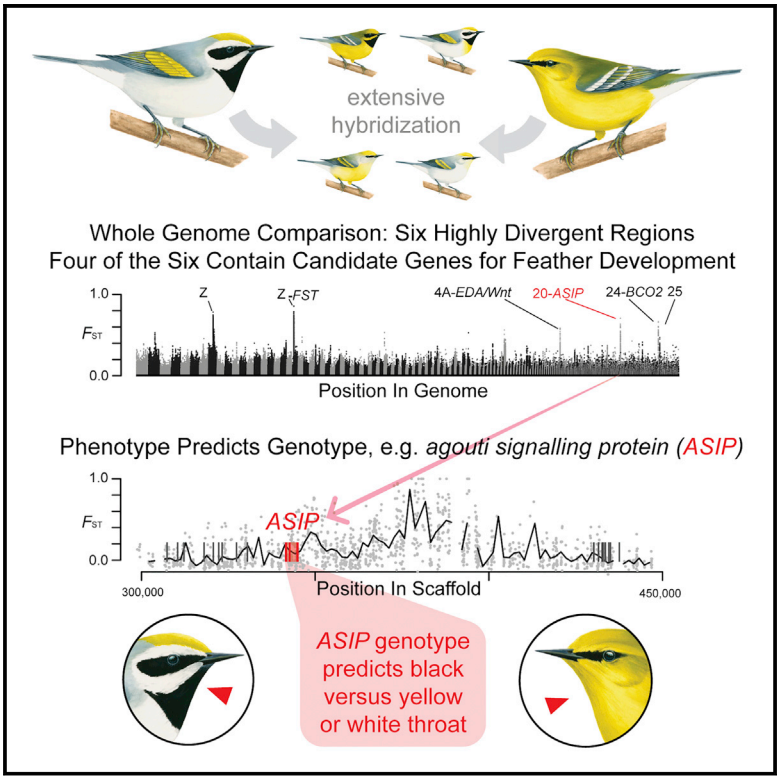
Short Article

Current Biology

Plumage Genes and Little Else Distinguish the Genomes of Hybridizing Warblers

Graphical Abstract



Highlights

- Only six genomic regions differ between golden-winged and blue-winged warblers
- Four of the regions are upstream of genes associated with plumage color and pattern
- A SNP in the presumed *agouti* promoter correlates perfectly with throat color
- Demographic models show hybridization is old and ongoing, against conventional wisdom

Authors

David P.L. Toews, Scott A. Taylor, Rachel Vallender, Alan Brelsford, Bronwyn G. Butcher, Philipp W. Messer, Irby J. Lovette

Correspondence

toews@cornell.edu

In Brief

Using whole genomes, Toews, Taylor et al. find six small genomic regions that are divergent between hybridizing wood warblers, which differ strongly in coloration. Most of these regions contain candidate genes for feather patterning and pigmentation. Hybridization appears to be a long-term component of their history, against conventional wisdom.

Toews et al., 2016, Current Biology *26*, 1–6
September 12, 2016
http://dx.doi.org/10.1016/j.cub.2016.06.034

CellPress

Please cite this article in press as: Toews et al., Plumage Genes and Little Else Distinguish the Genomes of Hybridizing Warblers, Current Biology (2016), http://dx.doi.org/10.1016/j.cub.2016.06.034

Current Biology
Short Article

Plumage Genes and Little Else Distinguish the Genomes of Hybridizing Warblers

David P.L. Toews,[1,2,6,*] Scott A. Taylor,[1,2,6,7] Rachel Vallender,[3] Alan Brelsford,[4] Bronwyn G. Butcher,[1] Philipp W. Messer,[5] and Irby J. Lovette[1,2]
[1]Fuller Evolutionary Biology Program, Cornell Lab of Ornithology, Cornell University, 159 Sapsucker Woods Road, Ithaca, NY 14850, USA
[2]Department of Ecology and Evolutionary Biology, Cornell University, Corson Hall, Ithaca, NY 14853, USA
[3]Canadian Wildlife Service, Environment and Climate Change Canada, 351 St. Joseph Boulevard, Gatineau, Québec K1A 0H3, Canada
[4]Department of Evolution, Ecology and Organismal Biology, University of California, Riverside, Batchelor Hall, Riverside, CA 92521, USA
[5]Department of Biological Statistics and Computational Biology, Cornell University, Weill Hall, Ithaca, NY 14853, USA
[6]Co-first author
[7]Present address: Department of Ecology and Evolutionary Biology, University of Colorado at Boulder, Ramaley N122, Boulder, CO 80309, USA
*Correspondence: toews@cornell.edu
http://dx.doi.org/10.1016/j.cub.2016.06.034

SUMMARY

When related taxa hybridize extensively, their genomes may become increasingly homogenized over time. This mixing via hybridization creates conservation challenges when it reduces genetic or phenotypic diversity and when it endangers previously distinct species via genetic swamping [1]. However, hybridization also facilitates admixture mapping of traits that distinguish each species and the associated genes that maintain distinctiveness despite ongoing gene flow [2]. We address these dual aspects of hybridization in the golden-winged/blue-winged warbler complex, two phenotypically divergent warblers that are indistinguishable using traditional molecular markers and that draw substantial conservation attention [3–5]. Whole-genome comparisons show that differentiation is extremely low: only six small genomic regions exhibit strong differences. Four of these divergence peaks occur in proximity to genes known to be involved in feather development or pigmentation: *agouti signaling protein* (*ASIP*), *follistatin* (*FST*), *ecodysplasin* (*EDA*), *wingless-related integration site* (*Wnt*), and *beta-carotene oxygenase 2* (*BCO2*). Throat coloration—the most striking plumage difference between these warblers—is perfectly associated with the promoter region of *agouti*, and genotypes at this locus obey simple Mendelian recessive inheritance of the black-throated phenotype characteristic of golden-winged warblers. The more general pattern of genomic similarity between these warblers likely results from a protracted period of hybridization, contradicting the broadly accepted hypothesis that admixture results from solely anthropogenic habitat change in the past two centuries [4]. Considered in concert, these results are relevant to both the genetic architecture of avian feather pigmentation and the evolutionary history and conservation challenges associated with these declining songbirds.

RESULTS AND DISCUSSION

The parulid warblers of North America are a well-known avian radiation in which species are distinguished by dramatic differences in plumage [6]. Unlike classic adaptive radiations, in which species show strong ecological differentiation driven by natural selection, the most conspicuous phenotypic differences across this family of warblers are presumably the result of divergent sexual selection on plumage characters [7]. In most species radiations (e.g., crater lake Cichlids, New Guinean birds of paradise), it remains challenging to link genes to phenotypes and thereby understand how selection shapes key traits at the molecular level. The genomic mixing that occurs from natural hybridization provides a powerful opportunity to identify the traits and associated genes that may function as reproductive barriers in such explosive radiations. Here we study two phenotypically distinct wood warblers that are known to be very similar throughout much of their nuclear genome, likely due to gene flow [3]. We leverage this genomic similarity to associate their pronounced plumage differences with candidate genes that likely underlie their coloration.

We compared whole-genome variation in golden-winged (*Vermivora chrysoptera*) and blue-winged (*V. cyanoptera*; Figure 1) warblers, which hybridize across a broad zone of eastern North America. These warblers have perplexed biologists since at least 1835, when pioneering ornithologist John James Audubon wrote about the paradox of their distinct plumages yet similar distributions, songs, and ecology [8]. Even though Audubon was unaware that hybridization occurred between them, he surmised that they might simply be strikingly different plumage variants of the same species [8]. Nearly a century later, hybridization was documented when a third species, then named ''Brewster's Warbler,'' was found to be a common hybrid of blue-winged and golden-winged warblers (Figure 1B). Numerous studies have documented extensive hybridization [3] with little or no

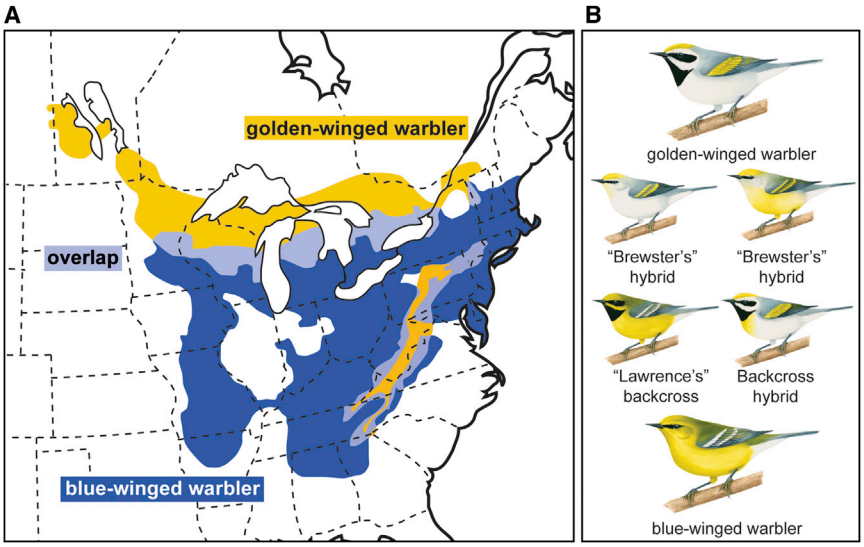


Figure 1. Geographic Variation across the Golden-Winged and Blue-Winged Warbler Complex

(A) The range of golden-winged (yellow) and blue-winged (blue) warblers. Areas of overlap (light blue) have both of the parental phenotypes, as well as birds of hybrid ancestry.

(B) Illustrations of the parental and several hybrid phenotypes (illustrations by Liz Clayton Fuller).

See also Figure S1.

detectable fitness reduction in hybrid individuals, which appear fully fertile [5]. Both taxa are declining, golden-winged warblers precipitously so, and are the focus of conservation efforts [4]. This decline is due in part to forest regeneration, which has reduced the availability of the early-successional habitats that both taxa rely on during breeding [4]. Golden-winged warblers are also thought to be threatened by displacement and hybridization due to expanding ranges of blue-winged warblers [5]: in many locations, golden-winged warblers have been replaced by hybrids and subsequently by blue-winged warblers [9].

Extreme Interspecific Genomic Similarity

Golden-winged and blue-winged warblers exhibit an unusual combination of substantial divergence in their mitochondrial genome (mtDNA) but very low divergence between their nuclear genomes [3, 13]. The two mtDNA lineages possessed by the warblers differ by approximately 3%, suggesting that they were separated for a substantial period at some time in the past [10]. The contemporary distribution of mtDNA lineages broadly corresponds to allopatric populations (Figure 1A); however, extensive mtDNA mixing occurs in locations where both forms interbreed [11, 12]. Divergence in the nuclear genome is much lower: assays of allozymes [12], microsatellites [3], introns [3], AFLPs [3], and reduced-representation genome sequencing (ddRAD, this study; Figure S1) have failed to identify fixed markers between the taxa.

Using whole-genome re-sequencing, we characterized the chromosomal regions at which these species differ, identified the genes that likely underlie their distinct plumage traits, and modeled the demographic history of introgression that has resulted in their combination of phenotypic differentiation and low genomic divergence. We first assembled the genome of a related warbler species (the yellow-rumped warbler, *Setophaga coronata*) as a high-quality reference. We then re-sequenced the genomes of ten male golden-winged and ten male blue-winged warblers to an average individual coverage of 4–5X.

We found 11.4 million single nucleotide polymorphisms (SNPs) among these 20 birds. Measures of differentiation across their genomes were extremely low (weighted-mean F_{ST} = 0.0045) compared to all other hybridizing avian species pairs for which similar data are available, such as Galápagos finches [14], *Ficedula* flycatchers [15], hooded and carrion crows [2], and subspecies of Swainson's thrush [16]. The level of nuclear differentiation is also much lower than would be expected based on distance estimates from non-recombining mtDNA, which would suggest approximately over one million years of independent evolution (Figure S2).

Genome-wide comparisons identified a number of strongly differentiated loci, the majority of which clustered within six very small regions of the genome (Figure 2A). For example, of the 362 SNPs with $F_{ST} > 0.9$, all but three were found in these small clusters, each with multiple divergent markers per scaffold. Most of the highly divergent loci (331 SNPs with $F_{ST} > 0.9$) occur on two scaffolds that map to the avian sex chromosome; these Z-linked regions also house the majority of fixed SNPs across the genome (61 of 74 SNPs with $F_{ST} = 1$). Within the six scaffolds with F_{ST} peaks, regions of high divergence were very small (median peak size = ~30 kb), with the largest peak occurring on one of the two Z-linked scaffolds (~180 kb). Together, these divergent sites represent <0.03% of the polymorphic regions of the genome. We also found that the divergent regions had, on average, higher levels of absolute divergence (e.g., d_{xy}; Figure S3). The disproportionate extent and number of divergence peaks associated with the Z chromosome are likely due to its reduced N_e and recombination rate but also may be due to reproductive barrier gene accumulation on the Z chromosome [17].

Admixture Reveals Genes for Feather Pigmentation and Development

To further test for associations between divergent genomic regions and the phenotypic traits that distinguish these warblers, we assayed SNPs in the six divergence peaks in a set of individuals sampled from across the ranges of both taxa (n = 346). These additional individuals came from both areas with active hybridization and allopatric regions and represented the full spectrum of phenotypes. In this assay, all six loci exhibited strong frequency differences between the phenotypes (Figure 2B; Table S1). We found a strong correlation between the additive score of the genotypes at these loci and a plumage score across a gradient of phenotypes (Figure 3; n = 132, R^2 = 0.94, df = 130, p = $2.2e^{-16}$), consistent with these genomic regions housing the causal genes that contribute to the phenotypic differences between these warblers (Table S2).

Please cite this article in press as: Toews et al., Plumage Genes and Little Else Distinguish the Genomes of Hybridizing Warblers, Current Biology (2016), http://dx.doi.org/10.1016/j.cub.2016.06.034

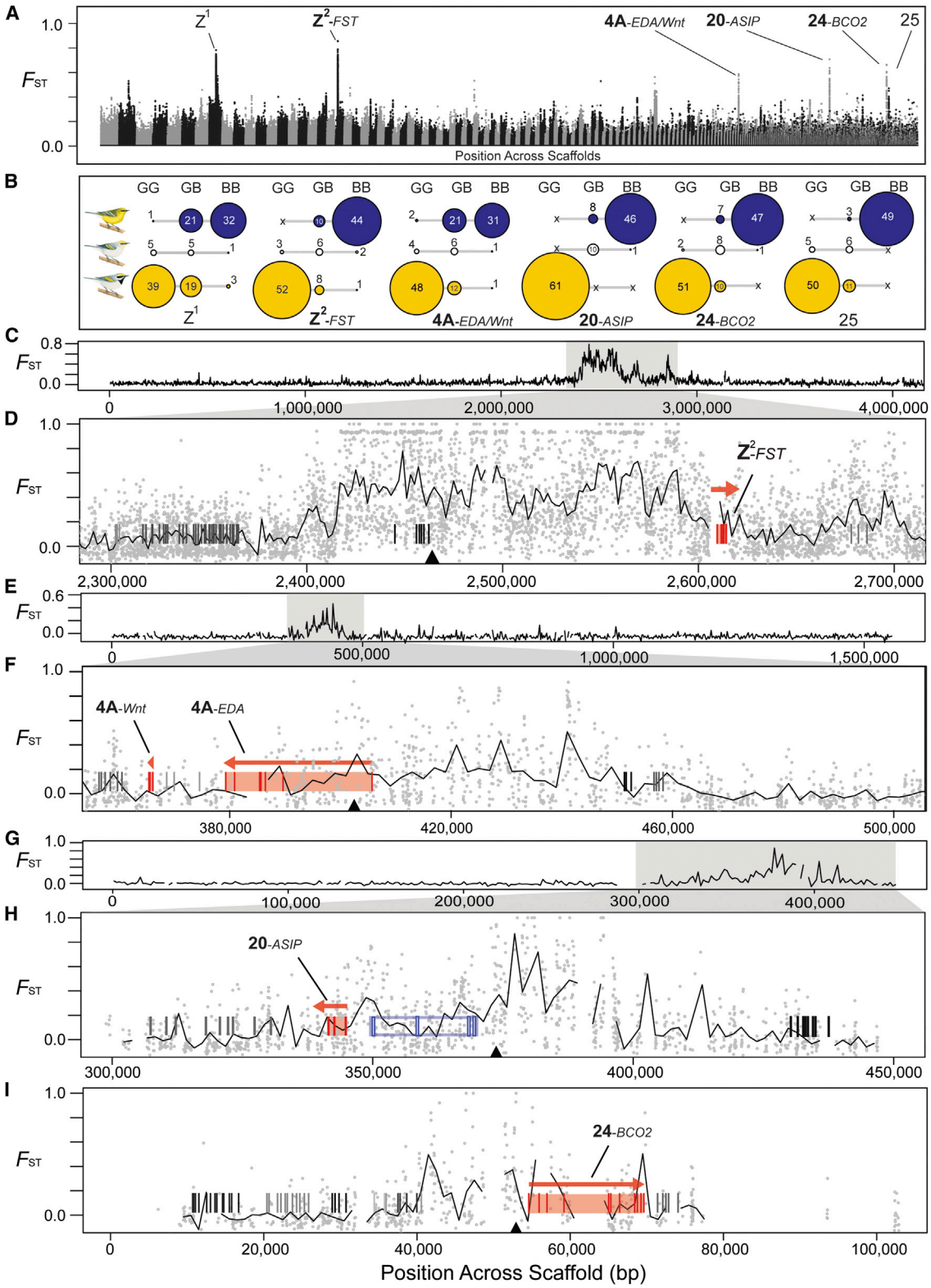


Figure 2. Genomic Patterns of Divergence between Golden-Winged and Blue-Winged Warblers

(A) Overlapping sliding windows of F_{ST}. Divergent scaffolds are identified by their position in the zebra finch (ZF) genome and the associated candidate genes. (B) Genotype frequencies for SNPs in each divergent region. The size of the circle corresponds to the sample size from 2015 sampling; rows are distinguished by phenotype (i.e., "blue-winged," "golden-winged," or "Brewster's").

(legend continued on next page)

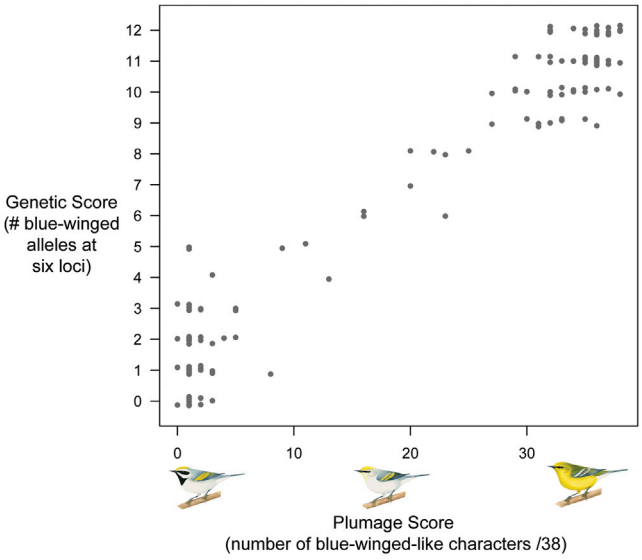


Figure 3. Correlation between Genotype and Phenotype

Sum of plumage score for individuals with complete genetic and plumage information (n = 132). The genetic score is the sum of the genotypes across the six divergent regions using an RFLP assay.

The very small size of the peaks allowed us to identify particular genes that are likely responsible for some of the observed plumage variation (Figure S4; Table S2), including *wingless-related integration site* (*Wnt*; Figures 2E and 2F), *ectodysplasin* (*EDA*; Figures 2E and 2F), *agouti signaling protein* (*ASIP*; Figures 2G and 2H), *beta-carotene oxygenase 2* (*BCO2*; Figure 2I), and *follistatin* (*FST*; Figures 2C and 2D). In each case, the divergent regions fall in the 5′ region that is directly upstream of the associated coding region. *Wnt* is expressed in developing feather tracts and buds in chicken embryos and is linked to defining feather bud polarity [18]. *EDA* is expressed in developing feather placodes in chicken embryos and involved in the feather patterning pathways [19]. *ASIP* interacts with *MC1R* in follicular melanocytes and has well-characterized function in pheomelanin (e.g., yellow/red pigment) synthesis and as an inhibitor of eumelanin (e.g., black/brown pigment) [20]. Several *ASIP* promoters have been characterized in chickens and are involved in feather dichromatism [21]; the chicken *ASIP* promoters align to positions very near the divergence peaks between the warblers (Figure 2H). *BCO2* is involved in the breakdown of carotenoids (red/yellow/orange pigments) to form precursors of vitamin A and other metabolites [22]; it is involved in yellow skin pigmentation in chickens but has not been directly implicated in differences in feather coloration [23]. Finally, *FST* is expressed during feather bud development in chickens [24]. *FST* has also been indirectly linked to possible plumage differences between *Ficedula* flycatchers [25]. Therefore, all five of these genes have strong evidence-based links to the specific aspects of plumage color and patterning that differ between these warblers. Finally, while warbler scaffold 653 holds only one uncharacterized gene (Figure S4G), it is within a protein family associated with feather kertains (Table S2). This region of chromosome 25 is also associated with the epidermal differentiation complex (EDC), a group of genes involved in integumentary development and very recently linked to possible ketocarotenoid metabolism in canaries [26]. This region and suite of genes is not yet well characterized [26]; additional functional annotation will be required before there can be any association with feather pigmentation or patterning in this or other systems.

Correlations of variation at particular feather tracts and specific genetic variants imply even stronger associations between genotype and phenotype. For example, the black throat of golden-winged warblers, absent in blue-winged warblers and F1 hybrids, was predicted to be a Mendelian recessive trait as long ago as 1908 [27]. This prediction is now validated by the perfect correlation between the *ASIP* region and throat coloration across every bird in our sample: all *ASIP*-heterozygous individuals have a yellow or white throat, whereas a black throat is found only in birds homozygous for the golden-winged variant (Figure 2B and Table S1). The recessive behavior of *ASIP* conferring a melanistic phenotype is consistent across other studies in other vertebrates [20]. The blue-winged variant of the *BCO2* region similarly has a high correlation with the extent of yellow in several feather tracts, consistent with its involvement in yellow carotenoid metabolism [22]. Moreover, the only individual in our sample that has the very rare "Lawrence's warbler" phenotype (i.e., a yellow body and upper parts, but with a distinct black throat and mask; Figure 1B) is also the only individual in our sample that is homozygous for both the blue-winged *BCO2* genotype and the golden-winged *ASIP* genotype. Parkes [28] suggested that the allele for white underparts is incompletely dominant over the allele for yellow and that heterozygotes could be both white and yellow. This is in general agreement with our data: birds heterozygous for the *BCO2* genotype exhibit a range of yellow feathers, particularly in the breast. More-specific genotype-phenotype connections, as well as more detail on the sequence variation and expression of these genes, will be important to identify causal mutations; however, these associations strongly implicate linked regulatory SNPs upstream of each gene.

Conservation in the Genomics Era

The genomic patterns we have characterized allow us to frame the historical context in which admixture has occurred. Hybridization with blue-winged warblers is considered one of the primary conservation threats to golden-winged warblers [4]. It has been widely posited that hybridization between these warblers has been caused by recent human-mediated habitat modification via the creation (and subsequent abandonment) of farmlands, which created new successional breeding habitats that brought the two species into wide sympatry [5, 9]. We tested whether this scenario is consistent with the genomic data or

(C–I) Patterns of F_{ST} across the warbler scaffolds that align to the ZF chromosomes: Z (C and D), 4A (E and F), 20 (G and H), and 24 (I). Gray points show per-locus F_{ST} estimates. Vertical lines indicate exons in coding regions; open vertical lines in (H) show the alignment of the chicken *ASIP* promoters (see Supplemental Experimental Procedures). Filled triangles show the locations of RFLP SNPs genotyped. Red lines identify the location and orientation of candidate genes involved in either feather development or pigmentation.

See also Figures S2–S4 and Tables S1, S2, and S4.

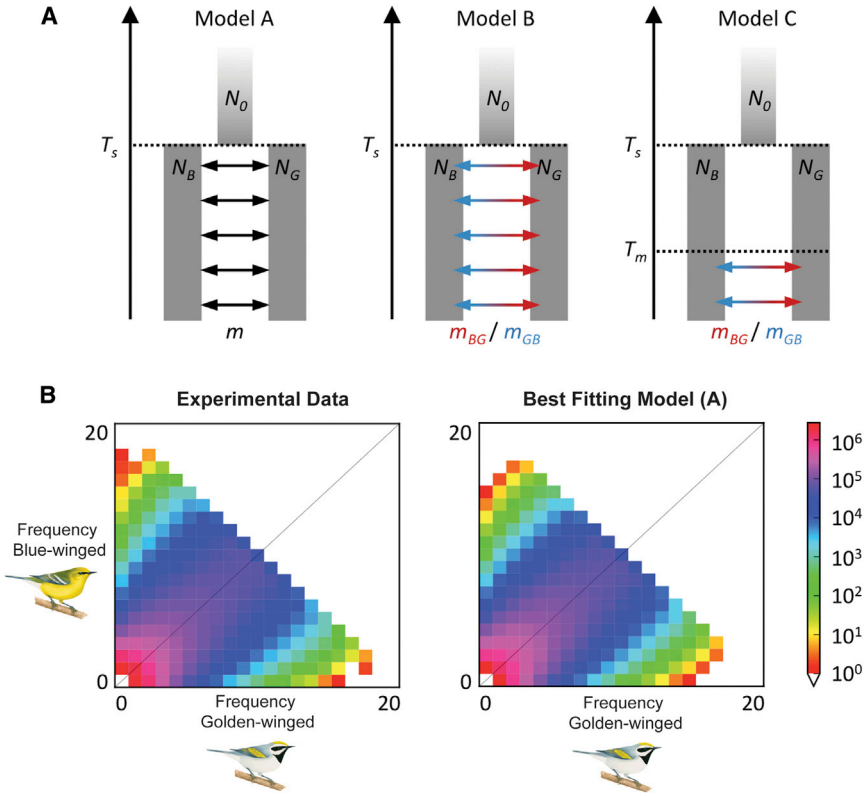


Figure 4. Demographic Modeling of Genomic Data

Using ∂a∂i, we tested several demographic models in order to estimate the timing of hybridization onset between golden-winged and blue-winged warblers.

(A) Model A specifies an ancestral population that, at time T_s, split into two subpopulations with gene flow (m) at a constant rate. Model B allows for the possibility of asymmetric migration rates (m_{BG} and m_{GB}). Model C allows for no gene flow until time T_m, with constant asymmetric migration rates between the two subpopulations afterward.

(B) Comparison of folded joint frequency spectrum from the observed data and the output from the best-fitting model (model A).

See also Table S3.

whether hybridization is an older and possibly ongoing feature of these warblers' evolutionary histories. The best model included an old and continuous movement of genes, as opposed to only recent gene flow (Figure 4; Table S3). This result is consistent with humans facilitating some recent admixture via 19th-century habitat change but also shows that hybridization has likely occurred over many millennia.

These genomic inferences have multifaceted implications for the conservation of these taxa. The golden-winged warbler is currently listed as "threatened" under the Canadian Species at Risk Act and is being considered for listing under the USA Endangered Species Act. On one hand, the low and restricted genomic divergence between these taxa—likely resulting from extensive hybridization—makes their classification as distinct species less certain; on the other hand, these distinct phenotypes appear to have persisted despite this extreme genomic similarity and despite a protracted period of hybridization. Both regulatory frameworks support the conservation of evolutionary significant variation within species, and we recommend focusing on managing the genetic and phenotypic diversity within the warbler complex as a whole. Range-wide monitoring at these divergent and functional loci will be a powerful tool for quantifying the extent, pace, and direction of hybridization in this dynamic hybrid zone mosaic.

General Conclusions

We have identified five candidate genes that are strongly correlated with plumage color and patterning in these distinct warblers. Identifying the genetic bases of coloration in vertebrates is important, particularly finding the genes involved in carotenoid metabolism and deposition, because they have eluded general characterization [22, 26, 29]. In the warblers, the finding that so few genes are likely responsible for such dramatic differences in coloration across multiple feather tracts is both unexpected and exciting. From a comparative evolutionary perspective, these findings will facilitate the study of molecular parallelism and convergence within the warbler family [6], as well as across other explosive avian radiations, such as the *Sporophilia* seedeaters [30], which both show similarly strong divergence in coloration across analogous feather tracts. In these particular hybridizing taxa, the mechanisms by which the divergent genomic regions and phenotypic differences are maintained with high levels of gene are still unknown. Nevertheless, these data have implications for our understanding of the hybridization dynamics in this system and pose a new challenge for how best to interpret and manage phenotypic distinctiveness in the face of striking genomic similarity.

ACCESSION NUMBERS

The accession number for the genotypes and plumage information, plumage scoring criteria, enzyme conditions and information for RFLP assays, variant calls for ddRAD dataset, and variant calls for resequencing dataset reported in this paper has been deposited to the Data Dryad database as Data Dryad: http://dx.doi.org/10.5061/dryad.kb610. The accession number for the genome assembly and raw reads reported in this paper is NCBI: PRJNA325157. The accession number for the raw reads from the resequencing dataset reported in this paper is NCBI: PRJNA325126.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2016.06.034.

AUTHOR CONTRIBUTIONS

D.P.L.T., S.A.T., and I.J.L. conceived the idea for the study. A.B. assembled the reference warbler genome. D.P.L.T. collected resequencing samples and generated resequencing libraries. D.P.L.T. and S.A.T. performed the RFLP assays, scored plumage characters, and worked on various parts of

the bioinformatics pipeline. D.P.L.T., S.A.T., and B.G.B. assisted with laboratory work, including DNA extraction and generating ddRAD libraries. P.W.M. performed the demographic modeling. R.V. collected initial samples for the study. D.P.L.T. and S.A.T. wrote the original version of the manuscript. D.P.L.T., S.A.T., A.B., P.W.M., R.V., B.G.B., and I.J.L. wrote and/or substantially edited portions of the paper.

ACKNOWLEDGMENTS

The authors thank A. Roth, S. Van Wilderberg, D. Buehler, K. Aldinger, S. Barker, R. Rohrbaugh, K. Rosenberg, R. Ricklefs, L. Soares, and J. Jankowski for sample collection, as well as N. Mason, P. Dean-Coe, L. Campagna, A. James, E. Larson, and two anonymous reviewers for comments on the manuscript; R. Burri and C. Linnen provided analytical input; Liz Clayton Fuller created the warbler illustrations. D.P.L.T. was supported by the Natural Sciences and Engineering Council of Canada. S.A.T. was supported by a Banting Postdoctoral Fellowship. Genome sequencing was partially supported by a Université Lausanne grant to T. Kawecki. Genome assembly computations were performed on the Vital-IT computing cluster of the Swiss Institute for Bioinformatics. All procedures were approved by Cornell's Institution Animal Care and Use Committee.

Received: February 26, 2016
Revised: April 13, 2016
Accepted: June 15, 2016
Published: August 18, 2016

REFERENCES

1. Seehausen, O. (2006). Conservation: losing biodiversity by reverse speciation. Curr. Biol. *16*, R334–R337.
2. Poelstra, J.W., Vijay, N., Bossu, C.M., Lantz, H., Ryll, B., Müller, I., Baglione, V., Unneberg, P., Wikelski, M., Grabherr, M.G., and Wolf, J.B. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science *344*, 1410–1414.
3. Vallender, R., Robertson, R.J., Friesen, V.L., and Lovette, I.J. (2007). Complex hybridization dynamics between golden-winged and blue-winged warblers (*Vermivora chrysoptera* and *Vermivora pinus*) revealed by AFLP, microsatellite, intron and mtDNA markers. Mol. Ecol. *16*, 2017–2029.
4. Buehler, D.A., Roth, A.M., Vallender, R., Will, T.C., Confer, J.L., Canterbury, R.A., Swarthout, S.B., Rosenberg, K.V., and Bulluck, L.P. (2007). Status and conservation priorities of golden-winged warbler (*Vermivora chrysoptera*) in North America. Auk *124*, 1439–1445.
5. Gill, F.B. (2004). Blue-winged warblers (*Vermivora pinus*) versus golden-winged warblers (*V. chrysoptera*). Auk *121*, 1014–1018.
6. Lovette, I.J., and Bermingham, E. (1999). Explosive speciation in the New World *Dendroica* warblers. Proc. Biol. Sci. *266*, 1629–1636.
7. Price, T., Lovette, I.J., Bermingham, E., Gibbs, H.L., and Richman, A.D. (2000). The imprint of history on communities of North American and Asian warblers. Am. Nat. *156*, 354–367.
8. Faxon, W. (1913). Brewster's warbler (*Helminthophila leucobronchlais*) a hybrid between the golden-winged warbler (*H. chrysoptera*) and the blue-winged warbler (*H. pinus*). Mem. Mus. Comp. Zoo. *40*, 311–316.
9. Gill, F.B. (1980). Historical aspects of hybridization between blue-winged and golden-winged warblers. Auk *97*, 1–18.
10. Gill, F.B. (1997). Local cytonuclear extinction of the golden-winged warbler. Evolution *51*, 519–525.
11. Shapiro, L.H., Canterbury, R.A., Stover, D.M., and Fleischer, R.C. (2004). Reciprocal introgression between golden-winged warblers (*Vermivora chrysoptera*) and blue-winged warblers (*V. pinus*) in eastern North America. Auk *121*, 1019–1030.
12. Dabrowski, A., Fraser, R., Confer, J.L., and Lovette, I.J. (2005). Geographic variability in mitochondrial introgression among hybridizing populations of golden-winged and blue-winged warblers. Conserv. Genet. *6*, 843–853.
13. Gill, F.B. (1987). Allozymes and genetic similarity of blue-winged and golden-winged warblers. Auk *104*, 444–449.
14. Lamichhaney, S., Berglund, J., Almén, M.S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C.J., Wang, C., Zamani, N., et al. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature *518*, 371–375.
15. Burri, R., Nater, A., Kawakami, T., Mugal, C.F., Olason, P.I., Smeds, L., Suh, A., Dutoit, L., Bureš, S., Garamszegi, L.Z., et al. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. Genome Res. *25*, 1656–1665.
16. Delmore, K.E., Hübner, S., Kane, N.C., Schuster, R., Andrew, R.L., Câmara, F., Guigó, R., and Irwin, D.E. (2015). Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. Mol. Ecol. *24*, 1873–1888.
17. Saether, S.A., Saetre, G.P., Borge, T., Wiley, C., Svedin, N., Andersson, G., Veen, T., Haavie, J., Servedio, M.R., Bureš, S., et al. (2007). Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. Science *318*, 95–97.
18. Chang, C.-H., Jiang, T.-X., Lin, C.-M., Burrus, L.W., Chuong, C.-M., and Widelitz, R. (2004). Distinct *Wnt* members regulate the hierarchical morphogenesis of skin regions (spinal tract) and individual feathers. Mech. Dev. *121*, 157–171.
19. Houghton, L., Lindon, C., and Morgan, B.A. (2005). The *ectodysplasin* pathway in feather tract development. Development *132*, 863–872.
20. Hoekstra, H.E. (2006). Genetics, development and evolution of adaptive pigmentation in vertebrates. Heredity (Edinb) *97*, 222–234.
21. Oribe, E., Fukao, A., Yoshihara, C., Mendori, M., Rosal, K.G., Takahashi, S., and Takeuchi, S. (2012). Conserved distal promoter of the agouti signaling protein (ASIP) gene controls sexual dichromatism in chickens. Gen. Comp. Endocrinol. *177*, 231–237.
22. Walsh, N., Dale, J., McGraw, K.J., Pointer, M.A., and Mundy, N.I. (2012). Candidate genes for carotenoid coloration in vertebrates and their expression profiles in the carotenoid-containing plumage and bill of a wild bird. Proc. Biol. Sci. *279*, 58–66.
23. Eriksson, J., Larson, G., Gunnarsson, U., Bed'hom, B., Tixier-Boichard, M., Strömstedt, L., Wright, D., Jungerius, A., Vereijken, A., Randi, E., et al. (2008). Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. PLoS Genet. *4*, e1000010.
24. Patel, K., Makarenkova, H., and Jung, H.S. (1999). The role of long range, local and direct signalling molecules during chick feather bud development involving the *BMPs, follistatin* and the Eph receptor tyrosine kinase *Eph-A4*. Mech. Dev. *86*, 51–62.
25. Lehtonen, P.K., Laaksonen, T., Artemyev, A.V., Belskii, E., Berg, P.R., Both, C., Buggiotti, L., Bureš, S., Burgess, M.D., Bushuev, A.V., et al. (2012). Candidate genes for colour and vision exhibit signals of selection across the pied flycatcher (*Ficedula hypoleuca*) breeding range. Heredity (Edinb) *108*, 431–440.
26. Lopes, R.J., Johnson, J.D., Toomey, M.B., Ferreira, M.S., Araujo, P.M., Melo-Ferreira, J., Andersson, L., Hill, G.E., Corbo, J.C., and Carneiro, M. (2016). Genetic basis for red coloration in birds. Curr. Biol. *26*, 1427–1434.
27. Nichols, J.T. (1908). Lawrence's and Brewster's warblers and Mendelian inheritance. Auk *25*, 86.
28. Parkes, K. (1951). The genetics of the golden-winged x blue-winged warbler complex. Wilson Bull. *63*, 5–15.
29. Mundy, N.I., Stapley, J., Bennison, C., Tucker, R., Twyman, H., Kim, K.W., Burke, T., Birkhead, T.R., Andersson, S., and Slate, J. (2016). Red carotenoid coloration in the zebra finch is controlled by a cytochrome P450 gene cluster. Curr. Biol. *26*, 1435–1440.
30. Campagna, L., Gronau, I., Silveira, L.F., Siepel, A., and Lovette, I.J. (2015). Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation. Mol. Ecol. *24*, 4238–4251.

Current Biology, Volume *26*

Supplemental Information

Plumage Genes and Little Else Distinguish the Genomes of Hybridizing Warblers

David P.L. Toews, Scott A. Taylor, Rachel Vallender, Alan Brelsford, Bronwyn G. Butcher, Philipp W. Messer, and Irby J. Lovette

SUPPLEMENTAL INFORMATION:

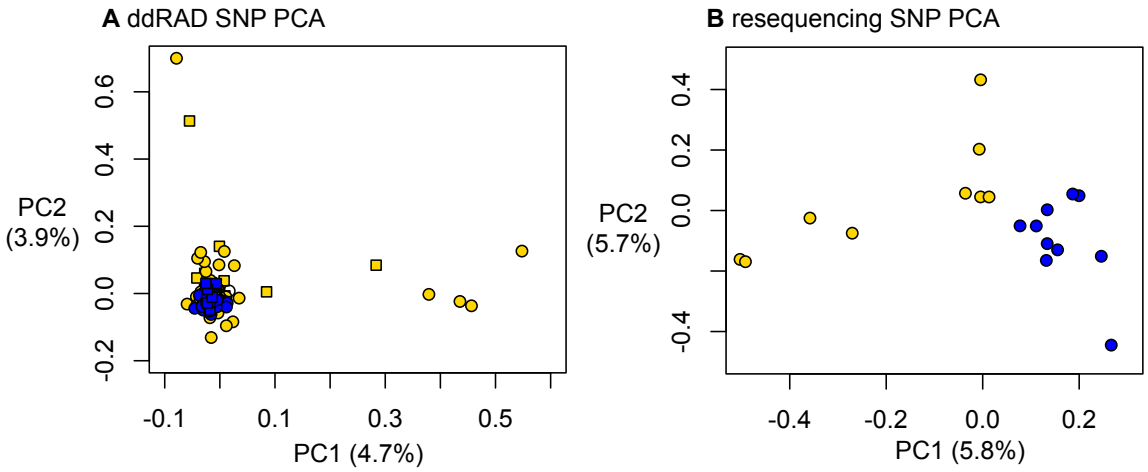


Figure S1. Principal components analysis of genome sequencing. Related to Figure 1. A) PCA of 16,103 SNPs derived from double-digest restriction site associated DNA sequencing (i.e. ddRAD; $n = 151$) of golden-winged and blue-winged warblers sampled across the range of both taxa. Yellow points are individuals classified by phenotype as golden-winged warblers; blue points are individuals classified as blue-winged warblers. Square symbols are birds sampled in the far allopatric populations (i.e. Missouri and Manitoba, see Figure 1) and circles are birds sampled from sites of range overlap. Most points are overlapping due to the extremely limited genetic structure and low resolution in this dataset across all PC axes. B) PCA of over 11 million SNPs derived from whole-genome resequencing from 20 individuals sampled in New York. This dataset fully resolves individuals of differing phenotypes. However, the differences between the two clusters are primarily being driven by the small proportion of SNPs occurring within the six divergent scaffolds discussed in the text.

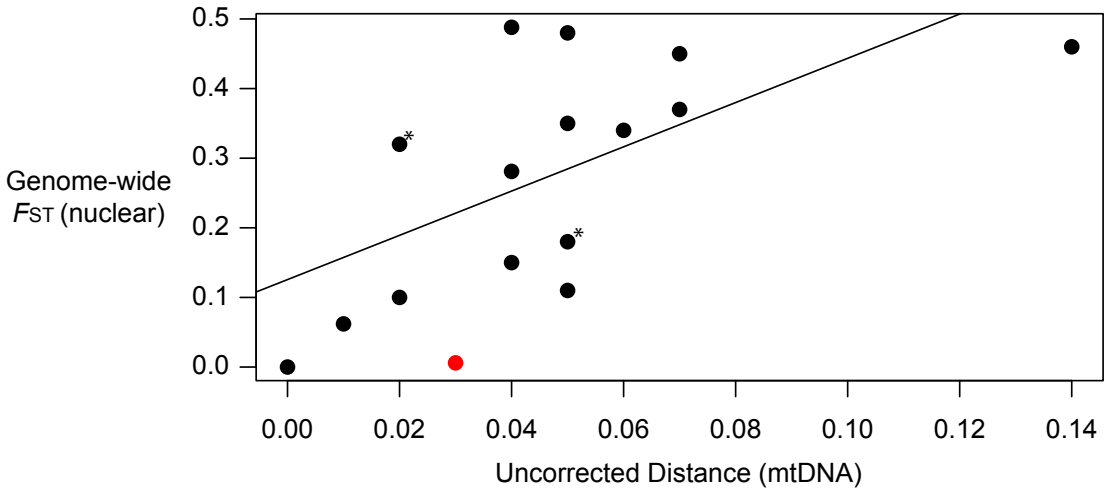


Figure S2. Comparison between mtDNA and genome-wide $F_{ST.}$ Related to Figure 2. Uncorrected mitochondrial DNA distance and genome-wide F_{ST} from 15 avian species-pair comparisons where there is a presumed history of hybridization (see Table S4 for information on each pair). F_{ST} estimates were taken from published studies that used comparable methods to the current study (i.e. used either reduced-representation or whole genome sequencing). The red point indicates the golden-winged and blue-winged warbler comparison (this study). In most cases *cytochrome b* (*cyt b*) sequences were used (* indicates where NADH sequences were analyzed instead). Using the calibrated avian molecular clock for *cyt b*, 3% difference roughly corresponds to 1.5 million years of isolation and divergence [S1]. The nuclear estimates of F_{ST} and mtDNA distance for golden-winged and blue-winged warblers are very low given this presumably long period of isolation. The discordance between the mitochondrial and nuclear genomes is consistent with substantial gene flow following isolation and secondary contact, mixing the nuclear but not the maternally inherited, non-recombining mitochondrial genome.

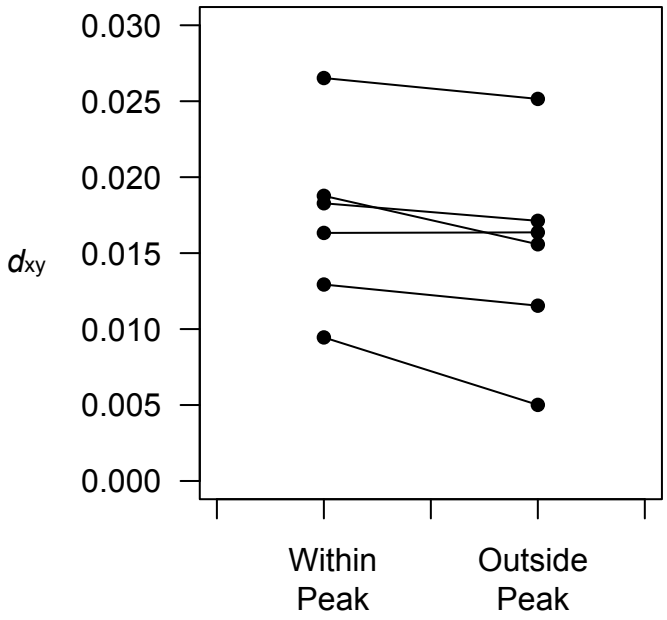


Figure S3. Estimates of d_{XY} within and outside peaks, along the same scaffolds. Related to Figure 2. Estimates of d_{XY} are on average higher within divergence peaks, although this difference is not statistically significant ($t = 0.53$, df = 9.8, $P = 0.6$). However, this pattern contrasts with most other studies that have tested for a relationship between d_{XY} and islands of divergence: most studies show strong d_{XY} valleys in regions of elevated F_{ST} [S2, S3], not something observed between golden-winged and blue-winged warblers.

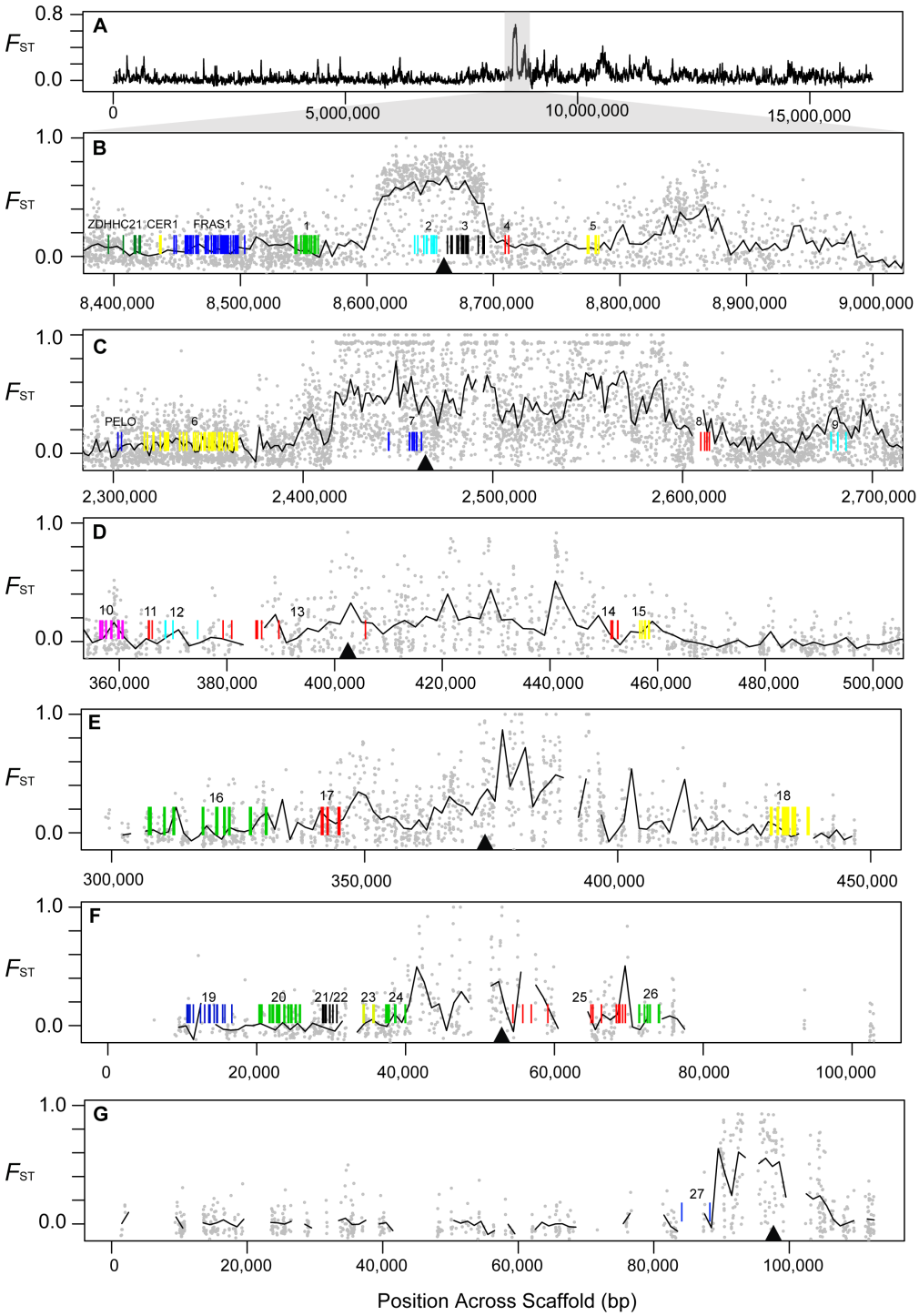


Figure S4. Patterns of F_{ST} differentiation for scaffolds showing strong divergence peaks. Related to Figure 2. Divergent scaffolds that align to the zebra finch Z chromosome (A-B, warbler scaffold 24; C, warbler scaffold 38), chromosome 4A (D, warbler scaffold 120), chromosome 20 (E, warbler scaffold 299), chromosome 20 (F, warbler scaffold 563), and chromosome 25 (G; warbler scaffold 653). Vertical lines indicate exons in coding regions identified by MAKER. Gene numbers correspond to the genes in Table S2. For genes outside of the 40k region surrounding divergent regions, gene abbreviations are given. Filled triangles indicate the locations of the SNPs used for RFLP genotyping. Gray points indicate per-locus F_{ST}-estimate and the solid lines are non-overlapping sliding window averages of F_{ST}.

Phenotype	GG Frequency	GB Frequency	BB Frequency	G allele Frequency	Sample Size
			Locus Z[1]		
Blue-winged	0.05	0.35	0.60	0.22	110
Brewster's	0.44	0.44	0.11	0.67	27
Golden-winged	0.66	0.30	0.04	0.81	184
			Locus Z[2]*-FST*		
Blue-winged	0.00	0.14	0.86	0.07	105
Brewster's	0.48	0.41	0.11	0.69	27
Golden-winged	0.92	0.07	0.01	0.95	187
			Locus 4A-EDA		
Blue-winged	0.08	0.32	0.61	0.24	104
Brewster's	0.48	0.41	0.11	0.69	27
Golden-winged	0.87	0.12	0.01	0.93	173
			Locus 20-ASIP		
Blue-winged	0.00	0.21	0.79	0.11	109
Brewster's	0.00	0.96	0.04	0.48	25
Golden-winged	1.00	0.00	0.00	1.00	186
			Locus 24-BCO2		
Blue-winged	0.03	0.13	0.85	0.09	118
Brewster's	0.39	0.54	0.07	0.66	28
Golden-winged	0.88	0.12	0.00	0.94	190
			Locus 25		
Blue-winged	0.04	0.14	0.82	0.11	114
Brewster's	0.56	0.44	0.00	0.78	27
Golden-winged	0.89	0.11	0.00	0.95	188

Table S1. Genotype frequencies, allele frequencies, and sample sizes for the RFLP assay. Related to Figure 2. Results from across all samples (i.e. "NY State" and "Range-wide" sample sets combined) across six divergent regions between golden-winged, blue-winged, and Brewster's warblers (the putative F1 hybrid phenotype). The locus name refers to the chromosome in the zebra finch that the scaffold aligned to and, in most cases, the candidate pigmentation gene that occurs near the divergence peak.

Zebra Finch - Warbler Scaffold	MAKER Evidence	Figure S4 #	Start Pos.	End Pos.	Protein Association	Other names for Protein
Z^1 - 24	c2g, Bx, p2g, tBx	1	8543300	8562890	ENSTGUP00000004708	tetratricopeptide repeat domain 39B
Z^1 - 24	c2g, Bx, p2g, tBx	2	8645031	8655212	ENSTGUP00000004679	small nuclear RNA activating complex, polypeptide 3, 50kDa
Z^1 - 24	c2g, Bx, p2g, tBx	3	8663669	8692802	ENSTGUP00000004673	PC4 and SFRS1 interacting protein 1
Z^1 - 24	c2g, Bx, p2g, tBx	4	8709399	8712142	ENSTGUP00000004643	Uncharacterized
Z^1 - 24	Bx, p2g	5	8774221	8781602	ENSTGUP00000016051	Uncharacterized
Z^2 - 38	c2g, Bx, p2g, tBx	6	2324762	2364413	ENSTGUP00000002480	integrin, alpha 1
Z^2 - 38	c2g, Bx, p2g, tBx	7	2454883	2458691	ENSTGUP00000002489	molybdenum cofactor synthesis 2
Z^2 - 38	**c2g, Bx, p2g, tBx**	**8**	**2609622**	**2614331**	**ENSTGUP00000002505**	**follistatin**
Z^2 - 38	Bx, p2g	9	2678171	2686212	ENSTGUP00000002512	NADH dehydrogenase (ubiquinone)
4A - 120	Bx, p2g	10	356382	360700	ENSTGUP00000002752	immunoglobulin (CD79A) binding protein 1
4A - 120	**c2g, Bx, tBx**	**11**	**365173**	**366134**	**ENSTGUP00000002757**	**Protein Wnt**
4A - 120	c2g, Bx, p2g, tBx	12	368422	374569	ENSTGUP00000002758	Uncharacterized
4A - 120	**c2g, tBx**	**13**	**379289**	**386680**	**ENSTGUP00000002770**	**ectodysplasin A**
4A - 120	Bx, p2g	14	451137	451610	ENSTGUP00000002772	Uncharacterized
4A - 120	Bx, p2g	15	456697	457758	ENSTGUP00000002775	family with sequence similarity 155, member B
20 - 299	c2g, Bx, p2g, tBx	16	307487	330612	ENSTGUP00000003784	adenosylhomocysteinase
20 - 299	**c2g, Bx, p2g, tBx**	**17**	**341540**	**345031**	**ENSTGUP00000003790**	**agouti signalling protein**
20 - 299	c2g, Bx, p2g, tBx	18	430277	437731	ENSTGUP00000003818	eukaryote initiation factor 2 beta-like
24 - 563	c2g, Bx, p2g, tBx	19	10632	16747	ENSTGUP00000000321	Uncharacterized
24 - 563	c2g, Bx, p2g, tBx	20	19803	25874	ENSTGUP00000000318	dihydrolipoamide S-acetyltransferase
24 - 563	c2g, Bx, p2g, tBx	21	28857	30766	ENSTGUP00000000308	chromosome 11 open reading frame 57
24 - 563	c2g, Bx, p2g, tBx	22	28878	30766	ENSTGUP00000000311	chromosome 11 open reading frame 57
24 - 563	c2g, Bx, p2g	23	34306	35789	ENSTGUP00000000305	succinate dehydrogenase complex subunit D-like
24 - 563	c2g, Bx, p2g, tBx	24	37343	38699	ENSTGUP00000000303	interleukin 18
24 - 563	**c2g, Bx, p2g, tBx**	**25**	**54481**	**69661**	**ENSTGUP00000000300**	**beta-carotene oxygenase 2**
24 - 563	Bx, p2g	26	71347	74121	ENSTGUP00000000296	6-pyruvoyl tetrahydrobiopterin synthase
25 - 653	c2g, Bx, p2g, tBx	27	84131	88282	ENSTGUP00000004415	Uncharacterized, Ensembl protein family ID: PTHR31203

Table S2. Evidence-based annotation information from MAKER. Related to Figure 2. Gene information shown for genes occurring within 40Kb of divergent regions between golden-winged and blue-winged warblers. Evidence sources: c2g = cdna2genome, Bx = blastx, p2g = protein2genome, and tBx = tblastx. We only included genes that were found on the appropriate zebra finch chromosome and that had at least two evidence sources from MAKER. Rows in bold indicate candidate genes involved in feather pigmentation or development.

	Model A	Model B	Model C1 ($T_m < T_s$)	Model C2 ($T_m < 300$)
Likelihood	**-378499**	-379457	-483767	-819984
N_0	183328	177825	180777	121003
N_B	183198	234962	149129	138093.0
N_G	189476	144661	256750	323774
T_s	4813653	3616473	7577825	430340
T_m	-	-	5011999	291
m_{BG}	6.763e-05	5.256e-05	6.011e-05	0.00764
m_{GB}	6.763e-05	9.670e-05	7.622e-05	0.00735

Table S3. Output of modeling with *∂a∂i*. Related to Figure 4. Model A yielded the highest overall likelihood, which infers similar effective population sizes and a constant symmetric migration rate. The strong fit between the 2D frequency spectra of this model and the data is shown in Figure 4 in the main text. Both variants of Model C that we tested resulted in much lower likelihoods. Importantly, the migration rates inferred by Model C1 are still of the same order of magnitude as those inferred by Models A and B, and the time at which gene flow started in Model C1 is also very similar to that inferred by Models A and B (where it coincides with the split-time). Given the high rates of continuous migration inferred by all of these models, *∂a∂i* has limited power to accurately time the population split. Therefore, estimates of T_s in all models should be treated with caution. Model C2, in which migration has occurred only recently, yields the lowest overall likelihood of all models tested and consistently places the inferred migration time T_m at the maximum allowed parameter bound ($T_m < 300$), suggesting that migration is not just a recent phenomenon.

Species 1	Accession	Species 2	Accession	Nuclear F_{ST} Reference	mtDNA_gene	mtDNA%	F_{ST}
Dendrocincla fulinosa atrirostris	KR781142	*D. f. rufoolivacea*	KR781141	[S4]	cyt b	0.05	0.35
Xiphorhynchus elegans elegans	KR781193	*X. spixii*	KR781195	[S4]	cyt b	0.05	0.48
Glyphorynchus spirurus inornatus	KR781151	*G. s. paraensis*	KR781147	[S4]	cyt b	0.06	0.34
Hypocnemis ochrogyna	KR781156	*H. striata striata*	KR781159	[S4]	cyt b	0.14	0.46
Rhegmatorhina hoffmannsi	KR706048	*R. gymnops*	KR781165	[S4]	cyt b	0.07	0.37
Lepidothrix nattereri	KR781218	*L. iris eucephala*	KR781213	[S4]	cyt b	0.04	0.15
Willisornis poecilinotus griseiventris	KR781191	*W. vidua nigrigula*	KR781181	[S4]	cyt b	0.07	0.45
Poecile atricapillus	KF134282	*P. carolinensis*	KF13438	[S5]	cyt b	0.05	0.11
Carduelis hornemanni	U83201	*C. flammea*	DQ192028	[S6]	cyt b	0	0
Ficedula hypoleuca	HM633303	*F. albicolus*	DQ674491	[S2]	cyt b	0.04	0.281
Carellina pussilla (western)	AF499572	*Carellina pussilla* (eastern)	AF499592	[S7]	cyt b	0.04	0.488
Manacus candei	KF228540	*Manacus vitellinus*	KF228542	[S8]	NADH	0.05	0.180
Corvus corone	HE805700	*Corvus cornix*	NC_024698	[S9]	cyt b	0.01	0.062
Catharus ustulatus swasinsoni	EU619775	*C. u. ustulatus*	EU619779	[S3]	cyt b	0.02	0.100
Malurus melanocephalus	FJ241901	*M. melanocephalus*	FJ241906	[S10]	NADH	0.02	0.320
Vermivora chrysoptera	AY216819	*V. cyanoptera*	GU932369	This paper	cyt b	0.03	0.006

Table S4. Details of studies of related avian taxa with evidence of hybridization and comparable genome-wide data. **Related to Figure 2.** In most cases we used uncorrected distance in *cytochrome b*, which has a well-calibrated molecular clock in birds. While there is variation in studies in terms of filters and number of loci, as expected there is a strong positive relation between mtDNA divergence and nuclear genome-wide F_{ST} (Figure S2). Golden-winged and blue-winged warblers, however, show much less nuclear divergence given the deep split in mtDNA.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES:

Sample collection: Sampling occurred over two time periods: 1) samples collected in May and June 2015 at breeding sites across New York state, in populations with a mix of phenotypically typical golden-winged warblers, blue-winged warblers, and phenotypically admixed birds (n = 166); and 2) samples collected prior to 2015 from breeding sites across the broader range of both taxa as part of the "Golden-winged Warbler Genetic Atlas" project, which includes regions of admixture and allopatry (n = 180). We refer to these datasets as "NY State" samples and "Range-wide" samples respectively, recognizing that some samples from the "Range-wide" set include birds captured in New York. Birds from the "Range-wide" set include warblers captured from Manitoba, Minnesota, Michigan, Wisconsin, West Virginia, Tennessee, and New York (see sampling information deposited to Data Dryad: http://dx.doi.org/10.5061/dryad.kb610). The primary differences in the data associated with these sampling efforts are: 1) the focus on golden-winged warblers in the "Range-wide" dataset and 2) detailed plumage information taken from the "NY State" birds. Prior to 2015, samples were categorically classified in the hand as: golden-winged warbler, blue-winged warbler, Brewster's warbler, Lawrence's warbler, or intermediate (i.e. some evidence of admixed plumage patterns). Brewster's warbler and Lawrence's warbler are the traditionally "named" hybrid phenotypes of these crosses [S1], but many phenotypic hybrids have other combinations of traits. In 2015 we took detailed photographs and plumage scores of each bird in the New York sample (see below). Across all sampling efforts, mostly territorial males were captured during May and June using song playback, although we include eight females from Michigan in the "Range-wide" sample. From each bird we collected blood samples from the brachial vein and stored them in Queen's lysis buffer, as described in [S11]. For the plumage scores, we followed the scoring criteria developed by [S12], with the addition of three characters: mask, eye-line, and moustache color (see plumage scoring criteria at DataDryad: http://dx.doi.org/10.5061/dryad.kb610). Birds were scored for the 11 characters in the hand as well as from digital photos, and all scores were quantified by SAT and DPLT blind to a bird's genotype data.

ddRAD-Seq library preparation and sequencing: We extracted genomic DNA from each sample using Qiagen® DNeasy kits (tissue protocol; Qiagen, Valencia, CA). We determined the final concentration of each extraction using Qubit Fluorometric Calibration (QFC; Invitrogen, Carlsbad, California). Blood samples and DNA extractions are archived at the Cornell Lab of Ornithology (Ithaca, NY).

We prepared ddRAD-Seq libraries using a modified version of the protocol outlined in [S13] and in [S14]. Following a standardizing dilution (all genomic DNA ~25ng/µl), we plated the samples and digested each with the restriction enzymes *SbfI* and *MspI* while ligating P1 (barcode) and P2 adaptor primers using 20 unique barcodes for each of three subsequent index groups (a total of 60 unique identifiers/plate) in three plates (a total of 180 samples). Individuals were randomized across plates. Each digestion reaction contained 300ng genomic DNA, 3µl 10x CutSmart buffer, 1µl of 250nM P1, 1µl of 25µM P2, 3ul 10mM ATP, 0.75µl (15U) each of 20 U/µl *SbfI-HF* and *MspI*, 0.75µl of 400U/µl T4 DNA ligase, and water to a total of 30µl. Next, samples were incubated at 37°C for 30min followed by one hour at 20°C, pooled in groups of 20, and cleaned with 1.5x volumes of AMPure beads (Beckman Coulter Inc) and two washes of 70% ethanol. The pooled samples were then eluted into 30µl of Qiagen EB buffer and quantified using QFC. We then size selected fragments of between 400 and 700 bp using Blue Pippin (Sage Science, MA, USA). For each of the three index groups within each plate (Illumina index primers 1, 6, 12), we set up six replicate PCRs containing 20ng DNA, 12.5µl 2x Phusion MM, 1.25µl of 5µM P1, 1.25µl of 5µM Index primer, and water to a total reaction volume of 25µl. The PCR temperature profile included a 30 second incubation at 98°C followed by 16 cycles of 98°C for 5 seconds, 60°C for 25 seconds, and 72°C for 10 seconds with a final extension step of 5 minutes at 72°C. The 6 replicate PCRs were pooled within each index group and visualized 4µl on a 1% agarose gel. Final elutions of each index group were analyzed using QFC and an Agilent Bioanalyzer 2100 (Agilent, Santa Clara, California). Finally, we diluted each index group to 2nM, combined all three in equal proportions for each plate of samples, and sequenced the three libraries on three lanes of an Illumina HiSeq 2000 (150 base pair (bp), single-end) at the Cornell University Life Sciences Core Laboratories Center (Ithaca, NY).

ddRAD-Seq locus assembly and SNP calling: With Trimmomatic [S15], we trimmed leading and trailing bases below a quality score of 3. We also used a 4bp sliding window to remove sequences with an average base quality across the window of less than 10. We then discarded reads with a minimum length of less than 30bp. We demultiplexed sequencing reads using the barcode-splitting program Sabre [S16]. We

allowed for one mismatch in the barcode plus enzyme cut-site sequence (the variable length barcodes we used differed by a minimum of 3bp). We used BOWTIE2 [S17] to map each of the individual reads to a build of the zebra finch genome [S18]. For this we used the "very sensitive local" set of alignment pre-sets. For SNP discovery and variant calling, we used the UnifiedGenotyper in GATK [S19] and followed the set of GATK "best practices" as a guideline. We removed possible variants that had "quality by depth" (QD) of < 2 and "mapping quality" (MQ) of < 30 (the full filtering expression we employed was: QD<2.0, FS>40.0, MQ<30.0, HaplotypeScore>12.0, MappingQualityRankSum<-12.5, ReadPosRankSum<-8.0). We applied additional filters using the program VCFtools [S20]. First, we coded genotypes with a Phred-scaled quality lower than 20 as missing data, which corresponds to a genotyping accuracy of at least 99%. We then removed the eight female warblers, and also 21 individuals with high levels of missing data. We excluded loci with more than 60% missing data and/or a minor allele frequency of less than 2%, resulting in 16,103 SNPs. To visualize the data and test for population structure we used a principal components analysis (Figure S1) using the SNPRelate package [S21] in R [S22].

Yellow-rumped warbler genome library construction: We sequenced the genome of a male yellow-rumped warbler of the eastern "myrtle warbler" subspecies (*Setophaga coronata coronata*) captured at Slave Lake, Alberta in May 2005 (sample 05-119 in [S23]). DNA was extracted from blood stored in Queens lysis buffer, using a standard phenol-chloroform procedure. A paired-end (PE) library with insert size 400bp was constructed using an Illumina TruSeq kit, and sequenced on three Illumina HiSeq 2500 lanes PEx100 cycles at the Genomic Technologies Facility of the University of Lausanne. Two mate-pair (MP) libraries, with 3 kbp and 8 kbp insert sizes, were constructed an Illumina Nextera library preparation kit. Each MP library was sequenced on a single HiSeq 2500 lane PEx100 cycles by the Weill Cornell Medical College Core Genomics facility.

Data pre-processing: We filtered raw Illumina reads using several publicly available tools. Reads failing the Illumina chastity filter were removed with a custom shell command, and PCR duplicates were removed with FastUniq [S24]. For PE sequences, we collapsed overlapping read pairs and removed adapters with PEAR [S25]. For MP sequences, we used NxTrim [S26] to separate true MP reads from contaminating PE reads, and removed adapter and linker sequences. Reads for which PE/MP status could not be determined are very likely to be MP reads [S26], and were therefore retained. We then trimmed low-quality bases (average phred-scaled quality <20 in a window of 10% read length) from all reads using Sickle [S27] retaining only reads >40 bp (PE) or >20 bp (MP) after quality trimming. Based on a 1.28 Gb genome [S28] the estimated genome coverage was: 96.3X for the 400bp fragment library; 14.0X for the 3kb library; and 8.4X for the 8kb library.

Genome assembly: We used a hybrid reference-based and *de novo* approach. We first mapped reads to the white-throated sparrow (*Zonotrichia albicollis*) genome to obtain a warbler consensus sequence. Separately, we performed de novo assembly of unmapped reads. Reference-based and *de novo* sequences were combined, and paired-end and mate-pair reads were used to build scaffolds. We used mate-pair reads to break this sequence at sites of potential assembly errors or rearrangements between the sparrow and warbler genomes. Scaffolds were rebuilt using more stringent parameter settings, and gaps filled.

We downloaded the white-throated sparrow genome from NCBI (Accession # PRJNA217032) and mapped the yellow-rumped warblers paired-end reads to this reference using BWA-mem 0.7.10 [S29]. We used Samtools 0.1.19 [S30] to generate a warbler consensus sequence based on the sparrow reference, incorporating an unpublished patch to the vcf2fq.pl script (http://sourceforge.net/p/vcftools/feature-requests/19/) to account for indels in the consensus sequence. Positions with sequence depth <10 were hard-masked. Positions near all indels identified by Samtools (within 4 bp of all indels 1-4 bp in length, within 8bp of all indels of 5-8 bp, etc.) were also hard-masked. Scaffolds with >50% uncalled bases were removed. Leading and trailing "N" characters were removed from each scaffold. We scanned the leading and trailing 500bp of each scaffold, removing a scaffold end if it contained >400 "N" characters, and removing any additional leading or trailing "N"s. This procedure was repeated for seven iterations, until no additional scaffold end trimming occurred.

Mapping, assembly of unmapped reads: We mapped single-end reads to the warbler consensus genome with BWA-mem. Reads that did not map were assembled into contigs using SOAPdenovo 2.04 [S31], with k values of 39, 47, 51, and 55. The k=51 assembly had the highest contig N50 after removal of contigs

<200bp, and was retained for possible incorporation into the reference-based assembly. All contigs were searched against the consensus genome with Blat 3.4 [S32]; contigs with >100 bp matching the consensus genome at >97% identity were discarded as redundant.

Scaffolding, breaking weak scaffolds, rescaffolding, filling gaps: We combined reference-based scaffolds and non-redundant *de novo* contigs, and scaffolded using SSPACE 1.0 [S33]. We used all three libraries (400 bp, 3 kbp, and 8 kbp inserts), requiring at least 3 links to join sequences. We split the resulting scaffolds at locations with insufficient support, using REAPR [S34] and the 8 kbp mate-pair data. The resulting fragments were re-scaffolded using SSPACE, requiring at least 5 links to join sequences. All heterozygous bases were converted to homozygous using seqtk randbase (https://github.com/lh3/seqtk). Finally, we used GapFiller [S35] to close gaps created by the scaffolding process and by masking positions near indels. Assembly and raw Illumina reads can be found at NCBI Bioproject: PRJNA325157.

Genome annotation: We annotated the warbler assembly with the MAKER pipeline 2.31.7 [S36], using the zebra finch (*Taeniopygia guttata*) Ensembl protein and cdna databases (downloaded October 27, 2015 from fhttp://useast.ensembl.org/Taeniopygia_guttata/Info/Index?redirect=no) to create gene models. Genes were predicted using SNAP, which was trained for the warbler using MAKER iteratively, as described in [54]. This produced 15,955 protein-coding gene annotations.

Genome resequencing library preparation*:* We used 24 blood samples from golden-winged and blue-winged warblers in the "NY State" sample set for whole genome resequencing. We included 10 individuals from each taxon chosen to represent the ends of the phenotypic spectrum (golden-winged sum plumage score: range = 0-2, mean = 0.9; blue-winged sum plumage score: range = 36-38, mean = 36.8). We also included four hybrid individuals: two "Brewster's" phenotypes (sum plumage score = 23, 22), and two backcrossed hybrid birds (sum phenotype score = 23, 30). Given our focus on identifying regions of genomic divergence between the distinct parental phenotypes, we did not use these four hybrid individuals in subsequent analysis although they are retained in the dataset. We used the Illumina TruSEQ PCR-Free Library Preparation kit to generate sequencing libraries. We followed the protocol for generating libraries of 350bp in size. We sequenced these 24 individuals on a single lane of a NextSEQ500 PEx150bp. Raw Illumina reads are available at NCBI SRA under BioProject ID PRJNA325126.

Data pre-processing, read mapping and variant calling: We first used AdapterRemoval [S37] to collapse overlapping paired reads, trim sequences of "Ns" along the 5' and 3' ends of reads, and remove sequences shorter than 20bp. Trimming and collapsing resulted in 441 million paired and collapsed reads across all of the samples. For the twenty non-hybrid individuals included in the analysis, this resulted in an average of 17.8 million reads per individual (min = 14.7 million reads, max = 20.9 million reads, standard deviation = 1.7 million reads).

We used BOWTIE2 [S17] to map each of the individual reads to the assembly of the warbler genome (see above). For this we used the "very sensitive local" set of alignment pre-sets. For SNP discovery and variant calling, we used the UnifiedGenotyper in GATK [S19]. We removed possible variants that had "quality by depth" (QD) of < 2 and "mapping quality" (MQ) of < 30 (the full filtering expression we employed was: QD<2.0, FS>40.0, MQ<30.0, HaplotypeScore>12.0, MappingQualityRankSum<-12.5,ReadPosRankSum<-8.0). Variant confidence is a measure of sequencing depth at a given variant site; mapping quality refers to the root-mean-square of the mapping confidence (from BOWTIE2) of reads across all samples.

We applied additional filters using the program VCFtools [S20]. First, we coded genotypes with a Phred-scaled quality lower than 8 as missing data. This is a permissive filter, necessary given our low-coverage data, which allows for genotype calls with as few as three reads at a given site for an individual. However, all of the subsequent analysis with this data was performed using population comparisons, therefore incorrect genotype calls should not bias the results in any direction. Moreover, this is one of the reasons we validated our results from the re-sequencing analysis with targeted RFLP genotyping (see below), where we could be much more confident in individual genotype calls. We excluded loci with more than 30% missing data and used a minor allele frequency (MAF) filter of 7%. This MAF filter means that, out of 48 chromosomes, a SNP would need to be observed on over three chromosomes to be included in subsequent analysis, a reasonable threshold set to exclude sequencing errors. This resulted in 11,432,965 SNPs. We used VCFtools [S20] to estimate F_{ST} between the ten golden-winged and ten blue-winged

warblers. We used patterns of divergence in F_{ST} estimates to identify six regions of the genome that were highly differentiated. To ensure that these highly differentiated markers were not an artifact generated from a combination of small sample size and a large number of SNPs we conducted 10 randomization tests. Across these tests the average number of SNPs with F_{ST} estimates >0.8, >0.9, and 1 were 13, 1.3, and 0.1 respectively. When the data are grouped by phenotype the average number of SNPs with F_{ST} estimates >0.8, >0.9, and 1 are 613, 362, and 74 respectively. To visualize the data we used a principal components analysis (Figure S1) using the SNPRelate package [S21] in R [S22]. As PCA can be sensitive to individual level missing data, we trimmed the dataset to 8,741,850 SNPs, removing those loci with the highest amount of missing data.

Identifying candidate pigmentation / development genes and the promoters of *ASIP*: To identify potential candidate genes involved in feather pigmentation / development, we performed a search including all of characterized genes within 40kb of six divergent F_{ST} regions. We used ontology information from the zebra finch Ensemble database as well as compared these genes to previously identified candidate gene lists (e.g. [S38-S42]). To identify the presumed promoters upstream of one of the candidate genes, *agouti signaling protein* (*ASIP*), we aligned non-coding *ASIP* exon sequences that have been identified in the domestic chicken (*Gallus gallus*). Chicken feather follicles express at least seven kinds of *ASIP* mRNA variants using three promoters [S39, S43]. To represent the range of sequences from these splice variants we aligned class 1d, class 1e, class 2, and class 3 mRNA to the warbler scaffold 299 (Genbank accession numbers: AB518064, AB518065, AB518066, and AB518067, respectively; [S39] and [S43]) using the pairwise alignment tool in Geneious (global alignment; 51% similarity; gap open penalty = 10; gap extension penalty = 3; [S44]). The distal class 1 promoter has been suggested to be the avian homolog of the mammal ventral-specific promoter, which is involved in countershading and sexual dichromatism in chickens [S39]. If the arrangement of promoters is conserved between mammals and birds, then the avian class 3 promoter – the most proximal promoter – may be the analogue to the mammalian hair-cycle *agouti* promoter (Figure 2h). The hair-cycle promoter has been characterized in mammals and is a good candidate for the plumage differences between blue-winged and golden-winged warblers. It is notable that its predicted location falls under a secondary divergence peak in the comparison between the warblers. However, its expression pattern has not been explicitly characterized in birds and is a candidate for future investigations.

***D_{xy}* estimates:** There are two different interpretations of very low estimates of genomic F_{ST}: 1) divergence is extremely recent and low F_{ST} values represent retained ancestral polymorphism between groups, while regions of elevated F_{ST} have been influenced mostly by selection and exhibit elevated F_{ST} that stands out against the genomic background; 2) regions of reduced F_{ST} have been influenced by a history of gene flow, while the regions of elevated divergence are less prone to introgression. Estimates of absolute divergence, such as d_{XY} (in contrast to relative measures, such as F_{ST}, that are influenced by linked selection and reduced diversity), have been suggested as an important statistic that may be able to distinguish between these alternatives: d_{XY} is less sensitive to selection, but is reduced in the face of gene flow [S45]. Therefore, if gene flow is responsible for shaping patterns of F_{ST}, we would also expect to observe elevated d_{XY} within divergence peaks. To date, however, few studies have found evidence of elevated d_{XY} in divergence islands [58]. In fact, many studies have found reduced d_{XY} compared to the rest of the genome, a pattern that is not fully understood (e.g. [S2]).

To estimate d_{XY} we used ANGSD [S46] to first calculate the minor allele frequencies using the "doMajorMinor" function, assuming the pre-specified major allele as the reference. We then calculated d_{XY} using the formula: $d_{XY} = A_1*B_2 + A_2*B_1$, with A_1 and B_1 being the allele frequencies of alleles A and B in population 1, and A_2 and B_2 being the frequencies of the same alleles in population 2. For the six scaffolds where we observed divergence peaks, we averaged per-SNP estimates of d_{XY} within the divergent region and outside the peaks (Figure S3).

RFLP genotyping: For the six divergent regions between golden-winged and blue-winged warblers, we validated our results with RFLP assays. We used SamTools [S30] to generate consensus sequences around SNPs with F_{ST} estimates >0.9 using the "mpileup" command. We used Primer3 [S47] to generate primers surrounding these regions that would result in single restriction enzyme cuts that differed between the golden-winged and blue-winged variants. It is important to note that we developed enzyme RFLPs prior to genome annotation, thus we were blind to the SNP's positions relative to subsequent candidate genes that

occur near the divergence peaks (see above). We present the SNP information, primer sequences, cut site information, and product size, in the DataDryad repository: http://dx.doi.org/10.5061/dryad.kb610. We performed 10μL PCR reactions with final concentrations: 1X $MgCl_2$-free Reaction buffer, 1.5mM $MgCl_2$, 0.2μM forward primer, 0.2μM reverse primer, 0.2μM dNTPs, 0.5U/μL Jumpstart *Taq* polymerase. We used the same thermocycling profile for each primer set: initial denaturing at 94ºC for 3 minutes, followed by 30 cycles of: 94ºC for 30 seconds, 55ºC for 30 seconds, and 72ºC for 1 minute; this was followed by a 5 minute extension of 72ºC. We digested each product in a 6μL digestion reaction with: 3μL of PCR product, 0.2μL of restriction enzyme, 1μL of the buffer identified by NEB as having 100% activity (either CutSmart or Buffer 3.1). We visualized the resulting fragments on a 2% agarose gel.

We calculated the genotype frequencies for each of the coarse phenotype classes (i.e. blue-winged, golden-winged, and Brewster's). We show the frequencies for "NY State" samples in Figure 2b; genotype frequencies for the entire sample (i.e. "NY State" and "Range-wide" sampling combined) can be found in Table S1. We used the sum of the genotype score for each of the six loci individuals to compare against the sum plumage score across all 11 plumage characters. Genotype and plumage information can be found at the DataDryad repository: http://dx.doi.org/10.5061/dryad.kb610.

Site Frequency Spectrum (SFS) and *∂a∂i*: Average nucleotide diversity per site is very similar in our samples of blue-winged warblers (π=0.00461) and golden-winged warblers (π=0.00460). This significantly limits the space of possible demographic scenarios without substantial gene flow between the two taxa. We tested several demographic models in order to estimate the timing of hybridization onset between golden-winged and blue-winged warblers. In particular, we were interested in the roll that anthropogenic habitat change may or may not have had on hybridization and admixture between these birds (i.e. whether hybridization is very recent, or is older and has been ongoing). We first imputed phase information and inferred missing data with BEAGLE [S48]. For input, we used the unfiltered variant calls from GATK and applied a lower minimum allele frequency cutoff of 3%. We restricted our analysis to those scaffolds that were longer than 1Mb (n = 160 scaffolds; this corresponds to >70% of the entire assembly). We used BEAGLE to impute the phases of each taxon separately, running the program with the default settings for calculating phase information from genotype likelihoods. Caution must be taken with computational phasing, particularly for low coverage data [S48]. However, our subsequent analyses with the SFS did not incorporate this phasing information; it only relied on the inferred missing data output from BEAGLE.

We ran *∂a∂i* [S49] on the 2D frequency spectrum of all SNPs in our data set (Figure 4), excluding only SNPs that showed high F_{ST} values and were located on warbler scaffolds 24 (ZF-Z^1), 38 (ZF-Z^2), and 120 (ZF-4A). Sample frequencies were estimated from the data and folded to prevent biases from misspecification of ancestral states. A total of 19,640,841 SNPs were included in the frequency spectra, covering 819 Mbp of genomic sequence in our scaffolds.

We tested three different demographic models of increasing complexity (Figure 4): Model A specifies an ancestral population of constant size N_0 that at time T_s split into two subpopulations of constant size N_B and N_G, evolving into contemporary blue-winged and golden-winged warblers, respectively. Gene flow between the two subpopulations is modeled to occur at a constant symmetric migration rate m. Model B additionally allows for the possibility of asymmetric migration rates (m_{BG} and m_{GB}) between the two subpopulations. Model C allows for the possibility that there was no gene flow until time T_m, with constant asymmetric migration rates between the two subpopulations afterwards. We tested two variations of this last model: in the first variant (Model C1), the onset of gene-flow can occur at any time after the split, whereas in the second variant (Model C2) gene-flow is only allowed to start more recently than 300 generations ago. With Models C1 and C2 we were testing whether the data were consistent with hybridization as a recent phenomenon (the conventional wisdom in this system), possibly being facilitated by significant land conversion due to anthropogenic modification. This could be due to fire regimes implemented by Native Americans, or agricultural practices following European settlement of North America. To convert the maximum-likelihood estimates from *∂a∂i* into units of generations and individuals, we assumed a single-nucleotide mutation rate of $\mu = 4.42*10^{-9}$ per generation [S50] and a warbler generation time of two years.

Supplemental References:

S1. Gill, F.B. (2004). Blue-winged warblers (*Vermivora pinus*) versus golden-winged warblers (*V. chrysoptera*). The Auk *121*, 1014-1018.

S2. Burri, R., Nater, A., Kawakami, T., Mugal, C.F., Olason, P.I., Smeds, L., Suh, A., Dutoit, L., Bureš, S., Garamszegi, L.Z., *et al.* (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. Genome Research *25*,1656-1665.

S3. Delmore, K.E., Hübner, S., Kane, N.C., Schuster, R., Andrew, R.L., Câmara, F., Guigó, R., and Irwin, D.E. (2015). Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. Mol. Ecol. *24*, 1873-1888.

S4. Weir, J.T., Faccio, M.S., Pulid-Santacruz, P., Barrera-Guzmán, A.O., and Aleixo, A. (2015). Hybridization in headwater regions, and the role of rivers as drivers of speciation in Amazonian birds. Evolution *69*, 1823-1834.

S5. Taylor, S.A., Curry, R.L., White, T.A., Ferretti, V., and Lovette, I.J. (2014). Spatiotemporally consistent genomic signatures of reproductive isolation in a moving hybrid zone. Evolution *68*, 3066-3081.

S6. Mason, N.A., and Taylor, S.A. (2015). Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a Holarctic songbird. Mol. Ecol. *24*, 3009–3025.

S7. Ruegg, K., Anderson, E.C., Boone, J., Pouls, J., and Smith, T.B. (2014). A role for migration-linked genes and genomic islands in divergence of a songbird. Mol. Ecol. *23*, 4757-4769.

S8. Parchman, T.L., Gompert, Z., Braun, M.J., Brumfield, R.T., McDonald, D.B., Uy, J.A.C., Zhang, G., Jarvis, E.D., Schlinger, B.A., and Buerkle, C.A., (2013). The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. Mol. Ecol. *22*, 3304-3317.

S9. Poelstra, J.W., Vijay, N., Bossu, C.M., Lantz, H., Ryll, B., Müller, I., Baglione, V., Unneberg, P., Wikelski, M., Grabherr, M.G., and Wolf, J.B. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science *344*, 1410-1414.

S10. Baldassarre, D.T., White, T.A., Karubian, J., and Webster, M.S. (2014). Genomic and morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical introgression of a sexual signal. Evolution *68*, 2644-2657.

S11. Vallender, R., Van Wilgenburg, S.L., Bulluck, L.P., Roth, A., Canterbury, R., Larkin, J., Fowlds, R.M., and Lovette, I.J. (2009). Extensive rangewide mitochondrial introgression indicates substantial cryptic hybridization in the Golden-winged Warbler (*Vermivora chrysoptera*). Avian Cons. and Ecol. *4*, 4.

S12. Gill, F.B. (1980). Historical aspects of hybridization between blue-winged and golden-winged warblers. The Auk *97*, 1-18.

S13. Campagna, L., Gronau, I., Silveira, L.F., Siepel, A., and Lovette, I.J. (2015). Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation. *Molecular Ecology 24*, 4238–4251.

S14. Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., and Hoekstra, H.E. (2012). Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS ONE *7*, e37135.

S15. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

S16. Sabre (2014). https://github.com/najoshi/sabre.

S17. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods *9*, 357-359.

S18. Warren, W.C., Clayton, D.F., Ellegren, H., Arnold, A.P., Hillier, L.W., Künstner, A., Searle, S., White, S., Vilella, A.J., Fairley, S., *et al.* (2010). The genome of a songbird. Nature *464*, 757-762.

S19. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research *20*, 1297-1303.

S20. 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., *et al.* (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156-2158.
S21. Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., and Weir, B.D. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics *28*, 3326-3328.
S22. Team, R.D.C. (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org.
S23. Brelsford, A., and Irwin, D.E. (2009). Incipient speciation despite little assortative mating: the yellow-rumped warbler hybrid zone. Evolution *63*, 3050-3060.
S24. Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., Chen, J., and Chen, S. (2012). FastUniq: a fast de novo duplicates removal tool for paired short reads. PLoS ONE *7*, e52249.
S25. Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics *30*, 614-620.
S26. O'Connell, J., Schulz-Trieglaff, O., Carlson, E., Hims, M.M., Gormley, N.A., and Cox, A.J. (2015). NxTrim: optimized trimming of Illumina mate pair reads. Bioinformatics *31*, 2035-2037.
S27. Joshi, N.A., and Fass, J.N. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. https://github.com/najoshi/sickle.
S28. Andrews, C.B., Mackenzie, S.A., and Gregory, T.R. (2009). Genome size and wing parameters in passerine birds. Proc. R. Soc. B. *276*, 55-61.
S29. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v1.
S30. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genomes Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078-2079.
S31. Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., *et al.* (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience *1*, 18.
S32. Kent, W.J. (2002). BLAT—the BLAST-like alignment tool. Genome Res. *12*, 656-664.
S33. Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. Bioinformatics *27*, 578-579.
S34. Hunt, M., Kikuchi, T., Sanders, M., Newbold, C., Berriman, M., and Otto, T.D. (2013). REAPR: a universal tool for genome assembly evaluation. Genome Biol. *14*, R47.
S35. Boetzer, M., and Pirovano, W. (2012). Toward almost closed genomes with GapFiller. Genome Biol. *13*, R56.
S36. Cantarel, B.L., Korf, I., Robb, S.M., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A.S., and Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Research *18*, 188-196.
S37. Lindgreen, S. (2012). AdapterRemoval: easy cleaning of next-generation sequencing reads. BMC Res. Notes *5*, 337.
S38. Hoekstra, H.E. (2006). Genetics, development and evolution of adaptive pigmentation in vertebrates. Heredity *97*, 222-234.
S39. Oribe, E., Fukao, A., Yoshihara, C., Mendori, M., Rosal, K.G., Takahashi, S., and Takeuchi, S., (2012). Conserved distal promoter of the *agouti signalling protein* (ASIP) gene controls sexual dichromatism in chickens. Gen. and Comp. Endocrinology *177*, 231-237.
S40. Walsh, N., Dale, J., McGraw, K.J., Pointer, M.A., and Mundy, N.I. (2012). Candidate genes for carotenoid coloration in vertebrates and their expression profiles in the carotenoid-containing plumage and bill of a wild bird. Proc. Roy. Soc. B. *279*, 58-66.
S41. Eriksson, J., Larson, G., Gunnarsson, U., Bed'hom, B., Tixier-Boichard, M., Strömstedt, L., Wright, D., Jungerius, A., Vereijken, A., Randi, E., *et al.* (2008). Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. PLoS Genetics *4*, e1000010.
S42. Lehtonen, P.K., Laaksonen, T., Artemyev, A.V., Belskii, E., Berg, P.R., Both, C., Buggiotti, L., Bureš, S., Burgess, M.D., Bushuev, A.V., Krams, I., *et al.* (2012). Candidate genes for colour and vision exhibit signals of selection across the pied flycatcher (*Ficedula hypoleuca*) breeding range. Heredity *108*, 431-440.
S43 Yoshihara, C., Fukao, A., Ando, K., Tashiro, Y., Taniuchi, S., Takahashi, S., and Takeuchi, S. (2012).

Elaborate color patterns of individual chicken feathers may be formed by the *agouti signaling protein*. Gen. and Comp. Endo. *175*, 495-499.

S44. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., *et al.* (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647-1649.

S45. Cruickshank, T.E., and Hahn, M.W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. *23*, 3133-3157.

S46. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. BMC Bioinformatics *15*, 356.

S47. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3—new capabilities and interfaces. Nuc. Acids. Res. *40*, e115-e115.

S48. Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. of Hum. Gen. *81*, 1084-1097.

S49. Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., and Bustamante, C.D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. *5*, e1000695.

S50. Nam, K., Mugal, C., Nabholz, C., Schielzeth, H., and Wolf, J.B. (2010). Molecular evolution of genes in avian genomes. Genome Biol. *11*, R68.