Title: History of divergence and gene flow shaping geographic variation in Andean warblers (*Myioborus*)

Laura N. Céspedes Arias[1,2*], Andrés M. Cuervo[3], Carlos Daniel Cadena[4], Elisa Bonaccorso[5], Christopher Witt[6], Irby Lovette[7,8] & Leonardo Campagna[7,8]

[1] Committee on Evolutionary Biology, The University of Chicago, Chicago, IL, USA

[2] Birds Collection, Field Museum of Natural History, Chicago, IL, USA

[3] Instituto de Ciencias Naturales, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia

[4] Departamento de Ciencias Biológicas, Universidad de los Andes, Bogotá, Colombia

[5] Laboratorio de Biología Evolutiva, Colegio de Ciencias Biológicas y Ambientales, Universidad San Francisco de Quito, Quito, Ecuador

[6] Museum of Southwestern Biology, University of New Mexico, Albuquerque, NM, USA

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

[8]Department of Ecology and Evolutionary Biology, Cornell University, 215 Tower Road, Ithaca, NY 14853, US

Corresponding authors: Laura N. Céspedes Arias (lncespedesarias@gmail.com)

Abstract

Studying how genetic variation is structured across space, and how it relates to divergence in phenotypic traits relevant to reproductive isolation, is important for our overall understanding of the speciation process. We used reduced-representation genomic data (ddRAD-seq) to examine patterns of genetic variation across the full distribution of an Andean warbler species complex (*Myioborus ornatus–melanocephalus)*, which includes a known hybrid zone between two strikingly different plumage forms. Genetic structure largely reflects geographic variation in head plumage, some of which corresponds to major topographic barriers in the Andes. We also

found evidence of isolation by distance shaping genetic patterns across the group's broad latitudinal range. We found that *chrysops* and *bairdi*, two taxa with marked plumage differences that have a known hybrid zone, were characterized by low overall genetic divergence. Based on our cline analyses of both plumage and genomic hybrid indices, this hybrid zone extends for approximately 250 km, where advanced generation hybrids are likely most common. We also identified a slight difference in the centers of the plumage and genomic clines, potentially suggesting the asymmetric introgression of *chrysops*-like plumage traits. By studying genetic variation in a phenotypically complex group distributed across a topographically complex area, which includes a hybrid zone, we were able to show how both geographic features and potentially sexually selected plumage traits may play a role in species formation in tropical mountains

Keywords: *phylogeography, hybrid zone, reduced representation genomic data, geographic variation, Parulidae.*

Introduction

Geographic variation in traits relevant to reproductive isolation is often viewed as a snapshot of the early stages of population divergence (Mayr, 1942; Price, 2008), making it a focus for understanding the process of speciation. To study this process, it is useful to evaluate how phenotypic variation in such traits is related to the underlying population genetic structure (Edwards et al., 2016; Zamudio et al., 2016) A common expectation is that phenotypic and genetic divergence are positively correlated: in the absence of strong divergent selection, extended periods of isolation are typically required for phenotypic differentiation to evolve (Nosil, 2008; Price, 2008; Winger & Bates, 2015). It follows that range disjunctions associated

with topographic and ecological barriers likely to have restricted historical gene flow are expected to be accompanied by genetic and phenotypic discontinuities (Edwards et al., 2016). However, geographic variation in phenotypic traits is not always concordant with genetic structure. This discordance could be driven, for example, by strong selection on some traits across environmental gradients over short time scales (Walsh et al., 2017), or complex histories of selection and gene flow leading to phenotypic convergence (Márquez et al., 2020). A better understanding of how species diversify across space requires studying such histories of isolation, and how these relate to divergence in traits, particularly those that might be important for mate choice and, consequently, reproductive isolation.

Patterns of plumage differentiation—thought to be important for reproductive isolation in birds (Price, 2008; Uy et al., 2018)— among allopatric populations within species are pervasive among Andean birds (Graves, 1988; Remsen, 1984). Although geographic variation in plumage and genetic structure in Andean birds is often associated with current topographic and ecological barriers (Gutiérrez-Pinto et al., 2012; Prieto-Torres et al., 2018), many of these barriers tend to be permeable and unstable. Dispersal over such barriers could potentially lead to introgressive hybridization that connecting otherwise diverging populations (Winger, 2017). Additionally, differentiated populations can come into renewed contact when the barriers themselves change over time. In the particular case of the Andes, climatic fluctuations (Ramírez-Barahona & Eguiarte, 2013) and, potentially, volcanic activity (Sanín et al., 2024) have promoted cycles of fragmentation and reconnection of elevational vegetation belts, creating opportunities for contact among periodically isolated populations (Vuilleumier, 1969). Examining these complex histories of divergence and gene flow in geographically variable and widely distributed Andean taxa

contributes to a more comprehensive understanding of how bird species are formed and maintained in this biodiversity hotspot (Rahbek et al., 2019; Sonne et al., 2022).

Here, we used a reduced representation genomic data set from an Andean bird species complex of *Myioborus* warblers (Parulidae) characterized by striking geographic variation in head color patterns to gain insights into differentiation in geographically isolated populations and the processes that occur after differentiated populations come into secondary contact. This young (<1 my; Cuervo, 2013) species complex inhabits the high Andes from Bolivia to Venezuela, and is characterized by differences in head color patterns despite shallow genetic differentiation in mitochondrial DNA (Céspedes-Arias et al., 2021; Pérez-Emán, 2005). Seven plumage groups have been described as subspecies belonging to two species, *Myioborus ornatus* and *M. melanocephalus*, distributed mostly allopatrically along different sections of the Andes (Cuervo & Céspedes Arias, 2023; Zimmer, 1949). According to this classification scheme, *M. ornatus* includes subspecies *ornatus* and *chrysops*, and *M. melanocephalus* includes subspecies *bairdi, griseonuchus, malaris, melanocephalus* and *bolivianus*. Recent taxonomic revisions, however, suggest that a four-species classification scheme might better capture what is known about genetic and plumage variation in this complex (Cuervo & Céspedes Arias, 2023). Regardless of the species delimitation scheme, analyses based on the ND2 mitochondrial gene have revealed extensive haplotype sharing among groups with conspicuous plumage differences, potentially suggesting a history of rapid plumage evolution or repeated mitochondrial introgression (Céspedes-Arias et al., 2021).

Our previous work in regions where the ranges of *chrysops* and *bairdi* abut, on the border between Colombia and Ecuador, revealed that these two taxa hybridize widely across more than 200 km where individuals with a variety of intermediate head color patterns occur (Céspedes-

Arias et al., 2021). Although hybridization there has been extensive, the existence of a sharp plumage cline, limited to a relatively narrow zone within these taxa's broader ranges (20–30%), suggests the action of some form of selection maintaining their differentiation (Barton & Hewitt, 1985; Brelsford & Irwin, 2009; Harrison, 1986). Here, we leveraged this previous systematic collection and detailed plumage characterization of the *chrysops-bairdi* hybrid zone (Céspedes-Arias et al., 2021) to gain a deeper understanding of hybridization dynamics using genomic data.

We examined patterns of genomic variation within this warbler species complex at two different spatial scales: the entire distribution of the species complex, which corresponds to a broad latitudinal band in the Andes (approximately 3800 km in geographic distance), and the hybrid zone existing between two markedly different plumage groups as outlined above. By focusing on these two different scales, we can address questions related to multiple processes relevant to the formation and maintenance of species. First, leveraging a broad geographic scale sampling of individuals representing all plumage forms in the complex, we asked: (1) How is genetic structure related to geographic distance, topography, and plumage variation along the Andes? At a finer spatial scale, in the hybrid zone between two plumage forms, we asked: (2.1) Does the geographic pattern of introgression mirror plumage variation across the hybrid zone? And (2.2) is there evidence for non-neutral processes such as selection against hybrids based on patterns of genomic variation? Our genome-wide data coupled with dense geographic sampling substantially improves our ability to reconstruct and understand the diversification history of these colorful montane warblers, which was previously limited to studies using only a single mitochondrial gene.

Methods

Study species and sampling

We used tissue samples (*n*=227, Figure 1) collected across the distribution of the *M. ornatus-melanocephalus* complex, by us (Céspedes-Arias et al., 2021), or previously deposited in museum collections (Supplementary Table 1). Samples encompassed all (seven) currently taxonomically recognized plumage forms and a denser sampling in the *bairdi-chrysops* hybrid zone area, along the border between Colombia and Ecuador (Figure 1A). We also obtained data for eight individuals as outgroups corresponding to three other species in *Myioborus*: *M. pictus* (n = 3), *M. miniatus* (n = 3) and *M. brunniceps* (n = 2) which vary in their phylogenetic and geographic proximity to the *M. ornatus-melanocephalus* complex (Lovette et al., 2010; Pérez-Emán, 2005).

The more conspicous differences among plumage groups in the *ornatus-melanocephalus* complex correspond to the presence/absence and extent of colored plumage patches in the head, as illustrated in Figure 1B. We note that there are different degrees of plumage differentiation and diagnosability based on our own assessments (Cuervo & Céspedes Arias, 2023). For example, the subspecies *malaris, bolivianus* and *melanocephalus* are very similar in plumage, and some of the described diagnostic traits do not seem to vary in a way that is discrete and corresponds with these groups (Cuervo & Céspedes Arias, 2023).

We also note that the geographic limits of some groups are also not clearly defined, specifically of the *bairdi* and *griseonuchus* groups which have only subtle differences in in the amount of yellow vs black on the sides of the face, as well as the presence and extent of a black nuchal band behind the rufous crown (Chapman, 1927; Cuervo & Céspedes Arias, 2023; Zimmer, 1949). Although birds from northern Peru (Dept. of Cajamarca) and central Ecuador are clearly distinct (Supplementary Figure 1), it is unclear whether a phenotypic transition is discrete

123 or clinal across intervening southern Ecuador and there is a lack of sampling across this area to
124 evaluate these possibilities. However, to inform our discussion on patterns of genetic variation,
125 we inspected most of the available specimens collected in this area, either directly or through
126 photographs, to describe variation in these two plumage traits (Supplementary Table 2 and
127 Supplementary Figure 1).
128
129 **Laboratory procedures**
130 ***DNA extraction and library preparation***
131 We extracted genomic DNA from pectoral muscle (n=232) and blood samples (n=3) employing
132 a modified phenol-chlorophorm protocol (Sambrook, 1987), using phase-lock tubes. Each DNA
133 extraction was subsequently purified using homemade Sera-Mag Magnetic Speed-beads for
134 magnetic nucleic acid extraction (MagNA). To obtain reduced representation genomic data, we
135 used double-digest restriction site-associated DNA sequencing (ddRAD) following Peterson et
136 al. (2012), and using modifications from Thrasher et al. (2018). The genomic libraries were then
137 sequenced on one HiSeq2500 Illumina lane (100 bp SE) by the Genomics Core Facility of
138 Cornell University.
139
140 **Bioinformatics pipeline**
141 ***Quality assessment, filtering and demultiplexing***
142 We assessed sequence quality using FastQC (Andrews, 2010). We used the *process_radtags*
143 function implemented in STACKS 2.66 (Rochette et al., 2019) to demultiplex the reads and
144 apply quality filters. Specifically, we filtered out all reads with uncalled bases (*--clean*), those
145 with low phred quality scores *(--quality*, *--score-limit* 20) and those that did not pass the Illumina

146 chastity/purity filter (*--filter-illumina*). After demultiplexing, we checked the number of reads
147 available per individual, and only kept those samples with 200,000 or more reads
148 (Supplementary Figure 2). Using this threshold, we discarded 11 individuals, mostly
149 corresponding to the *ornatus* group and tissues collected around ten years prior to this study
150 (range of number of reads for these individuals = 960–150,152). Despite discarding these
151 samples, 23 high-quality samples for the *ornatus* group remained in downstream analyses. After
152 excluding these individuals from downstream analyses, we kept a total of 242,204,894 reads
153 corresponding to 224 individuals (three represented twice in the library).
154
155 ***Integrated assembly and SNP calling***
156 After demultiplexing, we assembled the RAD reads *de-novo* using functions implemented in
157 STACKS 2.66 (Rochette et al., 2019), and later integrated position information for the loci in the
158 catalog from a reference genome (Paris et al., 2017; Rivera-Colón & Catchen, 2022). This
159 integrated approach performs better than directly mapping reads onto a reference genome in
160 some scenarios, for example, if the reference genome is relatively distant to the study species as
161 is our case (Paris et al., 2017). We used the chromosome-level assembly of the *Setophaga*
162 *coronata* genome (Baiz et al., 2021) as reference, the highest quality available genome within
163 Parulidae. The time to the most recent common ancestor between *Setophaga* and *Myioborus* is
164 estimated between 5–7 mya (Barker et al., 2015).
165
166 ***Optimization of de novo assembly parameters***
167 Before assembling the reads, we conducted an optimization of the parameters *M* and *n* used by
168 *denovo_map.pl* (Rochette et al., 2019), for three different assemblies: one including

169 representatives from the outgroups, one that only included individuals within the *ornatus-*
170 *melanocephalus* complex, and one including only samples relevant for the hybrid zone analyses
171 (*chrysops*, individuals in the hybrid zone, *bairdi,* and *griseonuchus*, Figure 1) (see
172 Supplementary Material for details).
173
174 ***Full de novo assembly run***
175 We then ran *denovo_map.pl* for the entire dataset with the chosen parameter values, using again
176 the three population maps (n = 224 full assembly, n = 216 ingroup only assembly, and n = 123
177 for the hybrid zone assembly). For the population maps corresponding to the three different
178 assemblies, we classified individuals into one of the following groups: *ornatus*, *chrysops*,
179 "hybrid zone", *bairdi, griseonuchus, malaris, melanocephalus*, and *bolivianus* (populations_2
180 column in Supplementary Table 1), plus *pictus, miniatus*, and *brunniceps* for the assembly
181 including the outgroups. For the population map purposes the hybrid zone group corresponds to
182 a geographic group encompassing populations in and near the hybrid zone, irrespective of the
183 plumage phenotype of the specimens (see the map in Figure 1).
184 After the assemblies were completed, we assessed the per sample coverage and
185 percentage of reads incorporated (Rivera-Colón & Catchen, 2022). The assembly corresponding
186 to only individuals within the *melanocephalus-ornatus* complex resulted in a catalog of 91,650
187 loci, with a mean effective coverage per sample of 59.1x (SD = 13.2x, min = 18.3x, max = 97.4).
188 The "hybrid zone" assembly resulted in a catalog of 70,281 loci, with a mean effective coverage
189 per sample of 61.6x (SD =12.1x, min = 22.5x, max = 99.4x). The assembly including outgroups
190 resulted in a catalog of 97,533 loci, with a mean effective coverage per sample of 59.3x (SD =
191 13.2x, min = 18.1x, max = 98.5x). Four individuals with a mean effective coverage lower than

192 20x in assemblies in which they were included initially were excluded from downstream
193 analyses (see Supplementary Table 1; Supplementary Figure 3).
194
195 ***Mapping to the reference genome***
196 After obtaining the *de novo* catalog, we mapped loci to the reference genome by first indexing
197 the *S. coronata* genome using *bwa index* (Li & Durbin, 2009). Then we used *bwa mem* to align
198 the *de novo* catalog to the reference genome. Finally, we used the *stacks-integrate-alignments*
199 script to generate a catalog with position information (Rivera-Colón & Catchen, 2022; Rochette
200 et al., 2019). After filtering based on mapping quality and alignment coverage (using default
201 values), 60,523 loci (66% of the total number of loci in *de novo* catalog) were retained for the
202 ingroup assembly, 46,036 (66% of the total number of loci in *de novo* catalog) were retained for
203 the "hybrid zone" assembly, and 65,537 (67% of the total number of loci in *de novo* catalog) for
204 the full assembly.
205
206 ***SNP datasets***
207 We obtained different SNP datasets to be used in downstream analyses using the *populations*
208 program from STACKS 2.66 (Rochette et al., 2019). For some analyses, as specified below, we
209 did further filtering of SNPs using *vcftools* (Danecek et al., 2011). For the ingroup assembly, we
210 obtained two different SNP datasets: one including one SNP per locus, and one that included all
211 variant and non-variant sites per locus (i.e., full haplotypes only for fineRADStructure, see
212 below). In both cases we applied the following filters: *--min-samples-overall* 0.8 (locus needs to
213 be present in 80% of all individuals to be retained), *--min-mac* 5 (minimum allele count of 5),
214 and *--max-obs-het* 0.7 (maximum observed heterozygosity of 0.7). For the full haplotype dataset,

215 we used the same parameters and included the *--filter-haplotype-wise* flag. We decided to apply
216 a minimum allele count (*--min-mac* 5) filter to exclude SNPs that may be artifacts of genotyping
217 errors, but kept a relatively low value to avoid excluding all low frequency SNPs which also
218 contain relevant information for population genetic analyses (Linck & Battey, 2019). Details on
219 how the SNPs datasets from the three different assemblies were obtained are in the
220 Supplementary information. In short, we used the same filtering parameters for the hybrid zone
221 assembly and for the assembly including outgroups we used *--min-mac 2* and *--min-samples-per-*
222 *pop* 0.5

223 Based on the assembly for the ingroup only, and after filtering, we obtained 5,054 loci,
224 composed of 73,786 sites, including 22,132 variant sites for the full haplotype dataset. For the
225 ingroup (one SNP per locus) dataset, we obtained the same number of loci and kept 4,800 variant
226 sites. The number of SNPs for the other two assemblies (i.e., including outgroups and for the
227 hybrid zone only), as well as for different filtering employed for specific analyses is fully
228 detailed in the Supplementary Table 2 [table needs updating].
229

230 **Analyses of diversification within the *ornatus-melanocephalus* complex**

231 We described and visualized genetic structure in order to test its relationship with plumage and
232 geography. To evaluate how genetic structure related to current taxonomic patterns of variation
233 in head color patterns (see Figure 1), we also inferred subspecies-level phylogenies and
234 admixture graphs. Additionally, we estimated the level of differentiation between plumage
235 groups in terms of F_{ST}. Lastly, we used a spatially explicit model to evaluate how genetic
236 variation within the *ornatus-melanocephalus* complex is structured across their range.
237

238 ***Phylogenetic and admixture graph inference***

239 With the goal of inferring the overall relationships among plumage groups, we employed two
240 methods to build phylogenies and infer admixture graphs. For this set of analyses, we excluded
241 all individuals from or adjacent to the hybrid zone (in white, Figure 1) and the *griseonuchus-*
242 *bairdi* transition area (arrows, Figure 1), and used the assembly that includes outgroups (*M.*
243 *pictus, M. miniatus, M. brunniceps*). For phylogenetic approaches, we included all outgroup
244 individuals (total n=110), while for the admixture graph inference, we only included one
245 outgroup (total n=105).

246

247 **raxml-ng:** We used RAxML-NG (Kozlov et al., 2019) to infer maximum-likelihood trees from
248 concatenated ddRAD loci, including both variant and invariant sites. Because individual
249 fragments are short (~150 bp), we concatenated all loci into a supermatrix (Leaché & Oaks,
250 2017). A VCF including invariant sites was generated with *populations* using the parameters --
251 *vcf-all --min-mac* 2 *--min-populations* 10 *--min-samples-per-pop* 0.5, then converted to Nexus
252 with *vcf2phylip*. (https://github.com/edgardomortiz/vcf2phylip). Then, we ran *raxml-ng* using
253 this nexus alignment and the GTR + G model of nucleotide substitution with no partitions. To
254 search for the tree with the highest likelihood, we ran 20 independent runs, 10 of them starting
255 from the maximum parsimony tree and 10 from random trees. From these searches, the tree with
256 the highest likelihood was selected among the 20 slightly different trees. We calculated branch
257 support with 100 bootstrap replicates.

258

259 **snapper:** To infer evolutionary relationships among plumage groups under a phylogenetic
260 framework, we used *snapper* (Stoltz et al., 2021). *Snapper* infers a "species" tree using diffusion

261 models and takes as input biallelic SNPs that are assumed to be unlinked, bypassing the necessity
262 of building individual gene trees, similar to site-based coalescent methods (Bryant et al., 2012).
263 We classified individuals into groups corresponding to the seven subspecies described in Figure
264 1. For this analysis, we generated a SNP dataset from the assembly including outgroups (*--min-*
265 *mac* 2, *--max-obs-het* 0.7, --min-populations 10, *--min-samples-per-pop* 0.5) filtered to only keep
266 biallelic SNPs using *vcftools* and then converted the resulting *vcf* files to nexus using the
267 *vcf2phylip* script. Using the nexus alignments as input, we prepared the *.xml* files in Beauti 2.7.6
268 (Bouckaert et al., 2019). We ran four independent chains with a length of 15,000,000 steps, and a
269 thinning interval of 1,500. We set a gamma prior for the birth rate (λ), and left default values for
270 the remaining priors. We increased the weights of two operators (gamma mover and rate mixer)
271 to 10. Stationarity was assessed using Tracer 1.7.1 (Rambaut et al., 2014), based on a visual
272 exploration of trace plots and ESS values (ESS > 500 was obtained for all parameters.) After
273 exploring trace plots, we set the burn-in at 10%.
274
275 **TreeMix (OrientAGraph):** As an alternative way to describe the historic relationships among
276 plumage groups, we used *OrientAGraph* (Molloy et al., 2021), an approach based on the
277 *TreeMix* algorithm, which implements a graph-based model that uses genome-wide allele
278 frequencies to infer population splits and admixture events (Pickrell & Pritchard, 2012). This
279 method detects evidence of historical introgression between plumage groups and can better
280 describe the history of populations when not fully bifurcating. For input, we applied the same
281 SNP filtering as in the *snapper* analysis, adjusting only *--min-populations* to 8 and excluding
282 SNPs with >30% missing data using *--max-missing* 0.7 in VCFtools. The filtered VCF was

283 converted to TreeMix format with *vcf2treemix.py*.

284 (https://github.com/CoBiG2/RAD_Tools/blob/master/vcf2treemix.py).

285 We ran *OrientAGraph* testing for a range of migration edges (*m* parameter), running 5

286 iterations per *m*, with *m* values ranging from 1 to 6. We used *M. miniatus* (n = 3) as outgroup to

287 root the tree and incorporated a *bairdi-chrysops* known migration event using the *-cor_mig* and -

288 *climb* flags, and a 0.5 migration weight. This was done to account for the high likelihood that

289 there are signs of introgression even away from the hybrid zone. After running all iterations for

290 this range of m values, we used *OptM* (Fitak, 2021) to evaluate the optimal value of *m* in each

291 case, using an approximation based on linear regressions between *m* values and likelihood. We

292 made a small modification to the *optm* function to be able to account for the known migration

293 event. Based on this linear regression criteria, we chose the optimal *m* values and included those

294 results here.

295

296 ***Genetic structure within the ornatus-melanocephalus complex***

297 **PCA:** To summarize genetic variation in the *ornatus-melanocephalus* complex, we conducted a

298 genomic PCA using functions implemented in *gdsmt* and *SNPrelate* (Zheng, 2013) for R. To

299 conduct this PCA, we used only one SNP per locus, including only biallelic SNPs. To account

300 for the effect of missing data on the PCA (Yi & Latch, 2022), we (1) visualized the amount of

301 missing data in our PC plots, to evaluate the possibility that high missing data samples are

302 clustered together and/or have values biased towards the origin, (2) ran an additional PCA only

303 including SNPs with a missing rate < 0.1, and (3) ran an additional PCA using a different *vcf* as

304 input with more strict filtering for missing data when running populations (*--min-samples-overall*

305 *0.9*). We also conducted a PCA for a subset of the individuals to describe in more detail genetic

306 variation in the hybrid zone and adjacent areas (*chrysops*, hybrid zone individuals, *bairdi*, and
307 *griseonuchus*).
308 **fineRADStructure:** As a complementary approach to visualize genetic structure within the
309 *ornatus-melanocephalus* complex, we used *fineRADStructure*, which estimates a coancestry
310 matrix and an associated distance tree based on full haplotype information from RADseq data
311 (Malinsky et al., 2018). This method has particularly high resolution in inference of recent
312 shared ancestry, and allows one to describe nested patterns of genetic structure (Malinsky et al.,
313 2018). To run *finestructure* (part of the *fineRADStructure* pipeline), we used the output from the
314 populations program containing both variant and invariant sites (i.e., full haplotypes). No
315 population prior was specified for the run, which was set for 200,000 iterations with a burn-in of
316 100,000 (-x 100,000, -y 100,000), and a thinning interval of 1,000 (-z 1000).
317
318 **Calculation of F_{ST}:** To estimate overall genetic differentiation among plumage groups, we
319 calculated pairwise F_{ST} between subspecies (i.e., plumage and geographic groups). We excluded
320 all individuals from the hybrid zone from this analysis. Given the uncertainty in the *bairdi-*
321 *griseonuchus* limits (see Results), we also excluded individuals from the southern Ecuadorian
322 provinces (as in the *snapper* analyses). We calculated the Weir and Cockerham estimator (Weir
323 & Cockerham, 1984) using *vcftools* (Danecek et al., 2011) in 25,000 bp windows. To test for
324 potential discordance between nuclear and mitochondrial divergence, we evaluated whether
325 SNPs-based pairwise F_{ST} values were correlated with those estimated from the ND2 gene
326 (Céspedes et al. 2021). To assess the correlation between these two matrices, we implemented a
327 Mantel Test based on a Pearson's correlation using the function *mantel* implemented in the R
328 package *vegan* (Oksanen et al., 2007)

329 To test for isolation-by-distance (IBD), we also calculated pairwise windowed F_{ST}
330 between localities as defined in Supplementary Table 1 and evaluated if these are correlated with
331 the pairwise geographic distances. Geographic distances between localities were calculated using
332 the function *rdist.earth* implemented in the R package *fields* (Nychka et al., 2025). We also
333 evaluated the correlation between these matrices using a Mantel test based on a Pearson's
334 correlation as implemented in *vegan*.
335
336 **Admixture:** We employed admixture (Alexander et al., 2009) to estimate individual genetic
337 ancestries using a maximum-likelihood approach at three different geographic scales: whole
338 complex, hybrid zone and adjacent populations (*bairdi, chrysops*, hybrid zone, *griseonuchus*),
339 and hybrid zone only (*bairdi*, *chrysops*, hybrid zone).
340 Prior to running admixture, we calculated relatedness between all individuals and then
341 excluded individuals that appeared closely related. Specifically, we used the *--relatedness2* flag
342 in *vcftools* to export a table of pairwise relatedness values using the KING kinship estimator
343 (Manichaikul et al., 2010). We then identified pairs of individuals with a relatedness higher than
344 0.177 (i.e. likely first-degree relatives) and excluded one of them from admixture analyses. We
345 then transformed the VCF files to *.bed* format by using plink2 (Chang et al., 2015). Additionally,
346 we performed a SNP window linkage-disequilibrium pruning using plink2 *(--indep-pairwise*,
347 followed by *--extract*), on windows of 50 SNPs and using a 0.3 threshold. This was done for both
348 scales of analyses, and then admixture was run on the thinned datasets.
349 For all admixture runs, we included the *–cv* flag, to perform the cross-validation
350 procedure to choose the optimal *k* (i.e. number of source populations) value. For the whole
351 complex analyses, we ran admixture for values of $k\{1\ldots10\}$, and for the other two scales, we

352 used values of $k\{1\ldots6\}$. For each value of k, we ran 10 iterations starting with a randomly
353 generated seed. We analyzed variation across runs using *pong* (Behr et al., 2016) and generated
354 figures using the *mapmixture* (Jenkins, 2024) package for R.
355
356 **FEEMS:** To more directly evaluate how genetic variation is shaped across the Andes, we used a
357 spatially explicit model based approach: *FEEMS* (Marcus et al., 2021)
358 (https://github.com/VivaswatS/feems). This approach is based on a model of IBD and builds
359 continuous migration surfaces based on a geographic grid, visualizing areas of inferred low and
360 high gene flow (Marcus et al., 2021). To run FEEMS, we prepared four input files: (1) *plink* file
361 with genotype data for all individuals in the *ornatus-melanocephalus* complex (same set as for
362 whole-complex admixture analyses), (2) text file with geographic coordinates for all specimens
363 with genotype data (Supplementary Table 1), (3) a global grid with triangular cells of
364 approximately 7,774 km^2, obtained using the package *dggridR* for R (res=8; Barnes et al. 2024),
365 and (4) outer extent boundary for the grid, which was generated as a polygon that roughly
366 delineates the montane areas occupied by *Myioborus* and encompasses all sampling sites. The
367 chosen resolution of the grid allowed for most individuals in the same sampling site to be
368 assigned to a unique deme, based on preliminary explorations. We used the cross-validation
369 approach (Marcus et al., 2021) to choose the penalty parameters to be used in the optimization
370 procedure (λ and λ_q; https://github.com/VivaswatS/feems/blob/main/docsrc/notebooks/cross-
371 validation.ipynb). After preliminary explorations, the migration surface presented here was
372 generated using $\lambda = 0.17$ and $\lambda_q = 100$, which minimized the cross-validation error (although
373 lower values appeared roughly as adequate, Figure 6).
374

375 **Analysis of genetic variation across the *bairdi* x *chrysops* hybrid zone**

376 ***Calculation of a genomic hybrid index and genetic composition of the hybrid zone***

377 We estimated a genomic hybrid index and interclass heterozygosity using functions implemented
378 in *triangulaR* (Wiens & Colella, 2025). As noted above, the southern parental population was
379 defined based on our own results and includes individuals from central to south Ecuador that do
380 not cluster genetically with *griseonuchus* (see Results). In the SNP dataset for this analysis, we
381 included all SNPs per locus.

382 To identify ancestry-informative SNPs, we chose an allele frequency difference of 0.5
383 between parental populations. Higher thresholds resulted in very few (i.e., less than 100) or no
384 SNPs being retained, and there were no alleles fixed in each parental population. Then, to
385 describe the genetic composition of the hybrid zone, we used the values of genome-wide hybrid
386 indices and observed interclass heterozygosity for all individuals, and plotted these in a triangle
387 plot (Cronemberger et al., 2020; Liu et al., 2022; Lopez et al., 2021).

388

389 ***Geographic cline analyses***

390 To describe the geographic extent of the hybrid zone and evaluate whether genetic and plumage
391 variation were concordant, we fitted geographic clines using *HZAR* (Derryberry et al., 2014). We
392 fitted and compared clines for genomic hybrid indices—corresponding to admixture proportions
393 and estimated using *gghybrid*—and plumage hybrid indices estimated previously for these
394 individuals (Céspedes-Arias et al., 2021). For each of these "traits", we fit a model that estimates
395 cline width (w) and center (c) but has fixed values for mean and variance of the trait values at the
396 cline extremes (μL, μR, $varL$, $varR$). Both models were run with a tune value of 1.5, and for each
397 of these we ran 3 chains of 2 million iterations discarding the first 10% as burn-in. All runs were

398 performed randomizing starting parameters. To evaluate whether plumage and genomic clines
399 are coincident and concordant, we compared the 95% confidence interval of the center and width
400 estimates, which were calculated aggregating the three independent chains for each model.
401

402 ## Results

403 **Evolutionary relationships among plumage groups as inferred from phylogenies and**
404 **admixture graphs**

405 We employed two different phylogenetic approaches to describe evolutionary relationships
406 among the seven plumage groups that define the study taxa. Using a maximum-likelihood
407 approach with all loci concatenated (21,304 variant sites), we found that most individuals of the
408 same subspecies clustered together (Figure 1, panel C), yet we did not observe complete
409 reciprocal monophyly in most groups. The *chrysops* group was the exception, as it appeared as a
410 clade nested within the *ornatus* population. Notably, however, support values were generally
411 low, implying that not many sites are phylogenetically informative as expected for a scenario of
412 recent divergence and/or ongoing gene flow. Generally, we observed that individuals from
413 geographically adjacent subspecies are intermixed in the tree.

414 The *snapper* subspecies-level tree (Figure 2A), inferred using 3,105 biallelic SNPs,
415 showed strong support for a monophyletic "southern" group, encompassing three *M.*
416 *melanocephalus* subspecies (*malaris, melanocephalus, bolivianus*) characterized by having
417 spectacles and a black crown. The taxon *griseonuchus*, from north of the Marañón River Valley,
418 characterized by having a reddish crown that is phenotypically very similar to *bairdi*, was
419 recovered as sister to a strongly supported clade including the more northerly groups from
420 Colombia and Ecuador. Within this expanded "northern" group, *chrysops* was recovered as sister

421 to *ornatus*, and these, in turn, are sister to *bairdi* (Figure 2A). Note that, unlike the maximum-
422 likelihood tree, we *a priori* assigned individuals to groups for these analyses, which can bias the
423 phylogenetic estimation given that the limits among these populations are not fully defined.
424 Overall, we found stronger signals of admixture events between the northern relative to
425 the southern group (Figure 2B). Depending on the *m* value, we recovered slightly different
426 topologies, though we note that these incongruencies were restricted to the relationships between
427 the northern group taxa, namely *griseonuchus, bairdi*, *chrysops*, and *ornatus* taxa (Figure 2B).
428 Here we present the inferences for m = 1 and m = 3 as these were the optimal values based on
429 the linear regressions fitting the likelihood vs *m* curves (Supplementary Figure 1). Both graphs
430 consistently recovered a monophyletic "southern" group (*malaris*, *melanocephalus* and
431 *bolivianus*), which was consistent with the topology of the *snapper* tree. In the m = 1 graph
432 (Figure 2B), *griseonuchus* and *bairdi* appeared as sister, and this clade was, in turn, sister to the
433 southern group. In contrast, in the m=3 graph, *bairdi* was sister to *chrysops*, and thus a migration
434 edge between these taxa was not estimated despite the known migration event parameter. Two
435 edges with high migration weight were inferred, between *griseonuchus* and *bairdi,* and
436 *griseonuchus* and the *bairdi-chrysops* node. An additional edge, with low migration weight, was
437 inferred between the outgroup and the *bolivianus-melanocephalus* group (Figure 2). We note,
438 however, that the outgroup has a very low sample size (n = 3) and thus do not find this inference
439 to be very reliable.
440
441 **Genetic structure within the *ornatus-melanocephalus* complex**
442 The PCA, based on 4,800 biallelic SNPs, showed that individuals generally cluster in relation to
443 plumage groups and geographic origin (Figure 3), but some phenotypically distinct groups were

very close in the PC1-2 space (particularly *chrysops, ornatus* and *bairdi*). As in the phylogenetic analyses, one can distinguish clearly the "southern" group, including all individuals from populations south of the Marañón River Valley (*bolivianus, melanocephalus, malaris,* Figure 3), which occupied a wide range of PC1 values (Figure 3). There was also an identifiable northern cluster (*chrysops, bairdi*, and hybrid zone individuals) that was, however, not fully distinct from a third cluster comprised of *griseonuchus* individuals. Within this northern cluster, the plumage groups *ornatus, chrysops, bairdi,* and individuals from the hybrid zone ("northern", hereafter) appeared very close in PC1-2 space. As expected, individuals from the hybrid zone (in gray, Figure 3) fell intermediate between their parental populations. The *ornatus* group appeared as a more defined cluster when considering PC3 (Supplementary Figure 5). Notably, some individuals putatively classified as *bairdi*, collected in the Ecuadorian provinces of Loja and outlining ridges in the Amazon, appeared to cluster with *griseonuchus* (Figure 2). We observed a similar pattern in the hybrid zone plus *griseonuchus* PCA (Supplementary Figure 6). We obtained similar results when using a SNP set of the whole complex obtained with a stricter filtering for the PCA (Supplementary Figure 7).

Two main clusters are supported in the coancestry matrix resulting from *fineRADstructure*: populations occurring from south Ecuador to Bolivia (*griseonuchus* plus "southern" group), and those occurring from Ecuador to Colombia ("northern" group) (Figure 3C), which contrasted with the topology inferred by *snapper*. Within the first group, *griseonuchus* and the southern group appear as differentiated clusters. The few exceptions to this pattern are a few putatively *bairdi* individuals that clustered with *griseonuchus*, but were collected in the southern edge of the distribution of this plumage group (where it is replaced by *griseonuchus,* Figures 1 and 3) in the Cordillera del Cutucú and Loja Province of Ecuador. We

467 detected two main groups within the "northern" group (Ecuador and Colombia): one
468 corresponded to *ornatus* (occurring in the Eastern Andes of Colombia) and the other to the
469 hybridizing taxa in the complex and individuals from the hybrid zone.

470 Admixture analyses including all individuals in the *ornatus-melanocephalus* complex
471 revealed genetic breaks along the distribution of these warblers that were overall consistent with
472 fineRADStructure (Figure 4, A). For example, when assuming K = 2 and K = 3 the "southern"
473 group appeared as an identifiable cluster. For K = 2, which is the optimal K value based on
474 cross-validation error values (Supplementary Figure 8, top panel), the group *griseonuchus* also
475 clustered with the southern group (Figure 4, A1). For K = 3, a "southern" plus *griseonuchus*
476 group can be also identified, and there was an additional genetic break roughly located in the
477 described *bairdi-chrysops* hybrid zone in southern Colombia (Figure 4, B1).

478 The results of an admixture analyses focused on the hybrid zone and adjacent
479 populations, when assuming K = 2, showed a genetic break that clearly coincides with the
480 *chrysops-bairdi* hybrid zone (Figure 4, B and C), as previously identified based on plumage
481 variation (see map in Figure 1). However, when also considering *griseonuchus*, the transition
482 between these genetic groups is less sharp and displaced to the south (Figure 4, panel B). In both
483 cases there was some east-west differentiation, with localities to the west of the Ecuadorian
484 Andes showing a higher mean admixture proportion corresponding to the southern cluster
485 (Figure 4, B1 and B3). When using K = 3 for the analyses including *griseonuchus* (a *K* that
486 corresponds to the number of identifiable plumage groups), two genetic clusters were identified
487 in Ecuador, and one in Colombia, and one of these genetic breaks did again roughly correspond
488 to the *bairdi-chrysops* hybrid zone (Figure 4, B2). In both cases the *K* value with the lowest
489 cross-validation is *K* = 1 (Supplementary Figure 7).

490 Values of pairwise F_{ST} between populations showed an overall correspondence with
491 geographic distance: subspecies that are far apart in space show higher genetic differentiation
492 (e.g., *ornatus* vs *bolivianus,* Figure 5). We observed higher pairwise differentiation between taxa
493 belonging to the "southern" versus "northern" groups, which is consistent with all other analyses
494 describing genetic structure. In contrast, there was very low differentiation among the three
495 subspecies occurring south of the Marañón River Valley ("southern" group), which all have
496 similar head coloration. However, there was also relatively low levels of differentiation within
497 the "northern" plus *griseonuchus* group, which encompasses striking plumage diversity,
498 particularly between the two hybridizing taxa, *bairdi* and *chrysops, and bairdi* and *griseonuchus*
499 (Figure 5). There was a positive correlation (Figure 5, Mantel statistic r = 0.75, p-value = 0.001)
500 between pairwise F_{ST} calculated based on our ddRAD SNPs and the ND2 mitochondrial gene
501 (Céspedes-Arias et al., 2021). Lastly, as expected if patterns of genetic variation reflect IBD, we
502 found a positive correlation between pairwise F_{ST} between localities and geographic distance
503 (Figure 5, Mantel statistic r = 0.65, p-value = 0.001). Using the same classification scheme for
504 individuals as that to calculate F_{ST}, we estimated genetic diversity statistics. We found that
505 nucleotide diversity is very similar among groups, ranging from 0.076 to 0.091 (Supplementary
506 Table 3).
507 The migration surfaces inferred by FEEMS suggest that there are areas within the
508 distribution of these warblers in which genetic differentiation cannot be fully explained by IBD
509 (i.e., areas of inferred low gene flow; Figure 6). These areas of low migration coincided mostly
510 with two geographic breaks that dissect the distribution of the complex (Figure 1, Figure 6),
511 specifically the Magdalena River Valley and the Marañón River Valley, and with the hybrid
512 zone. In contrast, most of the range south of the North Peruvian low showed mid to high values

513 of inferred effective migration, which can indicate that genetic structure along this area is well
514 explained by IBD alone. We acknowledge that low sampling density in certain regions,
515 particularly in central and southern Peru, may have influenced the inferred migration surface
516 (Figure 6). However, cross-validation results indicate that the selected regularization parameter
517 provides a good fit while accounting for unobserved nodes (Supplementary Figure 9). With that
518 caveat in mind, the FEEMS results are broadly consistent with the patterns of genetic structure
519 revealed by our other analyses.
520
521 **Hybrid zone**
522 To describe the geographic extent of hybridization and the genetic composition of the *bairdi* x
523 *chrysops* hybrid zone, we identified ancestry-informative markers to estimate genomic hybrid
524 indexes and interclass heterozygosity for individuals across the hybrid zone and parental *bairdi*
525 and *chrysops* populations. Using a relatively low threshold for allele frequency difference (0.5),
526 we identified only 29 sites that were ancestry informative from the total of 14,301, and none of
527 these were fixed between the parental populations. This result complicates the interpretation of
528 the triangle plots (see below) and highlights the low level of genetic differentiation between
529 these two hybridizing taxa. Because of the scarcity of ancestry informative sites in our ddRAD
530 data set, which can affect the accuracy of genomic hybrid index estimations, we opted to conduct
531 cline analyses with admixture proportions (with K=2) instead.
532 We found that the head coloration and genomic hybrid index, defined as the admixture
533 proportions in Figure 4 (panel C3) had different centers but similar widths (Figure 7, panel A).
534 The admixture proportion cline had a width of 281.6 km (95% confidence interval [CI]: 245.0–
535 319.8km), compared to 240.6 km for head coloration (95% confidence interval [CI]: 277.5–

536 311.1 km). Given the substantial overlap in confidence intervals, we do not interpret this 41 km
537 difference in width as significant. However, we found a 37 km difference in the estimates of the
538 center between the plumage and admixture proportion clines, for which the 95% confidence
539 intervals did not overlap. For admixture proportions the cline was centered around 348.8 km
540 (95% confidence interval [CI]: 335.4–362.7 km) in our sampling transect, while the plumage
541 cline was centered around 311.1 km (95% confidence interval [CI]: 300.7–322.6 km). Therefore,
542 based on our *HZAR* analyses, the plumage cline appeared to be displaced towards the south (i.e.,
543 towards Tungurahua, Ecuador, the origin of the sampling transect), implying that the *chrysops*-
544 like plumage traits may be introgressing into *bairdi*. However, there was an overall strong
545 correlation between the plumage hybrid indices and admixture proportion across the hybrid zone
546 and allopatric populations (Figure 7, panel B; linear model; *p-value* < 0.001, *m* = 0.9, *R-squared*
547 = 0.8). We identified the outliers of this linear regression as those individuals with residuals >2
548 SD units from the mean, finding that most were below the regression line (Figure 7).
549 Specifically, we identified six individuals that had a more *chrysops*-like plumage from the hybrid
550 zone area (i.e., lower) relative to what the model predicts based on the admixture proportion,
551 which was consistent with the cline analysis results that shows that *chrysops*-like plumage is
552 introgressing into *bairdi*.
553 To describe the genetic composition of the *bairdi-chrysops* hybrid zone, we built triangle
554 plots (interclass heterozygosity vs hybrid index) based on the 29 ancestry-informative markers
555 that we identified (Figure 7, panel C). The individuals from the hybrid zone are distributed
556 towards the center of the triangle, many below the line that defines what value combinations are
557 possible under Hardy-Weinberg Equilibrium which is likely a product of the low allele
558 frequency difference threshold. We note that individuals from populations away from the hybrid

559 zone (Figure 7, in green and orange) are not located in the bottom corners of the triangle plot,
560 which is expected given that the allele frequency difference threshold is lower than one. If
561 interpreted considering the caveats described above, the distribution of individuals in the triangle
562 is suggestive of a hybrid zone mostly composed of advanced generation hybrids, without many
563 F1s.

564 Discussion

565 We used reduced-representation genomic data to study genetic variation at two different spatial
566 scales: the full distribution of a geographically variable warbler complex (*M. ornatus-*
567 *melanocephalus*) and a known hybrid zone between two of these plumage forms. At both scales,
568 our set of thousands of SNPs allowed us to explore genetic structure with a resolution that
569 mtDNA loci could not provide in previous studies (Céspedes-Arias et al., 2021; Pérez-Emán,
570 2005). We found evidence of isolation-by-distance driving genetic structure across the
571 latitudinally extensive range of these warblers, and genetic divergence increased where ranges
572 are separated by known geographic barriers. These clear genetic breaks coincided with observed
573 differences in head plumage coloration among populations. Notably, one transition between
574 groups includes a geographically restricted hybrid zone between *bairdi* and *chrysops*, likely
575 comprised primarily of advanced generation hybrids, where differences in genetic ancestry and
576 head coloration are correlated.

577 Below we discuss how our findings relate to the history of divergence and gene flow
578 among these populations, with a focus on how genetic variation imperfectly covaries with
579 plumage traits, and how it is structured across geography. We then discuss how genetic variation
580 is structured across the hybrid zone, how it relates to patterns of variation in head coloration in

this area, and whether there is evidence of selection against hybrids maintaining this seemingly extensive hybrid zone.

Geographic variation in plumage color and genetic structure

Our qualitative assessments based on historical plumage descriptions and examinations of study skins suggest that for these *Myioborus*, patterns of geographically clustered phenotypic variation coincide with phylogeographic structure uncovered in the present study. First, we found that individuals belonging to the same subspecies usually form genetic clusters, and that all major genetic breaks within the species complex coincide with a transition between clearly differentiated plumage groups, although we note that we lack data quantifying the degree of plumage differentiation between subspecies (e.g. Winger and Bates 2015). In addition, subspecies with subtle plumage differentiation (Cuervo and Céspedes Arias 2023), specifically those occurring south of the Marañón River Valley (Figure 1), all characterized by black crown and spectacles, show little genetic differentiation from one another.

Previous examinations of genetic variation in mitochondrial loci (Céspedes-Arias et al., 2021; Pérez-Emán, 2005) found extensive haplotype sharing and little genetic differentiation among the subspecies *ornatus* and the hybridizing *bairdi* and *chrysops*, which are all characterized by striking plumage differences (see Figure 1). In contrast, in our analyses of ddRAD data, which encompass thousands of SNPs across the genome, these subspecies are each identifiable genetic clusters, albeit with shallow genetic differentiation. Taken together, we interpret these results as evidence of rapid and recent diversification of head plumage patches in the *ornatus-melanocephalus* complex, as seen in other Neotropical taxa (Campagna et al., 2012, 2017). Our results also highlight the usefulness of reduced representation genomic data to study

604 the evolutionary history of young radiations (Aguillon et al., 2018; Toews, Campagna, et al.,
605 2016), in which few loci might be insufficient to detect genetic differentiation and describe
606 evolutionary relationships among phenotypically distinct populations (Friis et al., 2016;
607 Stryjewski & Sorenson, 2017). We note that lack of differentiation in mtDNA between
608 phenotypically distinct groups could also be the result of mitochondrial introgression (Irwin et
609 al., 2009; Toews & Brelsford, 2012), although the strong positive correlation of F_{ST} for ND2 and
610 our ddRAD loci does not support a scenario of extensive mitochondrial introgression in the
611 group.

612 Genetic structure in the region of the Andes spanning from central Ecuador to northern
613 Peru (north of the Marañón River Valley) is complex and not clearly associated with a single
614 topographic feature. This area also broadly corresponds to the transition between *griseonuchus*
615 and *bairdi* plumage (see Figure 1, Supplementary Figure 1), two subspecies characterized by a
616 rufous crown and spectacles. Despite their overall similarity, these two groups show
617 differentiation in at least two head plumage traits: the presence of a black band behind the rufous
618 crown and the extent of black in the cheeks and lores (Zimmer 1949; Curson et al. 1994; World
619 and International 2019; Cuervo and Céspedes Arias 2023). Our qualitative assessments of
620 plumage variation (Supplementary Figure 1), however, suggest that the variation in these traits
621 might not be completely discrete. Moreover, we found that individuals from Loja and the
622 Cordillera del Cutucú (Ecuador), show substantial variation in both putatively diagnostic
623 plumage traits. Therefore, these individuals, which are genetically clustered with *griseonuchus,*
624 do not uniformly resemble the individuals from Cajamarca, Peru (Supplementary Figure 1).
625 Because of limited sampling in the area, we are unable to draw stronger conclusions, but our data
626 suggest that there is some discordance between plumage and genetic variation, and evidence of

627 introgression based on admixture graphs. Despite some differentiation in at least these two
628 plumage traits, our genetic results are concordant with substantial historical or contemporary
629 gene flow among populations in this area. This suggests that the geographic range of *bairdi* is
630 potentially outlined by two hybrid zones, one to the north with *chrysops* (Céspedes-Arias et al.,
631 2021), and to the south with *griseonuchus.* Evaluating this hypothesis of hybrid zones at the
632 range edges, as observed in barn swallows (Scordato et al., 2017), will require a systematic
633 collection of specimens across the *bairdi-griseonuchus* boundary.
634
635 **Genetic structure across space: the influence of geographic distance and topographic**
636 **features along the Andes**
637 Several of our analyses support a strong correlation between genetic variation and geographic
638 location of samples and support an important role of isolation-by-distance driving genetic
639 structure in *Myioborus* along the Andes. Patterns of IBD are close to ubiquitous in nature
640 (Aguillon et al., 2017; Seeholzer & Brumfield, 2018; Vos & Velicer, 2008), but see Meirmans,
641 2012), and are often taken as the null model for genetic differentiation if organisms have limited
642 dispersal given the scale of their range, as is the case for these small warblers with a
643 latitudinally-expansive distribution range. However, IBD does not account fully for the observed
644 patterns of genetic variation. Specifically, we found evidence of discrete genetic—and
645 phenotypic—clusters, and of reduced migration (relative to IBD) across some areas along the
646 Andes, which in turn correspond to lowland valleys such as the Marañón and the Magdalena
647 River Valleys. These deep, lowland gaps interrupt the continuity of humid high-elevation forest
648 and scrub (Graham et al., 2010) that these *Myioborus* inhabit in the present, and also coincide
649 with genetic and phenotypic breaks within other taxa that inhabit similar elevation bands

(Gutiérrez-Pinto et al., 2017; Martínez-Gómez et al., 2023; Winger & Bates, 2015). These two topographic features were identified in a comparative study as being associated with allopatric events across multiple Andean bird genera (Hazzi et al., 2018). These genetic breaks, clearly tied to topographic features, also correspond with striking discontinuities in head coloration traits (see Figure 1).

Other areas associated with clear genetic breaks observed in the inferred migration surface were not clearly associated with topographic features that interrupt the present-day continuity of humid, high-elevation forests. Specifically, low migration is inferred in a large area ranging from the border between Ecuador and Colombia down to southern Ecuador. We suggest that this pattern emerges results partially because this broad area includes the *chrysops-bairdi* hybrid zone, in which genetic differentiation relative to distance is expected to be higher than what is expected under pure IBD, as two somewhat differentiated subspecies come into contact. The extension of this area of low inferred migration to south Ecuador is consistent with our fineRADStructure, PCA and admixture results that suggest that there is substantial genetic differentiation between individuals in close proximity in this area. Specifically, we found that birds from isolated mountain spurs such as the Cordillera del Cutucú (southeastern Ecuador) and Cordillera del Condor (Peru-Ecuador border) clustered genetically with birds from the main Andes in Cajamarca (northwestern Peru), and Loja (southern Ecuador), instead of the more geographically proximate individuals from the eastern Ecuadorian Andes in Morona-Santiago (see Figure 2). The Cordillera del Cutucú and Cordillera del Cóndor are outlying mountain ranges in the Amazonian foothills, separated from the main Andes by the valleys of the rivers Upano and Zamora, respectively (Pozo-Zamora et al., 2022; Robbins et al., 1987). Our results suggest that these valleys, and surrounding lowland areas, have likely acted as dispersal barriers

673 for these higher-elevation warblers. On the other hand, the patterns of genetic variation described
674 here indicate that there has likely been substantial gene flow connecting the populations of
675 warblers inhabiting isolated peaks in these mountain ranges (separated by the Santiago River
676 Valley; Robbins et al. 1987; Pozo-Zamora et al. 2022), and the eastern slope of the Andes in
677 northern Peru.
678
679 **Hybrid zone dynamics**
680 Based on our reduced-representation genomic data set, the hybridizing *chrysops* and *bairdi* are
681 characterized by very shallow genetic divergence, which contrasts with their striking plumage
682 differentiation. Moreover, across thousands of SNP sites, we only identified 29 ancestry
683 informative sites with substantial allele frequency differences between parental populations. This
684 scarcity of ancestry informative markers limits our ability to draw definitive conclusions about
685 the genetic composition of this hybrid zone but highlights the low level of genetic differentiation
686 between *chrysops* and *bairdi*.
687 This hybrid zone likely formed by secondary contact of two formerly isolated plumage
688 forms. The disconnection and subsequent reconnection of habitat that could have shaped this
689 process may have followed elevational shifts in the cloud forest belt during the Pleistocene
690 (Flantua et al., 2014; Ramírez-Barahona & Eguiarte, 2013), or potentially been driven by
691 volcanic activity that modified topographic connectivity (Sanín et al., 2024). An investigation of
692 the evolutionary origins of this hybrid zones will require both a better understanding of the
693 patterns of connectivity of cloud forest in the area where the *chrysops-bairdi* hybrid zone is
694 centered, as well as an estimation of the time since secondary contact, which could be obtained
695 based on inferences of ancestry tract length in whole genome data (Pool & Nielsen, 2009).

Our geographic cline analyses using plumage and genomic data suggest that the hybrid zone extends for over 200 km, similar to the findings from its original description based on face color traits (Céspedes-Arias et al., 2021). We do not find statistically significant differences in the width of the admixture proportion (i.e., genomic) and plumage geographic clines. Taken together, plumage and genomic patterns of variation thus suggest that hybridization between *chrysops* and *bairdi* has been extensive, albeit still restricted to a fraction of the entire distribution of these taxa (approximately 30% of our sampling transect) which is continuous at least from central Ecuador to the north of the Central Andes of Colombia. Therefore, the geographic extent of hybridization is unlikely to be constrained by physical or ecological barriers that limit dispersal beyond the hybrid zone (e.g., DeRaad et al., 2023), raising the question of what factors are restricting the hybrid zone to this area.

The lack of information on natal dispersal distances as well as a the wide range of plausible times since secondary contact constrains our ability to generate a direct comparison of the observed cline widths with that expected under neutral diffusion (Barton and Gale 1993). With the available data we are thereby unable to evaluate whether the *chrysops* x *bairdi* hybrid zone corresponds to a stable tension zone, in which the width of clines is constrained by selection against hybrids. A second possibility is that the observed sharp genomic and plumage clines are not maintained by selection, but instead that this is a young and potentially expanding hybrid zone, and that the differentiation between *chrysops* and *bairdi* might become eroded over time (Kearns et al., 2018).

Using a range of plausible dispersal distances inferred from another *Myioborus* species (Céspedes-Arias et al., 2021; Mumme, 2015), we can approximate expectations under neutrality (Barton & Gale, 1993). For a cline width of 281 km, the neutral diffusion model predicts that

719 ~14,337 generations would be required assuming a dispersal distance of 935 m, or ~53,278
720 generations with a dispersal distance of 485 m. In principle, building upon our systematic
721 collection of specimens across the *chrysops* x *bairdi* hybrid zone, the alternative of an expanding
722 hybrid zone could be revisited by comparing plumage and genomic clines based on specimens
723 collected across different time points. However, given that the expansion of this hybrid zone
724 under neutral diffusion with plausible dispersal distances is expected to be slow (i.e. 0.5 km over
725 50 generations), time between sampling periods would need to be very long to detect changes.
726 We found that cline centers are not fully coincident, as that of the plumage cline is
727 displaced to the south of the genomic cline center, implying that the *chrysops*-like plumage is
728 introgressing into *bairdi.* Assuming that most of our ddRAD loci can be considered neutral
729 markers, our results indicate that head coloration might be under selection across this hybrid
730 zone (Gay et al., 2008). Specifically, the observed pattern would be consistent with a scenario in
731 which the *chrysops*-like plumage is favored across the hybrid zone, which would lead to the
732 asymmetric introgression of this phenotype towards the south. Asymmetric introgression of
733 plumage traits has been observed in other avian hybrid zones (Baldassarre et al., 2014; Lim et al.,
734 2024; Semenov et al., 2021), and in two of these cases, behavioral evidence supported the idea
735 that these traits provide fitness advantages in the context of mate choice (Baldassarre & Webster,
736 2013; Long et al., 2024; Stein & Uy, 2006).
737 Even though the difference in plumage and genomic cline center is statistically
738 significant, we found it to be relatively small (38 km; approximately 4% of the entire sampling
739 transect). This contrasts with much more marked differences in center estimations found in other
740 hybridizing birds like *Motacilla* wagtails (Semenov et al., 2021) and *Malurus* fairy-wrens
741 (Baldassarre et al., 2014). We acknowledge that this minor difference could be an artifact of the

742 way in which both of our hybrid indexes were calculated (e.g. with the scoring system giving
743 equal weight to pre-defined patches; Céspedes-Arias et al., 2021), and potentially low sampling
744 sizes in some of the localities. However, this slight difference could be the result of selection on
745 the head plumage patches. Even though the role of these color patches has not been studied in
746 these warblers yet, it is plausible that they may be relevant for intraspecific communication and
747 mate choice as seen in head plumage patterns in many other bird species (e.g., Marchetti 1992;
748 Andersson et al. 1998; Scott and Deag 1998; Rémy et al. 2010). Future behavioral studies could
749 evaluate these hypotheses of selection favoring *chrysops*-like head color patterns across this
750 hybrid zone.

751 Besides describing the geographic extent of hybridization, we used our reduced-
752 representation genomic data set to gain insights into the genetic composition of the *chyrsops x*
753 *bairdi* hybrid zone. However, as noted above, the scarcity of ancestry informative sites in our
754 dataset complicates the interpretation of these results, as it compromises the accuracy of genomic
755 hybrid indexes and interclass heterozygosity estimations (Gompert et al., 2024; Wiens & Colella,
756 2025). Setting a low allele frequency threshold, which was necessary in our case, can also result
757 in a tendency of individuals to cluster near the center of triangle plots (Wiens & Colella, 2025).
758 Bearing these caveats in mind, we interpret our triangle plot results as suggestive of a hybrid
759 zone composed mostly of advanced generation hybrids, with no F1s sampled, given the lack of
760 individuals with intermediate hybrid indexes and very high interclass heterozygosity (Capblancq
761 et al., 2020; DeRaad et al., 2023; Natola et al., 2021). This is consistent with the existence of a
762 high diversity of plumage phenotypes in the hybrid zone, as well as the lack of contemporary
763 geographic contact between parental *chrysops* and *bairdi* populations (i.e., we do not observe
764 “pure" individuals of both species in the same sites) (Céspedes-Arias et al., 2021). Given the

765 shallow genetic divergence between *chrysops* and *bairdi*, a more detailed examination of the
766 genetic composition of this hybrid zone will necessitate whole-genome sequencing, capturing
767 more ancestry informative sites. Specifically, as in other examples of hybridizing taxa with very
768 little genomic divergence, whole-genome sequencing may be needed to identify these regions
769 and sites that are highly differentiated and study hybridization dynamics in detail (Aguillon et al.,
770 2021; Poelstra et al., 2014; Toews, Taylor, et al., 2016).
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787

Figures

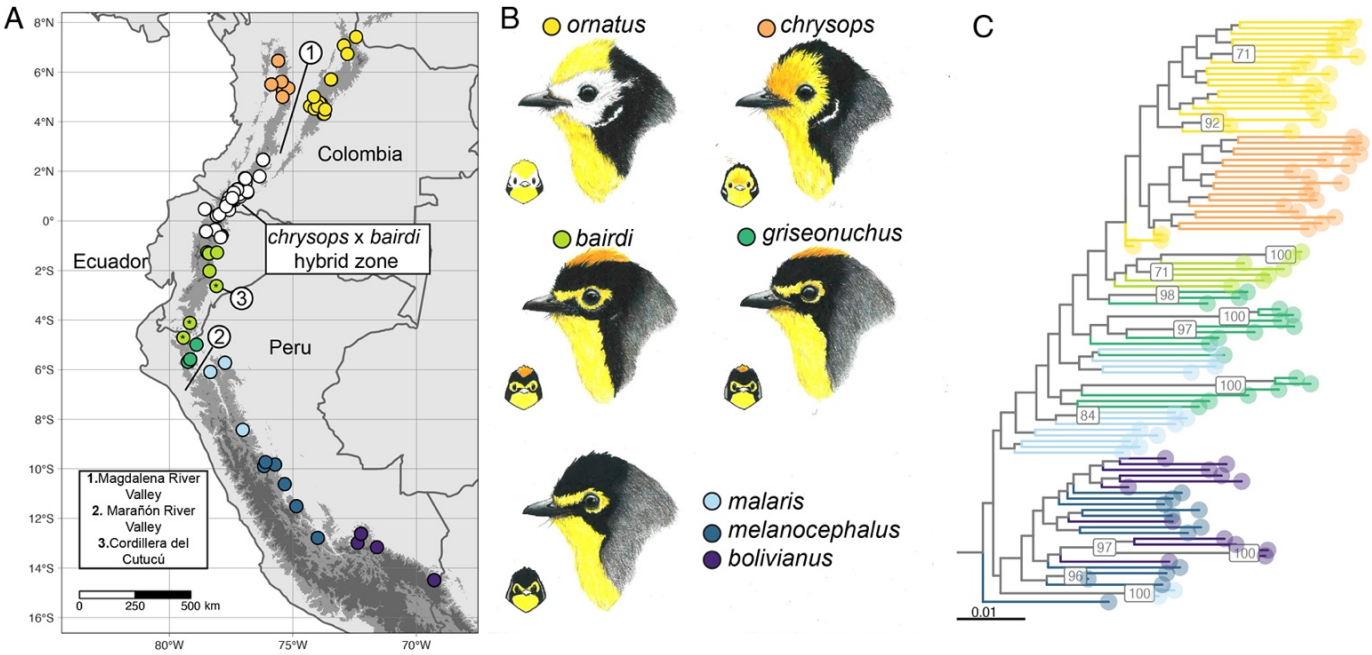

Figure 1: Sampling localities of specimens included in the study **(A)**, which encompass all major plumage groups **(B)**. The color of the dots corresponds to study taxa. The white dots correspond to areas around the *bairdi-chrysops* hybrid zone as identified based on plumage patterns (Céspedes-Arias et al., 2021). The subspecies *malaris, melanocephalus*, and *bolivianus* are here represented by one illustration. In the map, elevation is denoted by the gray scale in the background, with the elevation in which these *Myioborus* occur shown in an intermediate shade of gray. The localities marked with an asterisk correspond to those in Loja and outlining ridges in Morona-Santiago that are discussed in the main text, and numbers mark important topographic features. **(C)** Maximum-likelihood phylogeny estimated with our dataset of concatenated ddRAD loci, showing how overall relationships somewhat reflect geography (outgroups are cropped out of the figure here). The branches are colored by subspecies as in the map, with only bootstrap values over 70 shown (hybrid zone samples not included in the analysis). Illustrations: Andrés Montes Rojas.

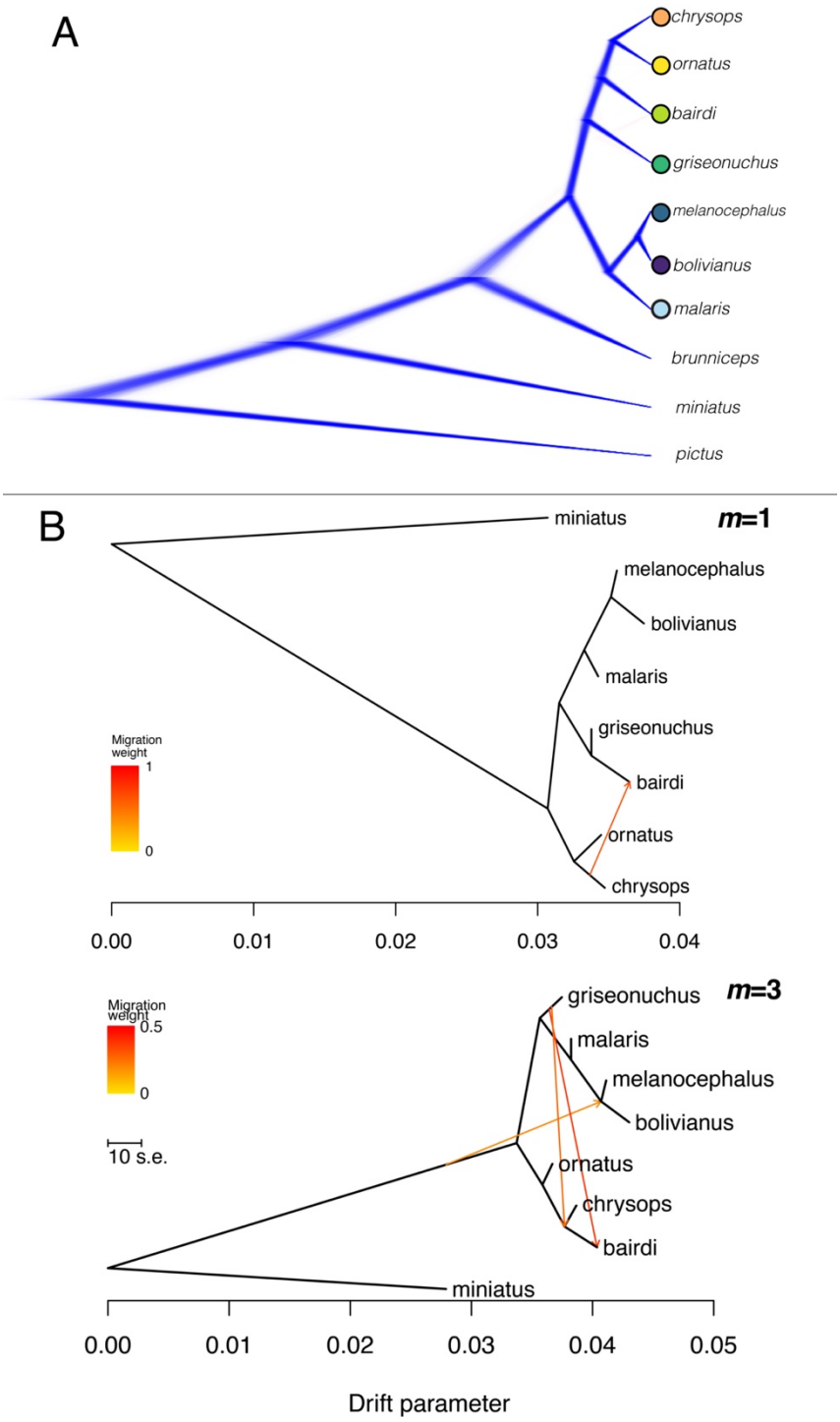

Figure 2. Evolutionary relationships within groups in the *Myioborus ornatus-melanocephalus* complex as inferred by snapper and *OrientAGraph*. **(A)** The main figure shows the phylogenetic relationships between taxa, with the blue topology corresponding to the most frequent in the posterior distribution, followed by red. **(B)** Admixture graphs from *OrientAGraph* with 1(top) and 3 (bottom) migration edges (*m*). The migration edges are shown as colored arrows, and the hue denotes the migration weight.

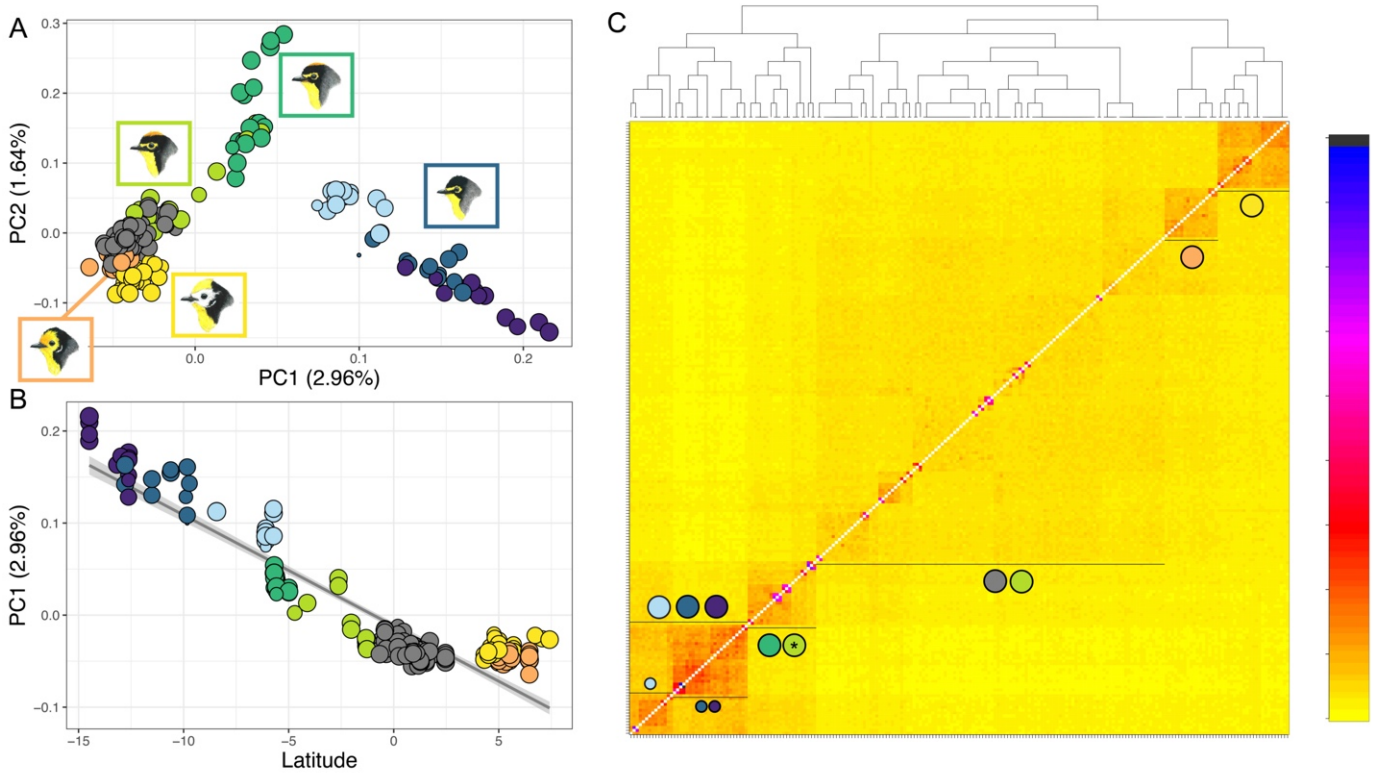

Figure 3. Genetic structure within the *ornatus-melanocephalus* complex. (**A)** The distribution of individuals in the PC1/PC2 space showing clustering based on plumage groups, which are denoted by the color of the dots. **(B)** Strong correlation between PC1 and latitude. In both cases (A and B) the size of the dot is inversely proportional to the proportion of missing data. **(C)** fineRADStructure coancestry matrix, and associated distance tree (above) showing clusters of individuals that overall correspond to plumage groups, which are denoted by the colored dots that correspond to groups as in Figure 1. The asterisk highlights a small number of putatively *bairdi* individuals, all from the south end of Ecuador, that genetically cluster with *griseonuchus* (from northern Peru). The color scale denotes values of coancestry, with blue and black tones corresponding to very high levels of coancestry.

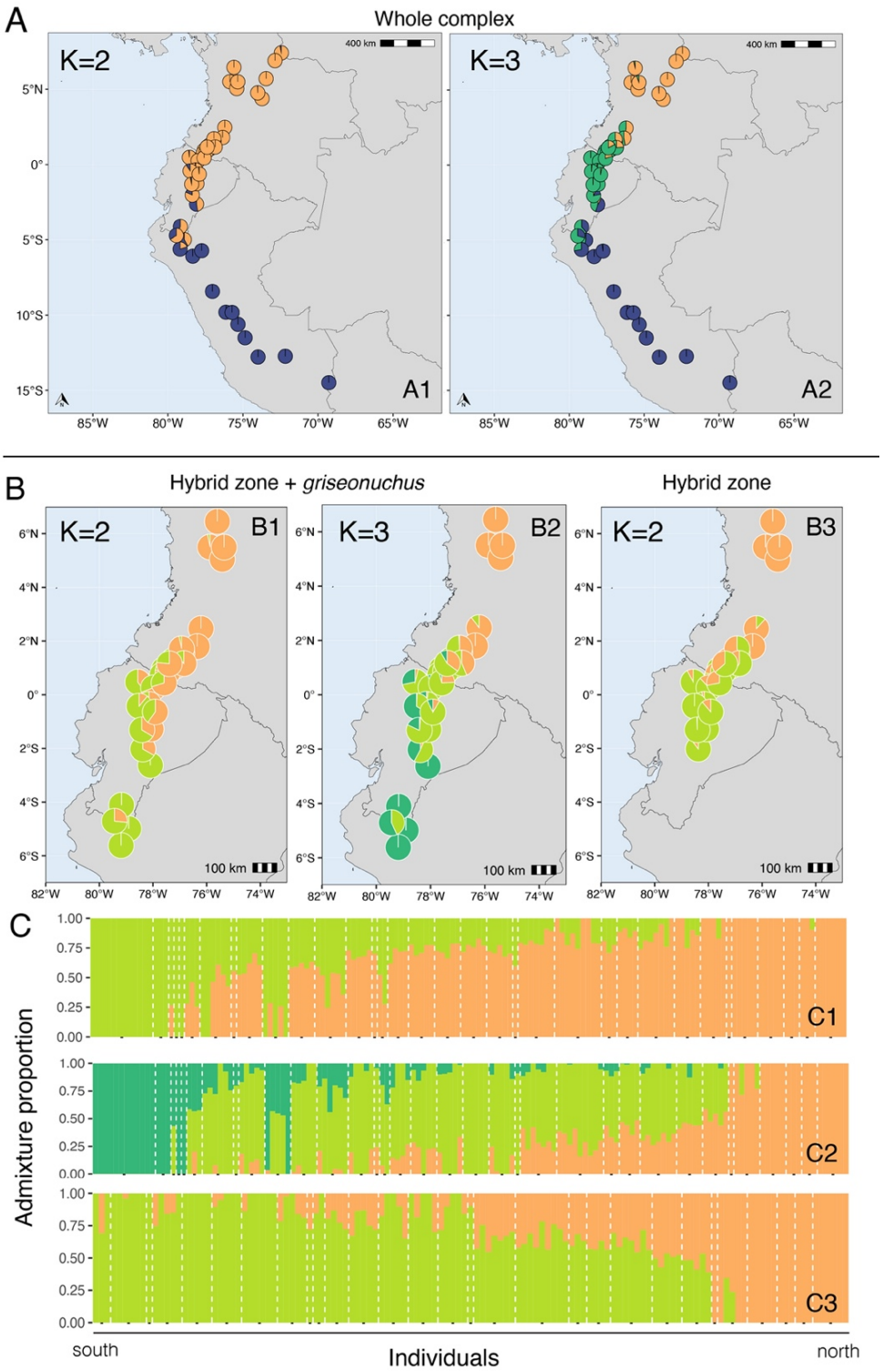

Figure 4. Genetic clusters (**A**) across the entire distribution of the *ornatus-melanocephalus* complex and (**B**) the hybrid zone and adjacent populations as shown by the mean admixture proportions per locality. The dataset and *K* value for each plot is specified at the top of each map. **(C) Admixture prop**barplots showing the admixture proportions per individual, which are organized into localities from south to north (localities are divided by white dotted lines).

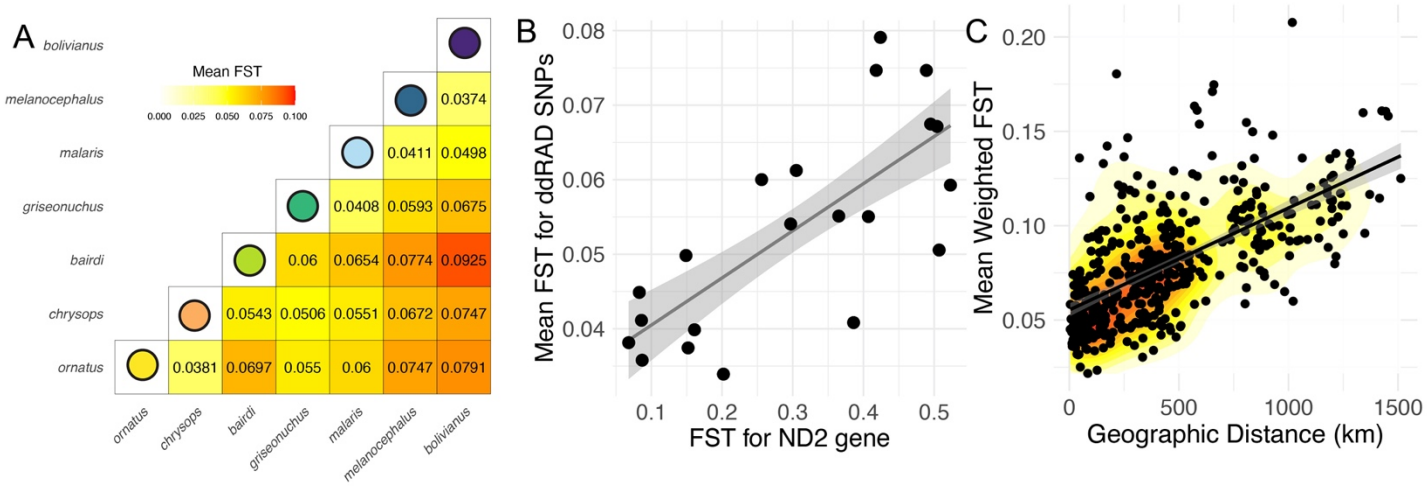

Figure 5. Genetic differentiation among Andean *Myioborus* taxa. **(A)** Mean pairwise F_{ST} values between subspecies (i.e., plumage and geographic groups). **(B)** Correlation between pairwise F_{ST} calculated with ddRAD SNPs (this study) and the ND2 mitochondrial gene from the same set of samples (Cespedes-Arias *et al* 2021). **(C)** Positive correlation between geographic distance and F_{ST} between localities supports a role of isolation-by-distance shaping genetic variation in the *ornatus-melanocephalus* complex.

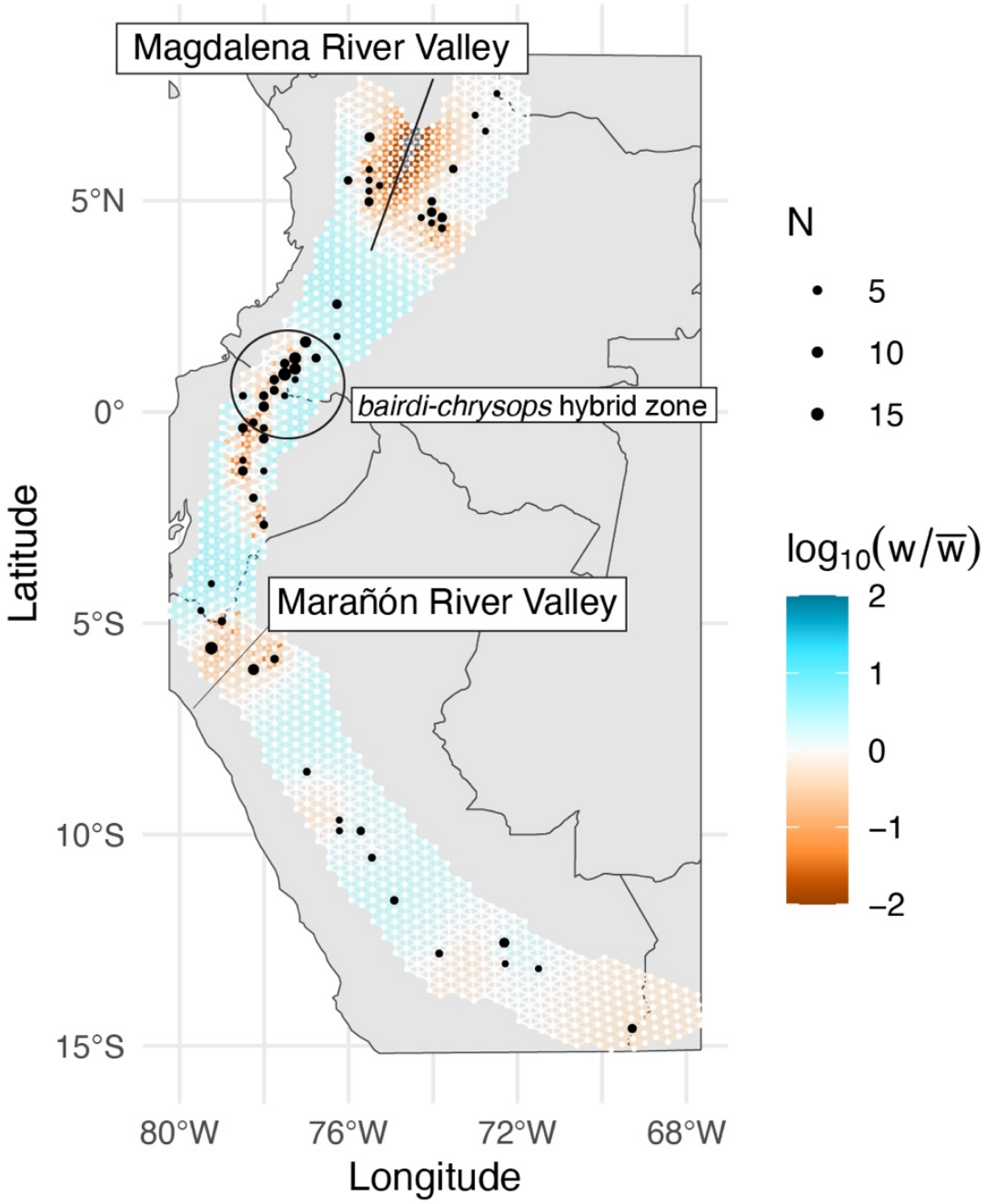

Figure 6. Estimated effective migration graph across the entire distribution of the *Myioborus ornatus-melanocephalus* complex, highlighting areas of restricted gene flow. The edges of the grid are colored by estimated effective migration, from burnt orange (low migration) to blue (high migration). In the grid, the black dots correspond to the sampled nodes and the size of the dot correspond to the sample size (*N*). We highlight three areas of low migration, two geographic barriers and the hybrid zone where genetically distinct taxa have come into contact.

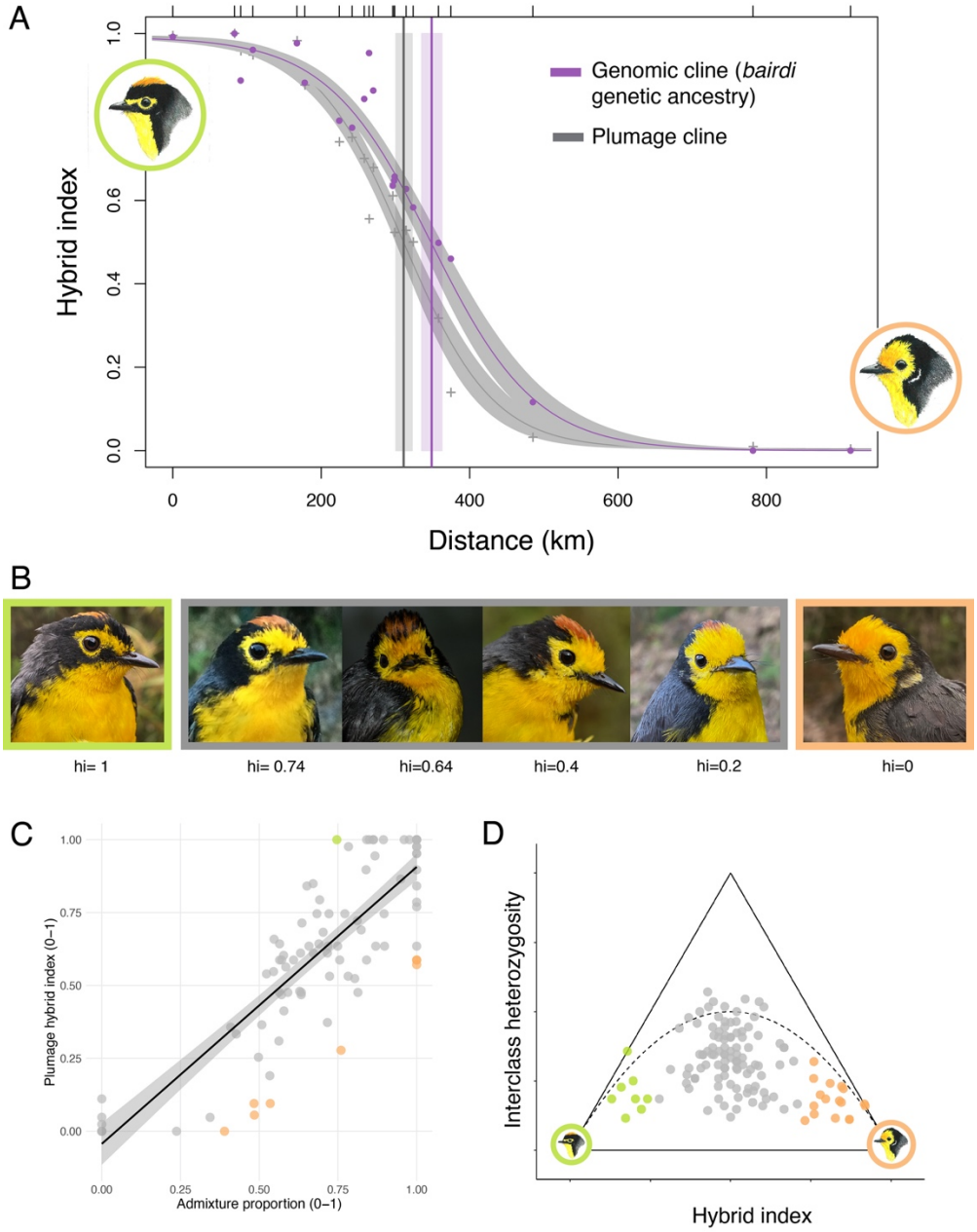

Figure 7. Genetic and plumage variation across the *bairdi x chrysops* hybrid zone. **(A)** The plumage and genomic clines have different estimated centers. **(B)** Photographs of specimens collected for which we calculated a plumage hybrid index, showing a sample of the phenotypic diversity across the hybrid zone, and "parental" phenotypes at both extremes. Photos: Andrés Cuervo, Paulo Pulgarín, David Ocampo, and Laura Céspedes. **(C)** We observed an overall a strong correlation between the genomic hybrid index (*bairdi* admixture proportion) and the plumage hybrid index. Most individuals that deviate from this relationship (outliers marked in color) are individuals with a more chrysops-like plumage than expected given the admixture proportion. **(F)** Triangle plot showing the genetic composition of the hybrid zone. The x axis corresponds to a genomic hybrid index as calculated using 29 ancestry informative markers. Parentals are shown in lime green and orange.

Author contributions: L.N.C.A, A.M.C., C.D.C. and L.C. conceived the study; L.N.C.A. and A.M.C. conducted fieldwork with substantial logistical support from C.D.C. and E.B.; L.N.C.A. conducted lab work with support from L.C. and I.L.; L.N.C.A. conducted the bioinformatic processing of data and data analyses with substantial input from L.C. L.N.C.A. wrote the original draft, with major contributions from L.C. All authors contributed to reviewing and editing the manuscript.

Funding statement: This project had the financial support of the Society of Systematic Biologists (Graduate Student Research Award to L.N.C.A), Association of Field Ornithologists (Bergstrom Award to L.N.C.A), Neotropical Ornithological Society (François Vuilleumier Fund to L.N.C.A), National Geographic Society (Young Explorers Grant WW-R014-17 to L.N.C.A. and, Support for Women + Dependent Care grant to L.N.C.A.) and the Fuller Evolutionary Biology program of the Cornell Lab of Ornithology. The participation of L.N.C.A. in the 2024 Workshop on Genomics, which was instrumental to conduct analyses presented here, was funded by The University of Chicago (Steiner Award and D&I Small Grants of the Biological Sciences Division).

Conflict of Interest: The authors declare no conflicts of interest.

Acknowledgements: We are grateful to Bronwyn Butcher for supporting our work at the Fuller Evolutionary Biology Lab (Cornell Lab of Ornithology). We thank the following people for contributing to fieldwork across the *bairdi-chrysops* hybrid zone which resulted in the collection of most specimens included here: P. Montoya, D. Céspedes-Arias, P. Pulgarín, F. González, E. Obando, M.A. Meneses, S. Herrera, P. Cardozo, and D. Ocampo. Our fieldwork was made possible by many landowners, local guides, nature reserve officials, and other members of local

communities who allowed us access to their land, contributed with invaluable logistical support and even welcomed us warmly into their homes. We are grateful to the Universidad de los Andes, particularly Mireya Osorio, and to the Universidad Tecnológica Indoamérica for their crucial support in obtaining export permits. Collecting and export from Ecuador **(**Contrato de Acceso a Recursos Genéticos MAE-DNB-CM-2015-0017 and N° 40-2018-EXP-CM-FAU-DNB/MA to the Universidad Tecnológica de Indoamerica); Colombia (Number 01374 to the Universidad de los Andes). See Cespedes *et al* 2021 for information on collecting permits.

We are also grateful to museum collections that provided tissue loans to conduct this study: Museum of Southwestern Biology, Kansas University Natural History Museum, Louisiana State University Museum of Natural Science, Pontificia Universidad Católica del Ecuador, Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Universidad de los Andes, and the Cornell University Museum of Vertebrates. We thank Nathan Rice, Jason Weckstein (ANSP), Sebastian C. Perez, Nicholas Mason (LSUMZ) and Santiago Burneo (QCAZ) for providing photographs of museum specimens. Finally, we thank Shannon Hackett, John Bates, Ben Marks (FMNH), Socorro Sierra and Gustavo Bravo (IavH) for facilitating the examination of study skins.

For computing cluster access and technical support, we thank the High-Performance Computing center at Cornell University (BioHPC) and The University of Chicago Center for Research Informatics. We thank Josie Paris for general advice on Stacks use, and Stepfanie Aguillon for suggestions on clustering and phylogenetic analyses. For helpful discussions L.N.C.A thanks the Lovette lab group, as well as Abhimanyu Lele, Louise Bodt, John Bates, Trevor D. Price, Marcus Kronforst, John Novembre and Pavitra Muralidhar. Lastly, we are very

grateful to Andrés Montes Rojas, who kindly provided the original illustrations used in the figures.

References

Aguillon, S. M., Campagna, L., Harrison, R. G., & Lovette, I. J. (2018). A flicker of hope: Genomic data distinguish Northern Flicker taxa despite low levels of divergence. *Auk*, *135*(3), 748–766. https://doi.org/10.1642/AUK-18-7.1

Aguillon, S. M., Fitzpatrick, J. W., Bowman, R., Schoech, S. J., Clark, A. G., Coop, G., & Chen, N. (2017). Deconstructing isolation-by-distance: The genomic consequences of limited dispersal. *PLoS Genetics*, *13*(8), 1–27. https://doi.org/10.1371/journal.pgen.1006911

Aguillon, S. M., Walsh, J., & Lovette, I. J. (2021). Extensive hybridization reveals multiple coloration genes underlying a complex plumage phenotype. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1943). https://doi.org/10.1098/rspb.2020.1805

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Andersson, S., Ornborg, J., & Andersson, M. (1998). Ultraviolet sexual dimorphism and assortative mating in blue tits. *Proceedings of the Royal Society of London B: Biological Sciences*, *265*(November 1997), 445–450.

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*.

Baiz, M. D., Wood, A. W., Brelsford, A., Lovette, I. J., & Toews, D. P. L. (2021). Pigmentation Genes Show Evidence of Repeated Divergence and Multiple Bouts of Introgression in *Setophaga* Warblers. *Current Biology*, *31*(3), 643-649.e3.

https://doi.org/10.1016/j.cub.2020.10.094

Baldassarre, D. T., & Webster, M. S. (2013). Experimental evidence that extra-pair mating drives asymmetrical introgression of a sexual trait. *Proceedings of the Royal Society B: Biological Sciences*, *280*(October), 1–7. https://doi.org/10.1098/rspb.2013.2175

Baldassarre, D. T., White, T. A., Karubian, J., & Webster, M. S. (2014). Genomic and morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical introgression of a sexual signal. *Evolution*, *68*(9), 2644–2657. https://doi.org/10.1111/evo.12457

Barker, K., Burns, K. J., Klicka, J., Lanyon, S. M., & Lovette, I. J. (2015). New insights into New World biogeography: An integrated view from the phylogeny of blackbirds, cardinals, sparrows, tanagers, warblers, and allies. *Auk*, *132*(2), 333–348. https://doi.org/10.1642/AUK-14-110.1

Barnes, R., Sahr, K., Evenden, G., Johnson, A., Warmerdam, F., Rouault, E., Song, L., & Krantz, S. (2024). *dggridR: Discrete Global Grids (v. 3.1.0)*.

Barton, N. H., & Gale, K. S. (1993). Genetic analysis of hybrid zones. In R. G. Harrison (Ed.), *Hybrid zones and the evolutionary process* (pp. 13–45). Oxford University Press.

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics. Vol. 16*, *16*, 113–148. https://doi.org/10.1146/annurev.es.16.110185.000553

Behr, A. A., Liu, K. Z., Liu-Fang, G., Nakka, P., & Ramachandran, S. (2016). Pong: Fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics*, *32*(18), 2817–2823. https://doi.org/10.1093/bioinformatics/btw327

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Müller, N.

F., Ogilvie, H. A., Du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., … Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *15*(4), 1–28. https://doi.org/10.1371/journal.pcbi.1006650

Brelsford, A., & Irwin, D. E. (2009). Incipient speciation despite little assortative mating: The yellow-rumped warbler hybrid zone. *Evolution*, *63*(12), 3050–3060. https://doi.org/10.1111/j.1558-5646.2009.00777.x

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & Roychoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, *29*(8), 1917–1932. https://doi.org/10.1093/molbev/mss086

Campagna, L., Benites, P., Lougheed, S. C., Lijtmaer, A., Giacomo, S. Di, Eaton, M. D., & Tubaro, P. L. (2012). Rapid phenotypic evolution during incipient speciation in a continental avian radiation. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 1847–1856. https://doi.org/10.1098/rspb.2011.2170

Campagna, L., Repenning, M., Silveira, L. F., Fontana, C. S., Tubaro, P. L., & Lovette, I. J. (2017). Repeated divergent selection on pigmentation genes in a rapid finch radiation. *Science Advances*, *3*(5), 1–11. https://doi.org/10.1126/sciadv.1602404

Capblancq, T., Després, L., & Mavárez, J. (2020). Genetic, morphological and ecological variation across a sharp hybrid zone between two alpine butterfly species. *Evolutionary Applications*, *13*(6), 1435–1450. https://doi.org/10.1111/eva.12925

Céspedes-Arias, L. N., Cuervo, A. M., Bonaccorso, E., Castro-Farias, M., Mendoza-Santacruz, A., Pérez-Emán, J. L., Witt, C. C., & Cadena, C. D. (2021). Extensive hybridization

between two Andean warbler species with shallow divergence in mtDNA. *Ornithology*, *138*(1), 1–28. https://doi.org/10.1093/ornithology/ukaa065

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 1–16. https://doi.org/10.1186/s13742-015-0047-8

Chapman, F. M. (1927). DESCRIPTIONS OF NEW BIRDS FROM NORTHWESTERN PERU AND WESTERN COLOMBIA. *American Museum Novitates*, *250*, 1–8.

Cronemberger, Á. A., Aleixo, A., Mikkelsen, E. K., & Weir, J. T. (2020). Postzygotic isolation drives genomic speciation between highly cryptic Hypocnemis antbirds from Amazonia. In *Evolution* (Vol. 74, Issue 11). https://doi.org/10.1111/evo.14103

Cuervo, A. M. (2013). *Evolutionary assembly of the Neotropical montane avifauna*. Lousiana State University.

Cuervo, A. M., & Céspedes Arias, L. N. (2023). The type of *Setophaga ruficoronata* (Kaup 1851) is a hybrid: implications for the taxonomy of Myioborus warblers (Passeriformes: Parulidae). *Zootaxa*, *5383*(4), 476–490. https://doi.org/10.11646/zootaxa.5383.4.3

Curson, J., David, Q., & Beadle, D. (1994). *Warblers of the Americas: An Identification Guide*. Houghton Mifflin Harcourt.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

DeRaad, D. A., Applewhite, E. E., Tsai, W. L. E., Terrill, R. S., Kingston, S. E., Braun, M. J., & McCormack, J. E. (2023). Hybrid zone or hybrid lineage: a genomic reevaluation of

Sibley's classic species conundrum in Pipilo towhees. *Evolution*, *77*(3), 852–869. https://doi.org/10.1093/evolut/qpac068

Derryberry, E. P., Derryberry, G. E., Maley, J. M., & Brumfield, R. T. (2014). HZAR: Hybrid zone analysis using an R software package. *Molecular Ecology Resources*, *14*(3), 652–663. https://doi.org/10.1111/1755-0998.12209

Edwards, S. V., Potter, S., Schmitt, C. J., Bragg, J. G., & Moritz, C. (2016). Reticulation, divergence, and the phylogeography-phylogenetics continuum. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(29), 8025–8032. https://doi.org/10.1073/pnas.1601066113

Fitak, R. R. (2021). OptM: Estimating the optimal number of migration edges on population trees using Treemix. *Biology Methods and Protocols*, *6*(1), 1–6. https://doi.org/10.1093/biomethods/bpab017

Flantua, S. G. A., Hooghiemstra, H., Boxel, J. H. Van, Cabrera, M., González Carranza, Z., & González-Arango, C. (2014). Connectivity dynamics since the Last Glacial Maximum in the northern Andes; a pollen-driven framework to assess potential migration. In W. D. Stevens, O. M. Montiel, & P. Raven (Eds.), *Paleobotany and Biogeography: A Festschrift for Alan Graham in His 80th Year* (pp. 98–123). Missouri Botanical Garden Press.

Friis, G., Aleixandre, P., Rodríguez-Estrella, R., Navarro-Sigüenza, A. G., & Milá, B. (2016). Rapid postglacial diversification and long-term stasis within the songbird genus Junco: phylogeographic and phylogenomic evidence. *Molecular Ecology*, *25*(24), 6175–6195. https://doi.org/10.1111/mec.13911

Gay, L., Crochet, P. A., Bell, D. A., & Lenormand, T. (2008). Comparing clines on molecular and phenotypic traits in hybrid zones: A window on tension zone models. *Evolution*, *62*(11),

2789–2806. https://doi.org/10.1111/j.1558-5646.2008.00491.x

Gompert, Z., DeRaad, D. A., & Buerkle, C. A. (2024). A next generation of hierarchical Bayesian analyses of hybrid zones enables direct quantification of variation in introgression in R. *BioRxiv*, 2003–2024. https://doi.org/10.1002/ece3.70548

Graham, C. H., Silva, N., & Velásquez-Tibatá, J. (2010). Evaluating the potential causes of range limits of birds of the Colombian Andes. *Journal of Biogeography*, *37*(10), 1863–1875. https://doi.org/10.1111/j.1365-2699.2010.02356.x

Graves, G. R. (1988). Linearity of geographic range and its possible effect on the population structure of Andean birds. *The Auk*, *105*(1), 47–52.

Gutiérrez-Pinto, N., Cuervo, A. M., Miranda, J., Pérez-Emán, J. L., Brumfield, R. T., & Cadena, C. D. (2012). Non-monophyly and deep genetic differentiation across low-elevation barriers in a Neotropical montane bird (*Basileuterus tristriatus*; Aves: Parulidae). *Molecular Phylogenetics and Evolution*, *64*(1), 156–165. https://doi.org/10.1016/j.ympev.2012.03.011

Gutiérrez-Pinto, N., Cuervo, A. M., Miranda, J., Pérez-Emán, J. L., Brumfield, R. T., Cadena, C. D., Gutiérrez-Pinto, N., Cuervo, A. M., Miranda, J., & Pérez-Emán, J. L. (2017). Non-monophyly and deep genetic differentiation across low-elevation barriers in a Neotropical montane bird ... *Molecular Phylogenetics and Evolution*, *64*(August), 156–165. https://doi.org/10.1016/j.ympev.2012.03.011

Harrison, R. G. (1986). Pattern and process in a narrow hybrid zone. *Heredity*, *56*(3), 337–349. https://doi.org/10.1038/hdy.1986.55

Hazzi, N. A., Moreno, J. S., Ortiz-Movliav, C., & Palacio, R. D. (2018). Biogeographic regions and events of isolation and diversification of the endemic biota of the tropical Andes. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(31),

7985–7990. https://doi.org/10.1073/pnas.1803908115

Irwin, D. E., Rubtsov, A. S., & Panov, E. N. (2009). Mitochondrial introgression and replacement between yellowhammers (*Emberiza citrinella*) and pine buntings (*Emberiza leucocephalos*) (Aves: Passeriformes). *Biological Journal of the Linnean Society*, *98*, 422–438.

Jenkins, T. L. (2024). mapmixture: An R package and web app for spatial visualisation of admixture and population structure. *Molecular Ecology Resources*, *24*(4), 1–6. https://doi.org/10.1111/1755-0998.13943

Kearns, A. M., Restani, M., Szabo, I., Schrøder-Nielsen, A., Kim, J. A., Richardson, H. M., Marzluff, J. M., Fleischer, R. C., Johnsen, A., & Omland, K. E. (2018). Genomic evidence of speciation reversal in ravens. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-03294-w

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Lim, H. C., Bennett, K. F. P., Justyn, N. M., Powers, M. J., Long, K. M., Kingston, S. E., Lindsay, W. R., Pease, J. B., Fuxjager, M. J., Bolton, P. E., Balakrishnan, C. N., Day, L. B., Parsons, T. J., Brawn, J. D., Hill, G. E., & Braun, M. J. (2024). Sequential introgression of a carotenoid processing gene underlies sexual ornament diversity in a genus of manakins. *Science Advances* , *10*(47). https://doi.org/10.1126/sciadv.adn8339

Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, *19*(3), 639–647. https://doi.org/10.1111/1755-0998.12995

Liu, S., Zhang, L., Sang, Y., Lai, Q., Zhang, X., Jia, C., Long, Z., Wu, J., Ma, T., Mao, K., Street, N. R., Ingvarsson, P. K., Liu, J., & Wang, J. (2022). Demographic History and Natural Selection Shape Patterns of Deleterious Mutation Load and Barriers to Introgression across Populus Genome. *Molecular Biology and Evolution*, *39*(2), 1–16. https://doi.org/10.1093/molbev/msac008

Long, K. M., Rivera-Colón, A. G., Bennett, K. F. P., Catchen, J. M., Braun, M. J., & Brawn, J. D. (2024). Ongoing introgression of a secondary sexual plumage trait in a stable avian hybrid zone. *Evolution*, *78*(9), 1539–1553. https://doi.org/10.1093/evolut/qpae076

Lopez, K. A., McDiarmid, C. S., Griffith, S. C., Lovette, I. J., & Hooper, D. M. (2021). Evaluating evidence of mitonuclear incompatibilities with the sex chromosomes in an avian hybrid zone. *Evolution*, *75*(6), 1395–1414. https://doi.org/10.1111/evo.14243

Lovette, I. J., Pérez-Emán, J. L., Sullivan, J. P., Banks, R. C., Fiorentino, I., Córdoba-Córdoba, S., Echeverry-Galvis, M., Barker, F. K., Burns, K. J., Klicka, J., Lanyon, S. M., & Bermingham, E. (2010). A comprehensive multilocus phylogeny for the wood-warblers and a revised classification of the Parulidae (Aves). *Molecular Phylogenetics and Evolution*, *57*(2), 753–770. https://doi.org/10.1016/j.ympev.2010.07.018

Malinsky, M., Trucchi, E., Lawson, D. J., & Falush, D. (2018). RADpainter and fineRADstructure: Population Inference from RADseq Data. *Molecular Biology and Evolution*, *35*(5), 1284–1290. https://doi.org/10.1093/molbev/msy023

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*, *26*(22), 2867–2873. https://doi.org/10.1093/bioinformatics/btq559

Marchetti, K. (1992). Dark habitats and bright birds illustrate the role of the environment in

species divergence. *Nature*, *362*, 149–152.

Marcus, J., Ha, W., Barber, R. F., & Novembre, J. (2021). Fast and flexible estimation of effective migration surfaces. *ELife*, *10*, 1–46. https://doi.org/10.7554/eLife.61927

Márquez, R., Linderoth, T. P., Mejía-Vargas, D., Nielsen, R., Amézquita, A., & Kronforst, M. R. (2020). Divergence, gene flow, and the origin of leapfrog geographic distributions: The history of colour pattern variation in Phyllobates poison-dart frogs (preprint). In *BioRxiv*. https://doi.org/10.1111/mec.15598

Martínez-Gómez, S. C., Lara, C. E., Remsen, J. V., Brumfield, R. T., & Cuervo, A. M. (2023). Unmasking hidden genetic, vocal, and size variation in the Masked Flowerpiercer along the Andes supports two species separated by Northern Peruvian Low. *Ornithology*, *140*(4), 1–14. https://doi.org/10.1093/ornithology/ukad028

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

Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, *21*(12), 2839–2846. https://doi.org/10.1111/j.1365-294X.2012.05578.x

Molloy, E. K., Durvasula, A., & Sankararaman, S. (2021). Advancing admixture graph estimation via maximum likelihood network orientation. *Bioinformatics*, *37*, I142–I150. https://doi.org/10.1093/bioinformatics/btab267

Mumme, R. L. (2015). Demography of Slate-throated Redstarts (*Myioborus miniatus*): a non-migratory Neotropical warbler. *Journal of Field Ornithology*, *86*(2), 89–102. https://doi.org/10.1111/jofo.12093

Natola, L., Curtis, A., Hudon, J., & Burg, T. M. (2021). Introgression between Sphyrapicus nuchalis and S. varius sapsuckers in a hybrid zone in west-central Alberta. *Journal of Avian*

Biology, *52*(8), 1–12. https://doi.org/10.1111/jav.02717

Nosil, P. (2008). Ernst Mayr and the integration of geographic and ecological factors in speciation. *Biological Journal of the Linnean Society*, *95*(1), 26–46. https://doi.org/10.1111/j.1095-8312.2008.01091.x

Nychka, D., Furrer, R., Paige, J., Sain, S., & Nychka, M. D. (2025). *Package 'fields.'* https://doi.org/10.5065/D6W957CT>.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M. H. H., Oksanen, M. J., & Suggests, M. A. S. S. (2007). *"The vegan package." Community ecology package*.

Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution*, *8*(10), 1360–1373. https://doi.org/10.1111/2041-210X.12775

Pérez-Emán, J. L. (2005). Molecular phylogenetics and biogeography of the Neotropical redstarts (*Myioborus*; Aves, Parulinae). *Molecular Phylogenetics and Evolution*, *37*(2), 511–528. https://doi.org/10.1016/j.ympev.2005.04.013

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, *7*(5). https://doi.org/10.1371/journal.pone.0037135

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics*, *8*(11). https://doi.org/10.1371/journal.pgen.1002967

Poelstra, J. W., Vijay, N., Bossu, C. M., Lantz, H., Ryll, B., Müller, I., Baglione, V., Unneberg, P., Wikelski, M., Grabherr, M. G., & Wolf, J. B. W. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*, *344*(6190),

1410–1414. https://doi.org/10.1126/science.1253226

Pool, J. E., & Nielsen, R. (2009). Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics*, *181*(2), 711–719. https://doi.org/10.1534/genetics.108.098095

Pozo-Zamora, G. M., Krabbe, N., Mena-Valenzuela, P., Nilsson, J., & Brito, J. (2022). Birds of the Cordillera del Kutukú, Morona Santiago, Southeastern Ecuador. *Revista Peruana de Biologia*, *29*(1). https://doi.org/10.15381/RPB.V29I1.20667

Price, T. (2008). *Speciation in Birds*. Robert and Co.

Prieto-Torres, D. A., Cuervo, A. M., & Bonaccorso, E. (2018). On geographic barriers and Pleistocene glaciations : Tracing the diversification of the Russet-crowned Warbler (*Myiothlypis coronata*) along the Andes. *Plos One*, *13*(3), e0191598.

Rahbek, C., Borregaard, M. K., Colwell, R. K., Dalsgaard, B., Holt, B. G., Morueta-Holme, N., Nogues-Bravo, D., Whittaker, R. J., & Fjeldså, J. (2019). Humboldt's enigma: What causes global patterns of mountain biodiversity? *Science*, *365*(6458), 1108–1113. https://doi.org/10.1126/science.aax0149

Rambaut, A., Suchard, M. A., Xie, D., & Drummond, A. J. (2014). *Tracer v1.6*.

Ramírez-Barahona, S., & Eguiarte, L. E. (2013). The role of glacial cycles in promoting genetic diversity in the Neotropics : the case of cloud forests during the Last Glacial Maximum. *Ecology and Evolution*, *3*(3), 725–738. https://doi.org/10.1002/ece3.483

Remsen, J. V. (1984). High Incidence of "Leapfrog" Pattern of Geographic variation in Andean Birds: Implications for the speciation process. *Science*, *224*, 171–173.

Rémy, A., Grégoire, A., Perret, P., & Doutrelant, C. (2010). Mediating male-male interactions: The role of the UV blue crest coloration in blue tits. *Behavioral Ecology and Sociobiology*,

64(11), 1839–1847. https://doi.org/10.1007/s00265-010-0995-z

Rivera-Colón, A. G., & Catchen, J. (2022). Population Genomics Analysis with RAD, Reprised: Stacks 2 (Chapter 7). In C. Verde & D. Giordano (Eds.), *Marine Genomics: Methods and Protocols*. Springer Nature- Humana Press.

Robbins, M. B., Ridgely, R. S., & Schulenberg, T. S. (1987). The avifauna of the Cordillera de Cutucut, Ecuador, with Comparisons to other Andean Localities. *Proceedings of The Academy of Natural Sciences of Philadelphia*, *139*(1), 243–259.

Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, *28*(21), 4737–4754. https://doi.org/10.1111/mec.15253

Sambrook, J. (1987). *Molecular cloning: a laboratory manual*.

Sanín, M. J., Cardona, A., Céspedes-Arias, L. N., González-Arango, C., Pardo, N., & Cadena, C. D. (2024). Volcanoes, evolving landscapes, and biodiversity in Neotropical mountains. *Frontiers of Biogeography*, *16*(1), 1–14. https://doi.org/10.21425/F5FBG61882

Scordato, E. S. C., Wilkins, M. R., Semenov, G., Rubtsov, A. S., Kane, N. C., & Safran, R. J. (2017). Genomic variation across two barn swallow hybrid zones reveals traits associated with divergence in sympatry and allopatry. *Molecular Ecology*, *26*(20), 5676–5691. https://doi.org/10.1111/mec.14276

Scott, G. W., & Deag, J. M. (1998). Blue Tit (*Parus caeruleus*) agonistic displays : A reappraisal. *Behaviour*, *135*(6), 665–691.

Seeholzer, G. F., & Brumfield, R. T. (2018). Isolation by distance, not incipient ecological speciation, explains genetic differentiation in an Andean songbird (Aves: Furnariidae: Cranioleuca antisiensis, Line-cheeked Spinetail) despite near threefold body size change

across an environmental gradien. *Molecular Ecology*, *27*(1), 279–296. https://doi.org/10.1111/mec.14429

Semenov, G. A., Linck, E., Enbody, E. D., Harris, R. B., Khaydarov, D. R., Alström, P., Andersson, L., & Taylor, S. A. (2021). Asymmetric introgression reveals the genetic architecture of a plumage trait. *Nature Communications*, *12*(1), 1–9. https://doi.org/10.1038/s41467-021-21340-y

Sonne, J., Dalsgaard, B., Borregaard, M. K., Kennedy, J., Fjeldså, J., & Rahbek, C. (2022). Biodiversity cradles and museums segregating within hotspots of endemism. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1981). https://doi.org/10.1098/rspb.2022.1102

Stein, A. C., & Uy, J. A. C. (2006). Plumage brightness predicts male mating success in the lekking golden-collared manakin, Manacus vitellinus. *Behavioral Ecology*, *17*(1), 41–47. https://doi.org/10.1093/beheco/ari095

Stoltz, M., Baeumer, B., Bouckaert, R., Fox, C., Hiscott, G., & Bryant, D. (2021). Bayesian inference of species trees using diffusion models. *Systematic Biology*, *70*(1), 145–161. https://doi.org/10.1093/sysbio/syaa051

Stryjewski, K. F., & Sorenson, M. D. (2017). Mosaic genome evolution in a recent and rapid avian radiation. *Nature Ecology and Evolution*, *1*(12), 1912–1922. https://doi.org/10.1038/s41559-017-0364-7

Thrasher, D. J., Butcher, B. G., Campagna, L., Webster, M. S., & Lovette, I. J. (2018). Double-digest RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: A proof of concept in a highly promiscuous bird. *Molecular Ecology Resources*, *18*(5), 953–965. https://doi.org/10.1111/1755-0998.12771

Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, *21*, 3907–3930. https://doi.org/10.1111/j.1365-294X.2012.05664.x

Toews, D. P. L., Campagna, L., Taylor, S. A., Balakrishnan, C. N., Baldassere, D. T., Deane-Coe, P. E., Harvey, M. G., Hooper, D. M., Irwin, D. E., Judy, C. D., Mason, N. A., McCormack, J. E., McCracken, K. G., Oliveros, C. H., Safran, R. J., Scordato, E. S. C., Stryjewski, K. F., Tigano, A., Uy, J. A. C., & Winger, B. (2016). Genomic approaches to understanding the early stages of population divergence and speciation in birds. *The Auk*, *133*, 13–30. https://doi.org/10.1642/AUK-15-51.1

Toews, D. P. L., Taylor, S. A., Vallender, R., Brelsford, A., Butcher, B., Messer, P. W., & Lovette, I. J. (2016). Plumage genes and little else distinguish the genomes of hybridizing warblers. *Current Biology*, *26*(17), 2313–2318. https://doi.org/10.1016/j.cub.2016.06.034

Uy, J. A. C., Irwin, D. E., & Webster, M. S. (2018). Behavioral isolation and incipient speciation in birds. *Annual Review of Ecology, Evolution, and Systematics*, *49*, 1–24. https://doi.org/10.1146/annurev-ecolsys-110617-062646

Vos, M., & Velicer, G. J. (2008). Isolation by Distance in the Spore-Forming Soil Bacterium Myxococcus xanthus. *Current Biology*, *18*(5), 386–391. https://doi.org/10.1016/j.cub.2008.02.050

Vuilleumier, F. (1969). Pleistocene speciation of birds living in the high Andes. *Nature*, *223*, 1179–1180.

Walsh, J., Lovette, I. J., Winder, V., Elphick, C. S., Olsen, B. J., Shriver, G., & Kovach, A. I. (2017). Subspecies delineation amid phenotypic, geographic and genetic discordance in a songbird. *Molecular Ecology*, *26*(5), 1242–1255. https://doi.org/10.1111/mec.14010

Weir, B. ., & Cockerham, C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358–1370.

Wiens, B. J., & Colella, J. P. (2025). triangulaR: an R package for identifying AIMs and building triangle plots using SNP data from hybrid zones. *Heredity*, *134*, 251–262. https://doi.org/10.1038/s41437-025-00760-2

Winger, B. M. (2017). Consequences of divergence and introgression for speciation in Andean cloud forest birds. *Evolution*, *71*(7), 1815–1831. https://doi.org/10.1111/evo.13251

Winger, B. M., & Bates, J. M. (2015). The tempo of trait divergence in geographic isolation: Avian speciation across the Marañon Valley of Peru. *Evolution*, *69*(3), 772–787. https://doi.org/10.1111/evo.12607

World, H. of the B. of the, & International, B. (2019). *Handbook of the Birds of the World and BirdLife International digital checklist of the birds of the world. Version 4.*

Yi, X., & Latch, E. K. (2022). Nonrandom missing data can bias Principal Component Analysis inference of population genetic structure. *Molecular Ecology Resources*, *22*(2), 602–611. https://doi.org/10.1111/1755-0998.13498

Zamudio, K. R., Bell, R. C., & Mason, N. A. (2016). Phenotypes in phylogeography: Species' traits, environmental variation, and vertebrate diversification. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(29), 8041–8048. https://doi.org/10.1073/pnas.1602237113

Zheng, X. (2013). *Package ' SNPRelate ': A High-performance computing toolset for relatedness and principal component analysis of SNP data*.

Zimmer, J. T. (1949). Studies of Peruvian birds. No. 54, The families Catamblyrhynchidae and Parulidae. *American Museum Novitates*, *1428*, 1–59.

Supplementary information- History of divergence and gene flow shaping geographic variation in color in Andean warbles

Supplementary information- Methods

DNA extraction and library preparation (full)

We extracted genomic DNA from pectoral muscle (n=232) and blood samples (n=3) employing a modified phenol-chlorophorm protocol (Sambrook 1987), using phase-lock tubes. Each DNA extraction was subsequently purified using homemade magnetic nucleic acid extraction (MagNA) made with Sera-Mag Magnetic Speed-beads. To obtain reduced representation genomic data, we used double-digest restriction site-associated DNA sequencing (ddRAD) following Peterson et al. (2012), and using modifications from Thrasher et al. (2018). Before preparing the genomic libraries, we measured DNA concentrations using a Qubit fluorometer (Life Technologies, Carlsbad, California, USA). For library preparation, we used approximately 500 ng of DNA per extraction, which had concentrations ranging from 6.86–55 ng/µL. Extractions of four individuals for which we had concerns about quality were included twice in the libraries with a different barcode (i.e., there were a total of 239 samples included, corresponding to 235 individuals). After demultiplexing, the reads corresponding to the same individual were concatenated into the same file (see below).

We prepared libraries as follows. First, DNA was digested by incubating the samples with the enzymes SbfI (8 base pair [bp] recognition site) and MspI (4 bp recognition site; New England BioLabs, Ipswich, Massachusetts, USA). Simultaneously, the P1 and P2 adaptors were ligated using T4 DNA Ligase (New England BioLabs). After digestion and ligation, samples with unique P1 barcodes were pooled together in 12 index groups of 20 individuals, each

identified with a unique P2 adapter. After pooling, we purified the DNA in each index group using homemade MagNA (1.5X volumes). Then DNA fragments between 450 and 600 bp were selected using Blue Pippin, by the Cornell University Biotechnology Resource Center (BRC). After the size selection step, we incorporated the index group and Illumina sequencing adapters by conducting a PCR (as in Thrasher et al. 2018). We then purified the reactions using MagNA, and pooled equimolarly to obtain a single library for Illumina sequencing.

Bioinformatics pipeline

Optimization of *de novo* assembly parameters (full)- Before assembling the reads, we conducted an optimization of the parameters *M* and *n* used by *denovo_map.pl* (Rochette et al. 2019). These parameters determine the number of mismatches allowed between stacks to merge into putative loci, and between loci when building the catalog, respectively (https://catchenlab.life.illinois.edu/stacks/comp-v1/denovo_map.php). Adequate values of these parameters are highly dependent on the dataset's characteristics, such as the genetic diversity and level of divergence among populations, and therefore it is necessary to perform this optimization step for *de novo* assemblies (Paris et al. 2017; Rivera-Colón and Catchen 2022). For this reason, we also conducted three separate optimizations and then, three separate full assemblies: one including representatives from the outgroups, one that only included individuals within the *ornatus-melanocephalus* complex, and one including only samples relevant for the hybrid zone analyses (*chrysops*, individuals in the hybrid zone, *bairdi,* and *griseonuchus*, Figure 1).
We subset the dataset retaining a small number of representative individuals for the three optimization runs (Rivera-Colón and Catchen 2022). Specifically, we included at least one sample per taxon within the complex, and only those that had an average number of reads after

demultiplexing (+/- 0.25 of the standard deviation). These led to population maps with 43 (the full, i.e. including outgroups), 40 (ingroup only), and 32 (hybrid zone only) assemblies. In these three optimization population maps, all of the individuals are grouped into one "population" (Rivera-Colón and Catchen 2022). To optimize *M* and *n* we ran *denovo-map.pl* using 36 combinations of these parameters following recommendations by Paris *et al* (2017) and Rivera-Colón & Catchen (2022): we varied *M* from 1 to 12, and then each value of M was run together with three *n* values (*n=M*, *n=M+1* and *n=M-1*). The parameter *m* was kept constant at the default value that has proven to work well in a diversity of biological scenarios (Paris et al. 2017). In each case we chose the *M/n* combination that maximized the number of polymorphic loci that were shared by at least 80% of the individuals (Paris et al. 2017; Rivera-Colón and Catchen 2022). Ultimately, we chose M=4/n=5 for the full assembly, M=3/n=4 for the only ingroup assembly, and M=3/n=4 for the hybrid zone assembly (Supplementary Table 2-denovo_optimization_results.xls).

SNP datasets (full)- We obtained different SNP datasets to be used in downstream analyses using the *populations* program from STACKS 2.66 (Rochette et al. 2019). For some analyses, as specified below, we did further filtering of SNPs using *vcftools* (Danecek et al. 2011). For the ingroup assembly, we obtained two different SNP datasets: one including one SNP per locus, and one that included all variant and non-variant sites per locus (i.e. full haplotypes only for fineRADStructure, see below). In both cases we applied the following filters: *--min-samples-overall* 0.8 (locus needs to be present in 80% of all individuals to be retained), *--min-mac* 5 (minimum allele count of 5), and *--max-obs-het* 0.7 (maximum observed heterozygosity of 0.7). For the full haplotype dataset, we used the same parameters and included the *--filter-haplotype-*

wise flag. We decided to apply a minimum allele count (*--min-mac* 5) filter to exclude SNPs that may be artifacts of genotyping errors, but kept a relatively low value to avoid excluding all low frequency SNPs which also contain relevant information for population genetic analyses (Linck and Battey 2019).

The assembly including outgroups was only used for phylogenetic (*raxml-ng*, and *snapper*) and admixture graph inference (see below). We obtained slightly different SNP datasets for each analysis, but for all we applied the following filters: *--min-samples-per-pop* 0.5, *--min-populations* {total number of populations in population map, either 10 or 8}, *--min-mac* 2, and *--max-obs-het* 0.7. In this case we lowered the *--min-mac* 2 value because outgroups are represented only by 2 or 3 individuals, and these likely introduce very low frequency alleles. Additional parameters used during filtering are specified below in the corresponding analyses. The "hybrid zone assembly" was used for analyses aimed at describing genetic variation across the *bairdi-chrysops* hybrid zone and adjacent areas. First, informed by our whole complex results (e.g. fineRADStructure, see results), we decided to conduct PCA and admixture analyses for this subset of individuals to further describe genetic structure in southern Ecuador, which is relevant for the patterns of genetic variation across the *bairdi x chrysops* hybrid zone. For this set of analyses, we obtained a dataset including one SNP per locus applying the following filtering: *-min-samples-overall* 0.8, *--min-mac* 5 and *--max-obs-het* 0.7. For hybrid zone-specific analyses (i.e. cline and triangulaR analyses, see below) we obtained a SNP dataset excluding all *griseonuchus* (from northern Peru) as well as individuals from southern Ecuador that genetically cluster with *griseonuchus* (see results). For these hybrid zone-specific analyses, we obtained SNP datasets both including one SNP and all SNPs per locus for different analyses. We used the same SNP filtering scheme as for the SNP dataset including *griseonuchus*.

Based on the assembly for the ingroup only, and after filtering, we obtained 5,054 loci, composed of 73,786 sites, including 22,132 variant sites for the full haplotype dataset. For the ingroup (one SNP per locus) dataset, we obtained the same number of loci and kept 4,800 variant sites. The number of SNPs for the other two assemblies (i.e., including outgroups and for the hybrid zone only), as well as for different filtering employed for different analyses is fully detailed in the Supplementary Table 2

References

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

Linck, E., and C. J. Battey. 2019. Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. Mol. Ecol. Resour. 19:639–647.

Paris, J. R., J. R. Stevens, and J. M. Catchen. 2017. Lost in parameter space: a road map for stacks. Methods Ecol. Evol. 8:1360–1373.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7.

Rivera-Colón, A. G., and J. Catchen. 2022. Population Genomics Analysis with RAD, Reprised: Stacks 2 (Chapter 7). P. *in* C. Verde and D. Giordano, eds. Marine Genomics: Methods and Protocols. Springer Nature- Humana Press.

Rochette, N. C., A. G. Rivera-Colón, and J. M. Catchen. 2019. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol. Ecol. 28:4737–

4754.

Sambrook, J. 1987. Molecular cloning: a laboratory manual. New York.

Thrasher, D. J., B. G. Butcher, L. Campagna, M. S. Webster, and I. J. Lovette. 2018. Double-digest RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: A proof of concept in a highly promiscuous bird. Mol. Ecol. Resour. 18:953–965.

Supplementary Figures

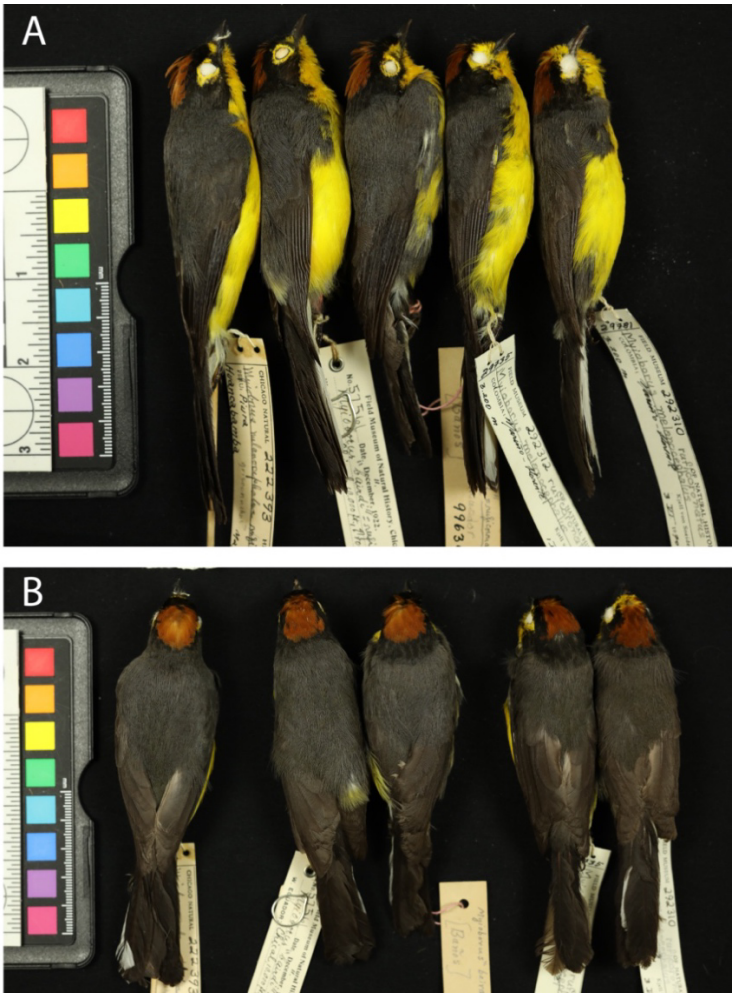

Figure S1. Sample of plumage variation from northern Peru (Piura) to southern Colombia (Nariño). The first specimen, from left to right, corresponding clearly to the *griseonuchus* group, shows a wide black band below the yellow spectacles that extend down to the cheeks, "broken" yellow spectacles and lacks a black band in the nape (i.e. the rufous crown is in direct contact with the grayish back color). The two specimens in the middle, collected in the known distribution of the *bairdi* group, also show a complete black band below the yellow spectacles, but vary on the extent that these extend down to the cheek (with the specimen to the left being more *griseonuchus*-like). In both cases the yellow spectacles are not broken. The two specimens also differ in the extent of the black nape, which is also one of the described differences between these plumage forms. The specimen closer to the left shows a very narrow black nape band, somehow resembling the *griseonuchus* specimen. The last two specimens, from left to right, were collected in the known *chrysops x bairdi* hybrid zone (see more in Figure 7 about plumage variation in the hybrid zone. The specimens shown from the left to the right: FMNH 222393 (Huancabamba, dept. Piura, Peru); FMNH 57561 (Chical , Prov. Cañar, Ecuador); FMNH 9963 (Baños, prov. Tungurahua, Ecuador); FMNH 292312 and 29310 (Llorente, depto. Nariño, Colombia).

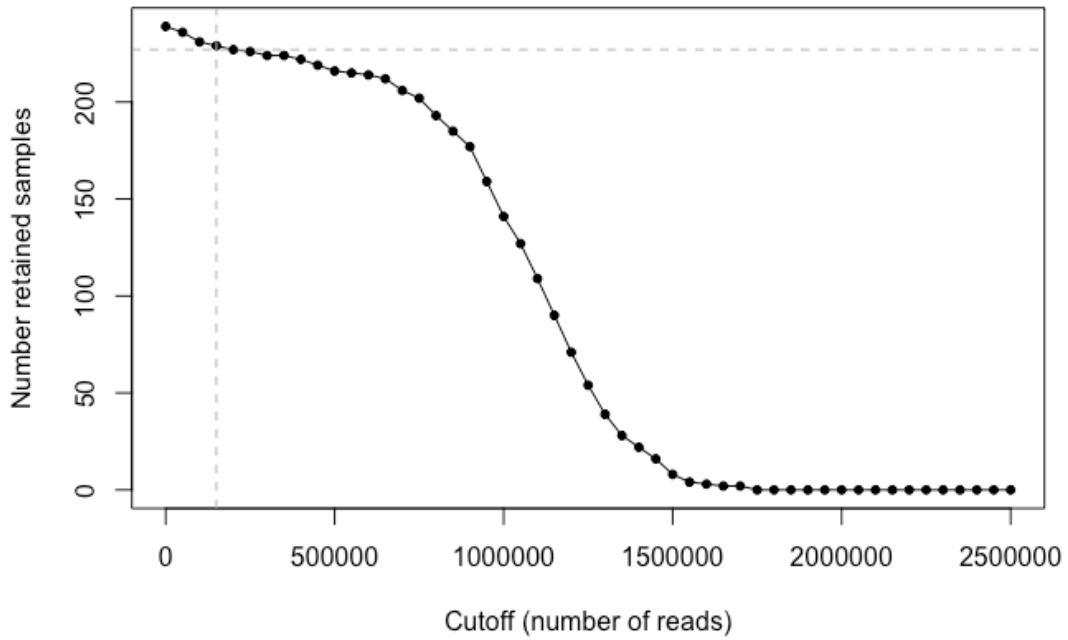

Figure S2. Number of retained samples based on different cutoffs of number of reads. We chose a cut off of 200,000 reads, shown in the dotted vertical line.

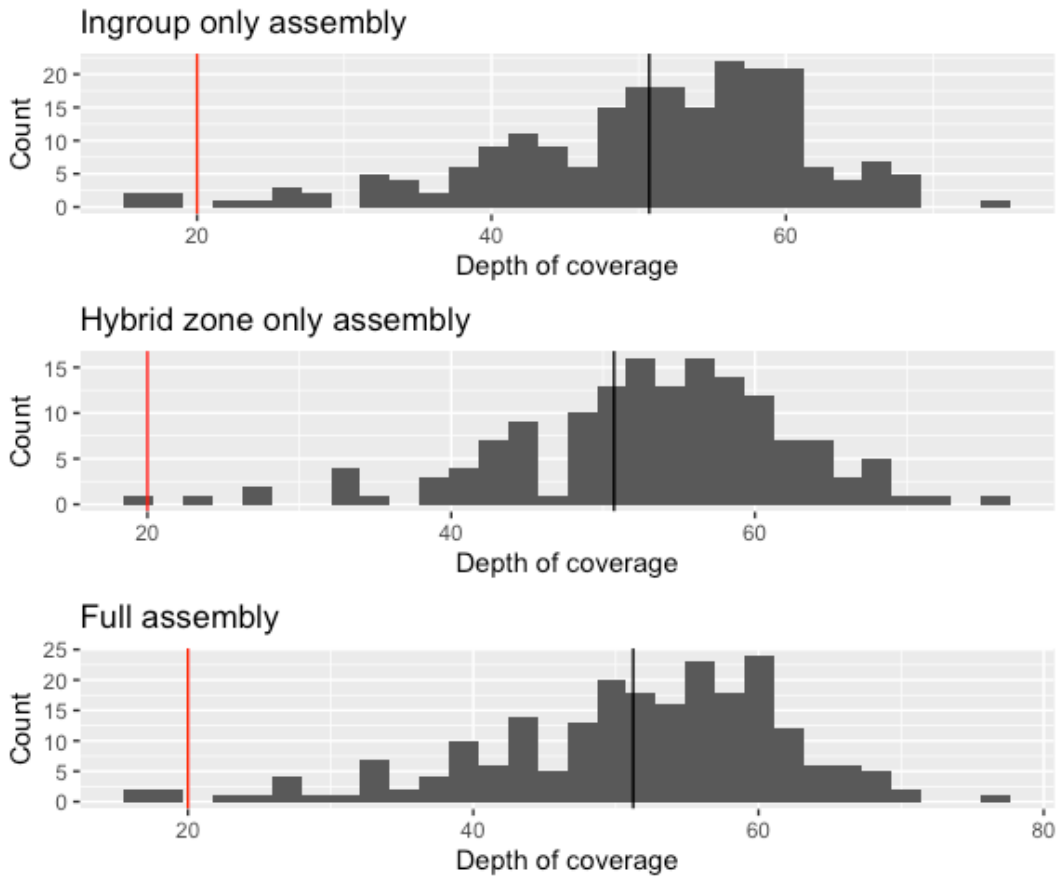

Figure S3. Distribution of coverage for the three assemblies. The red vertical line shows the lower threshold for coverage: all individual below this threshold were discarded from downstream analyses. The black vertical line shows the mean of depth of coverage.

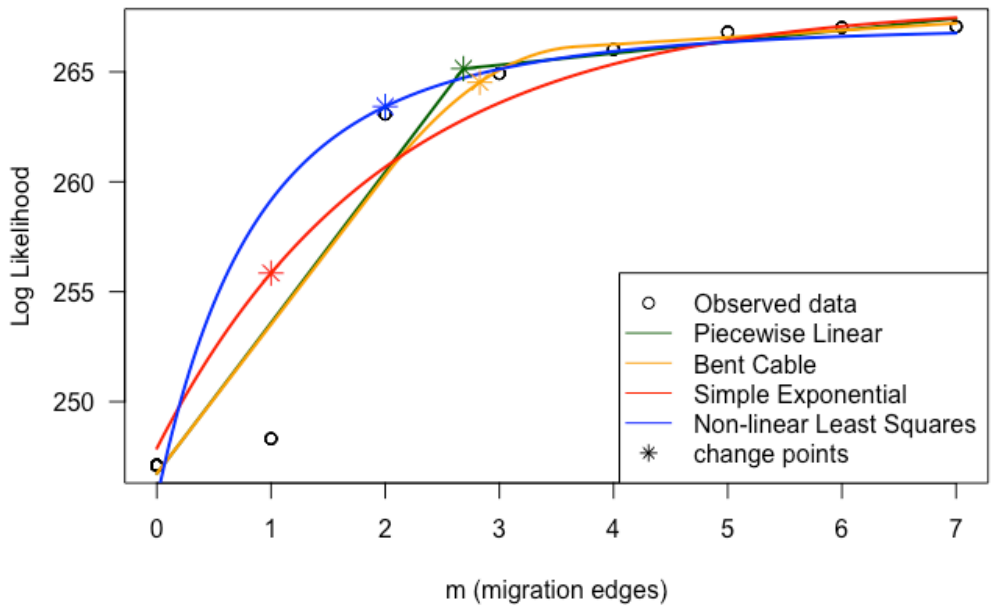

Figure S4. Log-likelihood across different values of migration edges (*OrientAGraph*). The asterisk shows inflection points that can be interpreted as the optimal m values. We show the results for m=1 and m=3, in Figure 2.

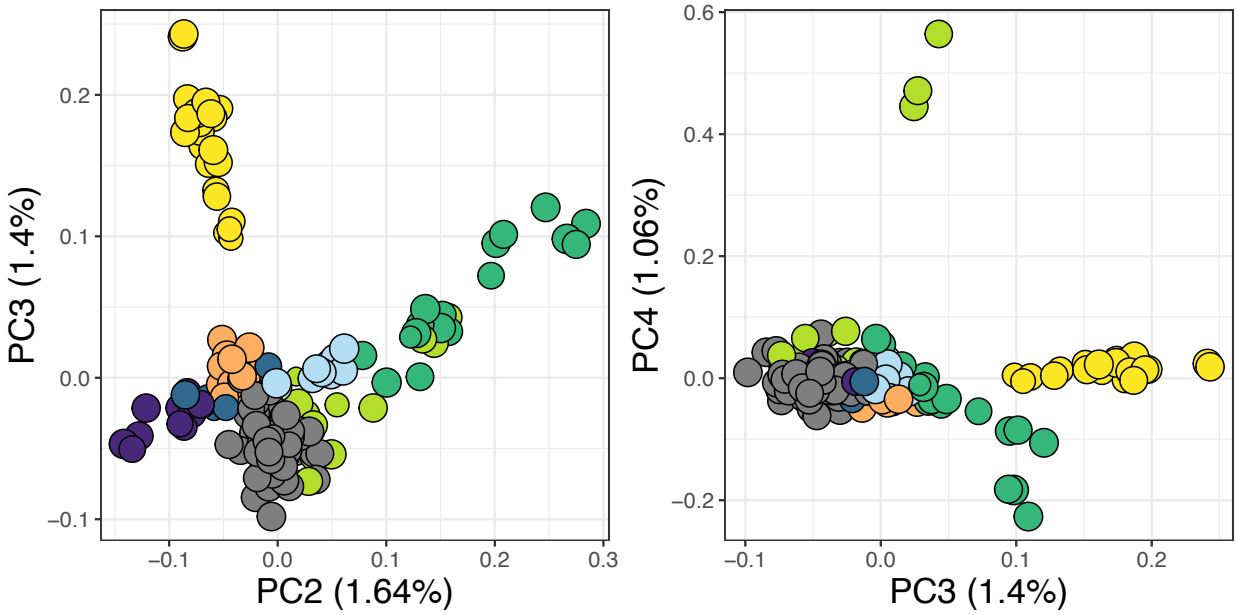

Figure S5. Distribution of individuals in the PC2/PC3 and PC3/PC4 spaces, showing clustering by plumage group which is encoded by color (as in main figures.)

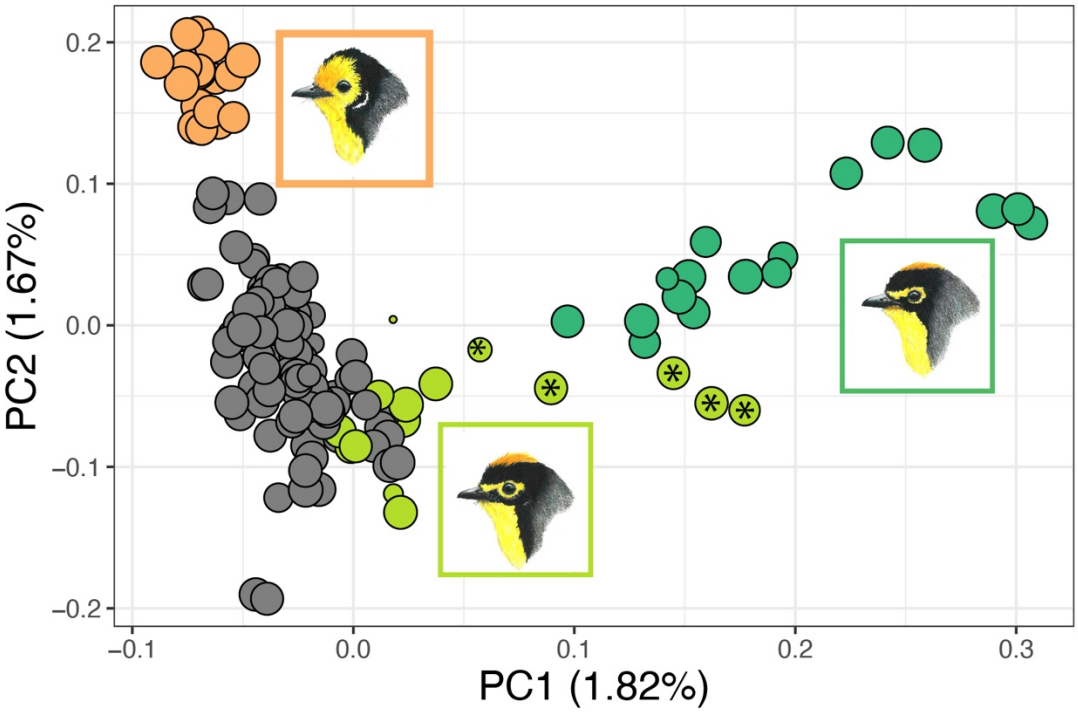

Figure S6. Genetic structure in the hybrid zone, parental populations and griseonuchus, as revealed by a PCA. The asterisks denote individuals from Loja (Ecuador) and outlining ridges in Morona-Santiago that were initially classified as *bairdi* but genetically cluster with *griseonuchus* in other analyses (e.g. fineRADStructure).

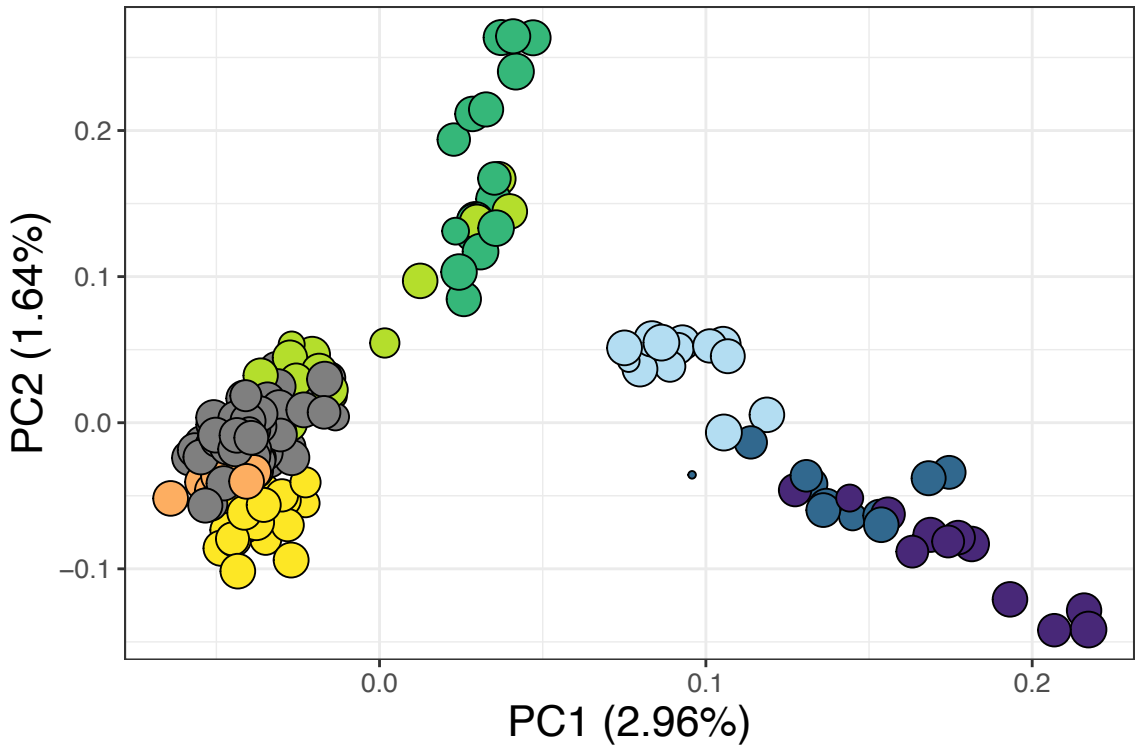

Figure S7. Distribution of individuals across the PC1/PC2 space, using a SNP dataset that excludes SNPs with high levels of missing data. Plumage groups are denoted by color.

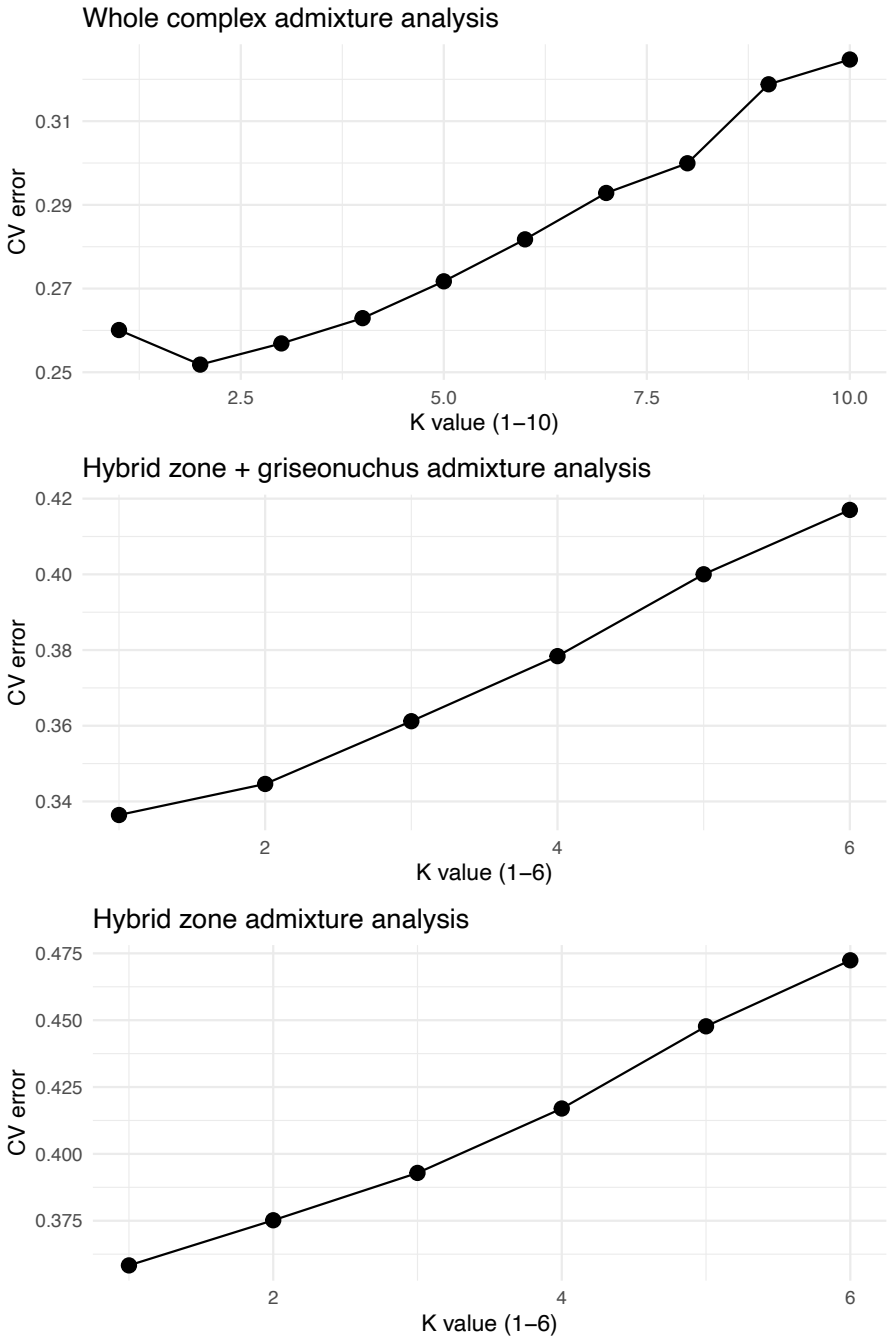

Figure S8. Cross-validation procedure results for admixture analyses.

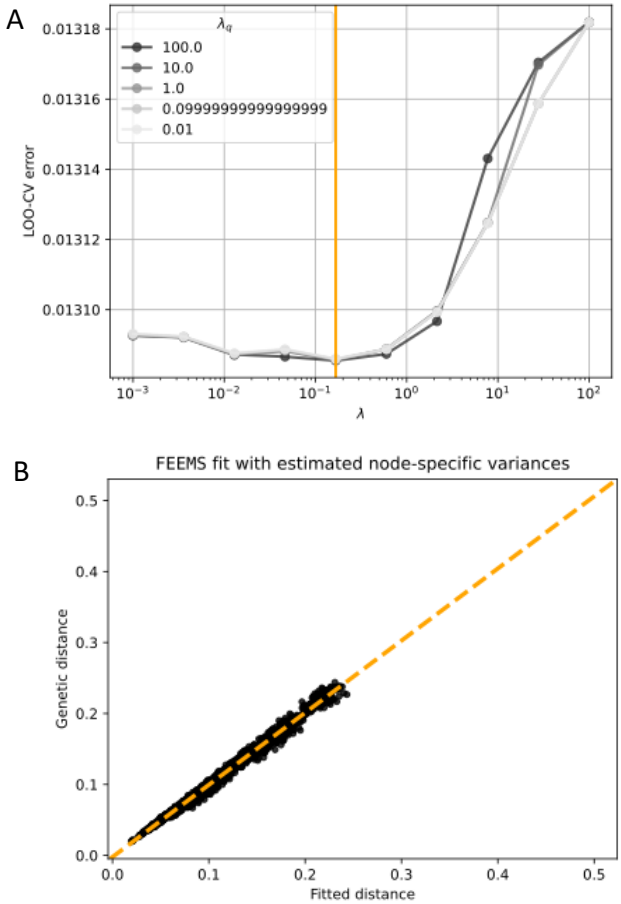

Figure S9. FEEMS model parametrization and fit.

Supplementary Tables

Table S1. Sample information

Table S2. Optimization results (for de novo assembly)