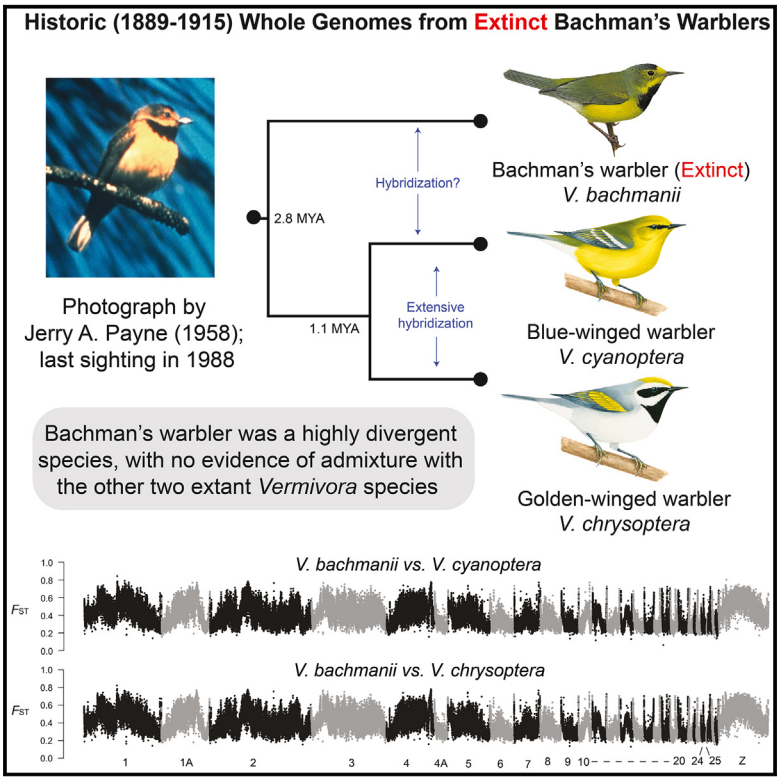
Report

Current Biology

Genomes of the extinct Bachman's warbler show high divergence and no evidence of admixture with other extant *Vermivora* warblers

Graphical abstract



Highlights

- Genomes show Bachman's warbler was a highly divergent, reproductively isolated taxon
- There is no evidence of admixture between *V. bachmanii* and extant *Vermivora*
- *Vermivora* have similar levels of ROH, implying small effective population sizes
- PBS in *V. chrysoptera* highlights *CORIN* as a novel color gene candidate in warblers

Authors

Andrew W. Wood, Zachary A. Szpiech, Irby J. Lovette, Brian Tilston Smith, David P.L. Toews

Correspondence

toews@psu.edu

In brief

Wood et al. sequenceed the genomes of extinct Bachman's warblers and find that it was likely a highly divergent, reproductively isolated species, with no evidence of admixture between *V. bachmanii* and other extant *Vermivora* warblers.

Wood et al., 2023, Current Biology *33*, 1–7
July 10, 2023
https://doi.org/10.1016/j.cub.2023.05.058

CellPress

Current Biology

CellPress

Report

Genomes of the extinct Bachman's warbler show high divergence and no evidence of admixture with other extant *Vermivora* warblers

Andrew W. Wood,[1,5] Zachary A. Szpiech,[1,2] Irby J. Lovette,[3] Brian Tilston Smith,[4] and David P.L. Toews[1,5,6,7,*]
[1]Department of Biology, Pennsylvania State University, 619 Mueller Laboratory, University Park, State College, PA 16802, USA
[2]Institute for Computational and Data Sciences, Pennsylvania State University, University Park, State College, PA 16802, USA
[3]Fuller Evolutionary Biology Program, Cornell Lab of Ornithology, Cornell University, 159 Sapsucker Woods Road, Ithaca, NY 14850, USA
[4]Department of Ornithology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA
[5]These authors contributed equally
[6]Twitter: @davetoews
[7]Lead contact
*Correspondence: toews@psu.edu
https://doi.org/10.1016/j.cub.2023.05.058

SUMMARY

Bachman's warbler[1] (*Vermivora bachmanii*)—last sighted in 1988—is one of the only North American passerines to recently go extinct.[2–4] Given extensive ongoing hybridization of its two extant congeners—the blue-winged warbler (*V. cyanoptera*) and golden-winged warbler (*V. chrysoptera*)[5–8]—and shared patterns of plumage variation between Bachman's warbler and hybrids between those extant species, it has been suggested that Bachman's warbler might have also had a component of hybrid ancestry. Here, we use historic DNA (hDNA) and whole genomes of Bachman's warblers collected at the turn of the 20th century to address this. We combine these data with the two extant *Vermivora* species to examine patterns of population differentiation, inbreeding, and gene flow. In contrast to the admixture hypothesis, the genomic evidence is consistent with *V. bachmanii* having been a highly divergent, reproductively isolated species, with no evidence of introgression. We show that these three species have similar levels of runs of homozygosity (ROH), consistent with effects of a small long-term effective population size or population bottlenecks, with one *V. bachmanii* outlier showing numerous long ROH and a F_{ROH} greater than 5%. We also found—using population branch statistic estimates—previously undocumented evidence of lineage-specific evolution in *V. chrysoptera* near a pigmentation gene candidate, *CORIN*, which is a known modifier of *ASIP*, which is in turn involved in melanic throat and mask coloration in this family of birds. Together, these genomic results also highlight how natural history collections are such invaluable repositories of information about extant and extinct species.

RESULTS AND DISCUSSION

> "Within an area of ten or fifteen acres there must have been nearly one thousand warblers, of which probably five per cent were Bachman's. It was comparatively easy to identify them."[1]

Few locally common North American passerines have gone extinct. Bachman's warbler (*Vermivora bachmanii*) is, unfortunately, one of the few that almost certainly has—the last confirmed sighting of this songbird was in 1988 in Louisiana.[2] Subsequently, many targeted search efforts have failed to locate any individuals, and it has not been observed despite the multifold increase in birding activity over recent decades. Although still listed as critically endangered (and possibly extinct) by the International Union for Conservation of Nature Red List,[3] it is widely accepted to be extinct in the ornithological community in North America and was officially declared so in 2021 by the United States Fish and Wildlife Service.[4]

The only extant members of the same genus, *Vermivora*, are the golden-winged warbler (*Vermivora chrysoptera*) and the blue-winged warbler (*Vermivora cyanoptera*), which are notable within their family (Parulidae) for their widespread and extensive hybridization.[5,6] This includes the "named" hybrid morphs "Lawrence's" and "Brewster's" warblers, which were originally each thought to be a distinct species before they were found to be hybrids.[7] Genomic analyses of *V. cyanoptera* and *V. chrysoptera* have since revealed that few and relatively small regions of their genomes are divergent between them. These regions house pigmentation genes, which underlie the markedly different plumage phenotypes of these species (Figure 1).[8,9] For instance, the distinctive black throat patch and face mask of *V. chrysoptera* appears in the Lawrence's warbler hybrid phenotype but is absent in the Brewster's type (Figure 1A), and this difference is associated with a 10-kb region upstream of the *Agouti signaling protein* (*ASIP*) gene, which has a gene product involved in the melanogenesis pathway.[8–10] This black throat patch is also one of the distinctive traits of Bachman's warbler,

Please cite this article in press as: Wood et al., Genomes of the extinct Bachman's warbler show high divergence and no evidence of admixture with other extant *Vermivora* warblers, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.05.058

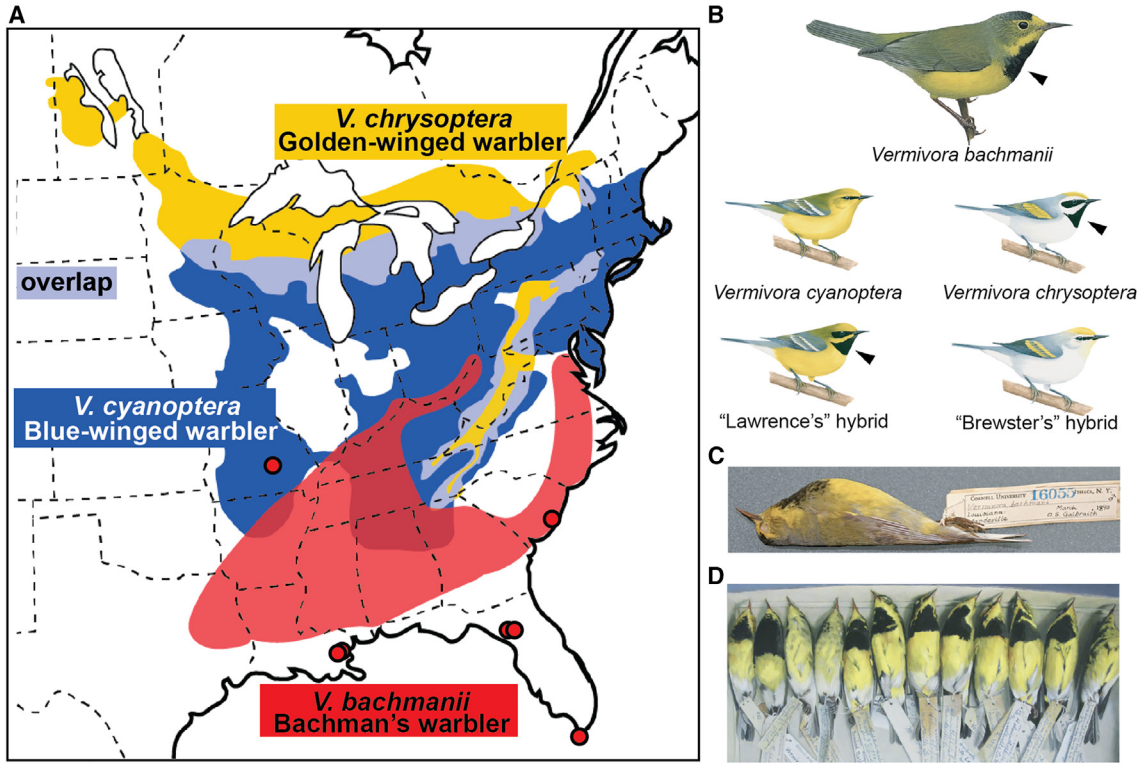


Figure 1. Distribution and plumage phenotypes of Vermivora warblers and their hybrids

(A) The range of *V. bachmanii* inferred from Hamel.[2] The red points illustrate the location where the Bachman's warblers were sampled. Because most of these were likely on migration, we do not know where within the red breeding range they held territories. (B) Bachman's warbler illustration used with permission from © Lynx Edicions. Other *Vermivora* illustrations by Liz Clayton Fuller. Photo in (C) from Cornell University Museum of Vertebrates, (D) is an oil painting by Isabella Kirkland, with reference from specimens at the AMNH.

See also Table S1.

making it phenotypically parallel to *V. chrysoptera* and the Lawrence's hybrid form.

The extensive hybridization between the two extant *Vermivora* species, combined with the similarity of plumage between *V. bachmanii* and contemporary *Vermivora* hybrids, caused Hamel[11] to ask whether *V. bachmanii* might have a component of hybrid ancestry. First, in plumage, in addition to the black throat patch that is common between *V. bachmanii* and the *Vermivora* hybrids, the broad distribution of body carotenoids on *V. bachmanii* and subtle olive coloration of its face make it similar to *V. cyanoptera*. Second, a song playback study[11] showed that individuals of each *V. cyanoptera* and *V. chrysoptera* responded to *V. bachmanii* song recordings as strongly as conspecific songs. This kind of response has been used to demonstrate weak isolating barriers based on song.[12]

Given the inferred distribution and habitat preference of *V. bachmanii* for bamboo,[13] it seems unlikely that it co-occurred in breeding sympatry with *V. chrysoptera*, and even perhaps not in tight local sympatry with *V. cyanoptera* (Figure 1A). However, extant *Vermivora* live in early successional habitat, which is dynamic on the landscape. Gill[5] documented the broad distributional change of both *V. cyanoptera* and *V. chrysoptera* over just 50 years, and therefore both may have undergone even greater changes over hundreds of generations. In addition, the breeding habitat of Bachman's warbler has "been the subject of intense speculation," [11] thus its precise historical distribution is unclear. Therefore, there may have been opportunity for local sympatry—and hybridization—for two or more of the species throughout their evolutionary history.

Genetic studies of the question of possible hybridization in *V. bachmanii* have not previously been possible. Prior to 2021, a small fragment of the mitochondrial ND2 gene obtained by Lovette et al.[14] was the only sequence available for *V. bachmanii*, and the maternal inheritance of mitochondrial DNA (mtDNA) makes this marker inadequate to detect genomic introgression. Likewise, a sequence-capture dataset from ultra-conserved elements produced insights into the demographic history of *V. bachmanii*[15] but is unsuitable to study past hybridization among extant *Vermivora* because their genomes are so uniformly similar.[8]

Here, we generated the data necessary to study the possibility of historical admixture among *Vermivora*. We used whole genome data for *V. bachmanii* produced from the toepad DNA of seven specimens collected in the southern USA between 1889 and 1915. We compared the extent and distribution of divergence among the species, as well as explicitly testing for introgression. We also explored genomic patterns relevant to the decline and conservation of small populations—runs of homozygosity (ROH)—among all three species; ROH are long stretches of identical-by-descent haplotypes within genomes

Please cite this article in press as: Wood et al., Genomes of the extinct Bachman's warbler show high divergence and no evidence of admixture with other extant *Vermivora* warblers, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.05.058

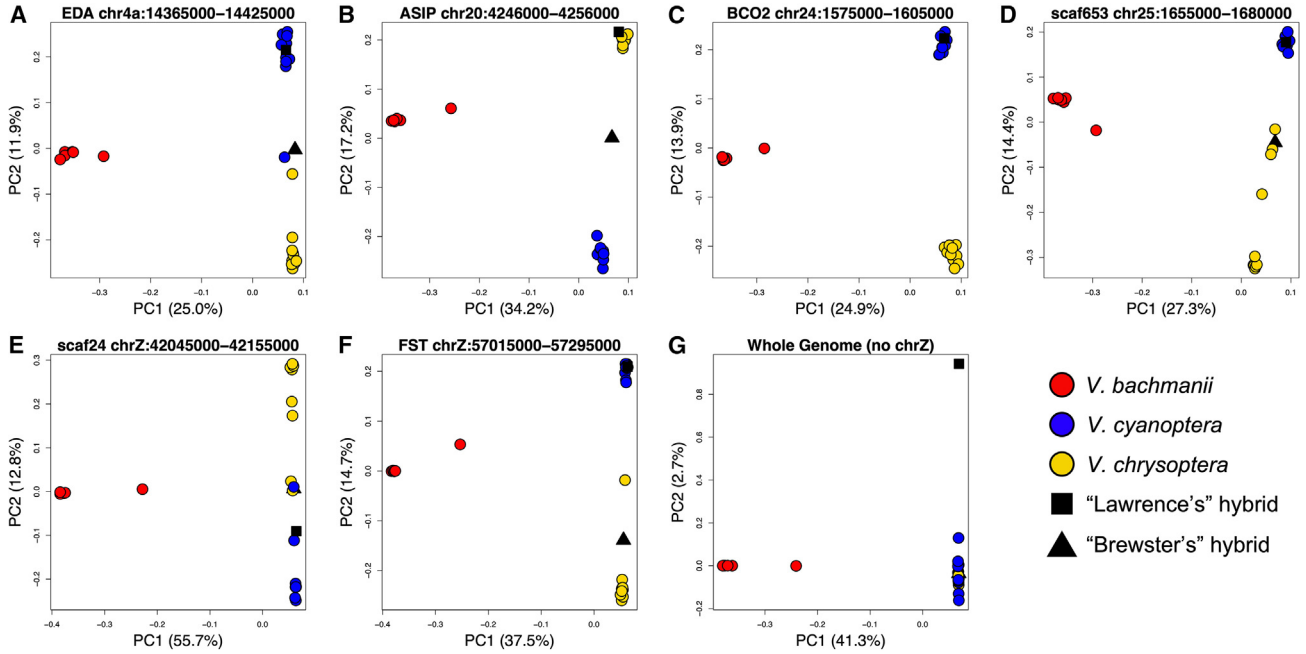


Figure 2. Principal components analysis using genotype likelihoods of all three *Vermivora* species

Data for six regions of divergence identified between *V. cyanoptera* and *V. chrysoptera* (A–F) as well as across the entire genome (G), excluding the Z (sex) chromosome. Chromosomal coordinates of regions are shown in plot titles. Candidate pigmentation genes as in Toews et al.[8] and, for regions without gene candidates, previous scaffold names are shown. (A) *EDA* (*ectodysplasin*) region (B) *ASIP* (*agouti signaling protein*) region, (C) *BCO2* (*beta-carotene oxygenase 2*) region, (D) unnamed region on chromosome 25, (E) unnamed region on chromosome Z, (F) *FST* (*follistatin*) region.

and are indicators of small population size, bottlenecks, and inbreeding.[16,17]

Genomes of Bachman's warblers

We generated genomes from *V. bachmanii* from toepads sampled from museum skins (Table S1; Wood et al.[18]) held at the Cornell Museum of Vertebrates (CUMV; n = 1) and the American Museum of Natural History (AMNH; n = 6). Most were collected during March or April and, given their location at the time of collection, these were likely individuals collected during migration. Thus, the geographic provenance of their breeding territories is unknown (Figure 1A). Given the limitations of the historical sampling, we were also not able to obtain *V. bachmanii* from sites with the most likely possibility of local sympatry, at least based on the contemporary distributions of the extant species. That said, the genomic tools employed here were aimed at describing evidence of historical admixture, not necessarily whether any of the individuals were early-generation hybrids themselves.

We combined our data from *V. bachmanii* with genomes of previously published extant *Vermivora*: ten *V. cyanoptera*, ten *V. chrysoptera*, and one each of the Lawrence's and Brewster's hybrid phenotype.[8,9] Previously published *Vermivora* were sequenced to 4–5× coverage. The *V. bachmanii* coverage for the present study was comparable for six samples (AMNH average coverage = 4.3×), with one sample having lower coverage (CUMV coverage = 0.8×; Table S1). For ABBA-BABA tests, we also generated 5× data from an outgroup, the ovenbird (*Seiurus aurocapilla*; n = 5), which is sister to all other wood warblers.[14]

Patterns of divergence among *Vermivora*

Using *cytochrome B* sequence extracted from mtDNA, represented at high coverage in toe pads (mean 404× for AMNH samples), we estimated the divergence time between *V. bachmanii* and the extant *Vermivora* to 2.82 million years ago, consistent with species-level divergence in birds.[19] We then focused on six previously identified genomic regions of divergence between *V. cyanoptera* and *V. chrysoptera* (Figures 2A–2F). This is because, outside of these regions, the genomes of *V. cyanoptera* and *V. chrysoptera* are effectively homogenized. *V. bachmanii* consistently fell into its own cluster along the first principal component (PC) axis, which reliably explained disproportionately more variation (>24%) than all other axes. PC2 clearly separated *V. chrysoptera* and *V. cyanoptera*, with the contemporary hybrids falling intermediate along this axis. *V. bachmanii* was also intermediate along PC2, which could be consistent with evidence of admixture and/or introgression at these gene regions. However, other data do not support this interpretation, especially when considering divergence evident from across the entire genome (Figure 2G), where the axes separate *V. bachmanii*, but not the other taxa.

The extent of divergence differed dramatically among the three species. As documented previously for *V. cyanoptera* and *V. chrysoptera*, F_{ST} was extremely low, save for the six small regions housing pigmentation genes (Figure 3C). By contrast, each pairwise comparison with *V. bachmanii* showed very high divergence (Figures 3A and 3B). This is likely due, in part, to the combined effects of linked selection and reduced recombination (reviewed in Burri[20]) and amplified by the effects of genetic drift in *V. bachmanii*.

Please cite this article in press as: Wood et al., Genomes of the extinct Bachman's warbler show high divergence and no evidence of admixture with other extant *Vermivora* warblers, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.05.058

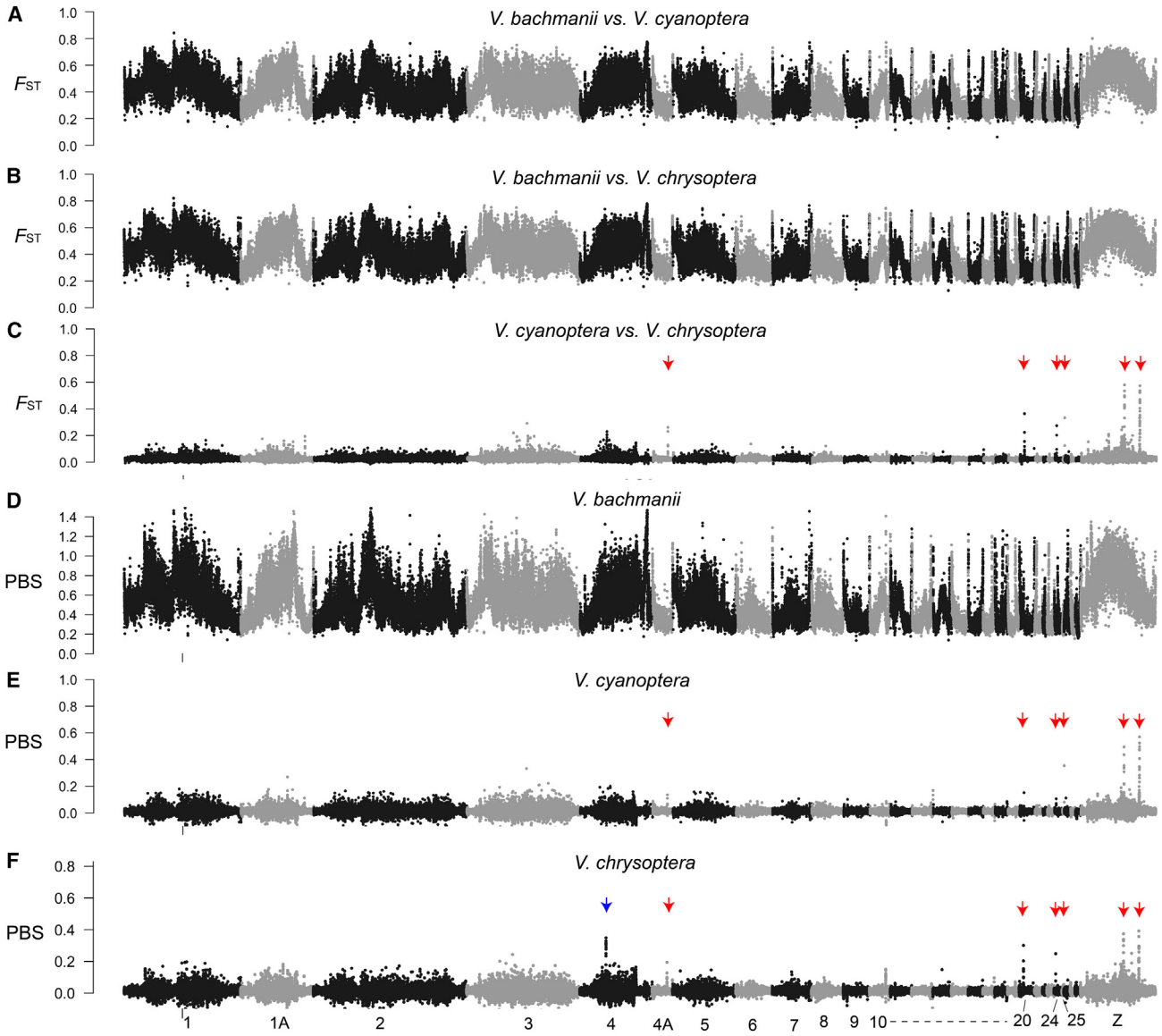


Figure 3. Patterns of genomic divergence among Vermivora warblers

Genome-wide F_{ST} and PBS measures (in non-overlapping 10 kbp windows) between each of the three species (A–C) as well as lineage-specific PBS estimates (D–F). The red arrows indicate the six divergent regions identified in Figure 2, the blue arrow in (F) indicates the location of the *CORIN* gene.
See also Figures S1–S4.

Having genomic data from all three taxa allowed us to estimate three-taxa population branch statistics (PBS). *V. bachmanii* showed very high lineage-specific evolution (Figure 3D), with high background PBS and large (i.e., >>5 Mb) regions of higher divergence clustered in chromosomal locations with presumed low recombination rates (e.g., the Z chromosome[21]). By contrast, and as expected based on low F_{ST} values, PBS values among *V. cyanoptera* and *V. chrysoptera* were very low (Figures 3E and 3F).

We found the ABBA-BABA test of introgression (D-statistic) was effectively zero and non-significant, consistent with a lack of historical admixture between *V. bachmanii* and *V. cyanoptera* or *V. chrysoptera* (across 12,201 blocks of 100 kb: D = −0.0004, p = 0.0624). Thus, taken together, while there are phenotypic similarities between *V. bachmanii* and some of the contemporary *Vermivora* hybrid phenotypes (e.g., Lawrence's warbler), our data suggest this similarity is not likely the result of historical introgression. Indeed, the genomic evidence is consistent with *V. bachmanii* having been a highly divergent, reproductively isolated species.

Evolution at pigmentation genes in *Vermivora*

That said, it is likely that parallel evolution at common gene targets is responsible for the phenotypic similarity across the genus. For instance, we compared PBS for *V. bachmanii* to the F_{ST} values contrasting *V. cyanoptera* and *V. chrysoptera* for two key pigmentation genes (*ASIP* and *beta-carotene oxygenase 2* [*BCO2*]), which have been the focus of previous work in

Please cite this article in press as: Wood et al., Genomes of the extinct Bachman's warbler show high divergence and no evidence of admixture with other extant *Vermivora* warblers, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.05.058

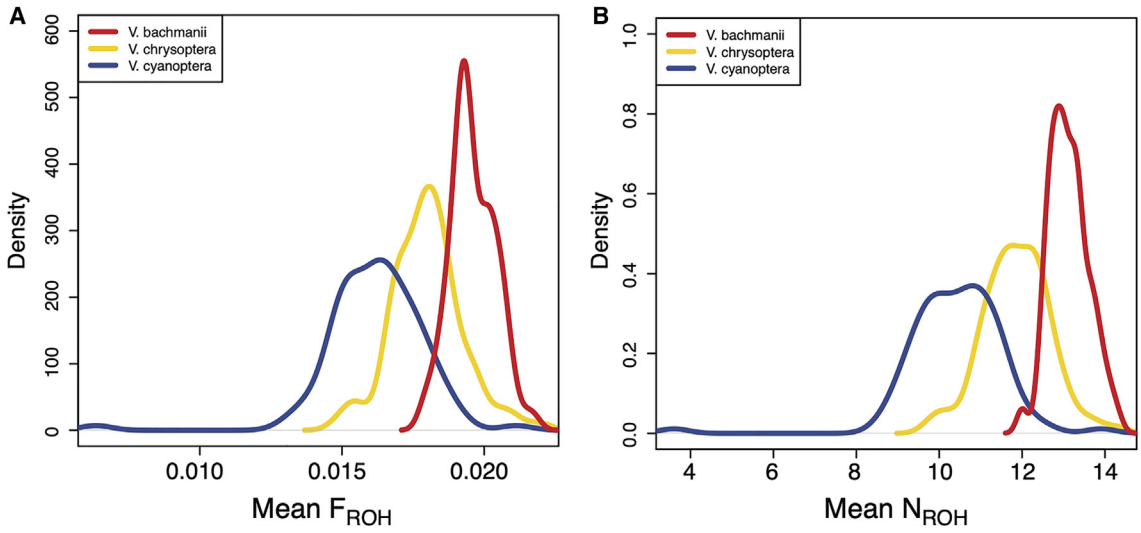


Figure 4. Distribution of runs of homozygosity across the three *Vermivora* species

Mean F_{ROH} (A), the fraction of the genome covered by runs of homozygosity (ROH), and (B) mean N_{ROH}, the total number of ROH segments, among *V. cyanoptera*, *V. chrysoptera*, and *V. bachmanii* individuals. The distributions show the estimated mean F_{ROH} and N_{ROH} values across 100 replicates, which down-sampled each species down to 52,783 sites across the autosomes. Each statistic was computed from individual ROH calls $\geq$ 1 Mbp long. Individual values can be found in Figure S3.

warblers (Figure S1).[8,10] Although there is high variability in PBS values for *V. bachmanii*, it has elevated divergence near *ASIP* (Figure S1A) at the same region where *V. cyanoptera* and *V. chrysoptera* are divergent. This is a common pattern with *ASIP*: multiple avian lineages, including other warbler species showing melanic differences, have clear evidence of repeated, independent divergence at *ASIP*.[10,22–25] We note, however, the distribution of melanin in *V. bachmanii* throat patches is qualitatively different from *V. chrysoptera*, with some *V. bachmanii* patches extending fully down the breast to the belly (Figure 1C), implying a different expression pattern.

The patterns of lineage-specific evolution at *BCO2* across the three *Vermivora* speaks to the broader evolution of this gene. First, unlike *ASIP*, there is no elevated PBS for *V. bachmanii* at *BCO2* (Figure S2). For the extant *Vermivora*, PBS values are approximately twice as high in *V. chrysoptera* as in *V. cyanoptera* for this region (maximum *V. chrysoptera* PBS for *BCO2* = 0.25, maximum PBS for *V. cyanoptera* = 0.11; Figures 3E and 3F). The evolution of *BCO2* is presumed to be linked to the distribution of body carotenoids, which is extensive for both *V. cyanoptera* and *V. bachmanii*, but not *V. chrysoptera*. The BCO2 protein degrades carotenoids, therefore its reduced action relates to an increase in observable yellow carotenoids in the feathers.[26] Taken together, these results suggest that the broad distribution of yellow coloration was an ancestral state in *Vermivora*, with lineage-specific evolution in *V. chrysoptera* representing the derived state (i.e., higher *BCO2* expression).

The inclusion of *V. bachmanii* in the PBS analysis also facilitated the discovery of an entirely novel pigmentation gene candidate in *Vermivora* (and parulids more generally). Specifically, the elevated lineage-specific evolution on chromosome 4 in *V. chrysoptera* (Figure S2; between 25,545,000 and 25,625,000 bp) was not previously identified as a strong outlier in F_{ST} with *V. cyanoptera* (Figure 3C). The maximum PBS in this region for *V. chrysoptera* is 0.35, whereas the maximum value is −0.04 in *V. cyanoptera*. This divergent region overlaps with only a single gene, *CORIN*. The gene product of *CORIN* is a modifier of the agouti pathway and, in mice, acts downstream of *ASIP* as a suppressor.[27] In *Vermivora* warblers, variation in the promoter region of *ASIP* has been linked to melanic variation in throat and mask color.[8,9] *CORIN* has also been implicated in the melanogenesis pathway of other birds, like crows[28] and white-eyes,[29] as well as pelage color variation in big cats.[30] In mice, binding of *ASIP* to *MC1R* reduces its activity and results in a shift from the production of eumelanin to the synthesis of pheomelanin.[27] This points to the possible action of *CORIN* suppressing agouti protein activity through proteolytic action. Thus, increased *CORIN* expression in *V. chrysoptera* could result in suppression of *ASIP* and, subsequently, higher melanin expression in the throat and mask.

Runs of homozygosity

Our analysis of long ROH suggested that contemporary *V. chrysoptera* and *V. cyanoptera* may have experienced a small effective population size or a population bottleneck and possibly inbreeding similar to historical *V. bachmanii* (Figure 4). The timing of a bottleneck or inbreeding is unclear, though demographic modeling of *V. bachmanii* showed that the species last exhibited a pattern of population expansion in the Late Pleistocene.[15] We restricted our ROH calls to only long ROH (>1 Mbp), as shorter ROH are unlikely to be accurately identified with sparse data. We also down-sampled our data, such that each species had the same number of SNPs analyzed and permuted 100 times. For each individual, we show F_{ROH}, which is the sum length of all ROH divided by the autosomal genome length, and N_{ROH}, which is the total number of ROH segments in that individual. For each replicate, we calculated mean F_{ROH} and mean N_{ROH} across individuals for each species (Figure 4). Although we found similar levels of total genomic ROH among these species, *V. bachmanii*

consistently had higher mean N_{ROH} and F_{ROH}. This was likely driven by one *V. bachmanii* sample (AMNH-759216) that had the highest N_{ROH} and F_{ROH} of all the individuals (Figure S3).

ROH can harbor increased levels of deleterious homozygotes,[31,32] and therefore their prevalence is important for studies of inbreeding depression in wild[33,34] and domesticated[35,36] animals, with potential implications for the long-term survival of the population/species. Whether the pattern among extant *Vermivora* indicates that *their* fate may be similar to that of *V. bachmanii* is unclear. At least among extant *Vermivora*, population declines of *V. chrysoptera* have been the most concerning, particularly across the Appalachians.[37,38] However, an important caveat to consider is that, given the low-sequencing coverage of these samples, ROH calling is likely to be noisy. As these samples were all sequenced to approximately the same coverage and processed with the same filters, we expect that these ROH are comparable among those analyzed here, but suggest that future, higher-coverage studies should carefully investigate these patterns.

Conclusions

Our study and others like it spotlight the invaluable stores of knowledge that rest in natural history collections, while also highlighting how the extinction of *V. bachmanii* represents the loss of an evolutionarily divergent and presumably fully reproductively isolated taxon. In 1890, C.S. Gailbraith, the researcher that collected one of Bachman's warbler sequenced here, likely did not anticipate its pending extinction. Moreover, he could never have imagined the tools available to extract new knowledge from its tissues as we have here. Investments in museums return healthy dividends in normal times; in the midst of the Anthropocene extinction the value of such investments are incalculable.

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
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Vermivora bachmanii
  - *Vermivora chrysoptera* and *V. cyanoptera*
  - Seiurus aurocapilla
- METHOD DETAILS
  - Sampling, DNA, Library Prep, and Sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2023.05.058.

ACKNOWLEDGMENTS

We thank both the CUMV and AMNH for access to specimens. We thank J. Walsh for assistance with the toepad extraction and comments on an earlier version. This work was supported by NSF-DEB (award# 2131469 to D.P.L.T.) and NSF-DBI (award# 2029955 to B.T.S.), the Huck Institutes of the Life Sciences and the PSU's Eberly College of Science (D.P.L.T. and Z.A.S.). Computations were performed using the PSU's Institute for Computational Data Sciences' Roar supercomputer.

AUTHOR CONTRIBUTIONS

A.W.W. and D.P.L.T. designed the study, with input from all authors; D.P.L.T. and B.T.S. collected samples, and D.P.L.T. and A.W.W. processed samples and prepared sequencing libraries; A.W.W., D.P.L.T., and Z.A.S. conducted data analysis; A.W.W. and D.P.L.T. wrote the manuscript, and all authors revised and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 23, 2023
Revised: April 25, 2023
Accepted: May 25, 2023
Published: June 16, 2023

REFERENCES

1. Brewster, W. (1891). Notes on Bachman's warbler (*Helminthophila bachmani*). Auk *8*, 149–157.
2. Hamel, P.B. (2020). Bachman's warbler (*Vermivora bachmanii*) version 1.0. In Birds of the World, A.F. Poole, ed. (Cornell Laboratory of Ornithology). https://doi.org/10.2173/bow.bacwar.01.
3. IUCN (2023). The IUCN red List of threatened species. Version 2022-2. https://www.iucnredlist.org.
4. Department of the Interior; Fish and Wildlife Service (2021). Federal register. Vol. 86, No. 187. Docket FWS–R4–ES–2020–0110. https://www.govinfo.gov/content/pkg/FR-2021-09-30/pdf/2021-21219.pdf.
5. Gill, F.B. (1980). Historical aspects of hybridization between Blue-winged and golden-winged Warblers. Auk *97*, 1–18.
6. 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.
7. Faxon, W. (1913). Brewster's warbler (Helminthophila leucobronchlais). A hybrid between the golden-winged warbler (H. chrysoptera) and the blue-winged warbler (H. pinus). Memoirs of the Museum of Comparative Zoology, at Harvard College*40* (Harvard University, Museum of Comparative Zoology), pp. 311–316.
8. 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.
9. Baiz, M.D., Wood, A.W., and Toews, D.P.L. (2021). Rare hybrid solves "genetic problem" of linked plumage traits. Ecology *102*. e03424.
10. Baiz, M.D., Wood, A.W., Brelsford, A., Lovette, I.J., and Toews, D.P.L. (2021). Pigmentation genes show evidence of repeated divergence and multiple bouts of introgression in *Setophaga* Warblers. Curr. Biol. *31*, 643–649.e3.
11. Hamel, P.B. (2018). Bachman's Warbler, A Species in Peril, Second Edition (U.S. Department of Agriculture Forest Service, Southern Research Station).
12. Freeman, B.G., and Montgomery, G.A. (2017). Using song playback experiments to measure species recognition between geographically isolated populations: a comparison with acoustic trait analyses. Auk *134*, 857–870.
13. Remsen, J.V. (1986). Was Bachman's warbler a bamboo specialist? Auk *103*, 216–219.

14. Lovette, I.J., Pérez-Emán, J.L., Sullivan, J.P., Banks, R.C., Fiorentino, I., Córdoba-Córdoba, S., Echeverry-Galvis, M., Barker, F.K., Burns, K.J., Klicka, J., et al. (2010). A comprehensive multilocus phylogeny for the wood-Warblers and a revised classification of the Parulidae (Aves). Mol. Phylogenet. Evol. *57*, 753–770.
15. Smith, B.T., Gehara, M., and Harvey, M.G. (2021). The demography of extinction in eastern North American birds. Proc. Biol. Sci. *288*. 20201945.
16. Ceballos, F.C., Joshi, P.K., Clark, D.W., Ramsay, M., and Wilson, J.F. (2018). Runs of homozygosity: windows into population history and trait architecture. Nat. Rev. Genet. *19*, 220–234.
17. Kardos, M., Luikart, G., and Allendorf, F.W. (2015). Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. Heredity *115*, 63–72.
18. Wood, A., Szpiech, Z.A., Lovette, I.J., Smith, B.T., and Toews, D.P.L. (2023). Genomes of the Extinct Bachman's warbler show high divergence and no evidence of admixture with other extant Vermivora warblers. Dryad Digital Repository. https://doi.org/10.5061/dryad.02v6wwq7h.
19. Price, T. (2008). Speciation in Birds (Roberts and Company).
20. Burri, R. (2017). Interpreting differentiation landscapes in the light of long--term linked selection. Evol. Lett. *1*, 118–131.
21. Irwin, D.E. (2018). Sex chromosomes and speciation in birds and other ZW systems. Mol. Ecol. *27*, 3831–3851.
22. 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. *4*, 502–515.
23. 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.
24. Semenov, G.A., Linck, E., Enbody, E.D., Harris, R.B., Khaydarov, D.R., Alström, P., Andersson, L., and Taylor, S.A. (2021). Asymmetric introgression reveals the genetic architecture of a plumage trait. Nat. Commun. *12*, 1019.
25. Campagna, L., and Toews, D.P.L. (2022). The genomics of adaptation in birds. Curr. Biol. *32*. R1173–R1186.
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. Enshell-Seijffers, D., Lindon, C., and Morgan, B.A. (2008). The serine protease Corin is a novel modifier of the Agouti pathway. Development *135*, 217–225.
28. Poelstra, J.W., Vijay, N., Bossu, C.M., Lantz, H., Ryll, B., Müller, I., Baglione, V., Unneberg, P., Wikelski, M., Grabherr, M.G., et al. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science *344*, 1410–1414.
29. Bourgeois, Y.X., Bertrand, J.A., Delahaie, B., Cornuault, J., Duval, T., Milá, B., and Thébaud, C. (2016). Candidate gene analysis suggests untapped genetic complexity in melanin-based pigmentation in birds. J. Hered. *107*, 327–335.
30. Xu, X., Dong, G.X., Schmidt-Küntzel, A., Zhang, X.L., Zhuang, Y., Fang, R., Sun, X., Hu, X.S., Zhang, T.Y., Yang, H.D., et al. (2017). The genetics of tiger pelage color variations. Cell Res. *27*, 954–957.
31. Szpiech, Z.A., Xu, J., Pemberton, T.J., Peng, W., Zöllner, S., Rosenberg, N.A., and Li, J.Z. (2013). Long runs of homozygosity are enriched for deleterious variation. Am. J. Hum. Genet. *93*, 90–102.
32. Szpiech, Z.A., Mak, A.C.Y., White, M.J., Hu, D., Eng, C., Burchard, E.G., and Hernandez, R.D. (2019). Ancestry-dependent enrichment of deleterious homozygotes in runs of homozygosity. Am. J. Hum. Genet. *105*, 747–762.
33. Stoffel, M.A., Johnston, S.E., Pilkington, J.G., and Pemberton, J.M. (2021). Mutation load decreases with haplotype age in wild Soay sheep. Evol. Lett. *5*, 187–195.
34. Nguyen, T.N., Chen, N., Cosgrove, E.J., Bowman, R., Fitzpatrick, J.W., and Clark, A.G. (2022). Dynamics of reduced genetic diversity in increasingly fragmented populations of Florida scrub jays, *Aphelocoma coerulescens*. Evol. Appl. *15*, 1018–1027.
35. Curik, I., Ferenčaković, M., and Sölkner, J. (2014). Inbreeding and runs of homozygosity: a possible solution to an old problem. Livest. Sci. *166*, 26–34.
36. Mooney, J.A., Yohannes, A., and Lohmueller, K.E. (2021). The impact of identity by descent on fitness and disease in dogs. Proc. Natl. Acad. Sci. USA *118*. e2019116118.
37. Rosenberg, K.V., Will, T., Buehler, D.A., Swarthout, S.B., Thogmartin, W.E., Bennett, R.E., and Chandler, R. (2016). Dynamic distributions and population declines of golden-winged warblers. Chapter 1. In Studies in Avian Biology, *49*, pp. 3–28.
38. Kramer, G.R., Andersen, D.E., Buehler, D.A., Wood, P.B., Peterson, S.M., Lehman, J.A., Aldinger, K.R., Bulluck, L.P., Harding, S., Jones, J.A., et al. (2018). Population trends in Vermivora Warblers are linked to strong migratory connectivity. Proc. Natl. Acad. Sci. USA *115*. E3192–E3200.
39. Klicka, J., Burns, K., and Spellman, G.M. (2007). Defining a monophyletic Cardinalini: a molecular perspective. Mol. Phylogenet. Evol. *45*, 1014–1032.
40. Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes *9*, 88.
41. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. BMC Bioinformatics *15*, 356.
42. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.
43. Szpiech, Z.A., Blant, A., and Pemberton, T.J. (2017). GARLIC: genomic autozygosity regions likelihood-based inference and classification. Bioinformatics *33*, 2059–2062.
44. Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Van der Auwera, G.A., Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D., et al. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. Preprint at bioRxiv. https://doi.org/10.1101/201178.
45. Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. Mol. Biol. Evol. *38*, 3022–3027.
46. Meisner, J., and Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. Genetics *210*, 719–731.
47. Okonechnikov, K., Conesa, A., and García-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics *32*, 292–294.
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. 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.
50. Weir, J.T., and Schluter, D. (2008). Calibrating the avian molecular clock. Mol. Ecol. *17*, 2321–2328.
51. Robinson, J.A., Ortega-Del Vecchyo, D., Fan, Z., Kim, B.Y., vonHoldt, B.M., Marsden, C.D., Lohmueller, K.E., and Wayne, R.K. (2016). Genomic flatlining in the endangered island fox. Curr. Biol. *26*, 1183–1189.
52. Purfield, D.C., Berry, D.P., McParland, S., and Bradley, D.G. (2012). Runs of homozygosity and population history in cattle. BMC Genet. *13*, 70.
53. Peripolli, E., Munari, D.P., Silva, M.V.G.B., Lima, A.L.F., Irgang, R., and Baldi, F. (2017). Runs of homozygosity: current knowledge and applications in livestock. Anim. Genet. *48*, 255–271.

Please cite this article in press as: Wood et al., Genomes of the extinct Bachman's warbler show high divergence and no evidence of admixture with other extant *Vermivora* warblers, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.05.058

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Bachman's warbler tissue samples	This paper	Table S1, Dryad https://doi.org/10.5061/dryad.02v6wwq7h
Ovenbird blood samples	This Paper	Dryad https://doi.org/10.5061/dryad.02v6wwq7h
Chemicals, Peptides, and Recombinant Proteins		
Qiagen DNeasy Blood and Tissue kit	Qiagen	Cat# 69504
QIAquick Spin Columns	Qiagen	Cat# QIA28115EA
Qubit Fluorometer	ThermoFisher Scientific	Cat#Q33239
TruSeq Nano DNA	Illumina	20015964
Deposited Data		
Reference *Setophaga coronata* genome	Baiz et al.[10]	NCBI Bioproject: PRJNA325157
Other *Vermivora* warbler DNA sequences	Toews et al.[8]	NCBI SRA: SAMN05223487, SAMN05223503, SAMN05223495, SAMN05223492, SAMN05223501, SAMN05223528, SAMN05223509, SAMN05223523, SAMN05223506, SAMN05223519, SAMN05223514, SAMN05223497, SAMN05223499, SAMN05223526, SAMN05223507, SAMN05223493, SAMN05223508, SAMN05223518, SAMN05223489, SAMN05223512, SAMN05223527, SAMN14431356
Bachman's resequencing data	This paper	NCBI SRA: SAMN34896771, SAMN34896772, SAMN34896773, SAMN34896774, SAMN34896775, SAMN34896776, SAMN34896777
Ovenbird resequencing data	This paper	NCBI SRA: SAMN34896778 SAMN34896779 SAMN34896780 SAMN34896781 SAMN34896782
Setophaga and *Vermivora cytochrome B* data	Lovette et al.,[14] Klicka et al.[39]	NCBI Genbank: GU932368, GU932369, EU815682, GU932365, EF529955
Software and Algorithms		
AdapterRemoval 2.1.1	Schubert et al.[40]	https://github.com/MikkelSchubert/adapterremoval
ANGSD 0.929	Korneliussen et al.[41]	http://www.popgen.dk/angsd/index.php/ANGSD
BowTie2 2.3.5.1	Langmead and Salzberg[42]	http://bowtie-bio.sourceforge.net/bowtie2/

(*Continued on next page*)

Please cite this article in press as: Wood et al., Genomes of the extinct Bachman's warbler show high divergence and no evidence of admixture with other extant *Vermivora* warblers, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.05.058

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
GARLIC v1.1.6a	Szpiech et al.[43]	https://github.com/szpiech/garlic/
GATK v4.2.6.1 HaplotypeCaller	Poplin et al.[44]	https://gatk.broadinstitute.org/hc/en-us
MEGA (11.0.13)	Tamura et al.[45]	https://www.megasoftware.net/
PCAngsd 1.10	Meisner and Albrechtsen[46]	https://github.com/Rosemeis/pcangsd
PicardTools 2.20.8	Broad Institute	https://broadinstitute.github.io/picard/
Qualimap	Okonechnikov et al.[47]	http://www.qualimap.conesalab.org/
Samtools 0.1.18	Li et al.[48]	http://www.htslib.org/
Seqtk 1.3	https://github.com/lh3	https://github.com/lh3/seqtk

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

- Raw genomic data have been deposited at the NCBI Short Read Archive (SRA) and are publicly available. Accession numbers are listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Vermivora bachmanii

Toepads of *V. bachmanii* were obtained from museum skins held at the Cornell Museum of Vertebrates (CUMV; $n = 1$) and the American Museum of Natural History (AMNH; $n = 6$). All were adult individuals and two of the seven were female. Location and sampling information can be found in Table S1 and Wood et al.[18]

Vermivora chrysoptera and *V. cyanoptera*

Sequence data from previously published adult male *Vermivora* was also included in the present paper. Those samples were studied under procedures approved in accordance with those outlined IACUC at Cornell University (#015-0065).

Seiurus aurocapilla

Blood samples were obtained from male, adult ovenbirds (*Seiurus aurocapilla*; $n = 5$) in New York and Pennsylvania. Detailed information on sampling location can be found in the Dryad Repository.[18] We collected blood samples from the brachial vein and stored them in Queen's lysis buffer under approved IACUC from Pennsylvania State University (#201900879).

METHOD DETAILS

Sampling, DNA, Library Prep, and Sequencing

Toepad specimens were extracted in dedicated labs for working on historical samples at the AMNH and CUMV. For the pilot study (the CUMV sample), we used a phenol-chloroform DNA extraction method. For the AMNH samples, we used the Qiagen DNeasy Blood and Tissue Kit with several modifications to improve DNA yields that included washing each toe pad sample with H_2O and EtOH, a longer digestion time, and using QIAquick PCR spin columns to improve size selection for highly fragmented DNA from historical museum specimens. The DNeasy spin columns size select for a fragment range of 50 kb to 100 bp, whereas the QIAquick PCR spin columns size select for a range of 10 kb to 100 bp.

The samples were sequenced across two separate NextSeq lanes. The CUMV sample was sequenced in 2018 as part of a pilot project conducted at the Cornell Institute for Biotechnology. The AMNH samples were sequenced in 2022 at the Penn State Genomics Core facility. We followed the protocol for the Illumina TruSeq Nano library preparation kit, with two important modifications.

Please cite this article in press as: Wood et al., Genomes of the extinct Bachman's warbler show high divergence and no evidence of admixture with other extant *Vermivora* warblers, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.05.058

First, given the DNA was already degraded, we did not mechanically shear the DNA. Second, we did not perform a SPRI bead size selection, with the justification that omitting size selection would result in smaller fragments, but would produce more resultant reads for analysis. Therefore, in place of the size selection step, we used undiluted SPRI beads. For the CUMV sample, it was multiplexed with 23 other individual birds from a variety of projects, which were individually indexed and pooled on an Illumnia NextSeq 500 lane using P2 paired-end 150bp sequencing chemistry. The six AMNH samples were run separately—the only samples included—on a single lane of a Illumina NextSeq2000 with P2 150bp paired end chemistry.

QUANTIFICATION AND STATISTICAL ANALYSIS

We first used the program AdapterRemoval[40] to trim and collapse overlapping paired reads with the following options: ''-collapse-trimns -minlength 20 -qualitybase 33.'' We then aligned all read data to the Myrtle Warbler (*Setophaga coronata coronata*) reference genome.[10] We used BowTie2[42] to align reads to the reference with the ''very-sensitive-local'' presets and set the option ''-X'' (the maximum fragment length for valid paired-end alignments) to 700 bp. We then used PicardTools (2.20.8;https://broadinstitute.github.io/picard/) to mark PCR duplicates. We quantified genomic coverage using ''qualimap''[47]. Because of the fragmented nature of the input DNA, most of the reads for each sample were merged

We used PCAngsd (version 0.981[46]) to conduct principal component analysis on the genotype likelihood data for each region separately and plotted the first two eigenvectors. Within ANGSD,[41] we first generated a ''beagle'' input file using the flags ''-GL 2 -doGlf 2 -doMajorMinor 1 -SNP_pval 1e-6 -doMaf 1'' and then ran PCAngsd with the default parameters.

To compare patterns of divergence among the three species across the entire genome we used both F_{ST} estimates and population branch statistics (PBS). To estimate F_{ST}, we used ANGSD,[41] 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'' in ANGSD. To calculate lineage-specific evolution (i.e., PBS) for each of the three *Vermivora* species, we used a 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 ''realSFS'' with the options ''-win 10000 -step 10000'' to calculate the non-overlapping sliding-window PBS statistic for each of the three species.

To calculate D-statistics for the ABBA-BABA test, we used the ''-doAbbababa2'' function, also in ANGSD, using the options ''-blockSize 100000 -do-Counts 1.'' For the analysis we used *V. cyanoptera* and *V. chrysoptera* as populations one and two (H1, H2, respectively), *V. bachmanii* samples as population three (H3), and *Seiurus aurocapilla* as the outgroup (H4) using the ''-useLast 1'' option. A significant positive D statistic in this case would indicate introgression between H2 and H3.

To estimate the divergence time among *Vermivora*, we first extracted high coverage *cytochrome B* (*cytB*) sequence from a single *V. bachmanii* individual (AMNH#759220, mean mtDNA coverage 344X). We used mpileup, in Samtools,[48] and seqtk (https://github.com/lh3/seqtk) to generate a fasta file for this region. We then combined this with cytB sequence from[14] of: *V. chrysoptera* (NCBI# GU932368), *V. cyanoptera* (GU932369), *Setophaga magnolia* (EU815682), and *Seiurus aurocapilla* (GU932365). We also included sequence from *Icteria virens* (EF529955) as an outgroup from Klicka et al.[39] In MEGA (11.0.13)[45] we first calculated pairwise distances among the six taxa (using the maximum composite likelihood method). We then built a maximum likelihood tree to feed into ''time-tree'' in MEGA. We then computed a timetree using the ''RelTime-ML'' method, specifying *I. virens* as the outgroup. Finally, we used a 5.8 mya estimate as a constraint at the node between *Setophaga magnolia* and *Seiurus aurocapilla* based on the estimate of Oliveros et al.[49] The estimated divergence time of *V. bachmanii* based on this method (2.82 MYA) is also consistent with a ~3.1 mya estimate based on the 2.1% *cytB* molecular clock[50] from the pairwise distances between *V. bachmanii* and *V. cyanoptera* / *V. chrysoptera*. This approach estimated the divergence time between *V. cyanoptera* / *V. chrysoptera* to 1.11 MYA.

We used GARLIC v1.1.6a to call ROH[43] as this method implements a model-based likelihood approach that incorporates genotype quality scores and handles arbitrary amounts of missing data. We first called variant sites in *V. chrysoptera* (n=10), *V. cyanoptera* (n=10), and *V. bachmanii* (n=6), excluding one *V. bachmanii* sample with mean coverage <1X (as well as the two hybrids) using GATK v4.2.6.1 HaplotypeCaller.[44]

As these samples have low average coverage (Table S1), we enforced a strict filtering scheme so that we retained only the most confidently called variants. For each species, separately, we first restricted our set of sites to only biallelic SNPs on autosomes, and we set any genotype with a read depth < 6 or a genotype quality < 20 to missing. Next, we removed any sites with a missing data rate > 0.5. Finally, we filtered sites with excess heterozygosity, as these sites are likely to be sequencing errors. However, with low sample sizes statistical tests to detect excess heterozygosity have no power, so we resort to filtering any site where >= 80% of samples have a heterozygote called (a similar strategy as in Robinson et al.[51]). For *V. chrysoptera* and *V. cyanoptera* samples this means any site with 8 or more heterozygotes, and for *V. bachmanii* this means any site with 5 or more heterozygotes. These filters result in 89,207, 179,495, and 60,124 variable sites for *V. chrysoptera*, *V. cyanoptera*, and *V. bachmanii*, respectively.

To then address possible biases in our ROH calling due to wide range of total number of variable sites per species, we subsampled our data to a constant number of loci across each species. We subsampled in the following way. First, for each chromosome we found the minimum number of sites observed across the three species and set this as our target number of sites for subsampling. Taking this approach resulted in 52,783 sites across all autosomes. Next, for each chromosome, we randomly selected sites (without replacement) until we reached the target number, resulting in a data set of 52,783 sites across all

autosomes for each species. Previous work has shown is it possible to call ROH at these SNP densities.[52,53] Each species was called separately with the following GARLIC parameters set: "–max-gap 1000000, –overlap-frac 0.05, –resample 100, and –auto-winsize". GARLIC chose the best window size for each species, resulting in a window size of 20 SNPs for each species. To estimate the variability introduced by this subsampling approach we then replicated this process 100 times and re-called ROH using the above parameters for each replicate of each species. We calculated mean F_{ROH} and N_{ROH} for each species for each replicate and plotted this distribution (Figure 4).