dispersal barriers in structuring genetic
420 variation. Mantel tests evaluate the correlation among two or more matrices, in our case
421 representing pairwise genetic and geographic distance. Significance is assessed by
422 permuting the rows and columns of one of these matrices (Mantel 1967). Partial Mantel
423 tests incorporate a third 'barrier' matrix, which contains information about environmental
424 or ecological distance. By evaluating the correlation between genetic and ecological
425 distance while controlling for geographic distance, a partial Mantel test can be used to
426 investigate the effect of a particular geographic barrier on genetic differentiation. We
427 follow the recommendations of Diniz-Filho et al. (2013), and only reject the null hypothesis
428 of no correlation if $p << 0.05$ (we use a very conservative 0.001 cutoff, ~10x more stringent
429 than a standard correction for multiple tests would require)(also see Legendre and Fortin
430 2010, Guillot and Rousset 2013).

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. These geographic distances were

435 paired with a matrix of pairwise genetic distances estimated in RAxML. We use
436 phylogenetic distance among individuals, rather than F_{st}, because our data is highly
437 hierarchically structured (though exploratory analyses based on F_{st} and other related
438 metrics were qualitatively similar). These procedures, including the non-intuitive
439 generation of an appropriate barrier matrix, are developed in a set of R functions we
440 provide as supplementary material which operate on adegenet GENIND objects (see
441 supplementary R code) and internally use the Mantel implementation from the ecodist R
442 package (Goslee and Urban 2007).

443 We also analyzed patterns of spatial variation in our data with a more sophisticated
444 population genetic model which relates effective migration rates across geographic space
445 to 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'. We view this method as complimentary to more
449 explicit hypothesis-testing approaches (such as Mantel tests), as it seeks to identify where
450 local IBD predictions implied by the data are strongly violated in geographic space, and
451 thereby can identify strong barriers to gene flow *de novo*. 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
454 genetic dissimilarity between two individuals, integrating over all possible migration
455 histories, and adjusting migration rates among graph edges to match the genetic
456 differences in the data. A migration surface is then interpolated across a region to indicate
457 where genetic similarities decay faster than what a pure 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.

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 distinct vocalization types based on our own survey of the available recordings, without regard to the geographic location of the recordings. Additional analytical details are reported as supplementary material.

Results

Genetic data collection

480 Our total genomic data set included an average of 1 ± 0.4 million sequences of 138
481 bp from 232 individuals of *Pseudopipra* and two *Ceratopipra rubrocapilla,* which were used
482 as an outgroup. Our reference-based assembly produced a catalogue with 47,046 RAD loci.
483 Filtered datasets ranged from 1,960 unlinked SNPs (dataset 1, one SNP per locus with 5%
484 MAF and up to 20% missing data filter), to 5,099 SNPs (dataset 3, all SNPs in all loci with
485 5% MAF and up to 80% missing data filter). Dataset 2 comprised the 1,960 SNPs from
486 dataset 1, plus 621 rare variants filtered out with a 5% MAF in dataset 1 (2,581 loci), and
487 up to 20% missing data filter, as described in the methods. We also obtained 168 mtDNA
488 ND2 sequences using PCR methods which broadly overlapped with our ddRAD dataset. Our
489 mtDNA dataset included two subspecies, *pygmaea* and *coracina*, from which we were not
490 able to obtain enough high-quality material for ddRAD sequencing.
491
492 *Phylogenetic analyses*
493 RAxML and SVDquartets analyses generated a well-supported phylogenetic
494 hypothesis (Figure 3). The monophyletic lineages of *Pseudopipra* identified have
495 distributions that are largely congruent with previously recognized avian areas of
496 endemism, and further reveal much finer scale patterns of local biogeographic
497 differentiation in certain areas. Using the two *C. rubrocapilla* individuals to root the
498 phylogeny, we detected support for at least five major clades across all phylogenetic
499 analyses of nuclear genomic data, with consistently high bootstrap support in both
500 concatenated and coalescent analysis (A-E in Figure 3, 5). We first describe the structure of
501 these major clades, followed by descriptions of well supported substructure within each of
502 these regional groupings. We report six bootstrap values for each clade with at least some

503 high support, from each of the three datasets analyzed in RAxML and SVDquartets (see
504 heatmaps, Figure 3). Figure 4 projects an idealized topology into geographic space, to show
505 how inferred population relationships are related to landscape features.

506 *Phylogenetic structure.*—The sister group to all the other *Pseudopipra* is a clade
507 including individuals from subtropical Central America and the northern Andes. The
508 northern Andes are represented here by a single individual from subtropical forests
509 (~1,000 m) of the eastern slopes of the Andes in San Martín, Peru (Clade A in Figure 3).
510 This subtropical San Martín specimen (identified as *P. pipra occulta*) was the only viable
511 sample in our genomic dataset from the entire northern Andes region (A1 in Figure 3).
512 Mitochondrial data from three additional specimens from the Andes of southern Ecuador
513 (subspecies *coracina*) cluster with the San Martín specimen (See Discussion). Thus, this
514 northern Andean sister lineage to the Central American lineage likely includes all montane
515 Andean populations of *Pseudopipra* in Ecuador, Colombia, and Venezuela. A well supported
516 Central American clade (A2 in Figure 3) was identified in all analyses, with Costa Rican (A3
517 in Figure 3) and Panamanian subclades (A4 in Figure 3) supported in RAxML.

518 The second successively nested clade in *Pseudopipra* includes individuals from
519 subtropical forests of the Andes of Central Peru, which are sister to all remaining lineages
520 (Clade B in Figure 3). The Central Peru clade is further subdivided into two very well
521 resolved clades: a northern clade (B2 in Figure 3), sampled here from the Cerro Azul in
522 southwestern Loreto, and a southern clade from Pasco, Junín, and Cusco (labeled B1 in
523 Figure 3). The Central Peru clade is the sister group to a monophyletic, lowland clade (F in
524 Figure 3) that is further subdivided into a southern Amazon clade + Atlantic Forest clade
525 (Clade C in Figure 3), and a Northern Amazonian + Guiana Shield clade (Clade D in Figure

3). In contrast, the analysis of mtDNA with better sampling of subspecies *pygmaea*, from the tropical forest of lower Rio Huallaga Valley, Peru, places this population as sister group to the entire lowland radiation (Clade F; Supplemental Appendix).

Within the broad lowland Northern Amazonian + Guiana Shield clade (Clade D), virtually no phylogenetic substructure is associated with strong bootstrap support. The maximum likelihood tree however does have substantial biogeographic coherence, which we use as additional justification for subdivision for descriptive statistics (Supplemental Appendix) when 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). We treat these geographic groups, as well as two other Guianan regions, as separate populations for descriptive population genetic analyses (Supplementary Appendix).

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). In contrast, however, population genetic analyses (below) indicated that this population is a distinct lineage of the southern Amazonian clade that has experienced substantial hybrid introgression from a northern Amazonian population (see Population Genetic analyses and Discussion).

Within the large lowland Southern Amazonia clade (Clade C in Figure 3), we recovered eight moderately supported and hierarchically nested clades, which are successive sister groups to each other, and are congruent in distribution with recognized

areas of endemism. First is a moderately supported clade (C1 in Figure 3) of individuals from the western Inambari area of endemism, from west of the Rio Ucayali to the Rio Purus. Although the basal lineages within this clade have low bootstrap support, a monophyletic subgroup of three montane (>1000m) individuals from the highlands between the Rio Huallaga and Rio Ucayali cluster with high bootstrap support (C2 in Figure 3).

The next successively nested clade includes individuals from eastern Inambari, between the Rio Purus and the Rio Madeira (C3 in Figure 3). The next successive clade includes two individuals sampled from the Rondônia area of endemism, between the Rios Madeira and Tapajos (C4 in Figure 3). The next successive clade includes a group of individuals from the Tapajós area of endemism between the Rio Tapajos and the Rio Xingu (C5 in Figure 3). Next, a clade comprised of individuals from the northern Xingu area of endemism, east of the Rio Xingu (C6 in Figure 3), is sister to a well-supported Brazilian Atlantic Forest clade (C7 in Figure 3), which is further subdivided into three successive clades (north to south) in Bahia, Espírito Santo, and Rio de Janeiro, respectively (C8, C9, and C10 in Figure 3). In summary, the subdivisions of the Southern Amazonian clade reflect a stepwise west-east progression, with a hierarchically nested structure:

(West Inambari (East Inambari (Rondônia (Tapajós (Xingu (Bahia (Espírito Santo (Rio de Janeiro))))))).

Cluster inference and patterns of genomic admixture

Broad-scale population structure and admixture.—Overall F_{st} was moderate with respect to *Pseudopipra*: 0.196 [95% CI: 0.188-0.204]. Pairwise population estimates of Weir and Cockerham's F_{st} ranged from essentially undifferentiated, to almost entirely distinct. At

572 the extremes: comparing the geographically proximate eastern Napo and Jaú populations
573 (both weakly resolved in phylogenetic analyses, but likely sister) – F_{st}: 0.0045 [95%CI:
574 0.0024 - 0.0067]. By contrast, comparing the Atlantic Forest Espírito Santo population to
575 Panama indicates an F_{st} of 0.81 [95%CI: 0.79 - 0.83], or almost entirely differentiated. See
576 Supplementary Figure 10 for a full pairwise F_{st} matrix. Hartl and Clark (1997) indicate F_{st} >
577 0.25 may be considered very great genetic differentiation.

578 K-means clustering of SNP data identified five clusters of genetically similar
579 individuals as having the lowest BIC score, with membership of one sample varying slightly
580 across datasets: Central America (Clade A2), South Andean Peru (Clade B), Atlantic Forest
581 (Clade C7), Southern Amazon including the western Napo population (the rest of Clade C +
582 Clade E), all Guiana Shield + Northern Amazon (Clade D) (Supplementary Figure 2). These
583 clusters are essentially identical to those detected by STRUCTURE at K=5 (Figure 4, 5,
584 below). In the most restricted dataset (#1, 1960 SNPs; 0.05 MAF), the montane sample
585 from San Martín (North Andean Peru, 5444.PE.MAR) clustered with Central American
586 populations (Clade A), as it does in all phylogenetic analyses. Standard PCoA explained 13-
587 17% of the variance in the SNP data on the first two axes (Supplementary Figure 2).

588 Visual inspection of $\Pr(X|K)$ from summarized STRUCTURE runs indicated that the
589 likelihood of each successive K from 1:20 plateaued at K = 5, with the standard deviation
590 across runs increasing rapidly after this point (Supplementary Figure 3). We describe
591 patterns of inferred admixture as they appear in Figure 5, reading from top to bottom. For
592 K5, individuals from Central America are unambiguously assigned to their own cluster
593 (Clade A1 in Figure 3, 5). The north Andean montane specimen from San Martín, (North
594 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 very weak 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 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 individuals from the Atlantic Forest (Clade C7) is first inferred in the Tapajos (Clade C5) area of Endemism. This admixture 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 (e.g. Bradburd et al. 2017)(See Supplementary Appendix), and not isolated introgression events.

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 these individuals (n=10) with 40.2% probability to the Guiana Shield population (SD 4.7%, average Bayesian credible interval (aBCI): [0.33-0.47]), 50.4% to the Southern Amazon population (SD 3.4%, aBCI: [0.43-0.58]), as well as limited 4.15% probability to Central Peru (SD 3.1%, aBCI: [0.01-0.08]), and 5.2% to Central America (SD 4.8%, aBCI: [0.02-0.1]). The three localities included in this clade– (1) north of Iquitos, on the east bank of the

618 confluence of the Rio Solimões and the Rio Napo, (2) southwest of Iquitos in Allpahuayo
619 Mishana Reserve (represented by three mtDNA samples), and (3) near San Jacinto,
620 Ecuador– indicate that this heavily admixed population (corresponding to subspecies
621 *discolor*) may occur across a large area of the western Napo region of endemism, and likely
622 does not represent a narrow hybrid zone (Figures 2-7).

623 Lastly, at K=5, all members of the Guiana Shield + Northern Amazon clade (Clade D)
624 are inferred to be virtually undifferentiated. Individuals in the far eastern Napo area of
625 endemism on the east bank of the Rio Putumayo (brown dots in Figure 3, 4) cluster with
626 other Guiana Shield groups in phylogenetic analyses but are detected to have limited
627 (~5%) admixture with pure individuals of Southern Amazon provenance. At higher values
628 of K, the broad-scale population assignments inferred at K=5 are mostly unchanged,
629 however additional admixture components are inferred for most groups. The introgressed
630 western Napo clade is eventually placed into its own cluster at K=9-10 (see Supplementary
631 Figure 4, Supplementary Figure 8).

632 *Fine-scale population structure and admixture.*—Admixture analysis using the RAD
633 locus optimized chromosome painting approach in RADpainter and then population
634 assignment with fineRADstructure (Lawson et al. 2012, Malinsky et al. 2018), was
635 congruent with other analyses, and also detected finer partitions of population structure.
636 The fineRADstructure program detected 15 groups of individuals which were each
637 supported with a posterior probability of 100% and which also overlapped with sampling
638 regions which we delimited for population genetic analyses based on clear geographic
639 barriers (Black dots in Figure 3, Supplementary Figure 5): These groups include: 1) Central
640 America (Costa Rica + Panama; Clade A2 in Figure 3), 2) North Andean Peru (grouped with

641 Central America; Clade A), 3) South Andean Peru (north clade; Clade B2), 4) South Andean
642 Peru (south clade: Clade B1), 5) Introgressed western Napo lineage (Clade E), 6) eastern
643 Guiana Shield (Suriname + Amapá; Clade D1), 7) eastern Napo, Jaú and western Guiana
644 (Clade D2+ other members of Clade D), 8) Eastern Inambari (Clade C3), 9) Western
645 Inambari (Clade C1), 10) Rondônia (Clade C4), 11) Tapajós (Clade C5), 12) Xingu (Clade
646 C6), 13) Atlantic Forest – Bahia (Clade C8), 14) Atlantic Forest - Espírito Santo (Clade C9),
647 15) Atlantic Forest – Rio de Janeiro (Clade C10). The only sampling regions which
648 fineRADstructure failed to strongly partition into separate groups consisted of a group of
649 individuals in the Jaú area of endemism, and a group of individuals in the eastern Napo area
650 of endemism (both of which are weakly resolved in phylogenetic analyses, but
651 unambiguously members of the broader Northern Amazon + Guiana Shield clade D). The
652 eastern Napo individuals are identified as a low support group (*pp* = 0.49), while the Jaú
653 individuals are lumped in with other western Guiana Shield individuals. See Supplementary
654 Figure 5 (fineRADstructure dendrogram), and the full dataset co-ancestry matrix plot
655 Supplementary Figures 6-7.

656 Patterns of inferred co-ancestry show a complex mosaic of shared ancestry among
657 populations (Supplementary Figures 6-7). In general, however, these patterns are
658 concordant with higher level patterns detected with STRUCTURE. Notably, the high level of
659 admixture initially inferred for our San Martín sample is recapitulated in RADpainter
660 analysis – this sample has relatively high co-ancestry across most populations. Lastly, the
661 Lawson et al. (2012) 'normalized PCA' approach provided with fineRADstructure captured
662 89% of the variance in the genetic data on the first four component axes (axis 1: 51.4%,
663 axis 2, 24.9%, axis 3: 7.77%, axis 4: 4.94%) indicating that the co-ancestry matrix reflects

664 substantially more information than standard PCoA/PCA of SNP data (above)
665 (Supplementary Figure 8).
666
667 *Phylogenetic reticulation in the western Napo lineage*
668 *Assessing hybrid introgression.*—Separate RADpainter and PhyloNet analysis of the
669 populations in the western Napo area of endemism and their putative progenitor lineages
670 supported the hypothesis that this region has a complex history of introgression. One-
671 tailed tests of mean co-ancestry indicated the null hypothesis that the difference in means
672 among groups is less than or equal to zero can be strongly rejected (ghlt $p < 2e-16$), and
673 that individuals from the western Napo population have greater median/mean co-ancestry
674 with each progenitor lineage than their progenitor lineages share with each other
675 (PHvGS/progenitors = ~1.078, PHvInambari/progenitors = ~1.16). These analyses
676 establish the asymmetry in introgression patterns detected by STRUCTURE. On average,
677 western Napo individuals share ~ 7.5% greater co-ancestry with Inambari individuals
678 (southern Amazon) than they do with other Northern Amazon individuals (glht $p < 2e-16$),
679 indicating that these populations are not likely to be composed of F1 hybrids. We also did
680 not observe clinal variation within our samples of this lineage.
681 fineRADstructure analysis of the co-ancestry matrix, as well as visualization of the
682 co-ancestry matrix, also indicated that the western Napo population is a differentiated form
683 (Figure 6a-d). Principle components analysis (Figure 6d) of the co-ancestry matrix indicate
684 that the western Napo lineage is intermediate on PC1 (compared to progenitor lineages),
685 and differentiated on PC2, as expected if the Napo lineage has had sufficient time in

reproductive isolation for sorting of ancestral alleles or for the evolution of new alleles (Barrera-Guzmán et al. 2018).

Bootstrapped estimates of F_{st} indicated the same general patterns: 95%CI; western Napo vs Eastern Napo (Northern progenitor, 0.07 - 0.090), western Napo v Inambari (Southern progenitor, 0.058 - 0.077): Eastern Napo vs Inambari: (between progenitors, 0.122 - 0.144). These patterns of F_{st} estimates were robust to alternative assignments of source linages (for example, including an additional adjacent population for both of the putative progenitor lineages). Thus, the introgressed western Napo lineage is less differentiated from *each* of its progenitor lineages than its progenitor lineages are to each other (i.e. it is intermediate, see methods), and it is *also* biased in its ancestry toward Inambari populations.

PhyloNet detected that an evolutionary reticulation is the best model to explain the origin of the western Napo population (Figure 6a). Computational limitations restricted us to evaluating a three-taxon case with a maximum of one reticulation. This included each putative progenitor lineage: (1) western Guiana Shield, northern Amazonian clade (clade D2), (2) a sample of northwestern Inambari individuals, part of the southern Amazonian clade (clade C1) —and (3) the introgressed western Napo population (clade E), each defined as 'species.' The RJMCMC search achieved an ESS of ~300 and detected two network topologies representing 98% of the posterior probability. The top ranked network (*pp* = 51%, Figure 6a) implies the MRCA of the northern Amazonian (Eastern Napo + Jaú) and southern (Inambari + Western Napo) lineages diverged first (clade F). This initial split was followed by a split within the latter lineage that led to the Inambari (S1) and progenitor western Napo (S2a) populations (MRCA of S1 and S2a: Figure 6a). Later,

709 genetically distinct genomes from the Northern lineage (clade D/D2) introgressed into S2a
710 (ancestral western Napo) to give rise to the contemporary western Napo population (S2).
711 The next best network (*pp* = 47%, Supplemental Data) implies that northern and southern
712 lineages of Clade F diverged, as in the top ranked network. Later, S1 and the NW
713 Amazonian lineages merged, forming a hybrid lineage. We base our subsequent discussion
714 on the top ranked network, as this hypothesis is consistent with patterns we detect in other
715 analyses, though we acknowledge that support for these two reticulation scenarios is
716 similar.

717 Taken in the context of our STRUCTURE (Figure 5) and fineRADstructure (Figure 6)
718 results, the introgression indicated in the top ranked network seems to have been
719 asymmetrical, leaving very little evidence of southern admixture among northern lineages
720 (see Figure 5, 'weakly resolved eastern Napo', which are detected to have very limited (<
721 ~5%) southern admixture). Demographic parameters estimated with G-PhoCS are also
722 consistent with this hypothesis: the only substantial (lower 95% CI > 1 migrant per
723 generation) migration rate was inferred to be from the eastern Napo (member of Guiana
724 Shield lineage) lineage into S2/S2a (Figure 6, Supplementary Table 5).

725 G-PhoCS also allows us to place a preliminary maximum age bound on the timing of
726 introgression into S2a. Assuming one generation per year and a constant mutation rate, our
727 analysis estimates the splitting time between S2a and S1 to be on the order of 500 Ka
728 (Figure 6, Supplementary Table 5). Though we emphasize that this is a rough
729 approximation, this result is consistent with prior divergence time estimates for several
730 groups in the genus based on mtDNA (Castro-Astor 2014); that study estimated ssp *pipra*
731 (Clade D; Northern Amazon + Guiana Shield lineage) to have branched off from its closest

732 relatives ~ 0.6717 [95% HPD: 0.3506-1.1287] Ma. Therefore, introgression into S2a must
733 have occurred more recently.
734 Lastly, patterns of admixture (Figure 5) also suggest some evidence of limited
735 introgression of Northern Amazon alleles into the southwestern Amazonian Inambari
736 population (Figure 5). This could be a consequence of limited gene flow into the Inambari
737 population *after* the primary introgression event which formed the introgressed western
738 Napo lineage. Ultimately, the product of this process was a differentiated and introgressed
739 lineage within the western part of Napo area of endemism.
740
741 *Spatial distribution of genetic variation*
742 Under isolation by distance (IBD), the empirical relationship between geographic
743 distance and genetic distance is related to offspring dispersal distance and population size
744 (Rousset 1997). Isolation by distance effects are thus predicted when individual dispersal
745 distances are smaller than a species' range (Teske et al. 2018). Fifteen out of eighteen
746 considered areas (see Population Genetics, Figure 4, and Supplemental Appendix for a
747 description of how these areas were defined) contained > 2 individuals, and thus could be
748 evaluated for isolation by distance effects. Within areas, only the western Inambari and
749 Atlantic Forest-Rio de Janeiro populations were detected to show significant signals of IBD
750 at the 0.001 alpha level (Supplementary Table 2a). When geographic distance was log
751 transformed, only the Jaú and Inambari areas were indicated to have significant IBD effects
752 (Supplementary Table 2a). Partial Mantel tests of specific barriers to gene flow indicated
753 that most evaluated river and physical barriers likely play significant roles in structuring
754 genetic variation between areas (Supplementary Table 2b). These patterns are expected if

755 individuals are dispersing within sampling regions but are generally unable to cross
756 barriers delimiting these areas; for our dataset, these barriers are predominantly the major
757 tributaries of the Amazon, or montane regions. Notably, the Rio Negro was the only
758 evaluated barrier which is not detected to exhibit a significant effect—in comparing of Jaú
759 and western Guiana Shield samples (Supplementary Table 2b).

760 Analysis with EEMS complimented and largely corroborated the specific hypotheses
761 implied by Mantel tests. For our analysis of 2000 fitted demes, the model fit between
762 demes was very high (R^2 = 0.871), and within demes somewhat less (R^2 = 0.579), indicating
763 that the EEMS model does a very good job of describing spatially structured variation in
764 this dataset (Supplementary Figure 11). EEMS also detected many strong signals (posterior
765 probability > 0.9) of barriers to gene flow and migratory corridors (Figure 7). As expected,
766 the Andes and the Amazon river were unambiguously detected as the most significant
767 barriers to gene flow and are clearly recovered as landscape features.

768 EEMS also detected signals of deviation from IBD that likely correspond to several
769 Amazonian tributaries in the Amazon basin, delimiting areas of endemism (Figure 7).
770 An additional strong barrier to gene flow was inferred in the south-central amazon (south-
771 central brown patch in Figure 7), corresponding to the Caatinga - Cerrado – Chaco 'dry-
772 diagonal' biomes. Strong signals of genetic connectivity are inferred in many localities (blue
773 in Figure 7), notably west of the Andes, connecting north Andean and Central American
774 lineages (despite a dearth of sampling in that region). This link likely reflects the common
775 ancestry among montane Andean lineages between Central America and Central Peru as
776 *Pseudopipra* does not presently occur on the western slopes of the Peruvian Andes. The
777 large Guianan + Northern Amazonian region (E. Napo + Jaú + Guiana) is detected to have

relatively weak population structure across a variety of barriers (including the Rio Negro). Several areas of endemism are also detected by EEMS in the Atlantic Forest, likely corresponding to the Coastal Bahia and Serra do Mar areas identified by da Silva et al. (2004) (Figure 7). Introgressed hybrids from the western Napo area of endemism are inferred as being genetically isolated from neighboring populations, with strong barriers to gene flow on all sides, perhaps corresponding to the effects of the Rio Marañón in the south and Rio Putumayo/Iça to the north.

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 populations. The other relatively high diversity group is the introgressed western Napo population. Areas of relatively low genetic diversity included the Atlantic Forest, the Peruvian Andes, and Central American lineages. These results are consistent with estimates of regional variation in allelic richness (Supplementary Appendix for details).

Vocal variation

We examined 198 recordings of *Pseudopipra* and identified 15 distinct vocalizations that can be diagnosed by ear or visual inspection of sonograms (labeled numerically (arbitrarily) in Figure 8 and 9). This set of vocalization types includes both advertisement vocalizations from males at leks, and calls by males and females in other contexts. By examining recordist notes, we generated a preliminary hypothesis which distinguishes 13 highly variable and distinct, male lekking vocalizations (Figure 8) from two broad ranging

call types (Figure 9). 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 lekking or call vocalizations can be statistically discriminated based on a combination of the frequency, duration, number of notes, and quality of the sound (buzzy vs pure tone). PCA and logistic regression on lekking vocalizations (n=114, Supplementary Table 3), and call vocalizations (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).

Assigning vocal types to well supported genetic groups is challenging in the absence of matching data (i.e., recordings from the same individuals from which we have genetic samples). Nevertheless, we found that most vocal types have distinct non-overlapping geographic ranges (Figure 8, Figure 9), and that the majority correspond to the geographic range of distinct lineages we identify (Taxonomic Summary, below). The complex phylogenetic history we have documented with genetic data has clearly had a strong impact on the evolution of male vocal advertisements in *Pseudopipra.* Because lek vocalizations are a focus of female choice in manakins and because suboscine passerines have largely innate vocalizations, the extensive vocal differentiation among populations of *Pseudopipra* strongly indicates the existence of many species with *Pseudopipra*.

Discussion:

In comparison to the great diversity of Neotropical organisms available for study, there have been far fewer 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 (Marks et al. 2002, Nyári 2007, Milá et al. 2012, Fernandes et al. 2013, Harvey and Brumfield 2015, Harvey et al. 2017). 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 *coracaina*, were represented by mtDNA. Our 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.

The predominant phylogenetic signal we detect in our dataset is of two successively nested montane clades as the sister groups to several major lowland Amazonian clades (see *Historical Biogeography*, below). This pattern implies that lowland populations of *Pseudopipra* expanded historically 'out of the Andes,' and that lowland clades of *Pseudopipra* are derived from ancient lineages endemic to the Andes (~ 2.5 Ma; Castro-Astor (2014)). The lowland clade within *Pseudopipra* originated (~1.085 [0.57-1.79] Ma) in southwestern Amazonia (Figure 4) and expanded, occupying Amazonia on both margins of the Amazon river, and eventually reached the Atlantic forest (see below). Therefore, to our knowledge, our study provides a unique example of a widespread suboscine passerine which originated in the Andean highlands.

Population genetic variation

847 We examined our dataset using a variety of population genetic approaches intended
848 to assess the number of natural genetic groupings that may be present in *Pseudopipra*.
849 Phenetic K-means clustering of SNP data and analysis with STRUCTURE estimated a
850 minimum of 5 broad genetic clusters (Figure 5, Supplementary Figure 2, Supplementary
851 Figure 4). K-means clustering of the co-ancestry matrix (which captured substantially more
852 variation than standard PCoA) more finely partitioned the dataset into at least ~8 broad
853 groups (Supplementary Figure 8), which generally recapitulate phenotypically delimited
854 subspecies boundaries (Figure 2). Phylogenetic (RAxML, SVDquartets) and model-based
855 population genetic (fineRADstructure) analyses of nuclear genomic data detected fifteen
856 clusters of individuals which coincided with sampling regions delimited by clear
857 geographic barriers (results, Figure 3, 4, Supplementary Appendix). Two additional
858 subspecies represented by mtDNA (results) bring the total number of distinct genetic
859 populations delimited by geographic barriers to at least seventeen. The close concordance
860 observed between population structure and landscape features implies that the
861 evolutionary history of *Pseudopipra* within the Neotropics is deeply connected to the South
862 American landscape, adding support to a rich body of literature endorsing this hypothesis
863 (Cracraft and Prum 1988, Brumfield 2012, Silva et al. 2019). Our analyses of isolation by
864 distance compliment and support these general conclusions (below).
865 Our analyses also lend themselves to some general conclusions about the sensitivity
866 of various methods to detecting population genetic structure. Overall, methods based on
867 clustering of SNP data were the least sensitive (prone to lumping) to population structure,
868 whereas phylogenetic analyses were the most sensitive to population structure (prone to

splitting). Thus, different phylogenetic and population genetic cluster analyses detected different levels of hierarchical structure in the data.

Overall, there was broad congruence in group assignments across methods, and no reciprocally incompatible clusters were recovered—with a notable exception: In phenetic K-means clustering of expanded datasets 2, 3 (Supplementary Figure 2), *and* of the normalized co-ancestry matrix (Supplementary Figure 8), the important and unique North Andean montane individual (5444.PE.MAR) clusters with other South Andean (Central Peruvian) lineages, contra phylogenetic analysis. But in all model based phylogenetic analyses (and fineRADstructure population assignment) 5444.PE.MAR clusters with Central American Lineages. Additional field sampling of these North Andean populations is required to further understand why this lineage may share signals of ancestry with both groups.

Our relatively dense sampling reveals a unique 'horseshoe-like' pattern of genetic differentiation across the Amazon basin (Figure 4). We recover 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 (Räsänen et al. 1987, Haffer 1992, Pupim et al. 2019). Given that the genus originated in the west, we hypothesize that these two 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

892 relatives—as are the Andes, which exhibit a disproportionate effect on divergence between
893 Peruvian foothills populations and Central American lineages (see Supplementary
894 Appendix for additional discussion).

895 Multiple analyses indicate that the western Napo population has a complex history
896 of initial and early differentiation from the broader Southern Amazonian clade, followed by
897 subsequent, asymmetric genetic introgression from Jaú and eastern Napo populations of
898 the Guianan Shield + Northern Amazon clade (Clade D; see discussion below). This result is
899 in conflict with other phylogenetic analyses. When the genetic data are analyzed under
900 either concatenated or coalescent phylogenetic frameworks which do not allow for
901 phylogenetic reticulation, the western Napo population is reconstructed as most closely
902 related to Guiana Shield + Northern Amazon populations (Figure 3, Figure 5, Clade D). In
903 contrast, phylogenetic network analysis strongly indicates that this population is more
904 closely related to individuals of Inambari (southern Amazon) provenance (Figure 6, also
905 summarized in Figure 4). This incongruence between methods is likely due to the
906 complexity of the constituent gene trees in this population-level differentiation induced by
907 asymmetric introgression from the Jaú and eastern Napo populations (Thom et al. 2018).
908 We hypothesize that introgression has introduced more recently derived variation from
909 the Guiana Shield + Northern Amazon clade D into an older, differentiated, population of
910 the Southern Amazonian clade, pulling the phylogenetic affinity toward the northern clade.
911

912 *History of an introgressed hybrid lineage*

913 Analysis of genomic admixture and phylogenetic reticulation revealed a population
914 of differentiated individuals restricted to the western Napo area of endemism which have

highly introgressed genomes and a particularly 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 a member of 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 species.

To date, five prior examples of hybrid species formation in birds have been evaluated with genomic data: (1) the golden crowned manakin (Barrera-Guzmán et al. 2018), (2), the Audubon's warbler (Brelsford et al. 2011), (3) the Italian sparrow (Hermansen et al. 2011), (4) the Hawaiian duck (Lavretsky et al. 2015), and (5) a putative species of Darwin's finches (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 (Jiggins et al. 2008, Counterman 2016), 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 genome composition derived from heavy historical admixture with a non-sister lineage. This historical scenario is not the same as those examples cited above, and yet it has produced a population genetic signature which is very similar (compare to Barrera-Guzmán et al. 2018).

Some authors have proposed strict population genetic criteria for identifying homoploid hybrid speciation, including demonstrating (1) evidence of reproductive isolation from parental species, (2) evidence of past hybridization, and (3) evidence that

938 intrinsic reproductive isolation was derived from the hybridization event (Schumer et al.
939 2014, Barrera-Guzmán et al. 2018).

940 The first two conditions are easily satisfied by the western Napo *Pseudopipra*
941 lineage. Phylogenetic and population genetic analysis indicate reproductive isolation. All
942 phylogenetic analyses, including concatenated and coalescent analyses (Figure 3 4, 6),
943 recover this lineage as a strongly supported reciprocally monophyletic group. PCA-based
944 and model-based population genetic cluster analyses (Figure 6b, 6c, Supplementary
945 Figures 4-9), also support the hypothesis that this lineage is both intermediate and
946 differentiated when compared its closest living relatives. Analysis of IBD also supports the
947 hypothesis that gene flow around the geographic extents of this population is at least
948 restricted and probably not occurring (Figure 7, Supplementary Figure 12). Estimates of
949 inbreeding coefficients also imply that introgression is probably not ongoing into the
950 western Napo lineage (Supplementary Appendix): The similar levels of inbreeding among
951 *Pseudopipra* populations in the western Amazon (F_{IS}, Supplementary Figure 9, and
952 Supplementary Appendix) imply that the western Napo population is not more outbred (i.e.
953 recently introgressed) than other inbred populations. Notably, the consistently positive
954 level of inbreeding (median ~0.16, 95% CI [0.11-0.21]) among most *Pseudopipra*
955 populations is perhaps attributable to a highly polygynous lek mating system (Smith 1979,
956 Waser et al. 1986, Stopher et al. 2012). In sum, these patterns imply contemporary
957 reproductive isolation of the Western Napo populations.

958 The second criteria, evidence of past hybridization, is directly supported by model-
959 based analysis of genomic admixture and phylogenetic reticulation (Figure 5, 6), and
960 indirectly supported by model based and phenetic cluster analysis (Supplementary Figures

4-9). We hypothesize that hybridization initially occurred between progenitor lineage S2a and western lineages of the Guiana Shield + Northern Amazon clade (Figure 6a). S2a then became extinct and was replaced by the novel introgressed lineage, which subsequently exchanged genes with neighboring populations. An intrinsic assumption of our analyses is that S1 (western Inambari) is an appropriate substitute for S2a (which we infer to have existed). We use S1 as proxy for S2a in population genetic analyses because it is the closest and only available relative.

The third criterion, evidence that intrinsic reproductive isolation was derived from the hybridization event, is the most difficult to demonstrate, and perhaps also the most tenuous and dependent on one's concept of species. This criterion is useful for framing the strength of evidence for or against the existence of a homoploid hybrid species if one adheres strictly to the biological species concept, as is frequently adopted for birds (Mayr 1942). However, if one adopts a more evolutionary definition of species, like the metapopulation lineage concept, in which the only necessary requirement is demonstration of separately evolving lineages (de Queiroz 2005), the importance of demonstrating intrinsic reproductive isolation to claim species status for a putative hybrid lineage (or any lineage at all) is less clear to us. Requiring hybridization to be the source of intrinsic reproductive isolation also excludes the possibility that extrinsic reproductive isolation may result from intrinsic changes that enable hybrids access to new niches, which may be also be geographically restricted relative to parental forms (Nieto Feliner et al. 2017). In sum, we cannot yet determine if the genetic signatures of reproductive isolation we observe for western Napo *Pseudopipra* are related to hybridization.

983 Intriguingly, the evolutionary scenario inferred by phylogenetic reticulation analysis
984 for the introgressed western Napo hybrids is consistent with observed patterns of lekking
985 vocalizations (Figure 8, Supplementary Appendix). The males from the Napo area of
986 endemism share vocalization type 2 with birds in the southeastern Amazon (Figure 8), and
987 with the restricted Huallaga valley *pygmaea* subspecies (the sister group to all lowland
988 Amazonian forms, according to mtDNA). Manakins are suboscine passerines, have innate
989 vocalizations (though see Saranathan et al. 2007). Therefore, the introgressed Napo lineage
990 inherited its vocalizations from the MRCA of groups S1 and S2 (Figure 6a) and retained it
991 through subsequent introgression with northern lineages that have a different vocal type
992 (Type 3 and/or 5). In this scenario, vocal type 7 in the western Inambari may represent
993 more recently derived variation in the southern Amazonian clade (Figure 8). The
994 observation that the introgressed western Napo hybrids share a greater proportion of their
995 genomes with their southern progenitors is consistent with this hypothesis, as is the
996 observation that the introgressed lineage has mtDNA haplotypes from the southern
997 Amazon (Supplementary Figure 16). This mitochondrial pattern implies that the
998 introgressed Napo lineage may be a product of progenitor female S2a and western males of
999 the Guiana Shield + Northern Amazon clade.

1000 Taking all of the available genetic and phenotypic evidence into account, we propose
1001 that the Western Napo lineage is an additional example of the formation of a hybrid species
1002 in birds, and one that was produced via an underappreciated historical mechanism. The
1003 patterns we observe appear to be most similar to 'hybrid trait speciation,' in which a hybrid
1004 species is formed after introgression from one species into a genomic background of a close
1005 relative (Jiggins et al. 2008, Counterman 2016, 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 no 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.

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

1029 analysis shows that large rivers delimit genetic clusters in as many as 23 groups of upland
1030 Amazonian forest birds, strongly arguing in favor of these rivers acting as barriers to gene
1031 flow (Silva et al. 2019). The somewhat distinct pattern found for *Pseudopipra* may be
1032 related to its recent expansion throughout the eastern Andean lowlands and its habitat
1033 tolerance.

1034 Within the manakins, a recent molecular phylogeny (Ohlson et al. 2013) placed
1035 *Pseudopipra* as sister to the genus *Ceratopipra*, which includes five well-recognized species
1036 that are extensively codistribtuted with *Pseudopipra.* The montane origin of *Pseudopipra* is
1037 congruent with the observation that *Pseudopipra* is the sister group to *Ceratopipra,* which
1038 is also broadly distributed in the Neotropical lowlands from Central America to the Atlantic
1039 Forest, with two secondary expansions into the subtropical forests in the tepuis (*C.*
1040 *cornuta*) and the southern Andes (*C. chloromeros*) (Ohlson et al. 2013). The breakpoints
1041 among these *Ceratopipra* species are highly concordant with the breakpoints among the
1042 genetic clusters within the *Pseudopipra* complex that we have presented here, implying
1043 that these closely related taxa have many components of their phylogeographic history in
1044 common. See the Supplemental Appendix for additional notes, as well as Castro-Astor
1045 (2014).

1046 A key feature of the phylogeny of *Pseudopipra* is that montane Andean lineages are
1047 sister groups to both the montane Central American and the lowland Amazonian-southeast
1048 Brazilian lineages. Ancestral character reconstruction with Bayesian stochastic mapping
1049 therefore unambiguously reconstructs the most recent common ancestor of *Pseudopipra* as
1050 montane (Figure 3, Supplemental Appendix). Thus, it appears that the most recent common
1051 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 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) (BS100) (Clade C2 in Figure 3) is recovered as nested within the well supported lowland western Inambari subclade (Cerro Azul, 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 Cerro Azul individuals are detected to have limited genomic background of derived lineages of the Guiana Shield + Northern Amazon clade (yellow in Figure 5), similar to proximate lowland relatives. The presence of genomic

1075 admixture from the Guiana Shield + Northern Amazon clade is consistent with a scenario of
1076 secondary invasion of the highlands from lowland Inambari lineages which were
1077 secondarily weakly introgressed from the western Napo (see below), which itself
1078 experienced significant initial introgression from the Northern Amazon + Guiana Shield
1079 clade (D).
1080
1081 *North Andean diversity*
1082 Our final genomic data set does not include any samples from Andean Ecuador or
1083 Colombia, a region that includes four previously recognized subspecies and at least four
1084 distinct vocalization types. Unfortunately, the three Andean Ecuadorian tissue
1085 (representing subspecies *coracina*) specimens in our original sample were not of high
1086 enough preservation quality to be viable for ddRAD sequencing (Supplementary Table 1).
1087 We were able to obtain mtDNA from these samples however, which unambiguously formed
1088 a monophyletic cluster with our single San Martín (North Andean Peru) specimen in
1089 mtDNA gene tree analyses (Supplemental Appendix, Supplementary Figure 16). A prior
1090 study by Castro-Astor (2014) based on mtDNA also included one Andean Ecuadorian
1091 sample and one Central American sample, which unambiguously clustered together
1092 (*pp*=100). Thus, it is likely that our unsampled Andean lineages of Ecuador and Colombia
1093 (including the subspecies *minimus, bolivari, unica,* and *coracina*) are likely members of the
1094 northern Andean clade represented in our genomic data set by our single San Martín
1095 specimen.
1096 As noted earlier, our San Martín specimen had approximately equal probability of
1097 assignment to multiple populations, which could indicate heavy admixture, or that this is a

single sample from a highly distinct population which has its own distinct history. These patterns may also be consistent with the hypothesis that the north Andean clade is the source of montane and lowland *Pseudopipra* diversity. 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.

Phenotypic 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.

Although we did not conduct a detailed analysis of plumage variation among populations of *Pseudopipra*, we note that 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 the entirely white crown feather bases are shared exclusively by *unica* and *minima,* and on the forecrown feathers of

bolivari, imply that this plumage character state is derived within *Pseudopipra*, and that *minima* and *unica* form a clade, with *bolivari* as their sister group. One lineage containing two subspecies– *separabilis* and *cephaleucos*- has evolved another unique, shared, derived plumage character– a distinctive, second, predefinitive male plumage which has 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, 9), and all but one was restricted to and diagnostic of a single previously recognized subspecies or a broader monophyletic group (See Expanded Taxonomic Summary in Supplementary Appendix for details). The only exception was vocalization type 2 (Fig. 8), which was 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*. 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.

Based on genetic variation, we also identified several additional well-supported clades for which we have limited vocal or plumage coloration data. For example, only one vocal record is known for the South Andean clade (Clade B), which is currently recognized as *P. p. comata.* This lineage is also composed of two well differentiated subclades: a northern clade (B1) from Cerro Azul in Loreto, and a southern clade (B2) from southern Huánuco (AMNH 820866, 820952), Pasco, Junín, and Cusco. Since the type of *comata* is within the southern clade, further investigation is necessary to determine whether the Cerro Azul populations should be recognized as a distinct, new taxon. Our phylogenetic analysis also identified a distinct montane clade from the highlands between the Rio Huallaga and Rio Ucayali (Clade C2) 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*. Although it is not certain that this montane clade represents a secondary expansion into the Andes, its phylogenetic distinction implies that it may also exhibit distinct plumage or vocal characters.

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 descriptions of cryptic diversity in *Pseudopipra*

are almost certainly an underestimate, 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 our limited genetic sampling of populations in the northern Andean clade from Ecuador and Colombia and our limited behavioral data. Three of these unsampled northern Andean subspecies– *coracina, minima,* and *occulta*– 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. The vocalization types of one other subspecies– *bolivari*– is currently unknown, but it may also be distinct species. Further research will be required to assess whether the unexpected vocal diversity among populations of *unica* (types 11a and 11b) from the central Andes of Colombia indicates the existence of additional, undescribed taxa. Plumage and vocal data from other newly identified clades may support the recognition of additional species including populations from the highlands between Rio Huallaga and Rio Ucayali (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.

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)

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)

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

Phylogenetic Position: (unsampled) Member of Clade A1

Lek Vocal Type: 9 (*reeee*)

Call Vocal Type: unknown

P. c. bolivari (de Schauensee 1950)

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)

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)

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 A1 (represented by

Lek Vocal Type: 1 (*trill-dink*) and 10 (*bree*)

Call Vocal Type: unknown

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

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**

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

Phylogenetic Position: Clade B

Lek Vocal Type: One record, statistically similar to type 1 (*trill-dink*)

Call Vocal Type: unknown

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

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

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

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

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

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 highlands between Rio Huallaga and Rio Ucayali

Phylogenetic Position: Clade C2

Lek Vocal Type: Unknown

Call Vocal Type: 13

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

***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

***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

***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

Call Vocal Type: 13

1304

1305 ***P. m. separabilis*** (Zimmer 1936)

1306 **Distribution**: Tropical forest of Rio Xingu east to central and southern Pará.

1307 **Phylogenetic Position:** Clade C6

1308 **Lek Vocal Type:** 2

1309 **Call Vocal Type**: 13

1310 ***Pseudopipra cephaleucos*** (Thunberg 1822) **Atlantic White-crowned Manakin**

1311 **Distribution**: Tropical forest from Bahia south to northern Rio de Janeiro, Brazil.

1312 **Phylogenetic Position:** Clade C7

1313 **Lek Vocal Type:** 6a (*zeeeee-tonk*)

1314 **Call Vocal Type**: 13

1315

1316 **Supplementary Material:** Supplementary material, including data files and online-only

1317 appendices, can be found in the Dryad Digital Repository:

1318 http://dx.doi.org/10.5061/dryad.[NNNN], and GenBank XXX-XXX.

1319

1320 **Funding:** J. S. B. was supported by a National Science Foundation Graduate Research

1321 Fellowship and NSF Doctoral Dissertation Improvement Grant [DGE-1650441, DEB-

1322 1700786]. L. C. was supported by NSF grant DEB-1555754 to I.J.L. T. J. F. was supported by

1323 an NSF Post-doctoral Fellowship in Biology [DBI-1523857]. C. C. R. was funded by CNPq

1324 (308927/2016-8), FAPEAM, FAPESP/NSF (NSF DEB-1241066, FAPESP, grant

1325 #2012/50260-6) and USAID (PEER AID-OAA-A-11-00012, cycle 5). J. S. B. and R. O. P. were

1326 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 were kindly received from Divisão de Aves do Museu de Zoologia da Universidade Estadual de Feira de Santana (MZFS), 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), University of Alaska Museum (UAM), University of Nevada, Las Vegas (UNLV), the Yale Peabody Museum of Natural History (YPM), and the Cornell Museum of Vertebrates (CUMV). 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, as well as **XXXX** anonymous reviewers, for comments on the research and the manuscript. The illustration depicted in Fig. 3 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 in 2011. 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. organized 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. proposed a revised taxonomy. The manuscript was written by J.S.B. with input from all authors.

Literature Cited

Agapow PM, Burt A. 2001. Indices of multilocus linkage disequilibrium. Molecular Ecology Notes, 1:101-102.

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 LG, Matos-Maraví P, Ritter CD, Sanmartín I, Silvestro D, Tejedor M, ter Steege H, Tuomisto H, Werneck FP, Zizka A, Edwards SV. 2018. Conceptual and empirical advances in Neotropical biodiversity research. PeerJ, 6:e5644.

Barrera-Guzmán AO, Aleixo A, Shawkey MD, Weir JT. 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 JM, Zink RM. 1994. Evolution into the Andes: Molecular Evidence for Species Relationships in the Genus Leptopogon. The Auk, 111:507-515.

Berv JS, Prum RO. 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.

1392 Bioacoustics_Research_Program. 2011. Raven Pro: Interactive Sound Analysis Software
1393 (Version 1.4) [Computer software]. Ithaca, NY, The Cornell Laboratory of
1394 Ornithology. Available from http://www.birds.cornell.edu/raven.

1395 Bollback JP. 2006. SIMMAP: Stochastic character mapping of discrete traits on phylogenies.
1396 BMC Bioinformatics, 7.

1397 Borges SH, Da Silva JMC. 2012. A New Area of Endemism for Amazonian Birds in the Rio
1398 Negro Basin. The Wilson Journal of Ornithology, 124:15-23.

1399 Botta F, Eriksen C, Fontaine MC, Guillot G, Yu D. 2015. Enhanced computational methods for
1400 quantifying the effect of geographic and environmental isolation on genetic
1401 differentiation. Methods in Ecology and Evolution, 6:1270-1277.

1402 Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A,
1403 Drummond AJ. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary
1404 Analysis. PLOS Computational Biology, 10:e1003537.

1405 Bradburd G, Coop G, Ralph P. 2017. Inferring Continuous and Discrete Population Genetic
1406 Structure Across Space. bioRxiv.

1407 Bradburd GS, Ralph PL, Coop GM. 2013. DISENTANGLING THE EFFECTS OF GEOGRAPHIC
1408 AND ECOLOGICAL ISOLATION ON GENETIC DIFFERENTIATION. Evolution,
1409 67:3258-3273.

1410 Brelsford A, Milá B, Irwin DE. 2011. Hybrid origin of Audubon's warbler. Molecular Ecology,
1411 20:2380-2389.

Brown AHD, Feldman MW, Nevo E. 1980. Multilocus Structure of Natal Populations of Hordeum Spontaneum. Genetics, 96:523-536.

Brumfield RT. 2012. Inferring the origins of lowland Neotropical birds. The Auk, 129:367-376.

Brumfield RT, Edwards SV. 2007. EVOLUTION INTO AND OUT OF THE ANDES: A BAYESIAN ANALYSIS OF HISTORICAL DIVERSIFICATION IN THAMNOPHILUS ANTSHRIKES. Evolution, 61:346-367.

Brumfield RT, Liu L, Lum DE, Edwards SV. 2008. Comparison of Species Tree Methods for Reconstructing the Phylogeny of Bearded Manakins (Aves: Pipridae, Manacus) from Multilocus Sequence Data. Systematic Biology, 57:719-731.

Burbrink FT, Gehara M. 2018. The Biogeography of Deep Time Phylogenetic Reticulation. Systematic Biology, 67:743-744.

Campagna L, Gronau I, Silveira LF, Siepel A, Lovette IJ. 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 JJH, Lougheed SC, Woods RW, Imberti S, Tubaro PL. 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.

1431 Capurucho JMG, Cornelius C, Borges SH, Cohn-Haft M, Aleixo A, Metzger JP, Ribas CC. 2013.
1432 Combining phylogeography and landscape genetics of Xenopipo atronitens (Aves:
1433 Pipridae), a white sand campina specialist, to understand Pleistocene landscape
1434 evolution in Amazonia. Biological Journal of the Linnean Society, 110:60-76.

1435 Castro-Astor IN. 2014. A Phylogeographic study of two neotropical bird species, the Red-
1436 Headed Manakin (Pipna rubrocapilla) and the White-crowned Manakin (Dixiphia
1437 pipra). Biology. CUNY Academic Works, CUNY.

1438 Castro-Astor IN, Alves MAS, Cavalcanti RB. 2007. DISPLAY BEHAVIOR AND SPATIAL
1439 DISTRIBUTION OF THE WHITE-CROWNED MANAKIN IN THE ATLANTIC FOREST
1440 OF BRAZIL. The Condor, 109:155-166.

1441 Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool
1442 set for population genomics. Molecular Ecology, 22:3124-3140.

1443 Chambers EA, Hillis DM. 2019. The Multispecies Coalescent Over-splits Species in the Case
1444 of Geographically Widespread Taxa. Systematic Biology.

1445 Chernomor O, Minh BQ, von Haeseler A. 2016. Terrace Aware Data Structure for
1446 Phylogenomic Inference from Supermatrices. Systematic Biology, 65:997-1008.

1447 Cheviron ZA, Hackett SJ, Brumfield RT. 2006. Sequence variation in the coding region of the
1448 melanocortin-1 receptor gene (MC1R) is not associated with plumage variation in
1449 the blue-crowned manakin (Lepidothrix coronata). Proceedings of the Royal Society
1450 B: Biological Sciences, 273:1613-1618.

1451 Cheviron ZA, Hackett SJ, Capparella AP. 2005. Complex evolutionary history of a
1452 Neotropical lowland forest bird (Lepidothrix coronata) and its implications for
1453 historical hypotheses of the origin of Neotropical avian diversity. Molecular
1454 Phylogenetics and Evolution, 36:338-357.

1455 Chifman J, Kubatko L. 2014. Quartet Inference from SNP Data Under the Coalescent Model.
1456 Bioinformatics, 30:3317-3324.

1457 Chifman J, Kubatko L. 2015. Identifiability of the unrooted species tree topology under the
1458 coalescent model with time-reversible substitution processes, site-specific rate
1459 variation, and invariable sites. Journal of Theoretical Biology, 374:35-47.

1460 Chou J, Gupta A, Yaduvanshi S, Davidson R, Nute M, Mirarab S, Warnow T. 2015. A
1461 comparative study of SVDquartets and other coalescent-based species tree
1462 estimation methods. BMC Genomics, 16:S2.

1463 Counterman BA. 2016. Hybrid Speciation. In: Kliman RM editor. Encyclopedia of
1464 Evolutionary Biology. Oxford, Academic Press, p. 242-248.

1465 Cracraft J. 1985. Historical Biogeography and Patterns of Differentiation within the South
1466 American Avifauna: Areas of Endemism. Ornithological Monographs:49-84.

1467 Cracraft J, Prum RO. 1988. PATTERNS AND PROCESSES OF DIVERSIFICATION: SPECIATION
1468 AND HISTORICAL CONGRUENCE IN SOME NEOTROPICAL BIRDS. Evolution, 42:603-
1469 620.

1470 Crother BI, Murray CM. 2011. Ontology of areas of endemism. Journal of Biogeography,
1471 38:1009-1015.

1472 Cunningham CW, Collins TM. 1994. Developing model systems for molecular biogeography:
1473 Vicariance and interchange in marine invertebrates. In: Schierwater B, Streit B,
1474 Wagner GP, DeSalle R editors. Molecular Ecology and Evolution: Approaches and
1475 Applications. Basel, Birkhäuser Basel, p. 405-433.

1476 Da Silva J, Novaes F, Oren D. 2002. Differentiation of Xiphocolaptes (Dendrocolaptidae)
1477 across the river Xingu, Brazilian Amazonia: recognition of a new phylogenetic
1478 species and biogeographic implications. BULLETIN-BRITISH ORNITHOLOGISTS
1479 CLUB, 122:185-193.

1480 da Silva JMC. 1995. Biogeographic analysis of the South American Cerrado avifauna.
1481 Steenstrupia, 21:49-67.

1482 da Silva JMC, de Sousa MC, H. M. Castelletti C. 2004. Areas of Endemism for Passerine Birds
1483 in the Atlantic Forest, South America. Global Ecology and Biogeography, 13:85-92.

1484 Da Silva JMC, Rylands AB, Da Fonseca GAB. 2005. The Fate of the Amazonian Areas of
1485 Endemism. Conservation Biology, 19:689-694.

1486 de Queiroz K. 2005. Ernst Mayr and the modern concept of species. Proceedings of the
1487 National Academy of Sciences, 102:6600-6607.

1488 Dickinson E. 2003. The Howard and Moore Complete Checklist of the Birds of the World
1489 Christopher Helm. London.

1490 Diniz-Filho JAF, Soares TN, Lima JS, Dobrovolski R, Landeiro VL, de Campos Telles MP,
1491 Rangel TF, Bini LM. 2013. Mantel test in population genetics. Genetics and Molecular
1492 Biology, 36:475-485.

1493 Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for
1494 visualizing STRUCTURE output and implementing the Evanno method. Conservation
1495 Genet Resour, 4:359-361.

1496 Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using
1497 the software structure: a simulation study. Molecular Ecology, 14:2611-2620.

1498 Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of Molecular Variance Inferred from
1499 Metric Distances among DNA Haplotypes: Application to Human Mitochondrial DNA
1500 Restriction Data. Genetics, 131:479-491.

1501 Fernandes A, Cohn-Haft M, Hrbek T, Farias I. 2015. Rivers acting as barriers for bird
1502 dispersal in the Amazon. Revista Brasileira de Ornitologia-Brazilian Journal of
1503 Ornithology, 22:363-373.

1504 Fernandes AM, Gonzalez J, Wink M, Aleixo A. 2013. Multilocus phylogeography of the
1505 Wedge-billed Woodcreeper Glyphorynchus spirurus (Aves, Furnariidae) in lowland
1506 Amazonia: Widespread cryptic diversity and paraphyly reveal a complex
1507 diversification pattern. Molecular Phylogenetics and Evolution, 66:270-282.

1508 Fjeldså J. 1992. Biogeographic patterns and evolution of the avifauna of relict high-attitude
1509 woodlands of the Andes. Steenstrupia:9-62.

1510 Francisco MR, Gibbs HL, Galetti M, Lunardi VO, Junior PMG. 2007. Genetic structure in a
1511 tropical lek-breeding bird, the blue manakin (Chiroxiphia caudata) in the Brazilian
1512 Atlantic Forest. Molecular Ecology, 16:4908-4918.

1513 Freile JF. 2014. First record of White-crowned Manakin Dixiphia pipra in western Ecuador.
1514 Bull. Brit. Orn. Club, 134:238-240.

1515 Gene_Codes_Corporation. 2010. Sequencher version 5.0 sequence analysis software. Ann
1516 Arbor Michigan.

1517 Gordon A, Hannon G. 2010. Fastx-toolkit. FASTQ/A short-reads preprocessing tools
1518 (unpublished) http://hannonlab. cshl. edu/fastx_toolkit, 5.

1519 Goslee SC, Urban DL. 2007. The ecodist package for dissimilarity-based analysis of
1520 ecological data. Journal of Statistical Software, 22:1-19.

1521 Gosselin T, Anderson EC, Ferchaud A-L. 2016. thierrygosselin/assigner: v.0.4.0 (Version
1522 0.4.0). Zenodo. http://doi.org/10.5281/zenodo.197418.

1523 Goudet J. 2005. hierfstat, a package for r to compute and test hierarchical F-statistics.
1524 Molecular Ecology Notes, 5:184-186.

1525 Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A. 2011. Bayesian inference of ancient
1526 human demography from individual genome sequences. Nature Genetics, 43:1031.

1527 Gubili C, Ritter CD, Motta J, Farias IP, Bates J, Canton R, Capurucho JMG, Cornelius C,
1528 Feldheim KA, Ribas CC. 2016. Isolation and characterization of polymorphic
1529 microsatellite DNA Markers from an Amazonian white-sand vegetation specialist

bird, Xenopipo atronitens (Aves: Pipridae). The Wilson Journal of Ornithology, 128:668-672.

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

Guillot G, Schilling R, Porcu E, Bevilacqua M. 2013. Validity of covariance models for the analysis of geographical variation. arXiv:1311.4136.

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

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

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

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

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

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

Hartl DL, Clark AG. 1997. Principles of population genetics. Sinauer associates Sunderland, MA.

1549 Harvey MG, Aleixo A, Ribas CC, Brumfield RT. 2017. Habitat Association Predicts Genetic
1550 Diversity and Population Divergence in Amazonian Birds. The American Naturalist,
1551 190:631-648.

1552 Harvey MG, Brumfield RT. 2015. Genomic variation in a widespread Neotropical bird
1553 (Xenops minutus) reveals divergence, population expansion, and gene flow.
1554 Molecular Phylogenetics and Evolution, 83:305-316.

1555 Hazzi NA, Moreno JS, Ortiz-Movliav C, Palacio RD. 2018. Biogeographic regions and events
1556 of isolation and diversification of the endemic biota of the tropical Andes.
1557 Proceedings of the National Academy of Sciences.

1558 Hermansen JS, Sæther SA, Elgvin TO, Borge T, Hjelle E, Sætre G-P. 2011. Hybrid speciation
1559 in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene
1560 flow. Molecular Ecology, 20:3812-3822.

1561 Hijmans RJ, Williams E, Vennes C. 2015. Geosphere: spherical trigonometry. R package
1562 version 1.3-11. URL: https://CRAN. R-project. org/package= geosphere.

1563 Hoang DT, Vinh LS, Chernomor O, Minh BQ, von Haeseler A. 2017. UFBoot2: Improving the
1564 Ultrafast Bootstrap Approximation. Molecular Biology and Evolution, 35:518-522.

1565 Hooghiemstra H, Van der Hammen T. 2004. Quaternary Ice-Age dynamics in the Colombian
1566 Andes: developing an understanding of our legacy. Philosophical Transactions of the
1567 Royal Society of London. Series B: Biological Sciences, 359:173-181.

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

Huelsenbeck JP, Nielsen R, Bollback JP. 2003. Stochastic Mapping of Morphological Characters. Systematic Biology, 52:131-158.

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

Jiggins CD, 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 BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods, 14:587.

Kamvar ZN, Tabima JF, Grünwald NJ. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ, 2:e281.

Kirwan GM, David N, Gregory SM, Jobling JA, Steinheimer FD, Brito GRR. 2016. The mistaken manakin: a new genus-group name for Parus pipra Linnaeus, 1758 (Aves: Passeriformes: Pipridae). Zootaxa, 4121:89-94.

1588 Kirwan GM, Green G. 2012. Cotingas and Manakins. Princeton, NJ, Princeton University
1589 Press.

1590 Kuhner MK. 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of
1591 population parameters. Bioinformatics, 22:768-770.

1592 Kumar S, Subramanian S. 2002. Mutation rates in mammalian genomes. Proceedings of the
1593 National Academy of Sciences, 99:803-808.

1594 Lamichhaney S, Han F, Webster MT, Andersson L, Grant BR, Grant PR. 2018. Rapid hybrid
1595 speciation in Darwin's finches. Science, 359:224-228.

1596 Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. PartitionFinder 2: New
1597 Methods for Selecting Partitioned Models of Evolution for Molecular and
1598 Morphological Phylogenetic Analyses. Molecular Biology and Evolution, 34:772-773.

1599 Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient
1600 alignment of short DNA sequences to the human genome. Genome Biology, 10:R25.

1601 Lavretsky P, Engilis A, Eadie JM, Peters JL. 2015. Genetic admixture supports an ancient
1602 hybrid origin of the endangered Hawaiian duck. Journal of Evolutionary Biology,
1603 28:1005-1015.

1604 Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of Population Structure using
1605 Dense Haplotype Data. PLOS Genetics, 8:e1002453.

1606 Leaché AD, Banbury BL, Felsenstein J, de Oca An-M, Stamatakis A. 2015. Short Tree, Long
1607 Tree, Right Tree, Wrong Tree: New Acquisition Bias Corrections for Inferring SNP
1608 Phylogenies. Systematic Biology, 64:1032-1047.

1609 Leaché AD, Zhu T, Rannala B, Yang Z. 2018. The Spectre of Too Many Species. Systematic
1610 Biology:syy051-syy051.

1611 Legendre P, Fortin M-J. 2010. Comparison of the Mantel test and alternative approaches for
1612 detecting complex multivariate relationships in the spatial analysis of genetic data.
1613 Molecular Ecology Resources, 10:831-844.

1614 Linck EB, Battey CJ. 2017. Minor allele frequency thresholds strongly affect population
1615 structure inference with genomic datasets. bioRxiv.

1616 Linder HP. 2001. On Areas of Endemism, with an Example from the African Restionaceae.
1617 Systematic Biology, 50:892-912.

1618 Long C, Kubatko L. 2017. The effect of gene flow on coalescent-based species-tree
1619 inference. arXiv:1710.03806.

1620 Luna LW, Souza TO, e Silva WAdG, Schneider H, Sampaio I. 2017. Genetic variation of the
1621 endangered Araripe Manakin (Antilophia bokermanni) indicates a history of
1622 demographic decline. Revista Brasileira de Ornitologia, 25:60-66.

1623 Lynch Alfaro JW, Boubli JP, Paim FP, Ribas CC, Silva MNFd, Messias MR, Röhe F, Mercês MP,
1624 Silva Júnior JS, Silva CR, Pinho GM, Koshkarian G, Nguyen MTT, Harada ML, Rabelo
1625 RM, Queiroz HL, Alfaro ME, Farias IP. 2015. Biogeography of squirrel monkeys

1626 (genus Saimiri): South-central Amazon origin and rapid pan-Amazonian
1627 diversification of a lowland primate. Molecular Phylogenetics and Evolution,
1628 82:436-454.

1629 MacQueen J. 1967. Some methods for classification and analysis of multivariate
1630 observations. Proceedings of the Fifth Berkeley Symposium on Mathematical
1631 Statistics and Probability, Volume 1: Statistics. Berkeley, Calif., University of
1632 California Press, p. 281-297.

1633 Malinsky M, Trucchi E, Lawson DJ, Falush D. 2018. RADpainter and fineRADstructure:
1634 Population Inference from RADseq Data. Molecular Biology and Evolution, 35:1284-
1635 1290.

1636 Mantel N. 1967. The detection of disease clustering and a generalized regression approach.
1637 Cancer research, 27:209-220.

1638 Marks BD, Hackett SJ, Capparella AP. 2002. Historical relationships among Neotropical
1639 lowland forest areas of endemism as determined by mitochondrial DNA sequence
1640 variation within the Wedge-billed Woodcreeper (Aves: Dendrocolaptidae:
1641 Glyphorynchus spirurus). Molecular Phylogenetics and Evolution, 24:153-167.

1642 Marques DA, Meier JI, Seehausen O. 2019. A Combinatorial View on Speciation and
1643 Adaptive Radiation. Trends in Ecology & Evolution.

1644 Mayr E. 1942. Systematics and the origin of species from the viewpoint of a zoologist.
1645 Columbia University Press.

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

Milá B, Tavares ES, Muñoz Saldaña A, Karubian J, Smith TB, Baker AJ. 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 RP, Robinson WD, Lovette IJ, Robinson TR. 2008. Experimental evidence for extreme dispersal limitation in tropical forest birds. Ecology Letters, 11:960-968.

Musher LJ, Ferreira M, Auerbach AL, 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 LN. 2011. Avian distribution patterns in the Guiana Shield: implications for the delimitation of Amazonian areas of endemism. Journal of Biogeography, 38:681-696.

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

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

1665 Noguera-urbano EA. 2016. Areas of endemism: travelling through space and the
1666 unexplored dimension. Systematics and Biodiversity, 14:131-139.

1667 Nyári ÁS. 2007. Phylogeographic patterns, molecular and vocal differentiation, and species
1668 limits in Schiffornis turdina (Aves). Molecular Phylogenetics and Evolution, 44:154-
1669 164.

1670 Nylander JAA, Olsson U, Alström P, Sanmartín I. 2008. Accounting for Phylogenetic
1671 Uncertainty in Biogeography: A Bayesian Approach to Dispersal-Vicariance Analysis
1672 of the Thrushes (Aves: Turdus). Systematic Biology, 57:257-268.

1673 Ohlson JI, Fjeldså J, Ericson PGP. 2013. Molecular phylogeny of the manakins (Aves:
1674 Passeriformes: Pipridae), with a new classification and the description of a new
1675 genus. Molecular Phylogenetics and Evolution, 69:796-804.

1676 Oliveira U, Vasconcelos MF, Santos AJ. 2017. Biogeography of Amazon birds: rivers limit
1677 species composition, but not areas of endemism. Scientific Reports, 7:2992.

1678 Paradis E. 2010. pegas: an R package for population genetics with an integrated–modular
1679 approach. Bioinformatics, 26:419-420.

1680 Paris JR, Stevens JR, Catchen JM, Johnston S. 2017. Lost in parameter space: a road map for
1681 stacks. Methods in Ecology and Evolution, 8:1360-1373.

1682 Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double Digest RADseq: An
1683 Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-
1684 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 JK, Stephens M, Donnelly P. 2000. Inference of Population Structure Using Multilocus Genotype Data. Genetics, 155:945-959.

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

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

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

Pupim FN, Sawakuchi AO, Almeida RP, Ribas CC, Kern AK, Hartmann GA, Chiessi CM, Tamura LN, Mineli TD, Savian JF, Grohmann CH, Bertassoli DJ, Stern AG, Cruz FW, 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.

R_Core_Team. 2018. R: A language and environment for statistical computing.

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

Räsänen ME, Salo JS, Kalliola RJ. 1987. Fluvial perturbance in the western Amazon basin: regulation by long-term sub-Andean tectonics. Science, 238:1398-1401.

Reichenbach HG. 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.

Revell LJ. 2012. Phytools: An R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution, 3:217-223.

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

Ribas CC, Moyle RG, Miyaki CY, 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 RS, Greenfield PJ. 2001. The birds of Ecuador: status, distribution, and taxonomy. Cornell University Press.

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

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

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

Salazar C, Baxter SW, Pardo-Diaz C, Wu G, Surridge A, Linares M, Bermingham E, Jiggins CD. 2010. Genetic Evidence for Hybrid Trait Speciation in Heliconius Butterflies. PLOS Genetics, 6:e1000930.

Santorelli S, Magnusson WE, Deus CP. 2018. Most species are not limited by an Amazonian river postulated to be a border between endemism areas. Scientific Reports, 8:2294.

Saranathan V, Hamilton D, Powell GVN, Kroodsma DE, Prum RO. 2007. Genetic evidence supports song learning in the three-wattled bellbird Procnias tricarunculata (Cotingidae). Molecular Ecology, 16:3689-3702.

Schmidt HA, Minh BQ, von Haeseler A, Nguyen L-T. 2014. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution, 32:268-274.

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

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

Sedano RE, Burns KJ. 2010. Are the Northern Andes a species pump for Neotropical birds? Phylogenetics and biogeography of a clade of Neotropical tanagers (Aves: Thraupini). Journal of Biogeography, 37:325-343.

Shafer ABA, Peart CR, Tusso S, Maayan I, Brelsford A, Wheat CW, Wolf JBW. 2017. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods in Ecology and Evolution, 8:907-917.

Silva SM, Peterson AT, Carneiro L, Burlamaqui TCT, Ribas CC, Sousa-Neves T, Miranda LS, Fernandes AM, d'Horta FM, Araújo-Silva LE, Batista R, Bandeira CHMM, Dantas SM, Ferreira M, Martins DM, Oliveira J, Rocha TC, Sardelli CH, Thom G, Rêgo PS, Santos MP, Sequeira F, Vallinoto M, Aleixo A. 2019. A dynamic continental moisture gradient drove Amazonian bird diversification. Science Advances, 5:eaat5752.

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

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

Smith RH. 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 DW. 1979. Tityrinae, Pipridae, Cotingidae. In: M. A. Traylor J editor. Check-list of birds of the world. Cambridge, Massachusetts, Museum of Comparative Zoology, Harvard Univ., p. 229-308.

Snow DW. 2004. Family Pipridae (Manakins). In: del Hoyo J, Elliot A, Christie DA editors. Handbook of the Birds of the World. Barcelona, Lynx Edicions, p. 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.

Standley DM, Katoh K. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution, 30:772-780.

Stopher KV, Nussey DH, Clutton-Brock TH, Guinness F, Morris A, Pemberton JM. 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.

1783 Sukumaran J, Knowles LL. 2017. Multispecies coalescent delimits structure, not species.
1784 Proceedings of the National Academy of Sciences, 114:1607-1612.

1785 Swofford DL. 2002. Phylogenetic analysis using parsimony (*and Other Methods). Version
1786 4. Sinauer Associates, Sunderland, Massachusetts.

1787 Tello JG, Moyle RG, Marchese DJ, Cracraft J. 2009. Phylogeny and phylogenetic classification
1788 of the tyrant flycatchers, cotingas, manakins, and their allies (Aves: Tyrannides).
1789 Cladistics, 25:429-467.

1790 Teske PR, Golla TR, Sandoval-Castillo J, Emami-Khoyi A, van der Lingen CD, von der Heyden
1791 S, Chiazzari B, Jansen van Vuuren B, Beheregaray LB. 2018. Mitochondrial DNA is
1792 unsuitable to test for isolation by distance. Scientific Reports, 8:8448.

1793 Thom G, Amaral FRD, Hickerson MJ, Aleixo A, Araujo-Silva LE, Ribas CC, Choueri E, Miyaki
1794 CY. 2018. Phenotypic and Genetic Structure Support Gene Flow Generating Gene
1795 Tree Discordances in an Amazonian Floodplain Endemic Species. Systematic
1796 Biology, 67:700-718.

1797 Thrasher DJ, Butcher BG, Campagna L, Webster MS, Lovette IJ. 2018. Double-digest RAD
1798 sequencing outperforms microsatellite loci at assigning paternity and estimating
1799 relatedness: A proof of concept in a highly promiscuous bird. Molecular Ecology
1800 Resources, 0.

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

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

van Els P, Norambuena HV, Etienne RS. 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.

Wahlund S. 1928. ZUSAMMENSETZUNG VON POPULATIONEN UND KORRELATIONSERSCHEINUNGEN VOM STANDPUNKT DER VERERBUNGSLEHRE AUS BETRACHTET. Hereditas, 11:65-106.

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

Waser PM, Austad SN, Keane B. 1986. When Should Animals Tolerate Inbreeding? The American Naturalist, 128:529-537.

1821 Weir BS, Cockerham CC. 1984. ESTIMATING F-STATISTICS FOR THE ANALYSIS OF
1822 POPULATION STRUCTURE. Evolution, 38:1358-1370.

1823 Weir JT. 2006. DIVERGENT TIMING AND PATTERNS OF SPECIES ACCUMULATION IN
1824 LOWLAND AND HIGHLAND NEOTROPICAL BIRDS. Evolution, 60:842-855.

1825 Wen D, Yu Y, Zhu J, Nakhleh L. 2018. Inferring Phylogenetic Networks Using PhyloNet.
1826 Systematic Biology:syy015-syy015.

1827 Zamudio KR, Bell RC, Mason NA. 2016. Phenotypes in phylogeography: Species' traits,
1828 environmental variation, and vertebrate diversification. Proceedings of the National
1829 Academy of Sciences of the United States of America, 113:8041-8048.

1830 Zhu J, Wen D, Yu Y, Meudt HM, Nakhleh L. 2018. Bayesian inference of phylogenetic
1831 networks from bi-allelic genetic markers. PLOS Computational Biology,
1832 14:e1005932.

1833

1834

1835

1836

1837

1838

1839

1840

Figure 1. Neotropical areas of endemism and field sampling. 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' (Haffer 1978, Cracraft 1985, Haffer 1985, Cracraft and Prum 1988, 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 277 *Pseudopipra* individuals sequenced for genomic analysis in the present study.

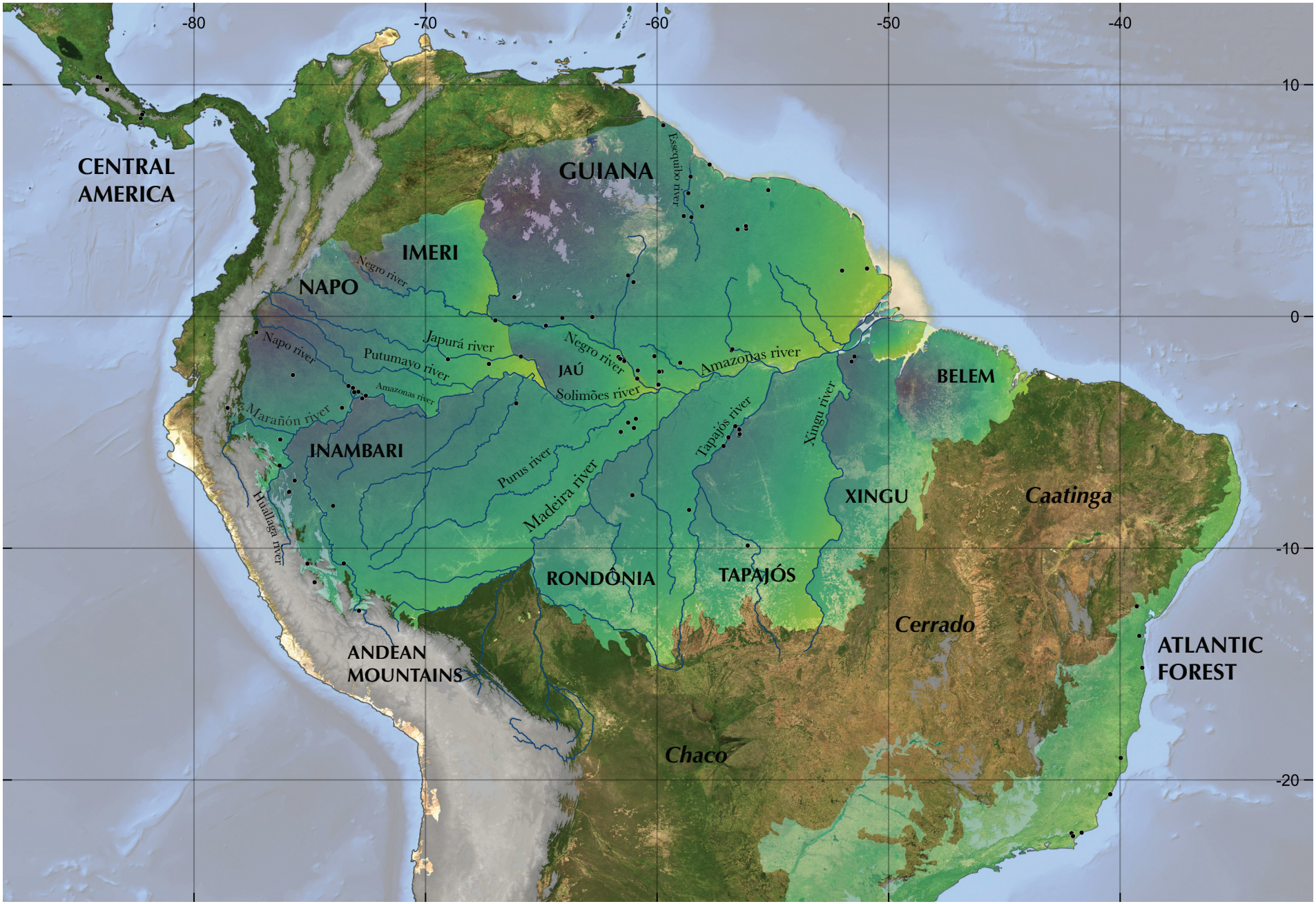
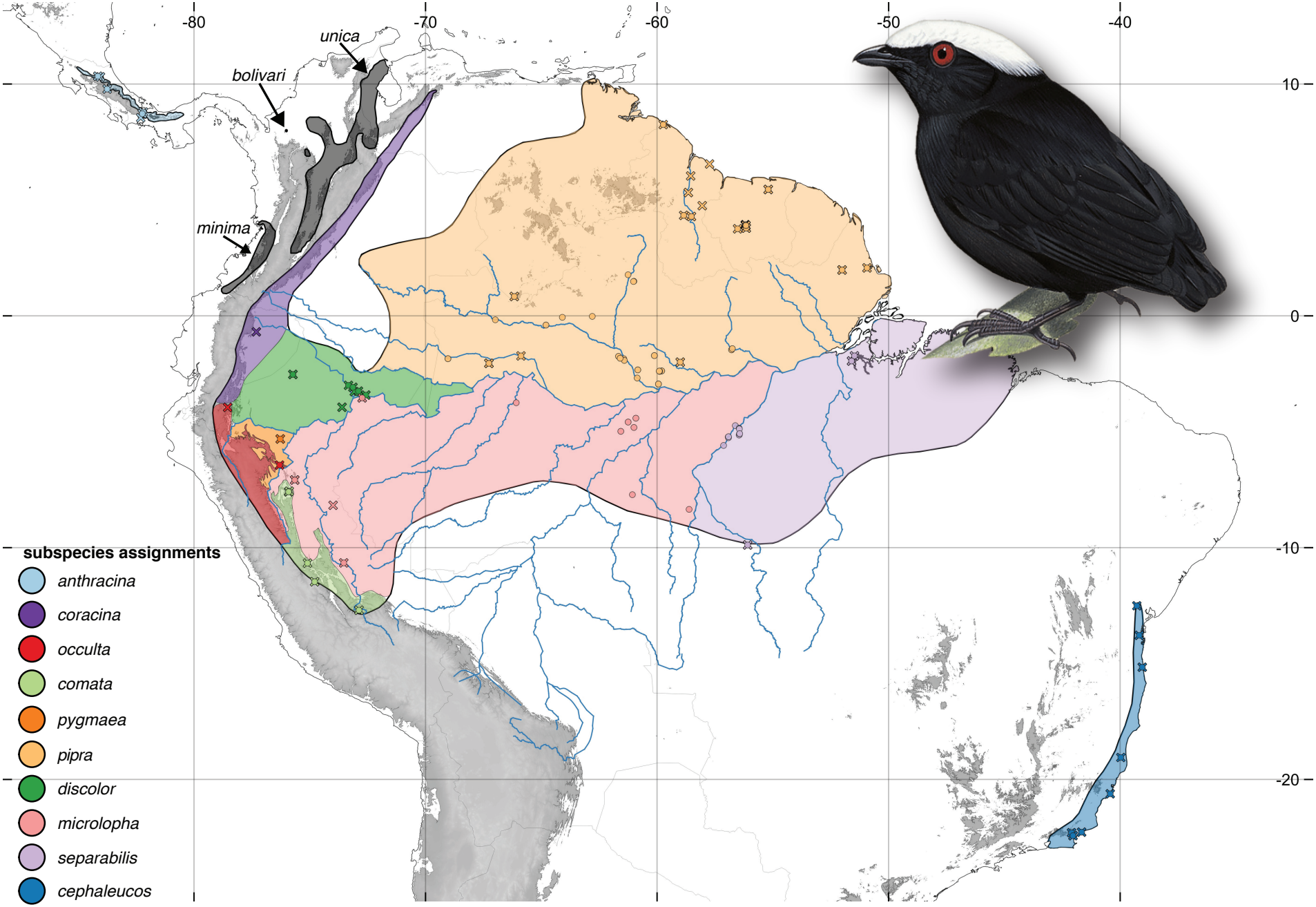
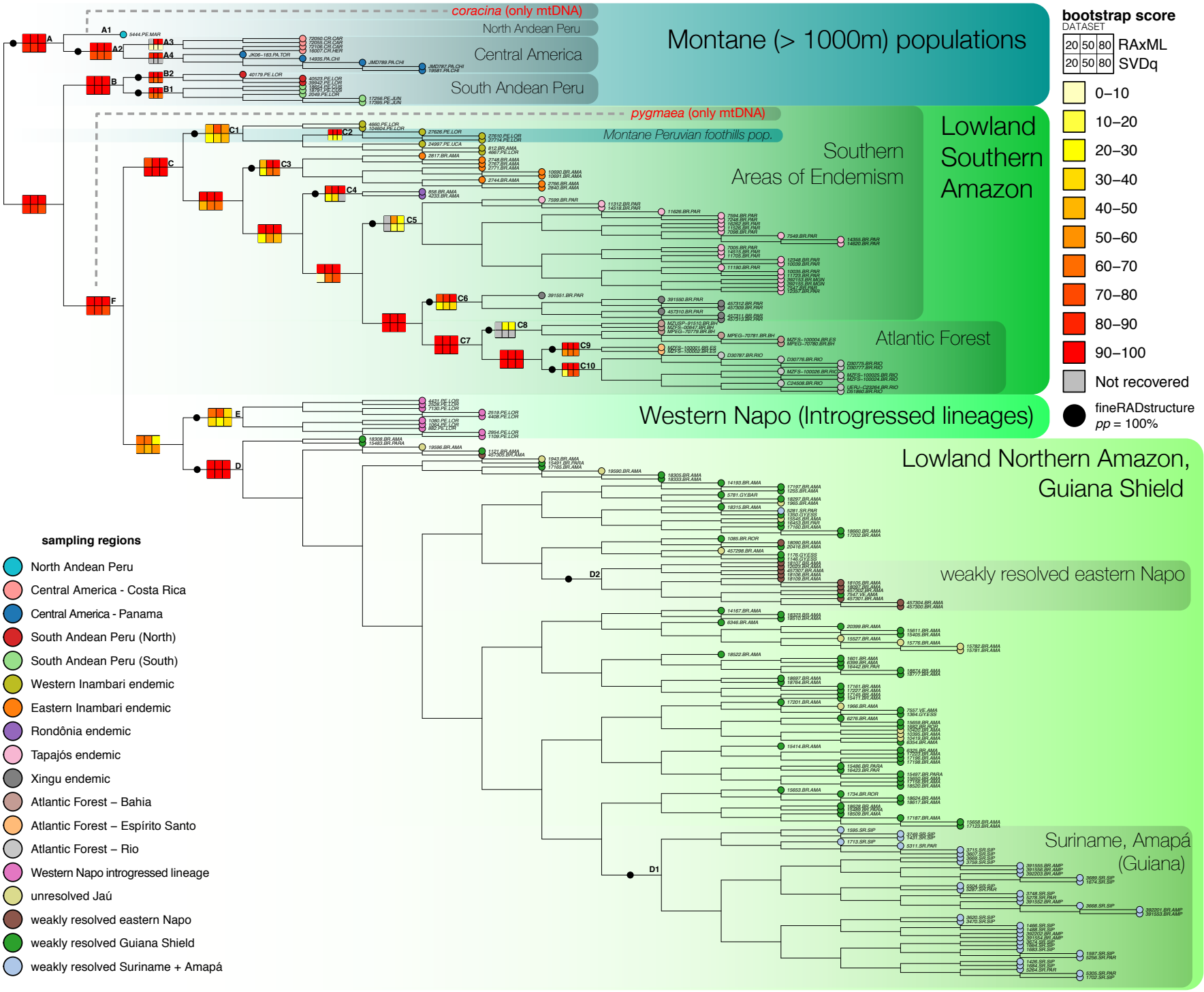

-80
-70
-60
-50
-40
10
0
-10
-20
CENTRAL AMERICA
GUIANA
Essequibo river
IMERI
Negro river
NAPO
Napo river
Japurá river
Putumayo river
Amazonas river
Negro river
JAÚ
Solimões river
Amazonas river
BELEM
Marañón river
INAMBARI
Purus river
Madeira river
Tapajós river
Xingu river
XINGU
Caatinga
Huallaga river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

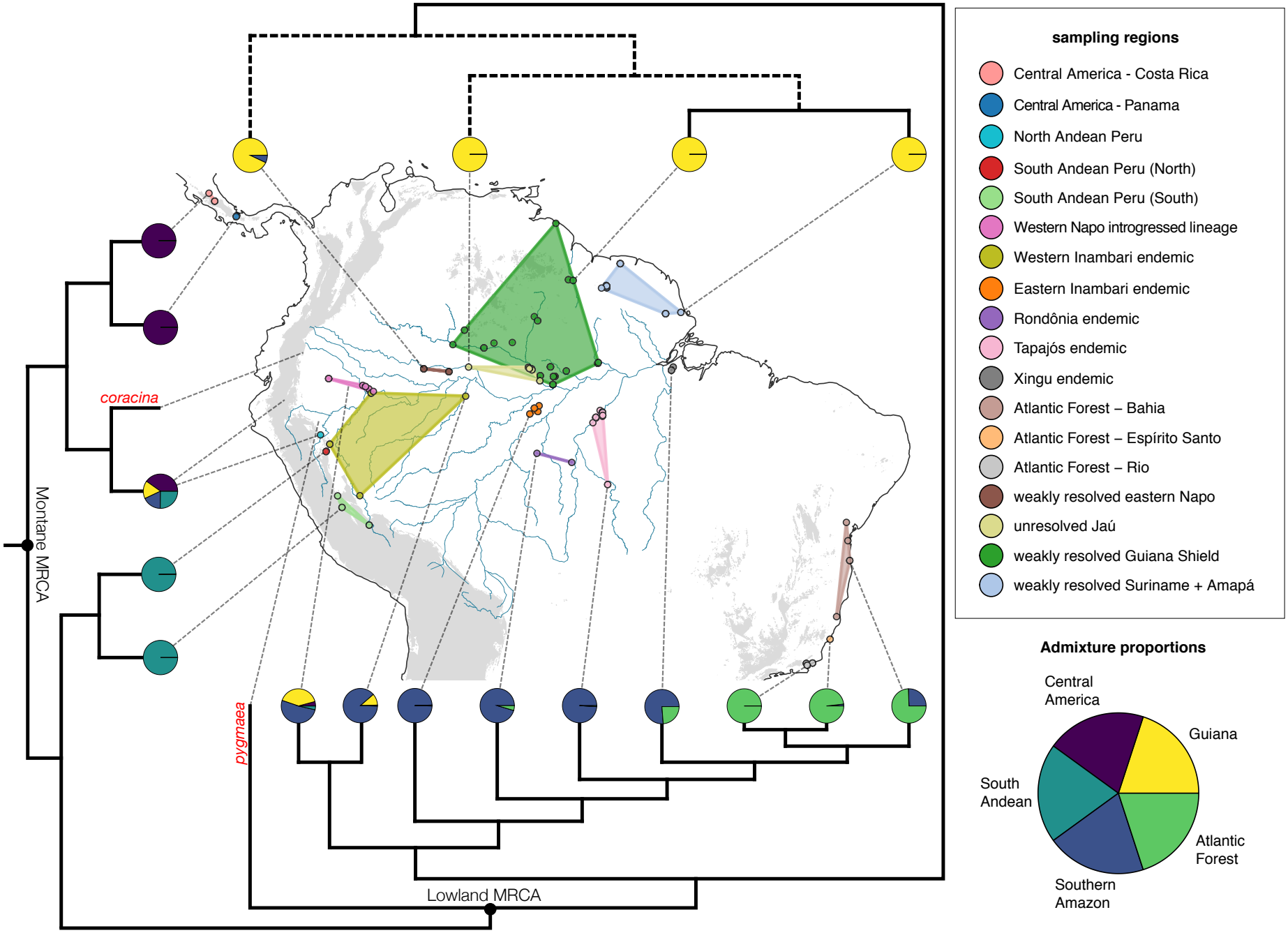
Figure 2. Subspecies sampling. Our sampling of *Pseudopipra pipra* intersects with the ranges of ten named subspecies. In this figure we color code our sampling localities as they overlap with described subspecies ranges. Areas greater than 1000m in elevation are shown in light grey; major Amazonian rivers are indicated in light blue; political boundaries in light grey. Outlined in black: genus range map reflecting the BirdLife International approximation updated to reflect current knowledge (BirdLife International Species factsheet: *Pseudopipra pipra*, 2018, also see Supplemental Figure 1). Hypothetical subspecies ranges are based on descriptions in Kirwan and Green (2012) and indicate maximal potential ranges to highlight proposed barriers delimiting subspecies. Circles indicate sampling for ddRADseq, X's indicate sampling for mtDNA ND2. Our sampling encompasses the majority of the lowland extents of the recognized range of *Pseudopipra*, as well as key montane populations in Peru, Ecuador, and Central America. Ssp *comata*, *anthracina,* and *occulta* are exclusively montane. 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
unica
bolivari
minima
subspecies assignments
anthracina
coracina
occulta
comata
pygmaea
pipra
discolor
microlopha
separabilis
cephaleucos

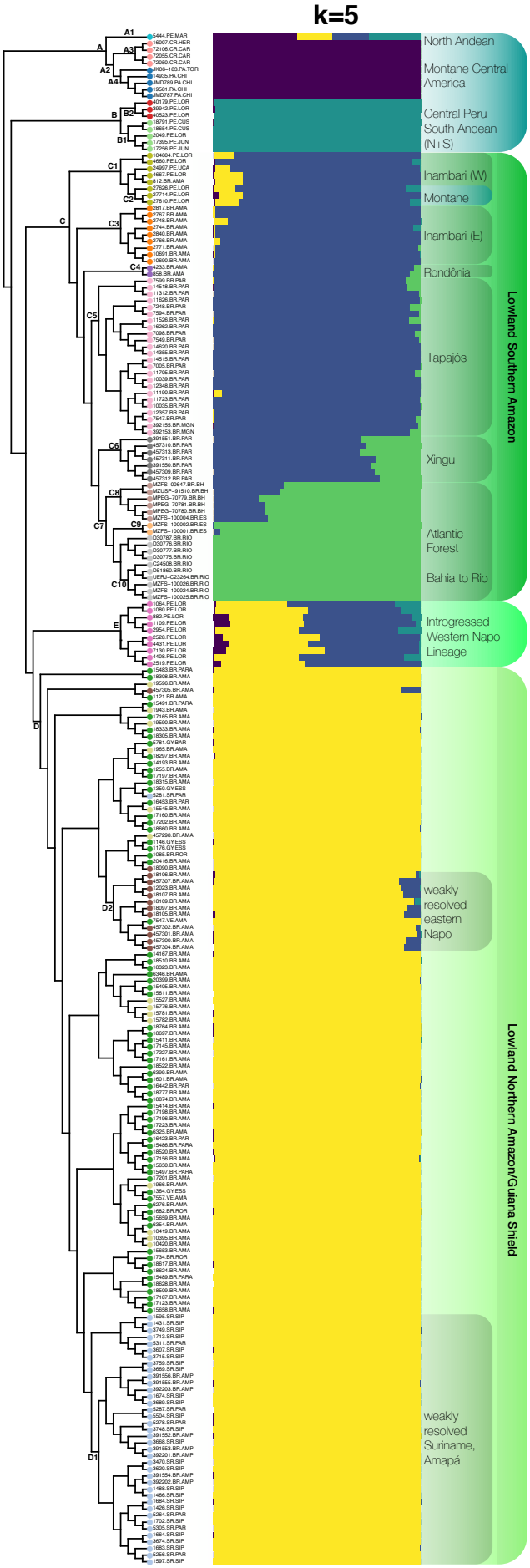
1885 **Figure 3.** Phylogenetic hypothesis for *Pseudopipra*. Phylogenetic analysis of short read
1886 ddRAD sequence haplotypes generated a well-resolved phylogenetic hypothesis which was
1887 largely congruent across datasets and analytical approaches. Montane clades, including a
1888 Northern Andean (A1) + Central American (A2) clade (A), and a central Peruvian clade (B),
1889 are recovered as nested sister groups to two wide ranging lowland clades (C and D),
1890 indicating that the lowland clades are descended from Andean lineages. Shown is the
1891 RAxML topology generated using the 50% haplotype dataset. Branch lengths are set to
1892 equal for graphical interpretability (see supplementary material for original newick
1893 formatted tree files). Colored circles at tips indicate group membership to one of eighteen
1894 population-areas (matching Figure 4 and 5). Heatmaps printed at well-supported nodes
1895 indicate bootstrap scores from each of six phylogenetic analyses for 20/50/80% datasets,
1896 as indicated in the legend. Black dots indicate populations which are identified with 100%
1897 posterior probability in fineRADstructure analysis. Most of the major and substructure is
1898 recovered by both RAxML and SVDquartets (see text), though RAxML recovers additional
1899 low-support substructure in the Northern Amazonian + Guiana Shield clade (D) which is
1900 coincident with several geographic features (clades D1 and D2). A clade of introgressed
1901 western Napo individuals (E) is recovered by all standard phylogenetic analyses as sister to
1902 the northern Amazonian + Guiana shield clade (D). As the history of this clade is
1903 characterized by a complex introgression scenario between northern and southern
1904 lineages, a purely bifurcating tree model is inadequate to describe its relations with other
1905 groups (see text) (Figure 6). Also shown are the well supported phylogenetic positions of
1906 two distinct subspecies lineages (*pygmaea* and *coracaina*) from which we were able to
1907 obtain mtDNA (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
Central America - Costa Rica
Central America - Panama
South Andean Peru (North)
South Andean Peru (South)
Western Inambari endemic
Eastern Inambari endemic
Rondônia endemic
Tapajós endemic
Xingu endemic
Atlantic Forest – Bahia
Atlantic Forest – Espírito Santo
Atlantic Forest – Rio
Western Napo introgressed lineage
unresolved Jaú
weakly resolved eastern Napo
weakly resolved Guiana Shield
weakly resolved Suriname + Amapá

1908 **Figure 4.** Phylogeospace mapping of topological relations in geographic space. Here, an
1909 idealized topology is plotted with tips attached to sampling coordinates. The colors of
1910 circles at sample localities matches those in Figure 3 and 5, and are enclosed by minimum
1911 convex hulls (i.e., hypotheses of minimum clade ranges) of the same colors. All of these
1912 colored hulls represent distinct non-overlapping geographic areas delimited by physical
1913 barriers (though this is somewhat visually confounded in the Northern Amazon + Guiana
1914 Shield clade (D), with two weakly resolved lineages (e.g. pale yellow Jaú and green Guiana
1915 Shield population groups appear to have some overlap in this figure, but they are delimited
1916 by the Rio Negro). Clade D bipartitions are shown with dashed branches (top) to indicate
1917 low support for phylogenetic structure (Figure 3, results). The ancestral habit of the genus
1918 is inferred to be montane (Supplementary Appendix), with a single origin of lowland
1919 lineages (*pp*=100). Also shown at the tips are median admixture proportions across
1920 sampling localities as inferred with STRUCTURE (Figure 5 for full detail).
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930

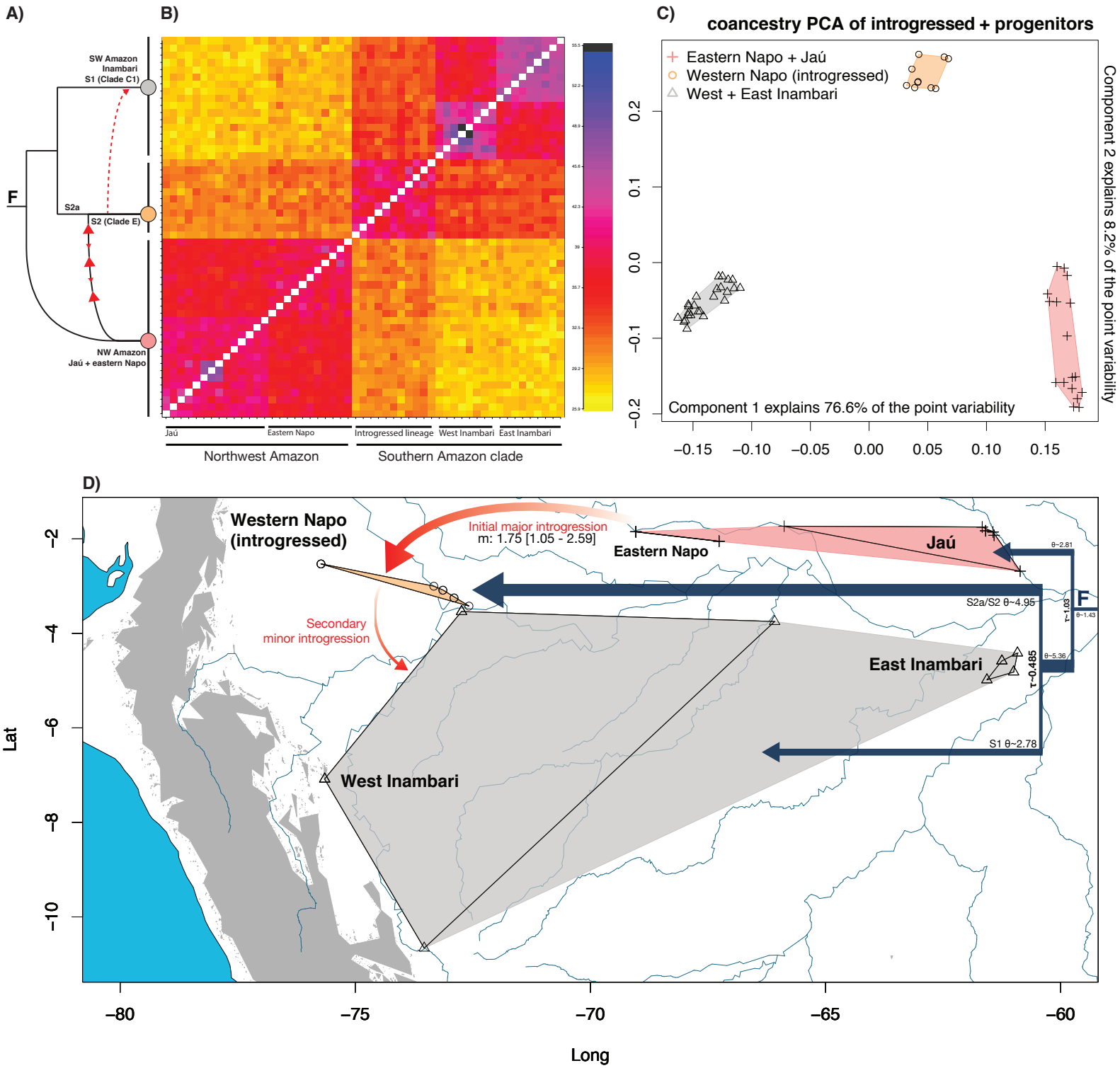

coracina
pygmaea
Montane MRCA
Lowland MRCA
sampling regions
Central America - Costa Rica
Central America - Panama
North Andean Peru
South Andean Peru (North)
South Andean Peru (South)
Western Napo introgressed lineage
Western Inambari endemic
Eastern Inambari endemic
Rondônia endemic
Tapajós endemic
Xingu endemic
Atlantic Forest – Bahia
Atlantic Forest – Espírito Santo
Atlantic Forest – Rio
weakly resolved eastern Napo
unresolved Jaú
weakly resolved Guiana Shield
weakly resolved Suriname + Amapá
Admixture proportions
Central America
Guiana
South Andean
Atlantic Forest
Southern Amazon

1931 **Figure 5.** STRUCTURE output for K-5. STRUCTURE analysis of dataset 1, depicting
1932 population assignment and admixture estimates for K=5. The likelihood of each evaluated
1933 number of K clusters from 1:20 plateaued at K = 5, with the standard deviation across runs
1934 increasing rapidly after this point (Supplemental Figure 2). The five optimized clusters
1935 broadly correspond to wide biogeographic Neotropical regions which coincide with
1936 lowland areas of endemism, Central American, and Peruvian montane regions (Figure 4).
1937 Given these five clusters, each labeled focal region is inferred to have a unique combination
1938 of admixture proportions. The admixture cline inferred from the Xingu to the southern
1939 Atlantic Forest appears to be a product of isolation by distance (Supplementary Appendix),
1940 while the signature of introgression inferred for western Napo lineages appears to be a
1941 product of substantial historical introgression from a northern lineage (likely from the
1942 poorly resolved eastern Napo group D2), into a distinct Southern Amazonian lineage (see
1943 Figure 6 and discussion). The tree to the left corresponds to the RAxML result using the
1944 50% haplotype dataset, with tip labels and colors indicating group membership to one of
1945 eighteen population-areas (matching those in Figures 3, 4). See Supplementary Figure 3 for
1946 K2-10 results.
1947
1948
1949
1950
1951
1952
1953


k=5
North Andean
Montane Central America
Central Peru South Andean (N+S)
Inambari (W)
Montane
Inambari (E)
Rondônia
Tapajós
Xingu
Atlantic Forest
Bahia to Rio
Lowland Southern Amazon
Introgressed Western Napo Lineage
weakly resolved eastern Napo
weakly resolved Suriname, Amapá
Lowland Northern Amazon/Guiana Shield
A
A1
A2
A3
A4
B
B1
B2
C
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
D
D1
D2
E

1954 **Figure 6.** Introgressed lineage in the western Napo area of endemism. All of the available
1955 genomic evidence is consistent with the hypothesis that individuals in the western Napo
1956 area of endemism are derived from a distinct southern Amazonian lineage (S2a in Panel A)
1957 which is not represented in contemporary samples. This lineage (S2a) experienced
1958 substantial introgression from the ancestral populations of what are today eastern Napo
1959 and Jaú restricted linages. Panel A shows the phylogenetic network (in black) inferred by
1960 PhyloNet, with arrows indicating hypothetical degree and direction of historical
1961 introgression as implied by patterns of admixture. The red arrows along the PhyloNet
1962 reticulation indicate inferred introgression from the northern Amazonian lineages (clade
1963 C1 in Figure 3) into this southern Amazonian lineage (S2a). The smaller red arrows
1964 indicate potential limited backcrossing into northern lineages implied by patterns of
1965 admixture. The dashed red line connecting S2 to S1 indicates minimal implied
1966 introgression into Inambari lineages, which may have occurred after the primary
1967 introgression event, and brought northern (Clade D) alleles into SW Amazonian
1968 populations (see text and Figure 5). Panel B displays the individual level co-ancestry matrix
1969 for introgressed Napo individuals and their source/progenitor lineages; darker colors
1970 indicate greater co-ancestry. Introgressed (PH: putative hybrid) hybrids have greater co-
1971 ancestry with each of their progenitor lineages (S2a and Northern Amazon+Guiana shield
1972 (abbreviated GS)) than their progenitor lineages share with each other
1973 (PHvGS/progenitors = 1.078, PHvInambari/progenitors = 1.17, see text). In the proposed
1974 scenario, the southern progenitor lineage (S2a) is either unsampled or extinct and has not
1975 persisted sympatrically with S2. Thus, the western Inambari southern lineage (S1)
1976 comprises the closest living relatives to S2a (see discussion). Panel C summarizes co-

ancestry variance as projected into component space after normalizing; the introgressed 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). G-PhoCS demographic parameter estimates are also indicated in panel D (m: migrants per generation, when lower 95% CI > 1, τ: splitting time in generations (10^6), θ: effective population size (10^6), Supplemental Table 5). In A-D, two subpopulations 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
Western Napo
(introgressed)
Initial major introgression
m: 1.75 [1.05 - 2.59]
Eastern Napo
Jaú
Secondary
minor introgression
S2a/S2 θ~4.95
F
East Inambari
S1 θ~2.78
West Inambari
Lat
Long

2000 **Figure 7.** Estimated Effective migration surface (EEMS). The estimated migration surface
2001 links the topography and drainage system of the Neotropics to spatial patterns of genetic
2002 diversity. Shown are results estimated from a model with 2000 demes (vertices). EEMS
2003 uses an MCMC approach to estimate expected genetic dissimilarity between all pairs of
2004 individuals, integrating over all possible migration histories across the grid, and adjusting
2005 migration rates among graph edges to match the genetic differences in the data. Bluer
2006 colors indicating areas where gene flow is more likely to be able to occur (i.e., isolation by
2007 distance is predicted), and browner colors indicate areas where barriers to gene flow are
2008 highly probable. In this analysis, the Andes and the Amazon River are the most clearly
2009 inferred barriers to gene flow. Major rivers, as well as a >1000m contour (light grey) are
2010 overlaid to aid interpretation. EEMS does a very good job of explaining spatial variance in
2011 this dataset (R^2 = 0.871, Supplementary figure 10, 11). There are areas where it performs
2012 poorly however, particularly in the western Amazon, where the inference of a major
2013 barrier to gene flow (the Andes) is inferred a few degrees too far east, pushing other
2014 western patterns eastwards. This is likely a consequence of uneven genetic sampling in this
2015 area. Remarkably, EEMS almost perfectly infers the spatial extents of the lowland
2016 Amazonian range (range map outlined in black), despite very limited genetic sampling
2017 toward the edge of the range. EEMS also correctly detects an isolated region of
2018 corresponding to the introgressed western Napo hybrid lineage. Also noted are likely
2019 geographic barriers that correspond to estimated barrier localities.
2020
2021
2022

Posterior mean migration rates

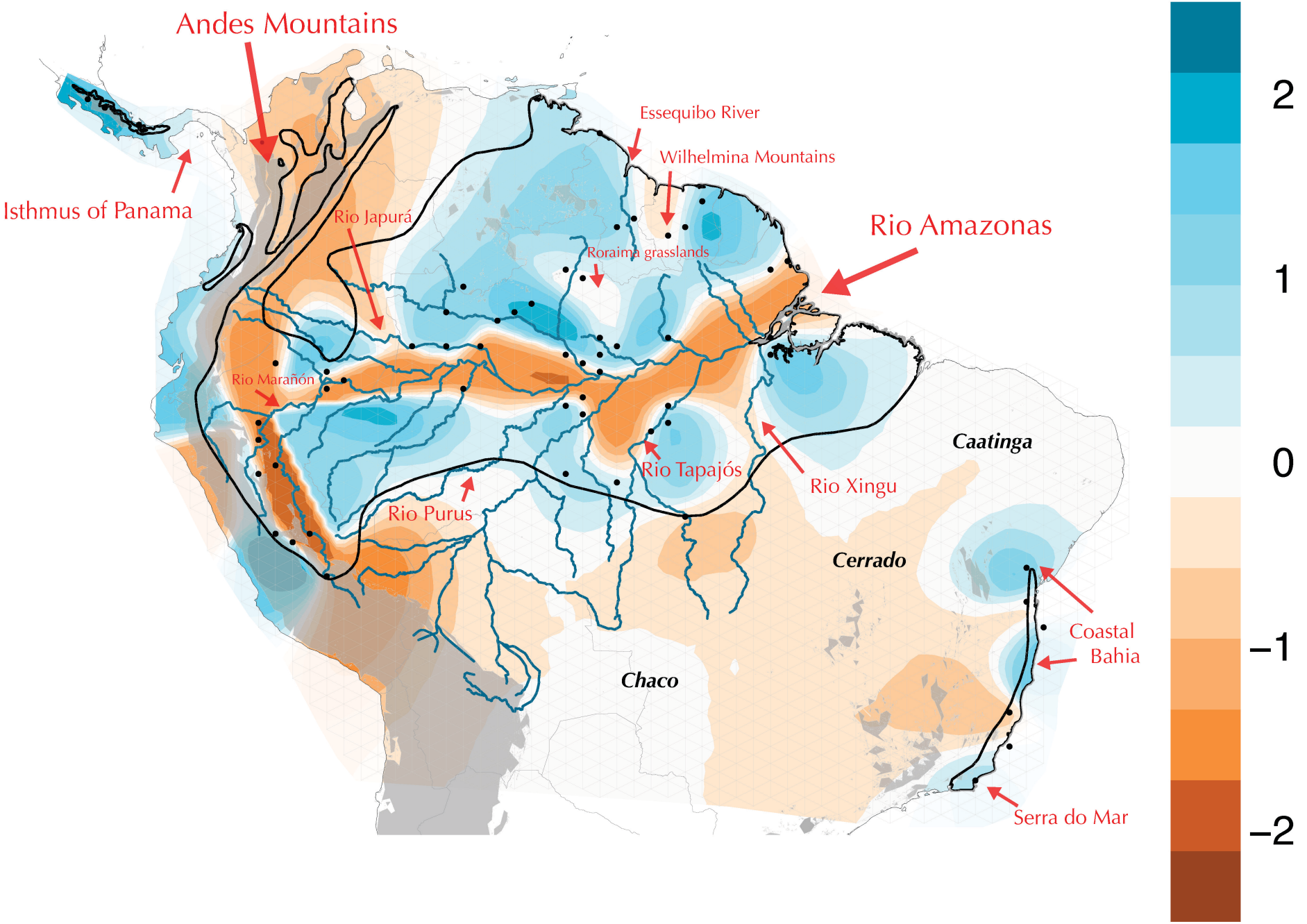


Figure 8. Summary of lek vocalization phenotypes. Our analysis identified 12 qualitatively distinct lek vocalization types that can be easily diagnosed by ear, or by visual inspection of sonograms (arbitrarily labeled numerically). See Supplemental Figure 13 for vocalization measures. The inset map indicates 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 pleisiomorphically, a pattern implied by one recording from Pará, eastern Brazil, which unambiguously clusters with other recordings from the western Napo area of endemism. Vocalization type 2 may therefore occur in other unsampled areas of the southern Amazon. Several vocalization types appear to be unique to a single monophyletic lineage based on overlapping geographic ranges. Vocalization type 4 coincides with the Central American lineage (Clade A2), type 3 with the Northern Amazon + Guiana Shield clade (Clade D) excluding eastern Napo, Jaú), type 7 with the Western Inamabri lineage (Clade C1), type 1 and 10 with the South Andean Peru clades (Clade B). In our initial assessment, we identified vocalization type 6, which has a geographic range which coincides with both the southern Tapajós clade and the Atlantic Forest clade. However, upon closer consideration, we found a subtle difference between the Atlantic Forest (6a) and Tapajós (6b) regions. Type 6a recordings, which are restricted to the Atlantic Forest, have a relatively constant pitch for both the buzz and tonal notes, whereas the type 6b recordings, which are restricted to the Tapajós regions, have a descending pitch for both notes. The sharing of a similar vocal phenotype (6a and 6b) clearly links the Atlantic forest populations to the Southern Amazon, in congruence with our genetic data (Clade C7), even though these southern Tapajós individuals have a genetic

affinity to other more northern Tapajós individuals. Lastly, we identified several lek vocalization types with restricted, non-overlapping geographic ranges in regions for which we lack genetic samples. Vocal type 8 was restricted to Subtropical forests of the eastern slope of the Andes from western Venezuela to Morona-Santiago, Ecuador. Vocal type 9 was restricted to Subtropical forests of western Cauca, Colombia south to Esmeraldas, Ecuador. Lastly, vocal types 11a and 11b were restricted to Subtropical forests of Magdalena Valley, Antioquia to Huila, Colombia. It is likely they fall within the range of variation delimited by the Central America + North Andean Peru clade.

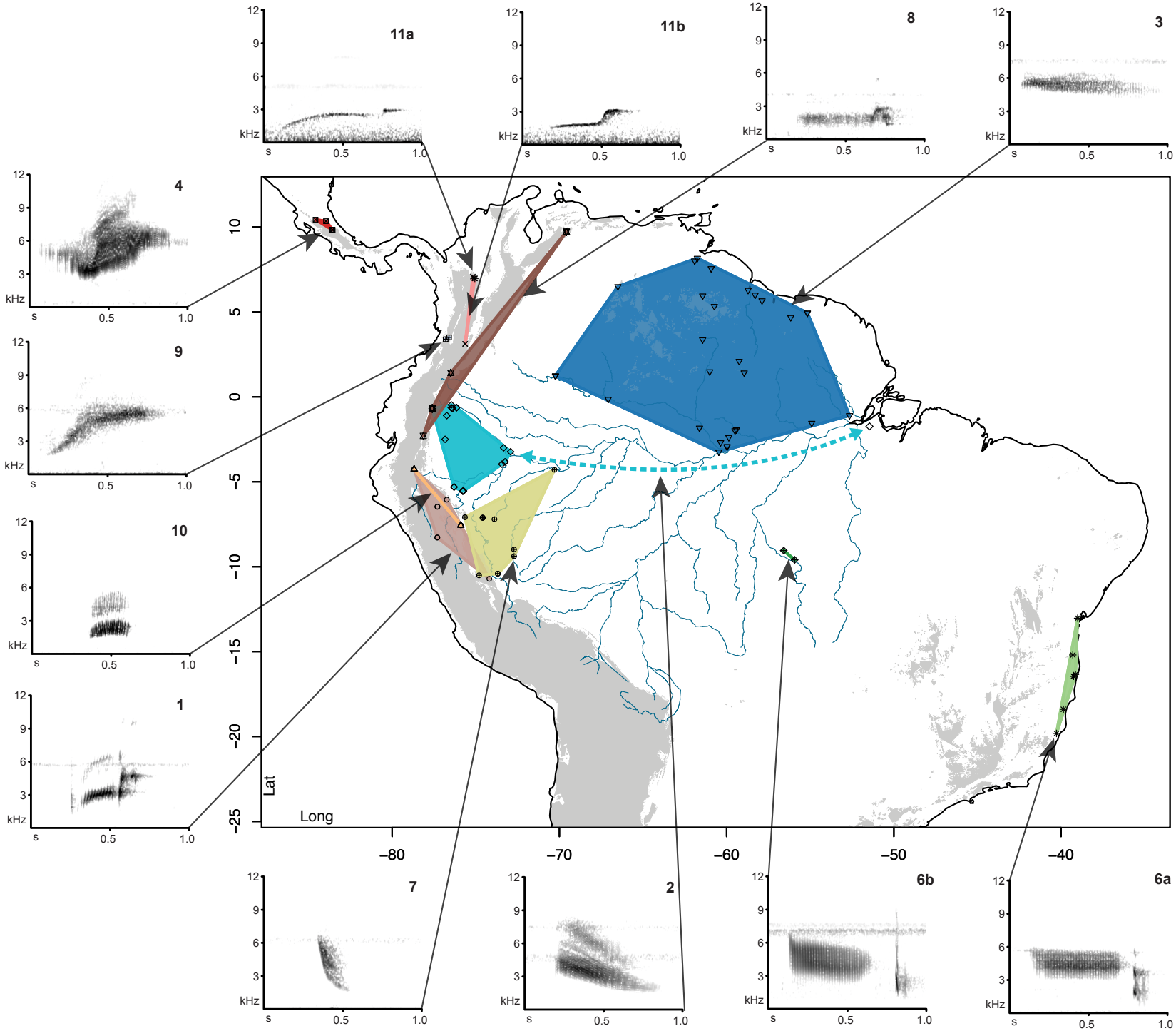
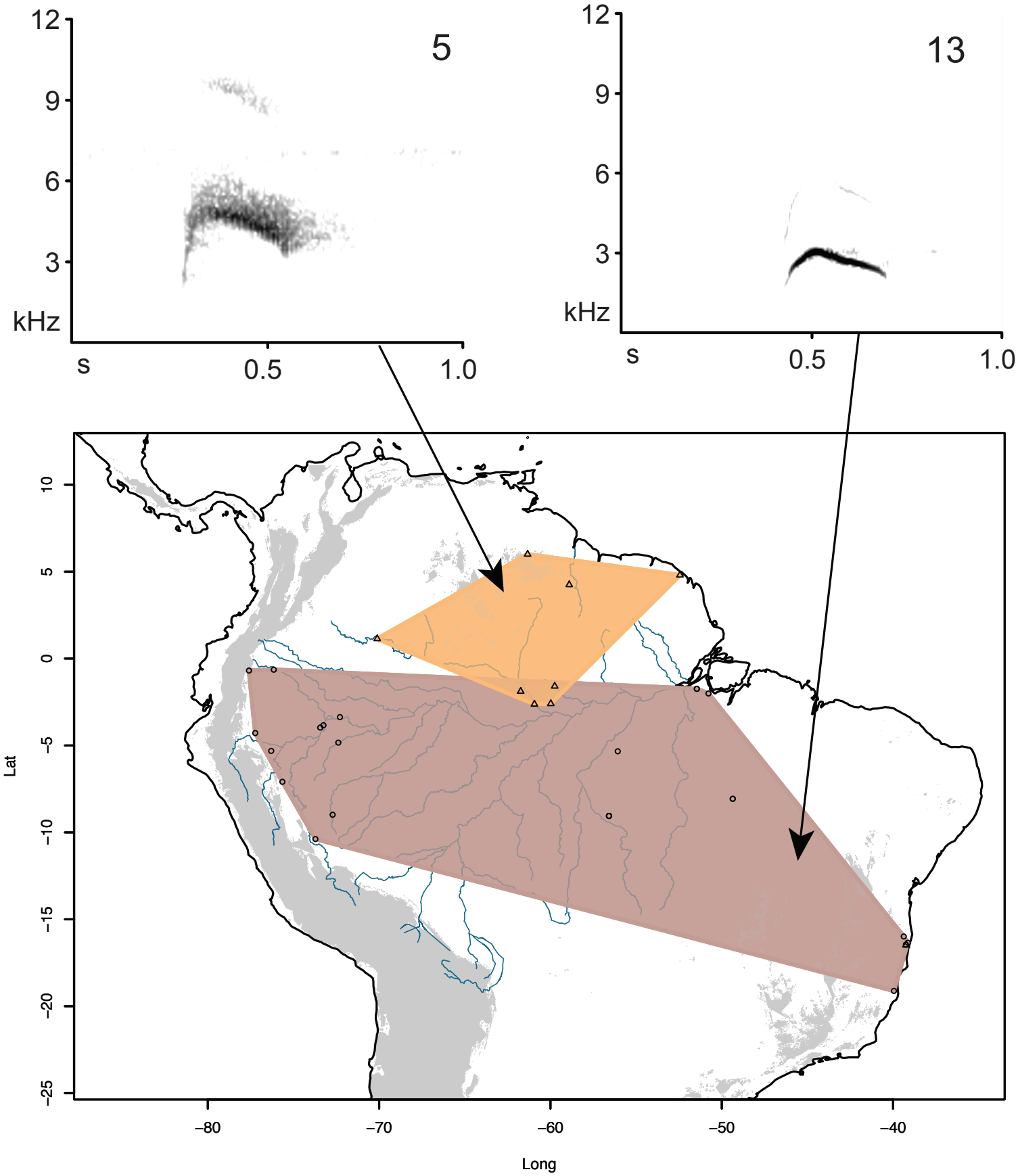

11a
11b
8
3
4
9
10
1
7
2
6b
6a
kHz
s
0.5
1.0
Lat
Long
10
5
0
−5
−10
−15
−20
−25
−80
−70
−60
−50
−40

Figure 9. Summary of call vocalization phenotypes. Our analysis identified 2 distinct but similar call vocalization types that can be diagnosed by ear, or by visual inspection of sonograms. See Supplemental Figure 13 for vocalization measures. The inset map indicates the minimum implied ranges for each identified call vocalization type, with colors and plotting symbols matching those in the PCA plot (Supplemental Figures 14, 15). 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 unbiased classification, one recording from the Atlantic Forest (XC427315) was assigned to type 5, and is denoted here with a triangle symbol, suggesting type 5 may be shared between the Guiana Shield clade and the Atlantic Forest clade. Call vocalization type 13 was entirely restricted to but shared across diverse lowland southern Amazonian forms (Clade C).


5
13
12
9
6
3
kHz
s
0.5
1.0
Lat
Long
10
5
0
−5
−10
−15
−20
−25
−80
−70
−60
−50
−40