Selection on a pleiotropic color gene block underpins early differentiation between two warbler species

Silu Wang[1]*, Sievert Rohwer[2], Devin R. de Zwaan[3], David P. L Toews[4], Irby J. Lovette[5], Jacqueline Mackenzie[1], Darren E. Irwin[1]

[1] Department of Zoology, 6270 University Blvd, University of British Columbia, BC, Vancouver, BC, V6T1Z4, Canada

[2] Department of Biology and Burke Museum, Box 353010, University of Washington, Seattle, WA, 98195, USA

[3] Department of Forest and Conservation Sciences, 2424 Main Mall, University of British Columbia, BC, V6T1Z4, Canada

[4] Department of Biology, 619 Mueller Laboratory, Pennsylvania State University, University Park, PA, 16802, USA

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

*Corresponding author

Email: wangsmt@zoology.ubc.ca

Abstract

When one species gradually splits into two, divergent selection on specific traits can cause peaks of differentiation in the genomic regions encoding those traits. Whether speciation is initiated by strong selection on a few genomic regions with large effects or by more diffused selection on many regions with small effects remains controversial. Differentiated phenotypes between differentiating lineages are commonly involved in reproductive isolation, thus their genetic underpinnings are key to the genomics architecture of speciation. When two species hybridize, recombination over multiple generations can help reveal the genetic regions responsible for the differentiated phenotypes against a genomic background that has been homogenized via backcrossing and introgression. We used admixture mapping to investigate genomic differentiation and the genetic basis of differentiated plumage features (relative melanin and carotenoid pigment) between hybridizing sister species in the early stage of speciation: Townsend's (*Setophaga townsendi*) and Hermit warblers (*S. occidentalis*). We found a few narrow and dispersed divergent regions between allopatric parental populations, consistent with the 'divergence with gene flow' model of speciation. One of the divergent peaks involves three genes known to affect pigmentation: ASIP, EIF2S2, and RALY (the ASIP-RALY gene block). After controlling for population substructure, we found that a single nucleotide polymorphism (SNP) inside the intron of RALY displays a strong pleiotropic association with cheek, crown, and breast coloration. In addition, we detect selection on the ASIP-RALY gene block, as the geographic cline of the RALY marker of this gene block has remained narrower than the plumage cline, which remained narrower than expected under neutral diffusion over two decades. Despite extensive gene flow between these species across much of the genome, the selection on ASIP-RALY gene block maintains stable genotypic and plumage difference between species allowing

further differentiation to accumulate via linkage to its flanking genetic region or linkage-disequilibrium genome-wide.

Key words: *speciation*, *admixture mapping, cline, plumage, RALY, ASIP, hybrid zone, Setophaga*

Introduction

Examining the genomic distribution of differentiation between two populations can reveal targets of divergent selection, advancing our understanding of the speciation process [1–6]. At the onset of divergence between sister taxa, differentiation at narrow regions of chromosomes, or "islands of differentiation", is expected [7–12]. Strong selection on a few genes of large phenotypic effects can result in a small number of highly differentiated regions. Physical linkage results in elevated divergence of the nearby neutral regions, a phenomenon known as divergence hitchhiking [2,7,11]. More differentiation can accumulate as speciation progresses, thus extending regions of high differentiation and forming "continents of divergence" [4]. Alternatively, speciation can be initiated via many regions of small effects on traits under multifarious selection (one selection targeting multiple traits) and gradually accumulate genome-wide differentiation via correlated differentiation across the genome [13–18]. A central debate in speciation research has therefore been whether speciation occurs through gradual differentiation throughout the genome or more rapidly at a few key regions. Genomic analyses of sister species in the early stages of speciation uncovers the evolutionary process of speciation, thus inform this debate.

Islands of divergence may be associated with divergent (and often diagnostic) traits that are involved in reproductive isolation [19–21]. Divergent phenotypic traits frequently play important roles in social interactions at geographic boundaries between closely-related species [22,23]. These features are commonly involved in mate-choice [19,20] or aggressive signaling [21,24] that are divergently selected [25–30] and facilitate species recognition and reproductive isolation [31,32]. We are just beginning to understand the genetic underpinnings of traits related to species divergence [26,27,29,33,34]. Integrating the genetic basis underlying pigmentation

with corresponding evolutionary forces will deepen our understanding of the formation and maintenance of species barriers.

Carotenoid and melanin pigmentation are commonly involved in species-diagnostic traits in avian signals [19,25–28,32,35]. At the boundaries between closely-related species, the intensity of coloration might not differ, but rather the differential patterning of the colors is involved in mate choice favoring conspecifics [32,35,36]. In such cases, the regulatory genes for melanin and carotenoid pathways are strong candidates for differentiation between species as these loci are likely to cause differences in plumage pattern.

Hybrid zones, locations where divergent lineages interbreed, provide opportunities to understand the relationship between species-diagnostic features and genomic differentiation. Social signals can be crucial for conspecific mate recognition and intrasexual competition for mating opportunities [19,20,37–39]. When the plumage signals established in isolated populations come into secondary contact, hybrids can be selected against due to their intermediate or mismatched parental signals interfering with mate recognition and competition [37,38].

The hybrid zone between *Setophaga townsendi* (abbreviated as *townsendi*) and *S. occidentalis* (abbreviated as *occidentalis*) along the Cascade mountains harbors extensive gene flow between these closely related species that started diverging an estimated 400,000 years ago [40–44]. Males of this species pair differ in several plumage features related to both carotenoid and melanin patterning on the crown, cheek, breast, and back. For example, *townsendi* has a melanin-based cheek patch (Figure 1A right), whereas *occidentalis* displays a completely carotenoid-based cheek (Figure 1A, left). Because apparent hybrids predominantly resemble *occidentalis* in crown and cheek coloration (Figure 1A center), with intermediate breast (the

extent of yellow plumage) and back colors (green plumage in the mantle), Rohwer & Wood [43] predicted that face coloration of Hermit Warblers and hybrids would be controlled by a single-locus dominant allele. In addition, whether the other carotenoid and melanin patterning differences between species are underpinned by the same genetic mechanism is an open question. These signals can be important in male-male competition for territories, as individuals in the hybrid zone demonstrate an aggression bias towards different plumage types [42,45].

We investigated whether speciation in this young species pair occurred through gradual widespread differentiation across the genome or via a small number of restricted regions. If selection for coloration phenotypes has targeted a small number of specific genomic regions, genomic islands of differentiation should be narrow, dispersed, and associated with loci involved in pigmentation pathways [3,7,8]. In addition, we asked the following questions regarding the genetic basis of species-specific traits: (1) is there an island of divergence pleiotropically underpins species-specific features (crown and cheek darkening, breast yellow, bib size, greenish back); (2) Is cheek darkening influenced largely by an allele of dominant effect, consistent with the prediction made two decades ago by Rohwer & Wood (1998); and (3) Is there stable selection maintaining differentiation in the gene region underlying key species diagnostic features?

Materials and Methods

Sampling

Whole specimens of *Setophaga* warblers were collected over two historical sampling sessions in the Washington Cascade hybrid zone [41,43]. The first sampling (N = 314 individuals; 35 sites) was carried out in 1987-1994 [43], while the second (N = 127; 11 sites) was in 2005-2008 [41] and covered a subset of the sites from the original sampling. We accessed these

specimens at the Burke Museum of Natural History and Culture (University of Washington, Seattle, Washington).

We carried out a third round of sampling using a catch-and-release approach during the breeding season (early May to mid-July) in 2015-16 (Figure 1B). Upon locating a territorial male by song, a mist net with a playback (of a locally recorded song) at the bottom was set up nearby. After capturing an adult, photographs and a blood sample were taken for further analysis. We re-sampled the sites that were sampled in 1986-94 [43]. For details, see previous work in this system [46].

Plumage measurements

Melanin- and carotenoid-based plumage traits allow identification of the two species (Figure 1B), but there is also variation within each species [43,47]. To quantify plumage variation within and between populations, we focused on five distinct plumage traits in males: the relative amount of black (melanin) and yellow (carotenoid) in the 1) cheek, 2) crown, and 3) breast and the 4) throat bib, as well as, 5) the intensity of green chroma on the back (Figure 1B).

For each of the 242 warblers captured in the field, we took three pictures from different angles: 1) frontal with head tilted up (for bib and breast measurements), 2) profile (cheek), and 3) from above (crown) (Figure S1). Plumage color metrics were measured using Adobe Photoshop CC in CIE (*Comission Internationale de l'Eclairage*) LAB color space. LAB color space is a 3-dimensional space consisting of 3 distinct, perpendicular axes: 1) Luminosity (L) ranging from 0 (black) to 100 (white), 2) 'a' ranging from green (negative) to red (positive), and 3) 'b' ranging from blue (negative) to yellow (positive; Adobe 2017). We chose this color space because it linearizes the variables of interest along three distinct axes: black ('L'), yellow (along the 'b' axis), and green ('a').

140 For the cheek and mantle, we selected a standardized area and averaged the pixels to
141 record the 'b' and 'a' values, respectively. For the cheek, we selected and averaged the entire area
142 from above the eye (but excluding the eye) to the throat badge, and from the base of the bill to the
143 mantle using the profile photos. Differences in ambient light conditions at the time a picture is
144 taken can confound comparison of color metrics among individuals. To address this, we used the
145 white-balance feature in Photoshop, using the white plumage of each individual's belly as a
146 standard, to correct for differences in ambient light among photos and standardize the color
147 metrics. We acknowledge that without spectral analysis, we do not incorporate UV reflectance
148 which is a ubiquitous aspect of signaling in avian systems [48]. However, our methods allow us
149 to estimate the relative intensity of melanin- and carotenoid-based plumage traits during the
150 breeding season.

151 To measure the size of the black bib, we used the program Analyzing Digital Images
152 (ADI;[49]). A scale was included in all photos to standardize size measurements among
153 individuals. We measured bib size (Figure 1A) by creating a polygon around the bib and
154 calculating the area (± 0.1 mm^2).

155 ***GBS pipeline***

156 The GBS pipeline details have been described in [46]. Following [50], we prepared
157 genotyping-by-sequencing (GBS) [51] libraries from individual DNA samples. In brief, genomes
158 were digested with restriction enzyme and ligated with barcode and adaptors, amplified with
159 PCR, and size selected (fragment length of 300 - 400 bp) for sequencing. Libraries were
160 sequenced at Genome Quebec with paired-end sequencing (read length = 125 bp) on an Illumina
161 HiSeq 2500 automated sequencer. The resulting sequences were processed following a GBS
162 pipeline [52]. All the sequence can be acquired through GenBank (accession number:

163 PRJNA573930, ID: 573930). We demultiplexed the reads with a custom script and trimmed them
164 using Trimmomatic 0.36 (Bolger et al 2014) [TRAILING:3 SLIDINGWINDOW:4:10
165 MINLEN:30], then we aligned reads to a *Taeniopygia guttata* reference version 3.2.4 [53] using
166 bwa (Li et al 2009) (default settings). We assumed synteny of *Setophaga* and *T. guttata* genomes
167 given the limited rearrangement in avian genomes [54,55], and the conclusions of this study
168 would be unlikely to be affected by a moderate number of rearrangements. We conducted SNP
169 calling with GATK (McKenna et al. 2010), which produced 4,097,089 SNPs. The SNP filtering
170 was done with VCFtools [56], which includes removing indels, genotype quality (GQ) > 20,
171 minor allele frequency (MAF) ≥ 0.05, removing loci with >30% missing data, and only including
172 biallelic SNPs, resulting in 21,852 SNPs remaining.

173 ***Whole Genome Sequencing (WGS)***

174 In addition, we selected 5 samples from *occidentalis* in Pinehurst, CA, U.S.A. (UWBM
175 66152, 66153, 66148-66150) and 5 samples from inland *townsendi* in Tok, AK, U.S.A. (UWBM
176 84816-84819, 84860) from the Burke museum for whole genome re-sequencing. For DNA
177 extraction, we used 2 mm^3 of tissue digested in Qiagen buffers—following the tissue extraction
178 procedure—and separated DNA using UPrep spin columns (Genesee). We standardized DNA
179 concentrations with after quantifying concentrations with a Qubit fluorometer, and then generated
180 sequencing libraries with the llumina TruSeq Nano kit—which includes an 8-cycle PCR
181 enrichment—selecting 350 bp insert sizes. We individually indexed each sample, and sequenced
182 the combined libraries across a single lane of an Illumina NextSeq using the paired-end 150 bp
183 sequencing chemistry. We combined these 10 samples with other wood warblers from other
184 projects, but consistently included 24 individuals per sequencing lane in order to generate
185 comparable coverage across samples.

186 ***Genomic differentiation***

187 To quantify the level of differentiation throughout the genome, we calculated F_{ST} [57]
188 with VCFtools [56] between allopatric *townsendi* (Montana and Idaho, USA; $N = 38$) and
189 *occidentalis* (Oregon and California, USA; $N = 23$) for each of the filtered SNPs.

190 The genomic architecture of differentiation from the GBS was compared to that of the
191 WGS data. The reads from the WGS data were aligned to the same *Taeniopygia guttata* reference
192 (version 3.2.4) [53] with bwa (Li et al 2009) (default settings). ANGSD [58] was employed to
193 calculate genotype frequencies accounting for genotyping uncertainty before estimating F_{ST} for
194 each of the non-overlapping (step size = 10kb) 10kb windows.

195 ***Admixture mapping***

196 To identify SNPs that are associated with variation in the five plumage traits, we
197 conducted admixture mapping with GenABEL package in R [59] for the subset (N = 242) of the
198 GBS data produced in [46] that had pigment quantification (see above): 189 birds sampled from
199 hybrid zone (around the center of the hybrid zone in Washington, U.S.A.), 46 from the
200 *occidentalis* zone (Southern Oregon and California, U.S.A.), and 7 from the *townsendi* zone
201 (Northeast Washington & Montana, U.S.A.). The GenABEL function *ibs*, which calculates
202 identity-by-state relatedness among individuals, was used to control for population structure.
203 Genomic control of inflation factor λ (which represents the effect of genetic structure and sample
204 size) was conducted [59]. Briefly, λ was estimated from the genomic data assuming that
205 randomly-selected markers from the genome are not associated with the trait (after controlling for
206 population substructure). λ was then used to correct the test statistic χ^2 of each association test, so
207 that the test is relative to the null hypothesis of no phenotype-genotype association [59]. Because

208 the cheek darkening and breast yellow intensity were left-skewed, we rank order transformed
209 these phenotype data.

210 To find the number of independent hypotheses for multiple hypothesis correction, we
211 estimated the total number of independent linkage blocks (a proxy for the number of independent
212 hypotheses) in the dataset. Principle Component Analysis calculating genotypic covariance
213 [60,61] was run until 99.5% of the variance was explained with the R package SNPRelate [62],
214 producing 210 PCs. We applied a Bonferroni correction to the independent blocks, such that
215 significance level alpha = 0.05/210. Because we conducted admixture mapping on five plumage
216 traits, we further applied Bonferroni correction across the 5 plumage traits, which gave rise to the
217 final alpha value of $0.05/(210\times5) \approx 4.76\times10^{-5}$. Candidate SNPs (i.e., those with *p*-values below
218 this alpha value) were annotated in Ensembl Zebra Finch (taeGut3.2.4;[63]).

219 ***Geographical cline analysis on candidate loci***

220 As one candidate gene, RALY (see Results), stood out in the above analysis as
221 particularly strongly associated with plumage variation, we investigated the spatial and temporal
222 variation in this locus relative to the plumage hybrid index and rest of the genome (using samples
223 described previously [46]). We fit the relationship between RALY allele frequency and location
224 using an equilibrium geographic cline model [64,65].

225 Geographical cline analysis followed [46]. Briefly, we collapsed the two-dimensional
226 sampling into a one-dimensional transect by measuring the location of each site to the 0.5 isocline
227 of the plumage hybrid index (HI) in the historical sampling [43], as follows. First, a Local
228 Polynomial Regression Fitting (LOESS) model in R [66] was used to fit variation in HI across the
229 hybrid zone, and this model was used to estimate the HI = 0.5 isocline in the 1987-94 sampling
230 [43]. Then for each site, the shortest distance to the 0.5 isocline was calculated with 'sp' package

231 [67]. The sites east or west of the isocline were specified as having positive versus negative
232 distance values, respectively. We added 1200 km to the distance score of each site so that all
233 distance values are above zero, while the relative distance of each site to the isocline is preserved.
234 Then the data was fit to the equilibrium sigmoidal cline model [64,65] $y = \frac{1}{1+e^{-\frac{4(x-c)}{w}}}$, in which
235 cline center (c) and width (w) was estimated (where y is HI and x is location with respect to the
236 HI = 0.5 isocline). To examine whether selection (i.e., divergent selection and/or selection against
237 hybrids) is acting on the candidate loci, we followed [46] and tested whether the increase in cline
238 width ($w^2_{2015\text{-}16}$ -$w^2_{1987\text{-}94}$) is significantly less than expected under the neutral diffusion model
239 [68]. The change of w^2 (between sampling periods) as opposed to w allows comparison of the
240 observed change in cline width to the neutral expectation [46]. To understand the RALY cline
241 width change relative to the plumage and genomic cline, we compared $w^2_{2015\text{-}16}$ -$w^2_{1987\text{-}94}$ relative
242 to the plumage and genomic HI cline [46], which was respectively based on the scores of the 8
243 plumage landmarks (with 0 representing pure *occidentalis* and 1 being pure *townsendi*), and the
244 scaled genomic PC1 (with 0 representing pure *occidentalis* and 1 being pure *townsendi*).

245 **Results**

246 Across the Cascade mountain range, the *occidentalis* plumage type was observed on the
247 southwest and the *townsendi* plumage type was on the northeast [43,46] (Figure 1 B). The overall
248 pattern of genomic differentiation was consistent with the plumage divergence: *occidentalis* and
249 *townsendi* plumage types were most different along genomic PC1. There was low genome-wide
250 weighted average F_{ST} [57] between allopatric *occidentalis* and *townsendi* populations (Weir and
251 Cockerham's F_{ST} = 0.03; Figure 2A). However, four high regions of differentiation (F_{ST} > 0.6)
252 were found (Figure 2A, Figure S2, gene association and functions summarized in Table S2) that

253 map to Zebra Finch chromosome (chr) 5 (nucleotide position 25064223-25875302 in the Zebra
254 Finch genome, mean F_{ST} = 0.75), chr 20 (1981369, mean F_{ST} = 0.90), and chr Z (66226657, F_{ST}
255 = 0.82). Such genomic architecture of divergence revealed by GBS data is consistent with the
256 pattern from the WGS data (Figure S3), where the genome is largely undifferentiated except a
257 few peaks at chr 1A, 4, 5, 20, and Z (Figure S3G).

258 The F_{ST} peak on chr 20 (Figure 2B) from the GBS data is within an intron of the gene
259 RALY, which encodes heterogeneous nuclear ribonucleoprotein, the cofactor for cholesterol
260 biosynthetic genes [69]. We hereafter refer to this SNP as the RALY SNP. This RALY SNP
261 demonstrated the highest F_{ST} (0.90) inside an "island" of relatively high F_{ST} on chr 20 between
262 *occidentalis* and *townsendi* parental populations across the genome (Figure 2B). The pure
263 *occidentalis* population mostly contains GG homozygotes (referred to hereafter as OO), while the
264 *townsendi* population contains CC homozygotes (referred to as TT). In the WGS data (Figure
265 S4), which has much higher density of markers across the genome than the GBS dataset, this
266 'island' of differentiation contains pigmentation genes ASIP, EIF2S2, and RALY (Figure S4BC),
267 thus we refer to it hereafter as the "ASIP-RALY" block. The RALY SNP revealed by the GBS
268 data (Figure 2B) conveniently becomes the marker representing this ASIP-RALY divergent gene
269 block surveyed across a large number of individuals (Figure S4BC).

270 Admixture mapping showed a strong pattern of few loci associated with divergent
271 plumage traits in the hybrid zone. In particular, the colors of the crown, cheek, and the breast
272 were each very strongly associated with the same RALY SNP (on chr 20) that was the outlier in
273 the F_{ST} analysis (crown: χ^2 = 57.89, r^2 = 0.57, p = 5.95 x 10^{-15}, Figure 3A, Figure 4AD; cheek: χ^2
274 = 44.76, r^2 = 0.55, p = 2.23 x 10^{-11}, Figure 3C, Figure 4BE; breast: χ^2 = 28.47, r^2= 0.57, p = 9.53 x
275 10^{-8}, Figure 3D, Figure 4CF). The RALY SNP shows a partial dominance pattern of the G allele

276 for the three plumage traits, with G/C heterozygotes tending to have similar phenotypes as G/G
277 homozygotes, although there appears to be some additivity as well (Figure 4). The RALY gene
278 (Figure 3B) is associated with yellow pigmentation in mice and quail [70,71], and is adjacent to
279 two other pigmentation genes -- ASIP (115941 bases away) and EIF2S2 (30223 bases away)
280 (Figure 3B, S4C).

281 A low-resolution sequencing approach such as GBS, which sequences less than 1% of the
282 genome, would be unlikely to detect a SNP that is directly causal for phenotypic variation. It is
283 more likely that the RALY SNP is closely physically linked to the causal DNA variant (SNP or
284 some other type of variant) for the phenotypic differences. It is also possible that multiple linked
285 SNPs in this region are responsible for the variation in different phenotypic traits, such that the
286 apparent pleiotropic effect (i.e., strong association with three plumage traits) of the RALY SNP
287 might be due to separate causal genes that are closely linked. The continuous divergence of the
288 ASIP-RALY gene block between allopatric species in the WGS dataset suggests that the ASIP-
289 RALY gene linkage block might be involved in plumage color divergence as a whole (Figure
290 S4).

291 Three additional SNPs were found to be significantly associated with the intensity of
292 yellow on the breast (Figure 3D). Two are on chr 4A at nucleotide location 5588139 and
293 5588235, both inside an ortholog of the mammal Immunoglobulin Binding Protein 1 (IGBP1)
294 gene. One is at nucleotide location 6623798 on chr 13, inside gene Annexin A6, which codes for
295 a phospholipid binding protein, Annexin VI [72]. We did not detect SNPs that significantly
296 explained variation in either bib size or green coloration on the back (Figure 3EF). The RALY
297 SNP explains 57% of variation in crown (Figure 4AD), 52% of the variation in cheek (Figure 4B,
298 E), and 43% variation in breast coloration (Figure 4CF).

299 We found evidence that the ASIP-RALY region is under divergent selection, as it is the
300 most extreme F_{ST} outlier, while most of the genome is not very differentiated (average = 0.03)
301 (Figure 5D). In addition, the RALY SNP (representing the ASIP-RALY gene block)
302 demonstrated a spatial cline that was stable in location over two decades (Figure 5A). The cline
303 center was at 1216.35 ± 4.0 (SE) km in 1987-94 (Figure 5A, blue curve), and did not significantly
304 shift in 2015-16 sampling (Figure 5A, yellow curve), occurring at 1213.140 ± 2.2 (SE) km.
305 Furthermore, the width of the RALY cline stayed narrower (95% CI of $w^2_{2015\text{-}16} - w^2_{1987\text{-}94}$: -
306 39004.70 to 1035.13 km^2) than under a model of neutral diffusion ($w^2_{2015\text{-}16} - w^2_{1987\text{-}94}$ = 7540
307 km^2) (Figure 5B, see [46] for explanation of method).
308

309 **Discussion**

310 We observed clear but narrow genomic 'islands of differentiation' between Townsend's
311 and Hermit Warblers on chr 1A, 5, 20, and Z (Figure 2). This agrees with the predicted pattern of
312 speciation in the face of gene flow: as most of the genome remains homogenized due to gene
313 flow and lack strong selection, restricted targets of divergent selection have become strongly
314 differentiated [3,7,11,73].

315 The RALY SNP (chr 20) within the ASIP-RALY gene block represents one of the
316 'islands of differentiation' and is strongly associated with key divergent traits: cheek, crown, and
317 breast coloration. The consistently narrow geographical RALY cline in the hybrid zone over
318 several decades, despite ongoing gene flow [46], further implicates selection (divergent selection
319 and/or selection against hybrids) around the RALY SNP as contributor to reproductive isolation
320 between these young sister species. The $w^2_{2015\text{-}16}$ -$w^2_{1987\text{-}94}$ estimate at the RALY cline was less
321 than the plumage cline and the genomic PC1 cline (Figure 5B), suggesting that a narrow region
322 centered at the RALY locus might be the direct target of selection maintaining the stable plumage
323 and genomic clines [46]. Such selection may further extend genomic differentiation by elevating
324 the differentiation at linked sites.

325 Although there is signature of genetic hitchhiking that underpins extended genetic
326 divergence around the ASIP-RALY targets of selection (Figure S5), it seems that the plumage
327 pigmentation effect is more narrowly underpinned by ASIP-RALY region. In the hybrid zone, the
328 recombination over multiple generations can break down the hitchhiking gene blocks (that tend to
329 co-segregate with the trait-determining loci in allopatric populations), revealing specific genetic
330 underpinnings of the differentiated phenotypes. The genetic underpinning of the plumage trait
331 should be very close to the narrow region around the RALY SNP.

332 The ‘divergence with gene flow’ pattern, in which closely related taxa hybridize
333 extensively and have little differentiation across much of the genome, along with evidence of a
334 narrow and highly differentiated genetic region that has large effect on divergent traits suggests a
335 simple genomic architecture of divergence. Such a strong target of selection could effectively
336 counteract gene flow and maintain a species boundary at secondary contact, allowing speciation
337 to progress [4,74,75]. There may also be weak multilocus divergent selection on a large set of
338 loci, yet such weak polygenic selection is inherently challenging to detect.

339 ***Pleiotropy***

340 This ASIP-RALY block underlying multiple plumage patches could provide a strong
341 pleiotropic mechanism (one region that affects multiple phenotypes [76]), linking divergent
342 selection and possibly reproductive isolation. The RALY SNP is associated with a suite of
343 melanin- and carotenoid-related plumage patches (crown, cheek, and breast coloration) that differ
344 diagnostically between species (Figure 3) and are likely under some form of selection. Previous
345 admixture mapping in *Setophaga coronata auduboni*/*S. c. coronata* revealed SCARF2 on chr 15
346 that is associated with multiple carotenoid and melanin traits [77]. Intriguingly, in this sister pair
347 within the same genus, drastically different color regions are involved in carotenoid versus
348 melanin plumage divergence. Such distinct pleiotropic genetic underpinnings of plumage
349 divergence may facilitate rapid speciation in *Setophaga* [78].

350 Melanin and carotenoid plumage traits are often involved in mate-choice and male-male
351 competition [79–82]. If carotenoid-related plumage traits are targets of female mate-choice or
352 male-male competition for mating opportunities in *Setophaga* warblers, strong selection on
353 pleiotropic genes can play a powerful role in limiting mating between species and facilitating
354 speciation. Pleiotropy is a more effective mechanism linking selection to reproductive isolation

355 than linkage or polygenic inheritance, although genetic evidence for pleiotropic genes that
356 influence both selected traits and reproductive isolation has been scarce (see review [83]). Among
357 the limited existing examples, the classic cases involve coloration: the wingless gene affecting
358 reproductive isolation and wing coloration in *Heliconius* butterflies [19], and the YUP locus
359 affects pollinator isolation and flower coloration in monkey flowers [10]. RALY could be an
360 additional example of a speciation gene. The stable and narrow cline of RALY SNP and extreme
361 differentiation at the RALY locus imply divergent selection and/or selection against hybrids at
362 this locus, which suggests that RALY could be associated with reproductive isolation in this
363 system. Future study is needed to validate the possibility by examining the role of RALY-ASIP-
364 regulated plumage signals in mate preference or male-male competition.

365 ***Dominance***

366 Our results support the prediction that Rohwer and Wood [43] made two decades ago that
367 the cheek coloration in hybrids is underpinned by a single dominant locus. Indeed, heterozygotes
368 of the RALY SNP exhibit similar yellow cheeks as the homozygous GG individuals (i.e., yellow
369 cheek), which were found predominantly in the *occidentalis* population (Figure 4D-F).
370 Dominance in the signal trait might reduce gene flow if it is in the opposite direction of
371 dominance in the receiver trait (i.e., heterozygous receivers are more sensitive to *townsendi*
372 signals), due to 'opposing genetic dominance' in receiver and signaler traits [83,84]. This is
373 because the F1 hybrids (heterozygous for both signaling and receiving) might for example,
374 demonstrate *occidentalis* signal and preference for *townsendi*, thus cannot mate with other F1
375 hybrids. Therefore the *occidentalis* RALY dominance effect on the cheek signal could suppress
376 gene flow if it is opposed by a *townsendi* dominant effect on receiving trait.

377 ***RALY SNP***

The RALY SNP is at the peak of divergence at allopatry (Figure 2B) and is tightly associated with plumage variation in the hybrid zone, although its neighboring SNPs are not significantly associated with plumage traits (after controlling for population substructure) (Figure 3, 6C). The roughly 180kb extended divergence around RALY in WGS data (Figure S2, S4) suggests the entire color gene block (ASIP-RALY) is involved in divergence.

The RALY gene encodes for heterogeneous nuclear ribonucleoprotein, the RNA binding protein that is known to regulate the expression of its downstream gene that codes for agouti signaling protein (ASIP) [70]. Deletion of RALY in Japanese quails and mice leads to novel transcripts of ASIP, which cause a yellow skin phenotype called "lethal yellow" [70,71]. In mice, a lethal yellow mutation (Ay) in ASIP exhibits dominant pleiotropic effects on other traits in addition to skin coloration, which includes obesity, diabetic condition, etc., that are unrelated to the ASIP gene [71]. ASIP itself is well known for its influence on skin pigment by binding to melanocortin receptors and competitively excluding its agonists, preventing black/brown pigmentation [85,86]. In birds, ASIP has been repetitively involved in genomic divergence between closely-related species in *Vermivora* [26], *Sporophila* [87], and *Munia* [88], in a mendelian inheritance pattern with the melanic form being the recessive phenotype. Genetic variants of RALY in warblers could result in differential expression ASIP leading to variations in carotenoid and melanin patterning. The pleiotropic dominance effect of the RALY SNP we identified in this natural hybrid zone is consistent with these known effects of RALY mutants in the lab. Future study should investigate the functional interaction of RALY and ASIP in plumage divergence, and whether there is any convergent mutation in ASIP between *Vermivora* and *Setophaga*.

On top of the strong pleiotropic ASIP-RALY association with plumage, there could be other regions of the genomes contributing to further refinement of species-specific coloration in this system. Specifically, other significant SNPs on chr 4A and 13 are associated with breast coloration. In addition, there might be narrow but important genetic regions that this GBS data did not cover. However, this is unlikely because the WGS data did not reveal additional differentiation peaks.

Conclusion

We identified a few narrow and dispersed genomic regions of differentiation between two hybridizing *Setophaga* warblers, consistent with divergent selection in the face of gene flow. A highly divergent peak spans among three closely linked pigmentation genes: ASIP, EIF2S2, and RALY. Admixture mapping within a hybrid zone between these species showed that this SNP was strongly associated with three distinct plumage traits that distinguish the species, revealing a major-effect region related to speciation. In contrast, we did not find genetic regions significantly associated with the other two species-diagnostic traits, implying many genes of small effects.

The extensive similarity in the rest of the genome raises a question regarding the distinctiveness of *townsendi* and *occidentalis* and their future. Key to this question is whether these regions of differentiation represent sufficient reproductive isolation that will maintain separation for further differentiation to build up. The genetic divergence of the ASIP-RALY gene region between allopatric populations and the stable narrow geographical cline of this region suggest that selection maintains such difference between species. Future studies should track the genomic differentiation in linkage disequilibrium with the ASIP-RALY region and investigate the mechanism of apparent selection. Overall, these results show that avian sister species can be

422 very similar across the great majority of their genomes, sometimes differing primarily at narrow
423 genomic regions under selection maintaining phenotypic differences between species.

424 **Reference**

425 1. Feder JL, Gejji R, Yeaman S, Nosil P. Establishment of new mutations under divergence
426 and genome hitchhiking. Philos Trans R Soc B Biol Sci. 2012;367: 461–474.
427 doi:10.1098/rstb.2011.0256
428 2. Via S. Divergence hitchhiking and the spread of genomic isolation during ecological
429 speciation-with-gene-flow. Philosophical Transactions of the Royal Society B: Biological
430 Sciences. 2012. pp. 451–460. doi:10.1098/rstb.2011.0260
431 3. Rice WR, Hostert EE. Laboratory Experiments on Speciation: What Have We Learned in
432 40 Years? Evolution (N Y). 1993;47: 1637. doi:10.2307/2410209
433 4. Nosil P, Harmon LJ, Seehausen O. Ecological explanations for (incomplete) speciation.
434 Trends in Ecology and Evolution. 2009. pp. 145–156. doi:10.1016/j.tree.2008.10.011
435 5. Irwin DE, Milá B, Toews DPL, Brelsford A, Kenyon HL, Porter AN, et al. A comparison
436 of genomic islands of differentiation across three young avian species pairs. Mol Ecol.
437 2018;27: 4839–4855. doi:10.1111/mec.14858
438 6. Wolf JBW, Ellegren H. Making sense of genomic islands of differentiation in light of
439 speciation. Nature Reviews Genetics. 2017. p. :87-100. doi:10.1038/nrg.2016.133
440 7. Nosil P, Funk DJ, Ortiz-Barrientos D. Divergent selection and heterogeneous genomic
441 divergence. Molecular Ecology. 2009. pp. 375–402. doi:10.1111/j.1365-
442 294X.2008.03946.x
443 8. Turner TL, Hahn MW, Nuzhdin S V. Genomic islands of speciation in Anopheles
444 gambiae. PLoS Biol. 2005;3: 1572–1578. doi:10.1371/journal.pbio.0030285

445 9. Wu CI. The genic view of the process of speciation. Journal of Evolutionary Biology.
446 2001. pp. 851–865. doi:10.1046/j.1420-9101.2001.00335.x
447 10. Bradshaw HD, Schemske DW. Allele substitution at a flower color locus produces a
448 pollinator shift in monkeyflowers. Nature. 2003;426: 176–178. doi:10.1038/nature02106
449 11. Via S. Natural selection in action during speciation. Proc Natl Acad Sci. 2009;106: 9939–
450 9946. doi:10.1073/pnas.0901397106
451 12. Schluter D, Conte GL. Genetics and ecological speciation. Proc Natl Acad Sci. 2009;106:
452 9955–9962. doi:10.17226/12692
453 13. Orr HA. The population genetics of speciation: The evolution of hybrid incompatibilities.
454 Genetics. 1995.
455 14. Orr HA, Turelli M. The evolution of postzygotic isolation: Accumulating Dobzhansky-
456 Muller incompatibilities. Evolution (N Y). 2001. doi:10.1111/j.0014-3820.2001.tb00628.x
457 15. Matsubayashi KW, Katakura H. Contribution of multiple isolating barriers to reproductive
458 isolation between a pair of phytophagous ladybird beetles. Evolution (N Y). 2009;63:
459 2563–2580. doi:10.1111/j.1558-5646.2009.00738.x
460 16. Lawniczak MKN, Emrich SJ, Holloway AK, Regier AP, Olson M, White B, et al.
461 Widespread divergence between incipient Anopheles gambiae species revealed by whole
462 genome sequences. Science (80-). 2010. doi:10.1126/science.1195755
463 17. Johnson NA, Porter AH. Rapid speciation via parallel, directional selection on regulatory
464 genetic pathways. J Theor Biol. 2000;205: 527–542. doi:10.1006/jtbi.2000.2070
465 18. Williams LM, Oleksiak MF. Ecologically and evolutionarily important SNPs identified in
466 natural populations. Mol Biol Evol. 2011;28: 1817–1826. doi:10.1093/molbev/msr004
467 19. Kronforst MR, Young LG, Kapan DD, McNeely C, O'Neill RJ, Gilbert LE. Linkage of

468 butterfly mate preference and wing color preference cue at the genomic location of
469 wingless. Proc Natl Acad Sci. 2006;103: 6575–6580. doi:10.1073/pnas.0509685103
470 20. Whibley AC, Langlade NB, Andalo C, Hanna AI, Bangham A, Thébaud C, et al.
471 Evolutionary paths underlying flower color variation in Antirrhinum. Science (80-).
472 2006;313: 963–966. doi:10.1126/science.1129161
473 21. Seehausen O, Schluter D. Male-male competition and nuptial-color displacement as a
474 diversifying force in Lake Victoria cichlid fishes. Proc R Soc B Biol Sci. 2004;271: 1345–
475 1353. doi:10.1098/rspb.2004.2737
476 22. Anderson CN, Grether GF. Interspecific aggression and character displacement of
477 competitor recognition in *Hetaerina* damselflies. Proc R Soc London, Ser B. 2010;277:
478 549–55. doi:10.1098/rspb.2009.1371
479 23. West-Eberhard M-J. Sexual selection, competitive communication, and species-specific
480 signals in insects. In: Lewis T, editor. Insect Communication. Academic Press; 1984. pp.
481 283–324.
482 24. Mikami OK, Kohda M, Kawata M. A new hypothesis for species coexistence: Male-male
483 repulsion promotes coexistence of competing species. Population Ecology. 2004. pp. 213–
484 217. doi:10.1007/s10144-004-0189-5
485 25. Uy JAC, Moyle RG, Filardi CE, Cheviron ZA. Difference in Plumage Color Used in
486 Species Recognition between Incipient Species Is Linked to a Single Amino Acid
487 Substitution in the Melanocortin-1 Receptor. Am Nat. 2009;174: 244–254.
488 doi:10.1086/600084
489 26. Toews DPL, Taylor SA, Vallender R, Brelsford A, Butcher BG, Messer PW, et al.
490 Plumage Genes and Little Else Distinguish the Genomes of Hybridizing Warblers. Curr

491 Biol. 2016;26: 2313–2318. doi:10.1016/j.cub.2016.06.034
492 27. Toews DPL, Brelsford A, Grossen C, Milá B, Irwin DE. Genomic variation across the
493 Yellow-rumped Warbler species complex. Auk. 2016;133: 698–717. doi:10.1642/AUK-
494 16-61.1
495 28. Barrera-Guzmán AO, Aleixo A, Shawkey MD, Weir JT. Hybrid speciation leads to novel
496 male secondary sexual ornamentation of an Amazonian bird. Proc Natl Acad Sci.
497 2018;115: E218–E225. doi:10.1073/pnas.1717319115
498 29. Knief U, Bossu CM, Saino N, Hansson B, Poelstra J, Vijay N, et al. Epistatic mutations
499 under divergent selection govern phenotypic variation in the crow hybrid zone. Nat Ecol
500 Evol. 2019;3: 570–576. doi:10.1038/s41559-019-0847-9
501 30. Poelstra J. The Genetics of Speciation and Coloration in Carrion and Hooded Crows.
502 Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science
503 and Technology. 2013.
504 31. Hill GE, Mcgraw KJ. Correlated changes in male plumage coloration and female mate
505 choice in cardueline finches. Anim Behav. 2004;67: 27–35.
506 doi:10.1016/j.anbehav.2003.02.002
507 32. Sætre GP, Moum T, Bureš S, Král M, Adamjan M, Moreno J. A sexually selected
508 character displacement in flycatchers reinforces premating isolation. Nature. 1997.
509 doi:10.1038/42451
510 33. Shriner D. Overview of admixture mapping. Curr Protoc Hum Genet. 2017;2017: 1.23.1-
511 1.23.8. doi:10.1002/cphg.44
512 34. Crawford JE, Nielsen R. Detecting adaptive trait loci in nonmodel systems: Divergence or
513 admixture mapping? Mol Ecol. 2013;22: 6131–6148. doi:10.1111/mec.12562

514 35. Hill GE. Sexiness, Individual Condition, and Species Identity: The Information Signaled
515 by Ornaments and Assessed by Choosing Females. Evol Biol. 2015;42: 251–259.
516 doi:10.1007/s11692-015-9331-x
517 36. Uy JAC, Moyle RG, Filardi CE, Cheviron ZA. Difference in Plumage Color Used in
518 Species Recognition between Incipient Species Is Linked to a Single Amino Acid
519 Substitution in the Melanocortin-1 Receptor. Am Nat. 2009. doi:10.1086/600084
520 37. Dijkstra PD, Border SE. How does male-male competition generate negative frequency-
521 dependent selection and disruptive selection during speciation? Curr Zool. 2018/01/02.
522 2018;64: 89–99. doi:10.1093/cz/zox079
523 38. Pfennig KS, Pfennig DW. Character displacement: ecological and reproductive responses
524 to a common evolutionary problem. Q Rev Biol. 2009;84: 253–276. Available:
525 https://www.ncbi.nlm.nih.gov/pubmed/19764283
526 39. van Doorn GS, Dieckmann U, Weissing FJ. Sympatric Speciation by Sexual Selection: A
527 Critical Reevaluation. Am Nat. 2004;163: 709–725. doi:10.1086/383619
528 40. Krosby M, Rohwer S. A 2000 km genetic wake yields evidence for northern glacial refugia
529 and hybrid zone movement in a pair of songbirds. Proc R Soc B Biol Sci. 2009;276: 615–
530 621. doi:10.1098/rspb.2008.1310
531 41. Krosby M, Rohwer S. Ongoing movement of the hermit warbler X Townsend's Warbler
532 Hybrid Zone. PLoS One. 2010;5: e14164. doi:10.1371/journal.pone.0014164
533 42. Pearson SF, Rohwer S. Asymmetries in male aggression across an avian hybrid zone.
534 Behav Ecol. 2000;11: 93–101. doi:10.1093/beheco/11.1.93
535 43. Rohwer S, Wood C. Three Hybrid Zones Between Hermit and Townsend ' S Warblers in
536 Washington and Oregon. Auk. 1998;115: 284–310.

537 44. Weir JT, Schluter D. Ice sheets promote speciation in boreal birds. Proc R Soc B Biol Sci.
538 2004;271: 1881–1887. doi:10.1098/rspb.2004.2803
539 45. Pearson SF. Behavioral asymmetries in a moving hybrid zone. Behav Ecol. 2000;11: 84–
540 92. doi:10.1093/beheco/11.1.84
541 46. Wang S, Rohwer S, Delmore KE, Irwin DE. Cross-decades stability of an avian hybrid
542 zone. J Evol Biol. 2019.
543 47. Owen-Ashley NT, Butler LK. Androgens, interspecific competition and species
544 replacement in hybridizing warblers . Proc R Soc B Biol Sci. 2004;271: S498–S500.
545 doi:10.1098/rsbl.2004.0230
546 48. Eaton MD, Lanyon SM. The ubiquity of avian ultraviolet plumage reflectance. Proc R Soc
547 B Biol Sci. 2003;270: 1721–1726. doi:10.1098/rspb.2003.2431
548 49. Snyder R. Analyzing Digital Images (ADI) Resources. STEM Digit. 2011;21.
549 50. Alcaide M, Scordato ESC, Price TD, Irwin DE. Genomic divergence in a ring species
550 complex. Nature. 2014;511: 83–85. doi:10.1038/nature13285
551 51. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust,
552 simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One.
553 2011;6: e19379. doi:10.1371/journal.pone.0019379
554 52. Irwin DE, Alcaide M, Delmore KE, Irwin JH, Owens GL. Recurrent selection explains
555 parallel evolution of genomic regions of high relative but low absolute differentiation in a
556 ring species. Mol Ecol. 2016;25: 4488–4507. doi:10.1111/mec.13792
557 53. Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A, et al. The
558 genome of a songbird. Nature. 2010;464: 757–762. doi:10.1038/nature08819
559 54. Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, et al. Comparative genomics reveals

560 insights into avian genome evolution and adaptation. Science (80-). 2014;346: 1311–1320.
561 doi:10.1126/science.1251385
562 55. Ellegren H. Evolutionary stasis: the stable chromosomes of birds. Trends Ecol Evol.
563 2010;25: 283–291. doi:10.1016/j.tree.2009.12.004
564 56. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant
565 call format and VCFtools. Bioinformatics. 2011;27: 2156–2158.
566 doi:10.1093/bioinformatics/btr330
567 57. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure.
568 Evolution (N Y). 1984;38: 1358. doi:10.2307/2408641
569 58. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation
570 Sequencing Data. BMC Bioinformatics. 2014. doi:10.1186/s12859-014-0356-4
571 59. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: An R library for genome-
572 wide association analysis. Bioinformatics. 2007;23: 1294–1296.
573 doi:10.1093/bioinformatics/btm108
574 60. Gao X, Starmer J, Martin ER. A multiple testing correction method for genetic association
575 studies using correlated single nucleotide polymorphisms. Genet Epidemiol. 2008;32:
576 361–369. doi:10.1002/gepi.20310
577 61. Johnson RC, Nelson GW, Troyer JL, Lautenberger JA, Kessing BD, Winkler CA, et al.
578 Accounting for multiple comparisons in a genome-wide association study (GWAS). BMC
579 Genomics. 2010;11: 724. doi:10.1186/1471-2164-11-724
580 62. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance
581 computing toolset for relatedness and principal component analysis of SNP data.
582 Bioinformatics. 2012;28: 3326–3328. doi:10.1093/bioinformatics/bts606

583 63. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018.
584 Nucleic Acids Res. 2018;46: D754–D761. doi:10.1093/nar/gkx1098
585 64. Szymura JM., Barton NH. Genetic Analysis of a Hybrid Zone Between the Fire-Bellied
586 Toads , Bombina bombina and B . variegata , Near Cracow in Southern Poland. Evolution
587 (N Y). 1986;40: 1141–1159.
588 65. Gay L, Crochet PA, Bell DA, Lenormand T. Comparing clines on molecular and
589 phenotypic traits in hybrid zones: A window on tension zone models. Evolution (N Y).
590 2008;62: 2789–2806. doi:10.1111/j.1558-5646.2008.00491.x
591 66. R Core team. R Core Team. R: A Language and Environment for Statistical Computing. R
592 Foundation for Statistical Computing , Vienna, Austria. ISBN 3-900051-07-0, URL
593 http://www.R-project.org/. 2015. pp. 275–286.
594 67. Pebesma EJ, Bivand RS. Classes and methods for spatial data in R. R News. 2005;5: 9–13.
595 doi:10.1159/000323281
596 68. Barton NH, Hewitt G. M. Analysis of Hybrid Zones. Annu Rev Ecol Syst. 1985;16: 113–
597 148.
598 69. Sallam T, Jones MC, Gilliland T, Zhang L, Wu X, Eskin A, et al. Feedback modulation of
599 cholesterol metabolism by the lipid-responsive non-coding RNA LeXis. Nature. 2016;534:
600 124–128. doi:10.1038/nature17674
601 70. Nadeau NJ, Minvielle F, Ito S, Inoue-Murayama M, Gourichon D, Follett SA, et al.
602 Characterization of Japanese quail yellow as a genomic deletion upstream of the avian
603 homolog of the mammalian ASIP (agouti) gene. Genetics. 2008;178: 777–786.
604 doi:10.1534/genetics.107.077073
605 71. Michaud EJ, Bultman SJ, Klebig ML, van Vugt MJ, Stubbs LJ, Russell LB, et al. A

606 molecular model for the genetic and phenotypic characteristics of the mouse lethal yellow
607 (Ay) mutation. Proc Natl Acad Sci. 1994;91: 2562–2566. doi:10.1073/pnas.91.7.2562
608 72. Chang N, Sutherland C, Hesse E, Winkfein R, Wiehler WB, Pho M, et al. Identification of
609 a novel interaction between the Ca(2+)-binding protein S100A11 and the Ca(2+)- and
610 phospholipid-binding protein annexin A6. Am J Physiol Cell Physiol. 2007;292: C1417-
611 30. doi:10.1152/ajpcell.00439.2006
612 73. Nosil P, Feder JL. Widespread yet heterogeneous genomic divergence. Molecular Ecology.
613 2012. pp. 2829–2832. doi:10.1111/j.1365-294X.2012.05580.x
614 74. Endler JA. Properties of Clines. Geographic Variation, Speciation, and Clines. Princeton,
615 New Jersey: Princeton University Press; 1977. pp. 62–65.
616 75. Gavrilets S. Fitness landscapes and the origin of species. Princeton University Press.
617 Princeton, New Jersey: Princeton University Press; 2004. Available:
618 https://books.google.co.uk/books?hl=en&lr=&id=duk9DwAAQBAJ&oi=fnd&pg=PR13&
619 dq=Fitness+Landscapes+and+the+Origin+of+Species&ots=qBMShqcSzt&sig=EXfli7MIc
620 LpN-IJqfmGZGrOyvTw#v=onepage&q=Fitness Landscapes and the Origin of
621 Species&f=false%0Ahttp://scholar.goog
622 76. Fisher RA. The genetical theory of natural selection. New York, NY: Oxford Univ. Press;
623 1930.
624 77. Brelsford A, Toews DPL, Irwin DE. Admixture mapping in a hybrid zone reveals loci
625 associated with avian feather coloration. Proc R Soc B Biol Sci. 2017;284: 20171106.
626 doi:10.1098/rspb.2017.1106
627 78. Lovette IJ, Bermingham E. Explosive speciation in the New World Dendroica warblers.
628 Proc R Soc B Biol Sci. 1999;266: 1629–1636. doi:10.1098/rspb.1999.0825

629 79. Kingma SA, Szentirmai I, Székely T, Bókony V, Bleeker M, Liker A, et al. Sexual
630 selection and the function of a melanin-based plumage ornament in polygamous penduline
631 tits Remiz pendulinus. Behav Ecol Sociobiol. 2008;62: 1277–1288. doi:10.1007/s00265-
632 008-0556-x
633 80. McGraw KJ. Carotenoid-based ornamentation and status signaling in the house finch.
634 Behav Ecol. 2000;11: 520–527. doi:10.1093/beheco/11.5.520
635 81. Pryke SR, Andersson S. Carotenoid-based epaulettes reveal male competitive ability:
636 Experiments with resident and floater red-shouldered widowbirds. Anim Behav. 2003;66:
637 217–224. doi:10.1006/anbe.2003.2193
638 82. Tarof SA, Dunn PO, Whittingham LA. Dual functions of a melanin-based ornament in the
639 common yellowthroat. Proc R Soc B Biol Sci. 2005;272: 1121–1127.
640 doi:10.1098/rspb.2005.3053
641 83. Nosil P. A Genetic Mechanism to Link Selection to Reproductive Isolation. Ecological
642 Speciation. Oxford: Oxford Univ. Press; 2012. pp. 109–138.
643 84. Forister ML. Independent inheritance of preference and performance in hybrids between
644 host races of Mitoura butterflies (Lepidoptera: Lycaenidae). Evolution (N Y). 2005;59:
645 1149–1155.
646 85. Lovett M, Cheng ZY, Lamela EM, Yokoi T, Epstein CJ. Molecular markers for the agouti
647 coat color locus of the mouse. Genetics. 1987;115: 747–754.
648 86. Wolf Horrell EM, Boulanger MC, D'Orazio JA. Melanocortin 1 receptor: Structure,
649 function, and regulation. Frontiers in Genetics. 2016. p. 95. doi:10.3389/fgene.2016.00095
650 87. Campagna L, Repenning M, Silveira LF, Fontana CS, Tubaro PL, Lovette IJ. Repeated
651 divergent selection on pigmentation genes in a rapid finch radiation. Sci Adv. 2017;3:

e1602404. doi:10.1126/sciadv.1602404

88. Stryjewski KF, Sorenson MD. Mosaic genome evolution in a recent and rapid avian radiation. Nat Ecol Evol. 2017;1: 1912–1922. doi:10.1038/s41559-017-0364-7

Figures

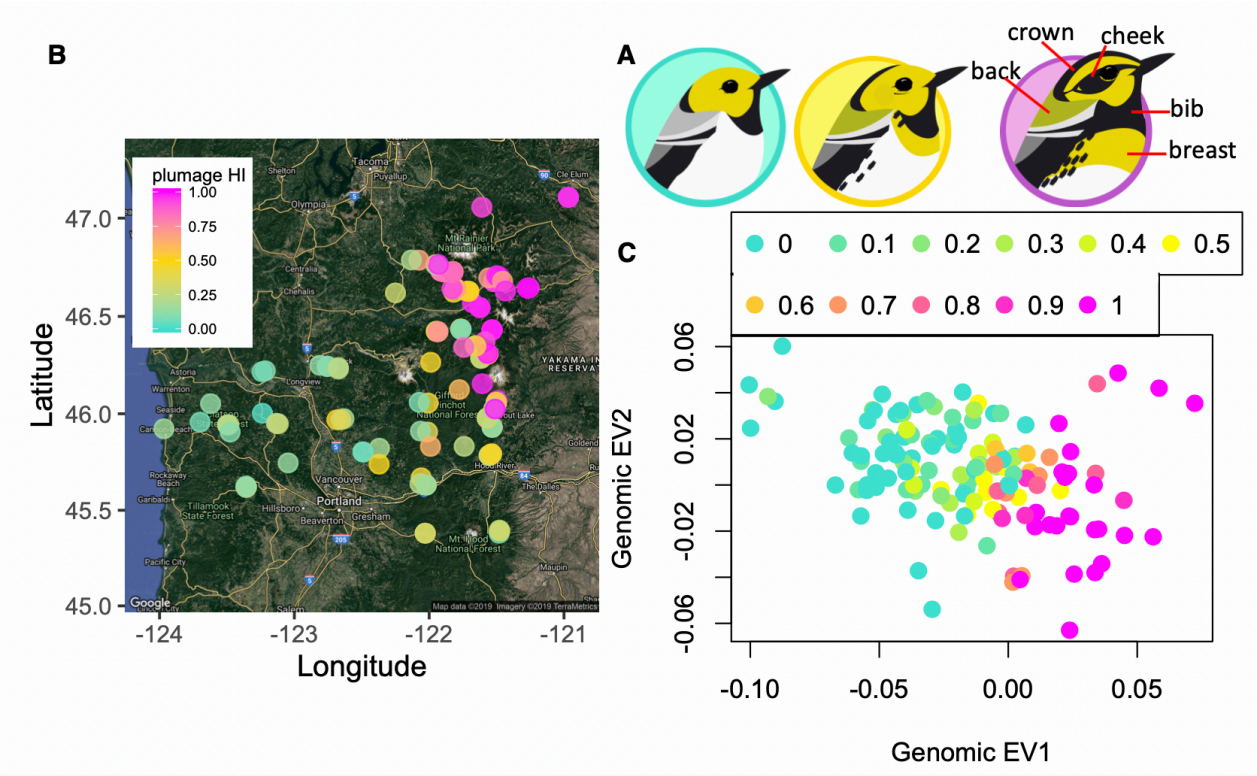

Figure 1 A, illustration highlighting the key plumage difference between *occidentalis* (left, turquoise), hybrids (center, yellow), and *townsendi* (right, magenta) (illustration by Gil Jorge Barros Henriques). **B**, a map showing site mean plumage hybrid index based on 8 plumage traits (0 for pure *occidentalis*, in turquoise; 1 for pure *townsendi*, in magenta). **C**, Genomic eigenvector 1 (EV1) and eigenvector 2 (EV2) with individual datapoints colored by plumage hybrid index. The genomic EV1 reflects the variation among individuals that are *occidentalis*-like (low genomic eigenvector EV1) versus *townsendi*-like (high genomic EV1).

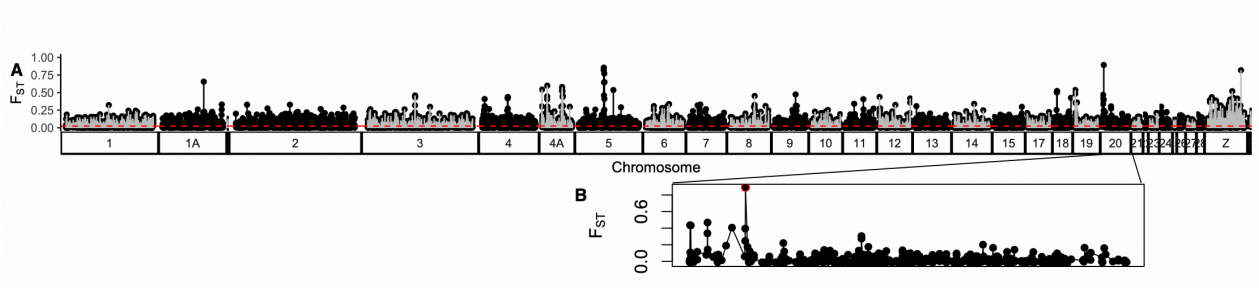

Figure 2 F_{ST} (between inland *townsendi* and *occidentalis*) scan is consistent with the isolation with gene flow model in which a few regions are under selection (divergent selection or selection against hybrids) while the rest of the genome is similar between species (Weir Cockerham Weighted average F_{ST} is indicated by the horizontal red dotted line) due to gene flow. A, F_{ST} genomic scan of the genome revealed an island of differentiation around the highest peak in chromosome 20 (B, zoom in chromosome 20).

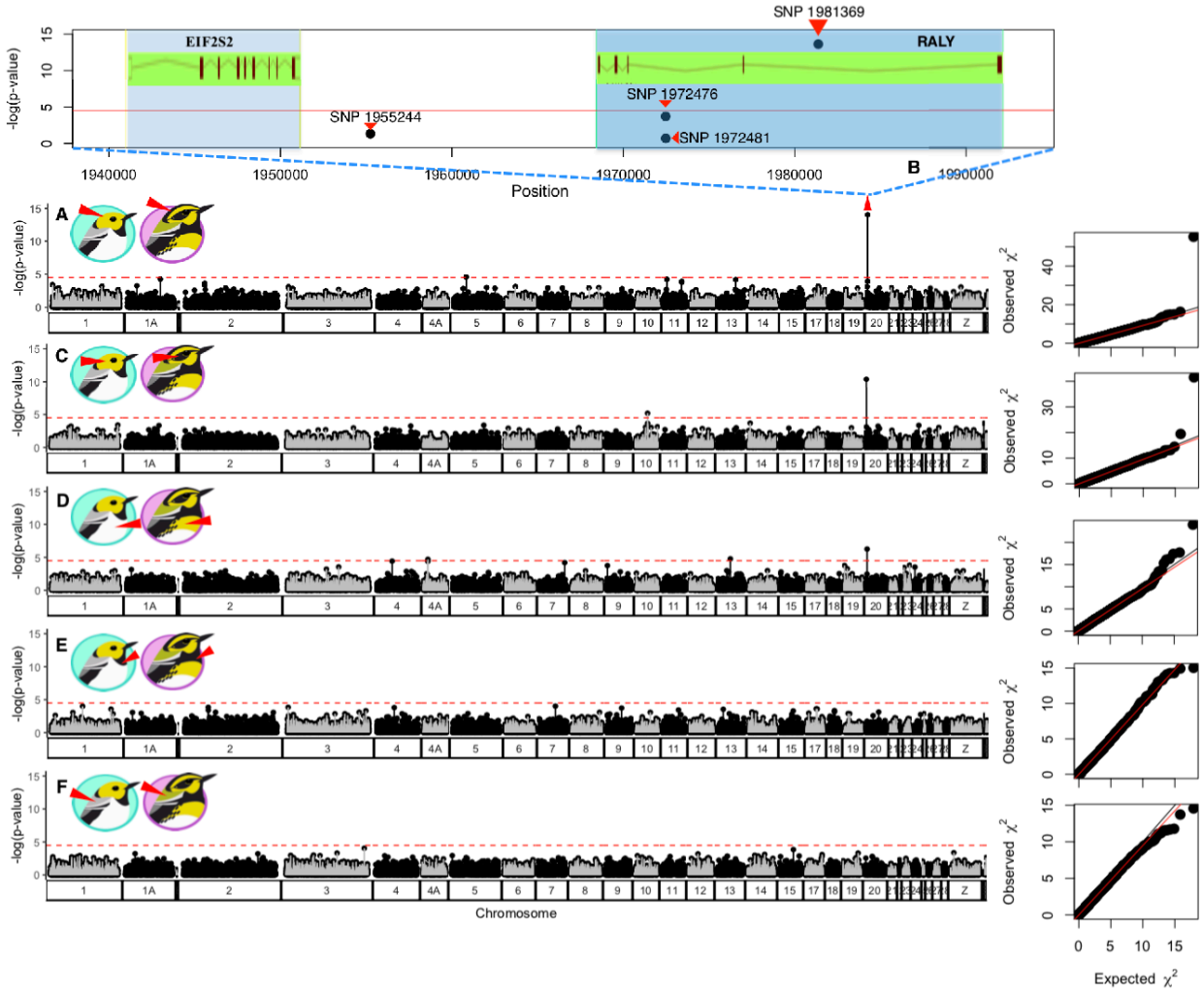

Figure 3 Admixture mapping revealed a SNP inside the RALY gene in chromosome 20 at position 1981369 that was significantly associated with crown (**A**, **B**) and cheek darkening (**C**), as well as the intensity of yellow on the breast (**D**), but not with bib size (**E**) nor with the color of the back (**F**). **A-F**, genomic scan of inflation factor-corrected –log (*p*-value) of genotype-phenotype association tests. The horizontal red lines represent the Bonferroni-corrected critical threshold. The plumage trait tested in each scan is indicated by the red triangle in the cartoons. **A-C**, there was a strong peak at chromosome 20 position 1981369 inside the RALY gene that indicates a pleiotropic effect on crown, cheek, and breast coloration. **B**, genetic map of around position 1981369 on chromosome 20 and its nearby SNPs in this study. The green boxes represent the stretches of the protein-coding genes, EIF2S2 and RALY, with the exons depicted as the red vertical bars connected by introns (red lines). **E-F**, no significant SNP was detected that is associated with bib size or greenish back.

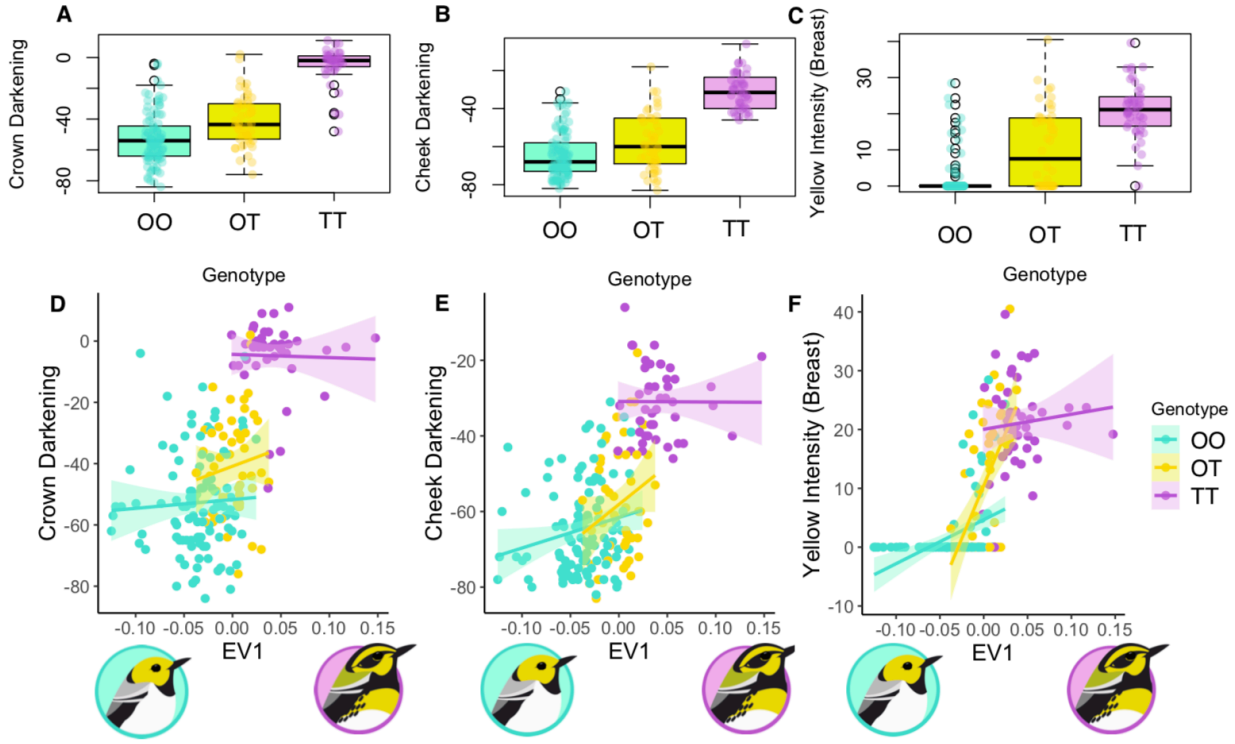

Figure 4 Association of the RALY SNP on crown (**A**, **D**), cheek (**B**, **E**), and breast (**C**, **F**) coloration, in which the pure *occidentalis* genotypes is denoted as "OO", pure *townsendi* genotype is denoted as "TT", and heterozygotes as "OT". **D-F**, The associations of the RALY genotype and cheek darkening among individuals are significant ($p < 10^{-7}$) after accounting for the underlying genomic ancestry. The data are consistent with a combination of dominance and additive effects on the three plumage traits.

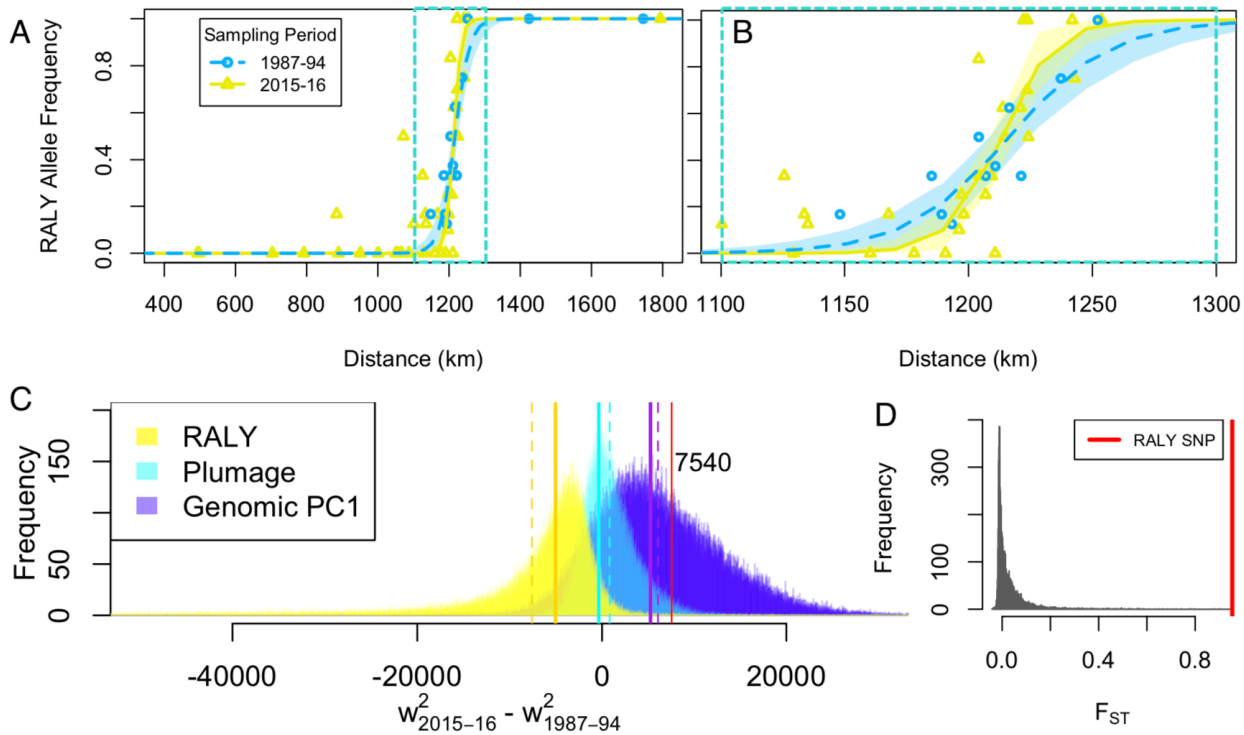

Figure 5 Evidence of selection on RALY locus or its linked region. The RALY SNP shows stable clines (sampled in 1987-94 and 2015-16) across the hybrid zone, which are extremely narrow (**A, B**, zoom in around 1100 to 1300 km), and has become significantly narrower in the recent sampling (**C**). The bootstrap distribution of $w^2_{2015-16} - w^2_{1987-94}$ (the change of squared cline width between sampling periods) of the RALY cline is less than expected under neutral diffusion, suggesting selection has been maintaining narrow clines of genomic PC1, plumage, and RALY SNP (**C**). The selection at RALY (greater deviation from neutral expectation than plumage and genomic PC1) might indirectly maintain the narrow plumage and genomic cline (**C**). The vertical lines are $w^2_{2015-16} - w^2_{1987-94}$ values (solid line: sample estimates; dotted line: bootstrap mean values) respectively for RALY cline (yellow), plumage cline (turquoise), and genomic PC1 cline (purple). (**D**), RALY is the extreme outlier in the distribution of F_{ST} between inland *townsendi* and *occidentalis* across the genome (RALY SNP F_{ST} value depicted by the red line).

Supplementary Information

We then investigated whether selection at the RALY marker representing ASIP-RALY gene block has led genetic hitchhiking of its flanking regions facilitating further differentiation around this target of selection. We tested whether the bootstrap (for 10,000 iterations) difference of F_{ST} between the candidate genetic block (300kb flanking region of the ASIP-RALY genetic block, excluding the RALY locus itself, colored in yellow Figure S5A) versus the rest of chr20 (colored in light blue Figure S5A) is significantly greater than 0. Indeed, we found the signature of hitchhiking, as ASIP-RALY candidate genetic block demonstrated significantly greater F_{ST} than the rest of the regions in chr20 (Figure S5AB, grey distribution of difference is greater than 0, 95% CI: 0.107-0.109, $t_{9999} = 254.31$, $p < 10^{-15}$).

Figure S1 Field photos of a hybrid male showing three different angles: 1) frontal with head tilted up showing throat badge and breast measurements), 2) profile showing the cheek, and 3) from above showing the crown.

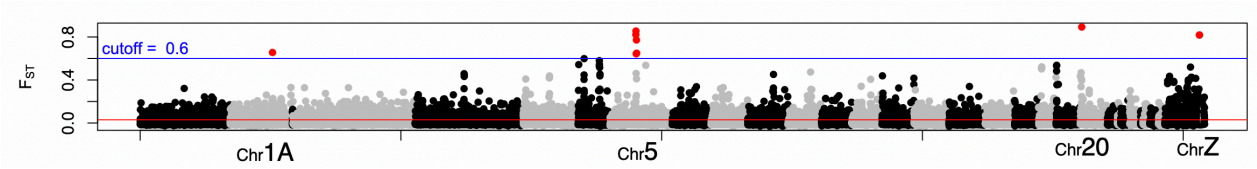

Figure S2 SNPs with $F_{ST} > 0.6$ are colored in red. High F_{ST} SNPS are distributed on chr1A, 5, 20, and Z.

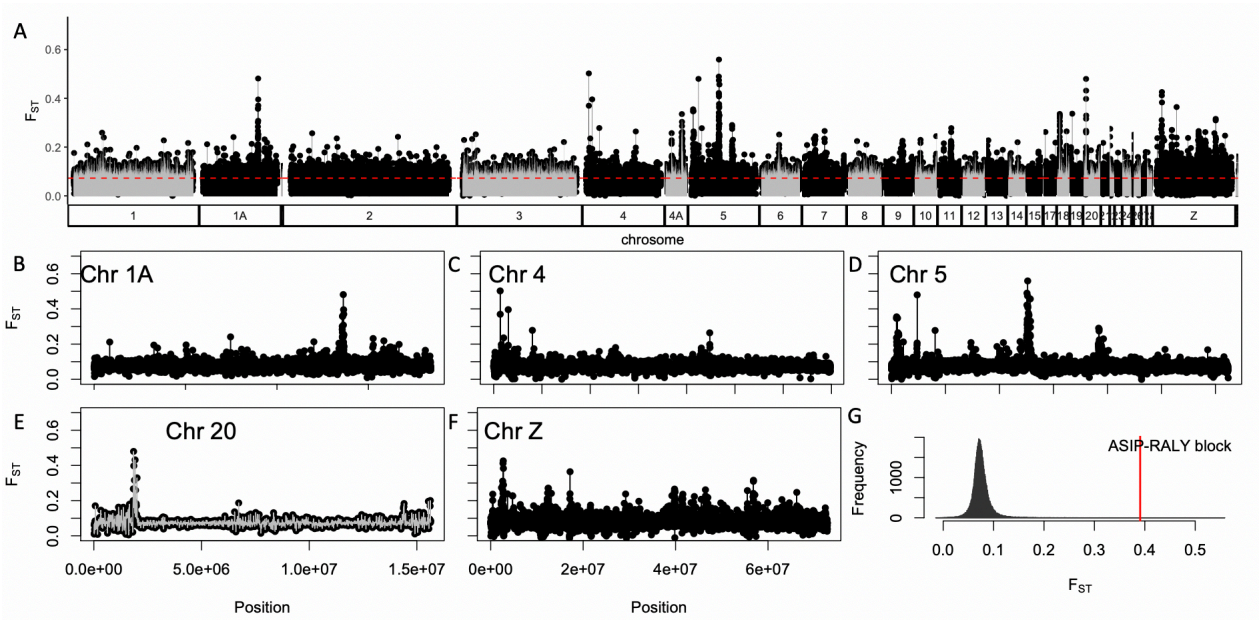

Figure S3 Weir Cockerham F_{ST} scan of with the WGS data, in which each dot represents a 10kb non-overlapping window (A), where peaks were found on chromosome 1A (B), 4 (C), 5 (D), 20 (E), and Z (F). G, the ASIP-RALY gene block demonstrates extremely high F_{ST} relative to the rest of the genome.

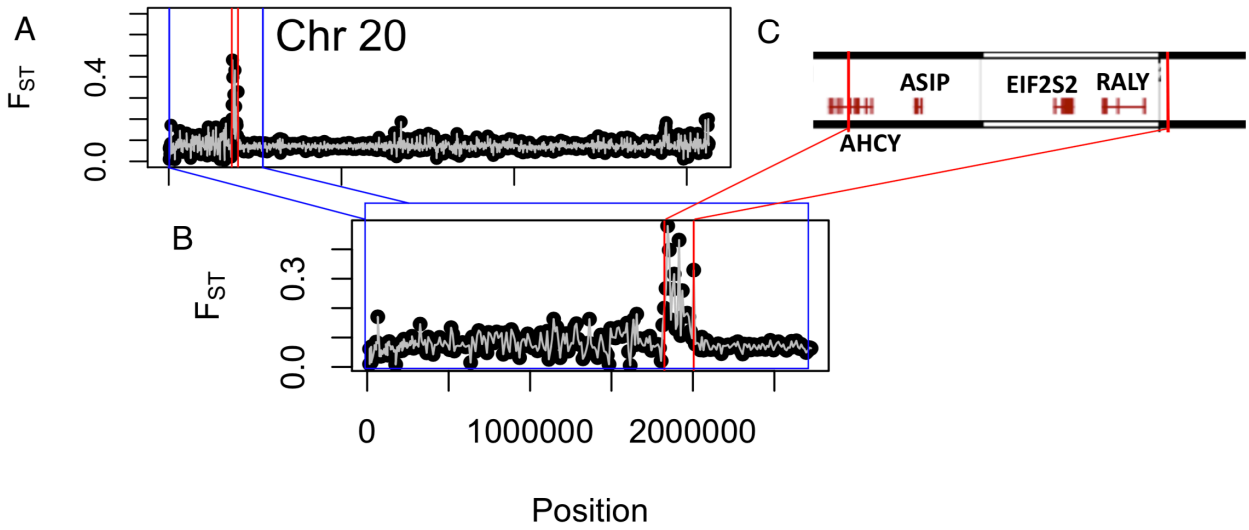

Figure S4 The F_{ST} peak on chromosome 20 (A) resides in the ASIP-RALY gene block (B, C). C, the position of the protein coding genes relative to the peak (bounded by the red vertical lines). C, within the genes, the vertical strikes are the coding regions flanked by the non-coding regions (horizontal lines).

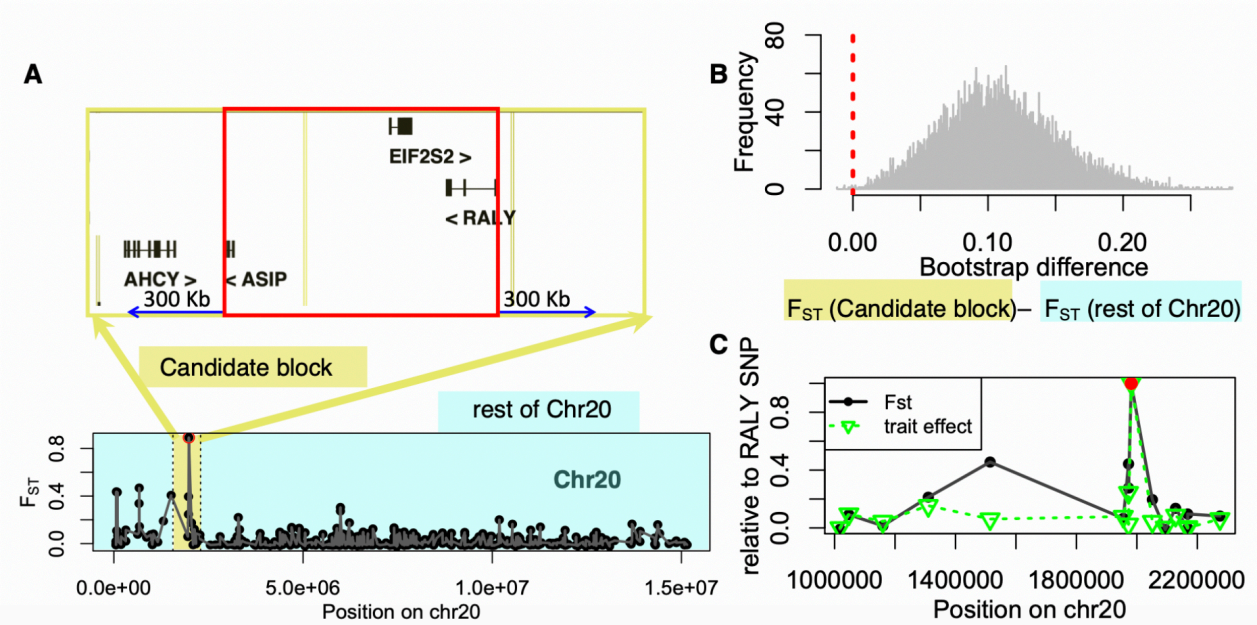

Figure S5 There is signature of hitchhiking around the RALY peak (elevated F_{ST} around RALY), while the effect on plumage coloration is more tightly associated with RALY. Thus the potential causal locus of this trait should be very close to RALY. **A, B** signature of hitchhiking: the candidate gene block (the pigmentation gene block and its 300kb flanking region excluding the RALY locus) has higher F_{ST} than regions in the rest of the chr20 (B, the bootstrap difference of F_{ST} is significantly greater than 0, indicated by the vertical dotted line, 95% CI: 0.107-0.109). **C**, in comparison to F_{ST}, the trait effect is more specific to the RALY peak: there is a steeper decay of scaled trait effect size (relative to RALY peak) than the scaled differentiation (relative F_{ST} to RALY peak).

Table S1 Allele frequencies of SNPs between RALY and ASIP in hybrid zone and parental zones. The causal genetic mechanism for plumage coloration might be narrower than the ASIP-RALY linkage block around the RALY SNP. The RALY SNP is in physical proximity to the two other pigmentation genes mentioned above, although 3 other SNPs (1955244, 1972476, and 1972481) that are 8888-26125 bp away from RALY SNP (Table S2; Figure 3B), physically closer to ASIP and EIF2S2, did not show association with phenotype (Figure 3B). Two of these SNPs showed low minor allele frequencies (0.05-0.09) in the hybrid zone, and thus are not expected to be highly associated with trait variation. However, the fact that SNP 1973476 (closer to the other genes than the candidate RALY SNP) demonstrated similar minor allele frequency as the significant RALY SNP and was not significantly associated with plumage coloration highlights the importance of the region around position 1981369 inside the RALY gene. Either way, the RALY SNP represents the ASIP-RALY gene block and appears to have an effect on multiple species-diagnostic coloration traits.

	POS	**allele1:freq**	**allele2:freq**
Hybrid zone			
	1955244	T:0.946	A:0.054
	1972476	A:0.72	C:0.28
	1972481	T:0.917	C:0.083
	1981369	C:0.375	G:0.625
townsendi zone			
	1955244	T:1	A:0
	1972476	A:0.5	C:0.5
	1972481	T:0.656	C:0.344
	1981369	C:1	G:0
occidentalis zone			
	1955244	T:0.95	A:0.05
	1972476	A:0.944	C:0.056
	1972481	T:0.972	C:0.028
	1981369	C:0.048	G:0.952

Table S2 SNPs with F_{ST} > 0.6 and their position, association to genes and molecular functions

Chromosome	Position	F_{ST}	Relation to genes	Gene(s)	Function
1A	54442413	0.656	intergenic	GRM8, ENSTGUG00000004218	G protein-coupled receptor activity, glutamate receptor activity, telomere maintenance, DNA binding
5	25064223	0.822	intron	UBR1	ubiquitin-protein transferase activity
5	25174918	0.855	coding	ENSTGUG00000011205	DNA binding
5	25746680	0.644	intergenic	DPF3, RGS6	histone acetyltransferase activity; G protein-coupled receptor signaling pathway, intracellular signal transduction
5	25783847	0.772			
5	25875302	0.649	intron	RGS6	G protein-coupled receptor signaling pathway, intracellular signal transduction
20	1981369	0.892	intron	RALY	DNA and RNA binding, cholesterol biosynthesis
Z	66226657	0.818	intron	BBOX1, TNPO1	carnitine biosynthesis, oxidation-reduction; protein import into nucleus, intracellular protein transport