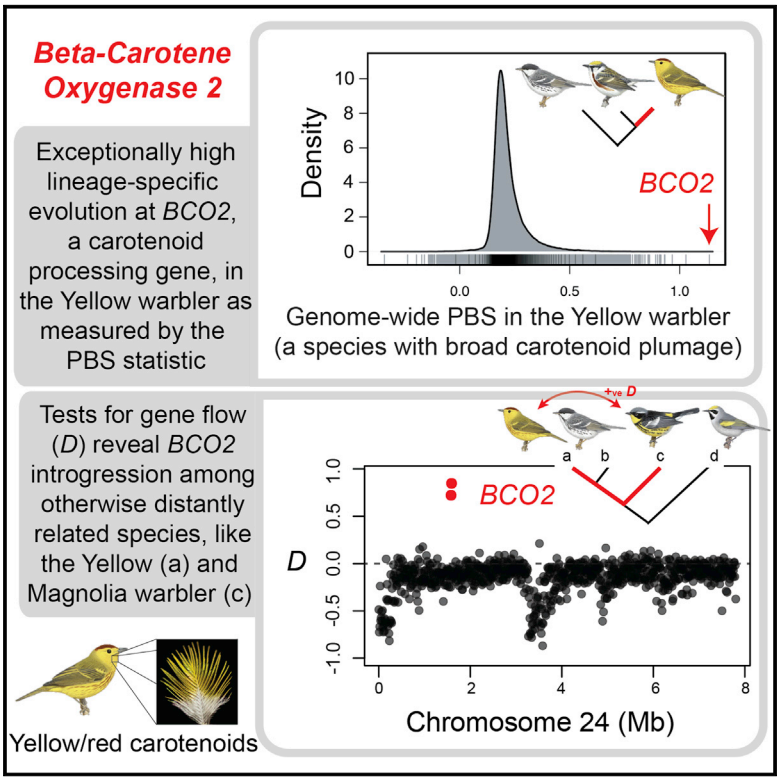
Report

Current Biology

Pigmentation Genes Show Evidence of Repeated Divergence and Multiple Bouts of Introgression in *Setophaga* Warblers

Graphical Abstract



Highlights

- Sister *Setophaga* species pairs share large F_{ST} peaks, especially on autosomes
- Divergence in *ASIP* suggests repeated mutation underlies melanic color differences
- Introgression at *BCO2* occurred among species with carotenoid-rich plumage
- The *BCO2* introgressed region includes a putatively adaptive amino acid substitution

Authors

Marcella D. Baiz, Andrew W. Wood, Alan Brelsford, Irby J. Lovette, David P.L. Toews

Correspondence

baizm@psu.edu (M.D.B.), toews@psu.edu (D.P.L.T.)

In Brief

Baiz et al. employ whole-genome sequencing across colorful *Setophaga* warblers to examine genome-wide differentiation between closely related species. Divergence in plumage coloration is mirrored by differentiation at two plumage pigmentation genes, with different evolutionary histories shaped by repeated mutation and introgression.

Baiz et al., 2021, Current Biology *31*, 1–7
February 8, 2021
https://doi.org/10.1016/j.cub.2020.10.094

CellPress

Current Biology

CellPress

Report

Pigmentation Genes Show Evidence of Repeated Divergence and Multiple Bouts of Introgression in *Setophaga* Warblers

Marcella D. Baiz,[1,4,*] Andrew W. Wood,[1] Alan Brelsford,[2] Irby J. Lovette,[3] and David P.L. Toews[1,4,5,*]
[1]Department of Biology, Pennsylvania State University, State College, PA, USA
[2]Evolution, Ecology, and Organismal Biology Department, University of California Riverside, Riverside, CA, USA
[3]Cornell Lab of Ornithology, Cornell University, Ithaca, NY, USA
[4]These authors contributed equally
[5]Lead Contact
*Correspondence: baizm@psu.edu (M.D.B.), toews@psu.edu (D.P.L.T.)
https://doi.org/10.1016/j.cub.2020.10.094

SUMMARY

Species radiations have long served as model systems in evolutionary biology.[1,2] However, it has only recently become possible to study the genetic bases of the traits responsible for diversification and only in a small number of model systems.[3] Here, we use genomes of 36 species of North, Central, and South American warblers to highlight the role of pigmentation genes—involved in melanin and carotenoid processing—in the diversification of this group. We show that *agouti signaling protein* (*ASIP*) and *beta-carotene oxygenase 2* (*BCO2*) are predictably divergent between species that differ in the distribution of melanin and carotenoid in their plumages, respectively. Among species, sequence variation at *ASIP* broadly mirrors the species' phylogenetic history, consistent with repeated, independent mutations generating melanin-based variation. In contrast, *BCO2* variation is highly discordant from the species tree, with evidence of cross-lineage introgression among species like the yellow warbler (*Setophaga petechia*) and magnolia warbler (*S. magnolia*) with extensive carotenoid-based coloration. We also detect introgression of a small part of the *BCO2* coding region (<3 kb) in *S. discolor* and *S. vitellina,* including an amino acid substitution that is unique to warblers but otherwise highly conserved across birds. Lateral transfer of carotenoid-processing genes has been documented in arthropods, but introgression of *BCO2* as demonstrated here—presumably adaptive—represents the first example of carotenoid gene transfer among vertebrates. These contrasting genomic patterns show that both independent evolution in a common set of genes and past hybridization have fueled plumage diversification in this colorful avian radiation.

RESULTS AND DISCUSSION

Evolutionary radiations—where lineages emerge at an exceptionally high rate—have long served as essential model systems in evolutionary biology.[1,2] What has only been possible recently, however, is studying the genetic bases of the traits associated with exceptional diversification.[3] With this information, we can ask: do the genomes of rapidly emerging species exhibit unique features conducive to divergence? Have the same genes been recruited repeatedly to create divergent phenotypes? Has hybridization fueled the diversification process?

Progress in this area is exemplified by the exceptional body of work on a small number of model systems, including *Heliconius*,[4] *Gasterosteus*,[5] cichlids,[6] and Darwin's finches.[7] Addressing these questions in other radiations remains challenging because a fulsome investigation necessitates (1) high species diversity, (2) a robust phylogenetic hypothesis, (3) independent evolution of similar phenotypes, and (4) admixture—naturally or in lab crosses—to facilitate genotype-phenotype associations. This latter requirement is the most challenging in most systems where natural hybridization is rare or absent or for which controlled crosses are intractable.[8] Here, we focus on an evolutionary radiation that fulfills all of these requirements.

The parulid warblers of the Americas are a well-known avian family with a rich legacy of study (Figure 1), including cornerstones of community ecology[9] and phylogenetic diversification.[10] Diverging from its sister family, Icteridae, approximately 10 mya, Parulidae radiated over the last 7 Ma.[11] Species accumulation in this group is exceptionally high compared to most other bird radiations; for example, among songbirds (order: Passeriformes), one of the largest shifts in diversification rate occurred in Parulidae.[11]

Unlike classic radiations where species show strong ecological differentiation driven by natural selection—i.e., adaptive radiations—the most conspicuous differences across Parulidae involve plumage.[13] In particular, coloration in this group derives from several distinct classes of pigments, primarily melanins (producing a range of brown, rufous, or black colors) and carotenoids (producing bright yellows, reds, and oranges).[14] The evolution of dramatic plumage color differences

Please cite this article in press as: Baiz et al., Pigmentation Genes Show Evidence of Repeated Divergence and Multiple Bouts of Introgression in *Setophaga* Warblers, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.10.094

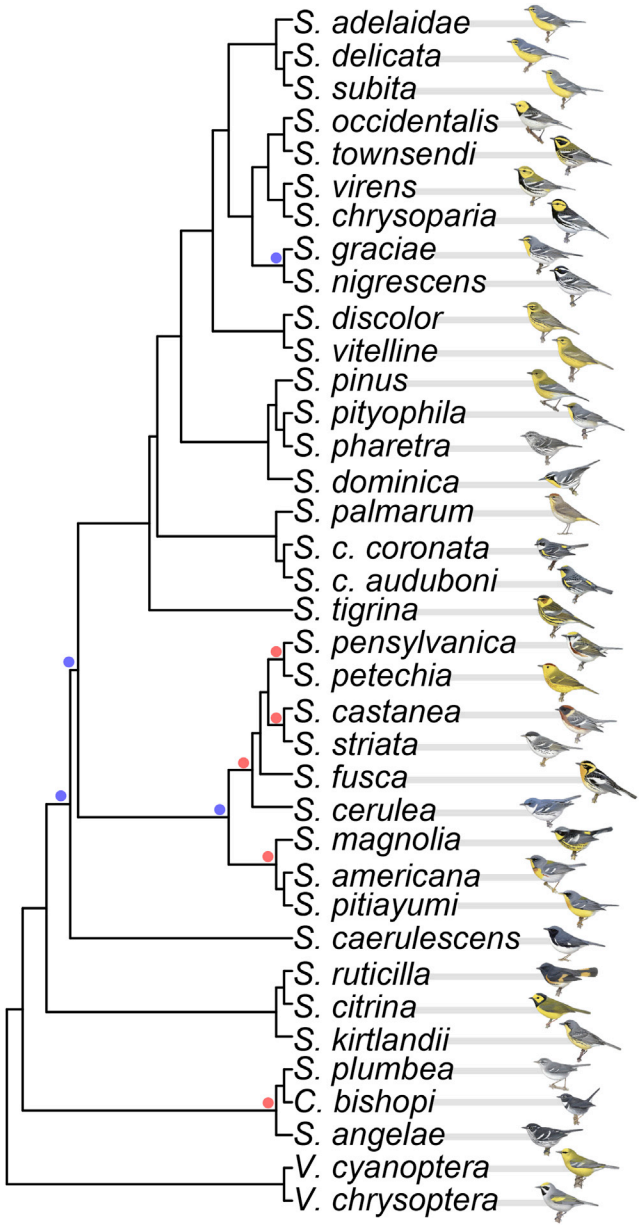


Figure 1. Phylogeny of *Setophaga*, Using *Vermivora* as an Outgroup
n = 34 Species, including the monotypic *Catharopeza bishopi*[12]. The tree is based on bioinformatically extracted UCE loci, with 1,000 bootstraps. All branches have 100% support, except those indicated by the colored dots (red, 50%–75% support; blue, 75%–98%). Warbler illustrations © Lynx Edicions.

in this group likely results from both natural and sexual selection: research on color-habitat associations[15] as well as on opsin gene (i.e., visual pigment) evolution[16] has provided a nuanced view of the joint role of both forces in driving adaptive plumage evolution.

Understanding the evolutionary history of genes involved in diversification first requires some knowledge of potential candidate genes. This is nearly universally unknown for non-model biological systems. In parulid warblers, however, detailed genomic studies of multiple, independent hybrid zones have generated a robust set of candidate genomic regions associated with species coloration differences.[17–20] The present study leverages these linkages for pigmentation-associated genes to ask whether these same genes have also been recruited in the evolution of similar phenotypes across the largest genus[12] in the family (*Setophaga*) and to test explicitly for evidence for historical introgression.

Whole-Genome Divergence across the Genus

We first produced a new chromosome-level reference genome for *Setophaga coronata*. We then generated whole-genome data at 4× to 5× coverage for multiple individuals—usually 5 per species—for all 34 species in the *Setophaga* genus (Figure 1). The only species omitted is the most recently described, *S. flavescens*.[21] We also included data from two of the *S. coronata* subspecies (*S.c. coronata*/*S.c. auduboni*)—where there is debate over species status—as well as published data from the two extant *Vermivora* species, which serve as outgroups in our phylogenetic analyses (n = 166 genomes).[22]

We first compared broad-scale patterns of differentiation across the genomes of nine sister species pairs (Figure S1). We identified a high level of sharing among large (>1-Mb) autosomal peaks of divergence between multiple pairs (Figure S2). This shared divergence is illustrated by two pairs: *S. graciae*/*S. nigrescens* and *S. petechia*/*S. pensylvanica* (Figures 2A and 2B). Both pairs have a highly heterogeneous pattern of divergence: overall low background F_{ST} punctuated by large, distinct peaks. Moreover, 43% of the 1-Mb peaks are shared between two or more of all the sister species pairs, with one—on chromosome 1a—shared across all but one of the nine pairs (Figure S2). We interpret this shared differentiation across independent sister species as a function of a common genomic architecture, linked selection, and reduced recombination.[23,24]

Much interest has been focused on the sex chromosomes and their role in speciation and divergence.[25] Our analyses show that (1) the Z chromosome is a hotspot of differentiation in *Setophaga* (33 large F_{ST} differentiation peaks among all pairs), but (2) patterns of differentiation across the Z appear more variable among taxa than across autosomes (only 18% of the large F_{ST} differentiation peaks are shared between 2 or more of the 9 pairs, compared to 53% of the autosomal differentiation peaks; Figures 3A–3D). For example, between *V. chrysoptera* and *V. cyanoptera* (Figure 3D), two of the six peaks in the genome occur on the Z, but these are not divergent in any of the other species pairs. Similarly, between *S. americana* and *S. pitiayumi* (Figure 3C), the only divergence peak across their genomes is restricted to a single, large, 20-Mb portion of the Z, which is possible evidence of a chromosomal inversion.

Pigmentation Gene Evolution Shows Repeated, Independent Evolution

Much smaller regions of divergence (i.e., <100 kb) included pigmentation genes. We focus on two with a clear connection to avian coloration: *agouti signaling protein* (*ASIP*) and *beta-carotene oxygenase 2* (*BCO2*). The *ASIP* gene product has a

Please cite this article in press as: Baiz et al., Pigmentation Genes Show Evidence of Repeated Divergence and Multiple Bouts of Introgression in *Setophaga* Warblers, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.10.094

CellPress

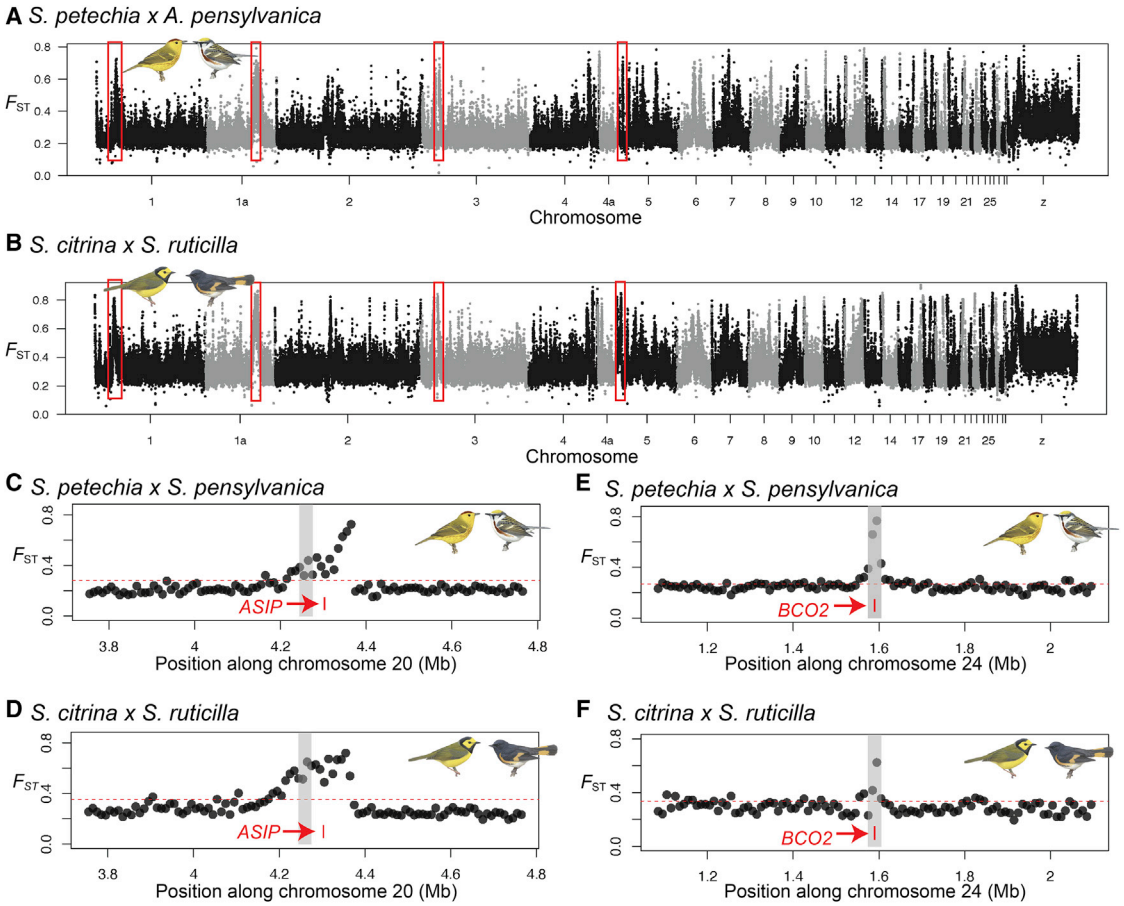


Figure 2. Differentiation between Two *Setophaga* Sister Species Pairs

(A and B) Genome-wide distribution of F_{ST} in 10-kb windows. The red boxes indicate a subset of divergent windows that are common between the pairs (see also Figure S1).

(C–F) Regions with *ASIP* (C and D) and *BCO2* (E and F). The gray boxes indicate locations that are divergent among *Vermivora*.[17] The vertical red line indicates the coding regions, and the horizontal dashed red line is the chromosome average F_{ST} (see also Figures S2 and S3). Warbler illustrations © Lynx Edicions.

well-characterized role in melanogenesis, and variation in its presumed promoter is associated with throat/mask plumage in *Vermivora*[17,20] and hybridizing *S. townsendi-S. occidentalis*.[19] *BCO2* is involved with carotenoid processing and plumage coloration in canaries[26,27] and is also associated with divergent plumage characteristics in *Vermivora*.[17]

We found that both *ASIP* and *BCO2* gene regions were consistently and predictably divergent between multiple *Setophaga* pairs that differ in melanic (*ASIP*) or carotenoid (*BCO2*) pigmentation (Figures 2C–2F and S3). For example, *S. petechia*, aptly known by its English common name of ''yellow warbler,'' has the most extensive carotenoid coloration of any species in *Setophaga*,[28] whereas its sister species, *S. pensylvanica*, has much less carotenoid-based coloration. Here, we found *BCO2* is highly divergent compared to the background level of differentiation between these two species (Figure 2E). Moreover, *S. pensylvanica* and *S. petechia* share a close common ancestor with a third species, *S. striata*, which also has little carotenoid-based plumage. We used these taxa in a three-population comparison to calculate population branch statistics (PBS) and estimate unique evolution associated with each. Notably, the region within the *S. petechia* genome that is the strongest outlier for PBS includes *BCO2* (Figure 3E).

We next compared gene trees specific to *ASIP* and *BCO2* to the inferred species tree to test for evidence of parallel divergence and/or introgression. We generated the species tree by bioinformatically extracting sequences from each of 4,008 ultra-conserved element (UCE) loci, identified in the *S. coronata* genome, across all the species (Figure 4A). For *ASIP*, we found that only two tips showed discordance with the presumed species branching pattern (Figure S4). Given that *ASIP* is divergent between many pairs and has been directly associated with melanic coloration in two warbler hybrid zones, our interpretation of this finding is that *ASIP* has been the target of repeated mutations that have contributed to melanic plumage differences across warblers. This is consistent with other avian pigmentation studies that have identified independent mutations in the *ASIP* region across a wide phylogenetic spectrum.[29–31]

Please cite this article in press as: Baiz et al., Pigmentation Genes Show Evidence of Repeated Divergence and Multiple Bouts of Introgression in *Setophaga* Warblers, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.10.094

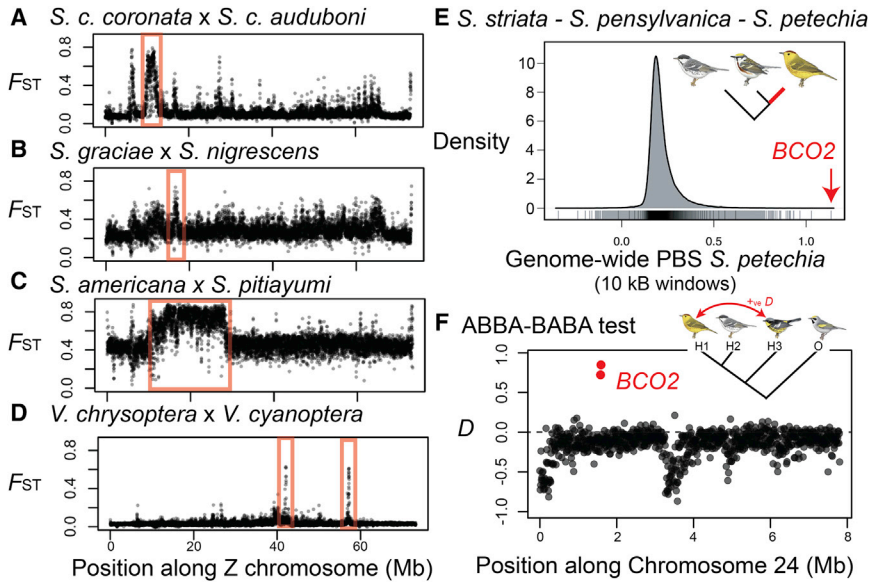


Figure 3. Divergence and Introgression among *Setophaga* Warblers

(A–D) Z chromosome differentiation among pairs of *Setophaga* (A–C) and (D) *Vermivora*.

(E) PBS for the three-species comparison. The 10-kb window with highest lineage-specific evolution in *S. petechia* includes the *BCO2* region (red arrow).

(F) D statistic from the ABBA-BABA test between *S. petechia* (H1), *S. striata* (H2), *S. magnolia* (H3), and *V. chrysoptera* (O) along chromosome 24. High *D* is consistent with introgression between *S. petechia* and *S. magnolia*. The autosomal windows with the highest *D* include the two 10-kb windows at *BCO2* (red points; see also Figure S2 for the whole genome comparison). Warbler illustrations © Lynx Edicions.

Repeated Instances of Historical Introgression at *BCO2*

BCO2 presents a strikingly different pattern from the inferred species tree, with clear evidence of introgression between otherwise highly divergent taxa (Figure 4A). As mentioned above, *S. petechia* forms a well-supported clade in the species tree with *S. pensylvanica*, *S. striata*, as well as *S. castanea*. However, at *BCO2*, *S. petechia* clusters with *S. fusca*, *S. magnolia*, and *S. ruticilla* (which we abbreviate the "PFMR group," using the first letter of their Latin species names; Figure 4A). None of these species are closely related to one another in the species-level topology, but at *BCO2*, there is strong support for their close relationship across many SNPs (bootstrap support is >94% for all internal nodes of this clade at *BCO2*). Importantly, all four PFMR species are well known for having extensive carotenoid-based coloration. Using *D*-statistics—an explicit test of gene flow—we found that the 10-kb *BCO2* region has the highest excess of shared mutations across all the autosomal windows between *S. magnolia* and *S. petechia* (Figures 3F and S2).

We also identified a more fine-scale pattern of introgression at *BCO2*. Across their genomes, *S. discolor* and *S. vitellina* are sister species that are in turn sister to the crown group of nine species, including *S. virens* and *S. graciae* (Figure 1). However, for a 3-kb region of *BCO2*, *S. discolor* and *S. vitellina*—species that also have extensive carotenoid-based body coloration—share several mutations with the PFMR group (Figures 4B and S2). Notably, this 3-kb region spans the 3rd and 4th exons of *BCO2* and includes an amino acid (aa) substitution in the 3rd exon, which is unique to these warblers, with the ancestral variant being highly conserved across other birds (Figure 4B). This substitution changes *BCO2* residue 39 (relative to *Taeniopygia guttata*)—an arginine (R) across all other *Setophaga* and *Vermivora* species—to a serine (S) in the PFMR group (plus *S. discolor* and *S. vitelline*). Nearly all avian genomes (66 of 68)[26] have either lysine (K) or R at this site, and no other S mutations have been characterized.

We interpret these findings at *BCO2* at three levels. First, it is consistent with at least two bouts of introgression, where a larger fragment containing *BCO2* moved among the species in the PFMR group after their ancestral lineages diverged. The high *BCO2* PBS in *S. petechia* (Figure 3E) is likely a signal of this introgression. Second, introgression also occurred into the common ancestor of *S. discolor* and *S. vitellina*. The resulting 3-kb fragment presumably represents only a small segment of the original introgressed region yet includes the putative functional mutation that underlies carotenoid coloration. Its much smaller size compared to the block that introgressed among the PFMR group suggests introgression with the ancestor of *S. discolor* and *S. vitellina* occurred earlier.[32] Finally, the PFMR group clade appears basal to the main *Setophaga* radiation in the *BCO2* tree (Figures 4 and S2). This could indicate this haplotype was originally from a warbler outside the genus, although additional data will be needed to fully test this scenario.

We predict that these changes in the 3rd exon of *BCO2* affect the function of the enzyme (e.g., Wu et al.[33]) and thus carotenoid feather pigmentation. Recent work in mosaic and urucum canaries found that downregulation[26] or disruption[27] of *BCO2* increased carotenoid deposition in feathers or bare parts, respectively, reflecting the carotenoid degradation properties of *BCO2*.[27] Thus, we speculate that the mutations in the 3rd exon of the PFMR group—plus *S. discolor* and *S. vitellina*—may change the function of *BCO2* and facilitate the broader deposition of carotenoids across the body feathers. We suggest future functional assays of these different *BCO2* variants in expressed cell lines, as has been done for *SCARB1*.[34] Movement of carotenoid genes between fungi and aphids is one of the most famous examples of horizontal gene transfer involving animals,[35] but introgression of *BCO2* among warblers would represent the best evidence so far for functional consequences of carotenoid gene movement among wild vertebrates.

Because parulid warblers have broadly overlapping ranges,[36] and different combinations of species likely came into and fell out of contact multiple times during the Pleistocene,[37,38] their biogeographic history has likely given them many opportunities to exchange genes compared to other groups that have been more consistently allopatric. Further, several species form hybrid zones (*Vermivora*,[17] *Setophaga*,[19,39] and *Myioborus*[40]), and multiple adult inter-generic hybrids have been confirmed.[41,42] Given

Please cite this article in press as: Baiz et al., Pigmentation Genes Show Evidence of Repeated Divergence and Multiple Bouts of Introgression in *Setophaga* Warblers, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.10.094

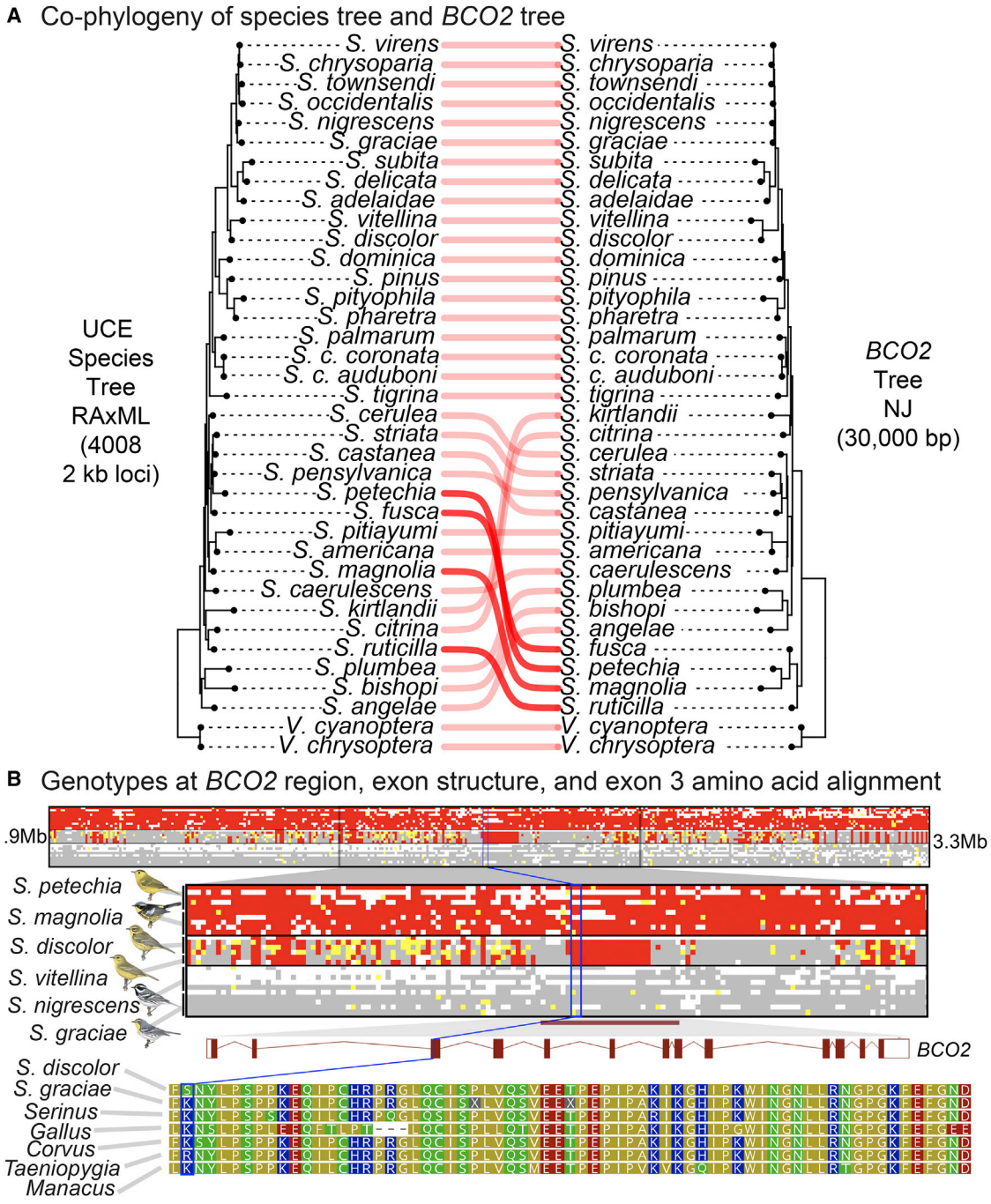


Figure 4. Multiple Bouts of Introgression at *BCO2*

(A) Co-phylogeny comparing the UCE species tree with the *BCO2* tree. The opaque red curves highlight discordant relationships among the PFMR group (see also Figure S4).

(B) Genotypes for high-F_{ST} SNPs for the *BCO2* region. Homozygous genotypes from two PFMR species are in red, gray genotypes represent homozygous genotypes for two non-PFMR species, and yellow shows heterozygous genotypes (white, missing). The middle rows include *S. discolor* and *S. vitellina*, which are homozygous for PFMR-group genotypes at this region (Figure S2). This introgressed region includes an AA substitution in exon 3, unique to the PFMR group plus *S. discolor* and *S. vitellina*. Warbler illustrations © Lynx Edicions.

this high frequency of hybridization, Parkes[43] speculated that many parulids have remained genetically compatible long after they evolved major phenotypic differences. The new molecular data presented here confirm that historical hybridization, as well as independent evolution in a common set of pigmentation genes, has likely fueled their evolution and diversity.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Sampling and DNA extraction
  - Reference genome
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Reference genome
  - Population genomic bioinformatic pipeline
  - Phylogenomic bioinformatic pipeline

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2020.10.094.

ACKNOWLEDGMENTS

For samples, we thank the Cornell University Museum of Vertebrates, the Museum of Comparative Zoology, the Museum of Southwestern Biology, the Museum of Vertebrate Zoology, the LSU Natural History Museum, the Smithsonian Migratory Bird Center, the Burke Museum, the Field Museum, Laurel Moulton, and Robert Ricklefs. We thank Bronwyn Butcher for molecular assistance and Leonardo Campagna for comments on a previous version. Funding was supported by the Cornell Lab of Ornithology, Pennsylvania State University, and startup funds from the Eberly College of Science and the Huck Institutes of the Life Sciences.

AUTHOR CONTRIBUTIONS

Conceptualization, M.B.D., D.P.L.T., and I.J.L.; Investigation, M.B.D., A.W.W., A.B., and D.P.L.T.; Resources, I.J.L. and D.P.L.T.; Data Curation, M.D.B. and D.P.L.T.; Writing – Original Draft, M.D.B. and D.P.L.T.; Writing – Review and Editing, all authors; Visualization, M.D.B. and D.P.L.T.; Supervision and Project Administration, D.P.L.T.; Funding Acquisition, A.B., I.J.L., and D.P.L.T.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 18, 2020
Revised: October 27, 2020
Accepted: October 29, 2020
Published: November 30, 2020

REFERENCES

1. Schluter, D. (2000). The Ecology of Adaptive Radiation (OUP Oxford).
2. Rundell, R.J., and Price, T.D. (2009). Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. Trends Ecol. Evol. *24*, 394–399.
3. Berner, D., and Salzburger, W. (2015). The genomics of organismal diversification illuminated by adaptive radiations. Trends Genet. *31*, 491–499.
4. Edelman, N.B., Frandsen, P.B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R.B., García-Accinelli, G., Van Belleghem, S.M., Patterson, N., Neafsey, D.E., et al. (2019). Genomic architecture and introgression shape a butterfly radiation. Science *366*, 594–599.
5. Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Jr., Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J., et al. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. Science *327*, 302–305.
6. Terai, Y., Mayer, W.E., Klein, J., Tichy, H., and Okada, N. (2002). The effect of selection on a long wavelength-sensitive (LWS) opsin gene of Lake Victoria cichlid fishes. Proc. Natl. Acad. Sci. USA *99*, 15501–15506.
7. 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.
8. Smith, S.D., Pennell, M.W., Dunn, C.W., and Edwards, S.V. (2020). Phylogenetics is the new genetics (for most of biodiversity). Trends Ecol. Evol. *35*, 415–425.
9. MacArthur, R.H. (1958). Population ecology of some warblers of Northeastern coniferous forests. Ecology *39*, 599–619.
10. Rabosky, D.L., and Lovette, I.J. (2008). Density-dependent diversification in North American wood warblers. Proc. Biol. Sci. *275*, 2363–2371.
11. Oliveros, C.H., Field, D.J., Ksepka, D.T., Barker, F.K., Aleixo, A., Andersen, M.J., Alström, P., Benz, B.W., Braun, E.L., Braun, M.J., et al. (2019). Earth history and the passerine superradiation. Proc. Natl. Acad. Sci. USA *116*, 7916–7925.
12. Lovette, I.J., Pérez-Emán, J.L., Sullivan, J.P., Banks, R.C., Fiorentino, I., Córdoba-Córdoba, S., et al. (2010). A comprehensive multilocus phylogeny for the wood-warblers and a revised classification of the Parulidae (Aves). Molecular Phylogenetics and Evolution *57*, https://doi.org/10.1016/j.ympev.2010.07.018.
13. 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.
14. Hill, G., and McGraw, K. (2006). Bird Coloration, Volume 1: Mechanisms and Measurements (Harvard University).
15. Simpson, R.K., Mistakidis, A.F., and Doucet, S.M. (2020). Natural and sexual selection shape the evolution of colour and conspicuousness in North American wood-warblers (Parulidae). Biol. J. Linn. Soc. *130*, 89–100.
16. Bloch, N.I. (2015). Evolution of opsin expression in birds driven by sexual selection and habitat. Proc. Biol. Sci. *282*, 20142321.
17. Toews, D.P.L., Taylor, S.A., Vallender, R., Brelsford, A., Butcher, B.G., Messer, P.W., and Lovette, I.J. (2016). Plumage genes and little else distinguish the genomes of hybridizing warblers. Curr. Biol. *26*, 2313–2318.
18. Brelsford, A., Toews, D.P.L., and Irwin, D.E. (2017). Admixture mapping in a hybrid zone reveals loci associated with avian feather coloration. Proc. Biol. Sci. *284*, 20171106.
19. Wang, S., Rohwer, S., de Zwaan, D.R., Toews, D.P.L., Lovette, I.J., Mackenzie, J., and Irwin, D. (2020). Selection on a small genomic region underpins differentiation in multiple color traits between two warbler species. Evol. Lett. Published online October 19, 2020. https://doi.org/10.1002/evl3.198.
20. Baiz, M.D., Kramer, G.R., Streby, H.M., Taylor, S.A., Lovette, I.J., and Toews, D.P.L. (2020). Genomic and plumage variation in *Vermivora* hybrids. The Auk *137*, ukaa027.
21. McKay, B.D., Reynolds, M.B.J., Hayes, W.K., and Lee, D.S. (2010). Evidence for the species status of the Bahama yellow-throated warbler (*Dendroica "dominica" flavescens*). The Auk *127*, 932–939.
22. Baiz, M.D., Wood, A.W., Brelsford, A., Lovette, I.J., and Toews, D.P.L. (2020). Data from: pigmentation genes show evidence of repeated divergence and multiple bouts of introgression in *Setophaga* warblers. DataDryad. https://doi.org/10.5061/dryad.n02v6wwvv.
23. Battey, C.J. (2020). Evidence of linked selection on the Z chromosome of hybridizing hummingbirds. Evolution *74*, 725–739.
24. Henderson, E.C., and Brelsford, A. (2020). Genomic differentiation across the speciation continuum in three hummingbird species pairs. BMC Evol. Biol. *20*, 113.

25. Irwin, D.E. (2018). Sex chromosomes and speciation in birds and other ZW systems. Mol. Ecol. *27*, 3831–3851.
26. Gazda, M.A., Araújo, P.M., Lopes, R.J., Toomey, M.B., Andrade, P., Afonso, S., Marques, C., Nunes, L., Pereira, P., Trigo, S., et al. (2020). A genetic mechanism for sexual dichromatism in birds. Science *368*, 1270–1274.
27. Gazda, M.A., Toomey, M.B., Araújo, P.M., Lopes, R.J., Afonso, S., Myers, C.A., Serres, K., Kiser, P.D., Hill, G.E., Corbo, J.C., and Carneiro, M. (2020). Genetic basis of de novo appearance of carotenoid ornamentation in bare parts of canaries. Mol. Biol. Evol. *37*, 1317–1328.
28. Gray, D.A. (1996). Carotenoids and sexual dichromatism in North American passerine birds. Am. Nat. *148*, 453–480.
29. Uy, J.A.C., Cooper, E.A., Cutie, S., Concannon, M.R., Poelstra, J.W., Moyle, R.G., and Filardi, C.E. (2016). Mutations in different pigmentation genes are associated with parallel melanism in island flycatchers. Proc. Biol. Sci. *283*, 20160731.
30. Campagna, L., Repenning, M., Silveira, L.F., Fontana, C.S., Tubaro, P.L., and Lovette, I.J. (2017). Repeated divergent selection on pigmentation genes in a rapid finch radiation. Sci. Adv. *3*, e1602404.
31. Stryjewski, K.F., and Sorenson, M.D. (2017). Mosaic genome evolution in a recent and rapid avian radiation. Nat. Ecol. Evol. *1*, 1912–1922.
32. Harris, K., and Nielsen, R. (2013). Inferring demographic history from a spectrum of shared haplotype lengths. PLoS Genet. *9*, e1003521.
33. Wu, L., Guo, X., Wang, W., Medeiros, D.M., Clarke, S.L., Lucas, E.A., Smith, B.J., and Lin, D. (2016). Molecular aspects of β, β-carotene-9′, 10′-oxygenase 2 in carotenoid metabolism and diseases. Exp. Biol. Med. (Maywood) *241*, 1879–1887.
34. Toomey, M.B., Lopes, R.J., Araújo, P.M., Johnson, J.D., Gazda, M.A., Afonso, S., Mota, P.G., Koch, R.E., Hill, G.E., Corbo, J.C., and Carneiro, M. (2017). High-density lipoprotein receptor SCARB1 is required for carotenoid coloration in birds. Proc. Natl. Acad. Sci. USA *114*, 5219–5224.
35. Moran, N.A., and Jarvik, T. (2010). Lateral transfer of genes from fungi underlies carotenoid production in aphids. Science *328*, 624–627.
36. Lovette, I.J., and Hochachka, W.M. (2006). Simultaneous effects of phylogenetic niche conservatism and competition on avian community structure. Ecology *87* (7, Suppl), S14–S28.
37. Lovette, I.J. (2005). Glacial cycles and the tempo of avian speciation. Trends Ecol. Evol. *20*, 57–59.
38. Weir, J.T., and Schluter, D. (2004). Ice sheets promote speciation in boreal birds. Proc. Biol. Sci. *271*, 1881–1887.
39. Toews, D.P.L., Brelsford, A., and Irwin, D.E. (2011). Hybridization between Townsend's *Dendroica townsendi* and black-throated green warblers *D. virens* in an avian suture zone. J. Avian Biol. *42*, 434–446.
40. Céspedes-Arias, L.N., and Cuervo, A.M. (2018). Extensive hybridization between two Andean warbler species with little divergence in mtDNA. Masters thesis (Universidad de los Andes).
41. Toews, D.P.L., Streby, H.M., Burket, L., and Taylor, S.A. (2018). A wood-warbler produced through both interspecific and intergeneric hybridization. Biol. Lett. *14*, 20180557.
42. Toews, D.P.L., Kramer, G.R., Jones, A.W., Brennan, C.L., Cloud, B.E., Andersen, D.E., Lovette, I.J., and Streby, H. (2020). Genomic identification of intergeneric hybrids in New World wood-warblers (Aves: Parulidae). Biol. J. Linn. Soc. *131*, 183–191.
43. Parkes, K.C. (1978). Still another Parulid intergeneric hybrid (*Mniotilta* × *Dendroica*) and its taxonomic and evolutionary implications. The Auk *95*, 682–690.
44. Hahn, C., Bachmann, L., and Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads–a baiting and iterative mapping approach. Nucleic Acids Res. *41*, e129.
45. Lindgreen, S. (2012). AdapterRemoval: easy cleaning of next-generation sequencing reads. BMC Res. Notes *5*, 337.
46. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.
47. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics *15*, 356.
48. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
49. Faircloth, B.C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics *32*, 786–788.
50. Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. *3*, 217–223.
51. Faircloth, B.C., Branstetter, M.G., White, N.D., and Brady, S.G. (2015). Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. Mol. Ecol. Resour. *15*, 489–501.

Please cite this article in press as: Baiz et al., Pigmentation Genes Show Evidence of Repeated Divergence and Multiple Bouts of Introgression in *Setophaga* Warblers, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.10.094

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Warbler blood and tissue samples	[22]	https://doi.org/10.5061/dryad.n02v6wwvv
Chemicals, Peptides, and Recombinant Proteins		
QIAGEN DNeasy Blood and Tissue kit	QIAGEN	Cat# 69504
UPrep Spin Columns	Genesee Scientific	Cat# 88-143
Qubit Fluorometer	ThermoFisher Scientific	Cat# Q33239
TruSeq Nano DNA	Illumina	20015964
Deposited Data		
Reference *Setophaga coronata* genome	This paper	NCBI# PRJNA325157
Whole genome re-sequencing *Setophaga* data	This paper	NCBI# PRJNA630247
Software and Algorithms		
mitoBIM 1.9.1	[44]	https://github.com/chrishah/MITObim
AdapterRemoval 2.1.1	[45]	https://github.com/MikkelSchubert/adapterremoval
BowTie2 2.3.5.1	[46]	http://bowtie-bio.sourceforge.net/bowtie2/
PicardTools 2.20.8	Broad Institute	https://broadinstitute.github.io/picard/
ANGSD 0.929	[47]	http://www.popgen.dk/angsd/index.php/ANGSD
Samtools 0.1.18	[48]	http://www.htslib.org/
Seqtk 1.3	https://github.com/lh3	https://github.com/lh3/seqtk
Geneious 11.0.3	Geneious	https://www.geneious.com/
Phyluce 1.6	[49]	https://phyluce.readthedocs.io/en/latest/
Phytools 0.6-60	[50]	https://cran.r-project.org/web/packages/phytools/

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, David P. L. Toews (toews@psu.edu)

Materials availability

This study did not generate new unique reagents.

Data and code availability

All code used in the bioinformatic pipeline to analyzed re-sequencing data can be found in[22]. The "Version 2.0" *Setophaga coronata* genome can be found at NCBI #PRJNA325157. Re-sequencing data is deposited at NCBI# PRJNA630247.

METHOD DETAILS

Information about sample origin and sample ID can be found in [[22]; available at DataDryad https://doi.org/10.5061/dryad.n02v6wwvv]. All procedures were performed in accordance with procedures outlined in the approved IACUC at Cornell University (#015-0065) and Pennsylvania State University (#201900879). Bird banding and blood sampling was approved by the USGS (Master Banding Permit to DPLT #24043) and NYDEC (banding permit to DPLT #156).

Sampling and DNA extraction

The majority of samples (117 of 166) were DNA extracted from tissues loaned from several museums[22]. For DNA extraction, we used 2 mm^3 of tissue digested in buffer ATL and AL and washed with AW1 and AW2 of the QIAGEN DNeasy Blood and Tissue kit—following the tissue extraction procedure—and separated DNA using UPrep spin columns (Genesee). For the remaining 49 samples, we used blood samples. Most of these were from species of significant conservation concern, or endemic to Caribbean Islands, so obtaining tissue samples is logistically challenging, and thus relied on previously collected blood in buffer.

Please cite this article in press as: Baiz et al., Pigmentation Genes Show Evidence of Repeated Divergence and Multiple Bouts of Introgression in *Setophaga* Warblers, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.10.094

We used 75 uL of blood and followed the same QIAGEN DNA extraction procedure as above, except using the blood extraction protocol.

We standardized DNA concentrations after quantifying with a Qubit fluorometer, and then generated sequencing libraries with the Illumina TruSeq Nano kit, targeting 350 bp insert sizes. We individually indexed each sample and sequenced the libraries with Illumina NextSeq 500. We sequenced 24 individuals together at a time on a single lane, with paired-end 150bp chemistry, and used seven NextSeq lanes in total. Data from the 10 *Vermivora* individuals have been published previously[17].

Reference genome

We generated a new chromosome-level reference genome by combining the fragment libraries (SRA accession number SRX2019496) and mate-pair libraries (SRA accession numbers SRX2019494 and SRX2019495) from the original *Setophaga coronata* assembly, with chromosome conformation capture data, generated by Dovetail Genomics, from a newly sampled individual. This new *S. coronata* individual (USGS band #284029498) was sampled in Centre County, PA (latitude: 40.723591, longitude: −77.791108) on June 12th, 2019.

QUANTIFICATION AND STATISTICAL ANALYSIS

Reference genome

To generate the new assembly (NCBI accession #PRJNA325157), we used the Dovetail "Meraculous" pipeline. The new scaffolding data came from Hi-C and Chicago libraries using the Dovetail "Hi-Rise" pipeline. We manually adjusted the fourth largest scaffold outputted from the pipeline, as the Hi-C report and subsequent alignments suggested that various segments of it each matched a different micro-chromosome in the Zebra Finch genome. To generate the full mitochondrial genome sequence, we used the *S. coronata* tissue sample "CUMV4915" and ran mitoBIM[44] using the "-quick" option and the complete mitochondrial genome of a Canada warbler (*Cardellina canadensis*; NCBI accession number MK033135) as bait.

Population genomic bioinformatic pipeline

We generated resequencing data for 156 individuals across 34 Setophaga species (range 1-15, mean = 4.45, individuals per species). We first used the program AdapterRemoval[45] to trim and collapse overlapping paired reads with the following options: "-collapse -trimns -minlength 20 -qualitybase 33." We then used BowTie2[46] to align reads to the new reference genome with the "very-sensitive-local" presets and set the option "-X" (the maximum fragment length for valid paired-end alignments) to 700 bp. We used PicardTools (2.20.8; https://broadinstitute.github.io/picard/) to mark PCR duplicates.

To estimate F_{ST}, we used ANGSD[47] version 0.929, which accounts for genotyping uncertainty in low-coverage data, and estimated F_{ST} for non-overlapping 10-kb windows. For each population, we calculated the site allele frequency likelihoods using the "-dosaf 1" command. We then calculated the two-population site frequency spectrum and resulting F_{ST} estimates using "realSFS." To calculate population branch statistics (PBS) for *S. petechia*, *S. pensylvanica*, and *S. striata*, we used as similar approach, calculating site allele frequency likelihoods for each species, and then calculating all the pairwise two population site frequency spectra. These were then used by "realSFS" with the options "-win 10000 -step 10000" to calculate the sliding-window PBS statistic for each of the three species. To calculate *D*-statistics, we used the "-doAbbababa2" function, also in ANGSD, using the options "-blockSize 10000 -doCounts 1." For the analysis of *S. petechia*, we used it and *S. striata* and *S. magnolia* as populations 1, 2, and 3, respectively, and *Vermivora chrysoptera* as the outgroup (with the "-useLast 1" option).

To identify shared clusters of F_{ST} differentiation between independent sister species pairs, we focused our analysis on nine pairs that the UCE tree identified as sister taxa (Figure S1 shows the nine pairs included in the analysis). Given the species-status uncertainty between *S. c. coronata* and *S. c. auduboni,* we only include the nominate form (*S. c. coronata*) in this analysis and compare it to *S. palmarum*. Some possible sister taxon comparisons were excluded due to their limited sample sizes[22].

We first binned the reference genome into 991, 1-Mb windows, excluding the mitochondrial genome, as well as small scaffolds not assigned to chromosomes. Then, for each pair, we used the output from 10-kb non-overlapping F_{ST} window analysis and identified the top 1% of windows (referred throughout as the 'top 1%' windows). This resulted in 880-1108 of the most divergent windows, depending on the species pair. We then ran a permutation sampling of 991 10-kb windows—with 10,000 replicates—to determine how many of these 'top 1%' windows would be expected to fall in the same larger 1-Mb window by chance. For each replicate we calculated the maximum number of 10-kb windows in each of these larger window. Across these 10,000 replicates the maximum number of 'top 1%' windows occurring in a single 1-Mb window by chance was 7, and therefore used 8 'top 1%' windows as our threshold to delimit the highly divergent, large F_{ST} peaks.

Phylogenomic bioinformatic pipeline

To delineate regions analyzed for *ASIP* and *BCO2* we first identified the divergent regions near these genes between *Vermivora chrysoptera* and *V. cyanoptera* in the new *S. coronata* reference genome. *V. chrysoptera* and *V. cyanoptera* are extremely similar in their nuclear genome, except for six small regions (including *ASIP* and *BCO2*) that stand out against this background[17]. For *ASIP*, the divergent region spans nearly 100 kb, including a large portion in the presumed promoter region as well as the full coding region. For *BCO2*, the divergent region between *Vermivora* species spans 30 kb, which includes the upstream portion of the gene, as

Please cite this article in press as: Baiz et al., Pigmentation Genes Show Evidence of Repeated Divergence and Multiple Bouts of Introgression in *Setophaga* Warblers, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.10.094

well as the first four exons. The locations used for as *ASIP* in the new reference assembly are on chromosome 20, between 4,215,000-4,315,000 bp; for *BCO2,* these are on chromosome 24, between 1,575,000-1,605,000 bp.

For the phylogenetic analyses, we combined data for all the individuals in each species to produce a single sequence file. We did this to generate sequence data with sufficiently high coverage to avoid large amounts of missing data. Therefore, to generate fasta files, we used the "mpileup" command in samtools version 0.1.18[48], the "call -c" command in bcftools version 1.10.2, and the "vcf2fq" option to generate fastq files. Finally, we used the program seqtk (version 1.3; https://github.com/lh3/seqtk) to translate into fasta format. Any gaps without sequence coverage were manually closed in the alignments using Geneious (version 11.0.3). To generate trees for *ASIP* and *BCO2*, we used the Geneious Tree Builder to generate Neighbor-Joining trees using the HKY distance model.

To generate a species tree, we used ultra-conserved element (UCE) loci. To locate UCEs, we aligned the "Tetrapods-UCE-5Kv1" probe set[51] to the *S. coronata* reference genome using BowTie2. We removed all but the P1 probes, as well as 175 UCEs that did not align, 7 UCEs that did not align to a chromosome (i.e., they aligned to small, unplaced scaffolds), and 42 UCEs that aligned to > 1 position in the reference assembly. We then trimmed the probes such that they were at least 3,000 bp apart from any subsequent UCE locus. This resulted in 1,000 base pairs from either side (i.e., 2000 bp total) from a final set of 4,008 UCE loci.

We utilized the phyluce phylogenomics pipeline (version 1.6) to produce the species tree[49]. We extracted UCE loci from each species with seqtk by filtering the fasta files for the regions identified above as UCE loci. We aligned all loci using the MAFFT aligner option and retained only those alignments that included all target species (4,008 UCE loci). We then concatenated UCE loci for all species into a phylip file using the command "phyluce_align_format_nexus_files_for_raxml" and ran RAxML (-m GTRGAMMA) with 1000 bootstrap replicates. To compare the gene trees for *ASIP* and *BCO2* to the species tree, we used the "co.phylo" function in phytools version 0.6-60[50].