Selection on *VPS13A* linked to migration in a songbird

David P. L. Toews[a,1], Scott A. Taylor[b], Henry M. Streby[c], Gunnar R. Kramer[c], and Irby J. Lovette[d]

[a]Department of Biology, The Pennsylvania State University, University Park, PA 16802; [b]Department of Ecology and Evolutionary Biology, University of Colorado Boulder, Boulder, CO 80309; [c]Department of Environmental Sciences, University of Toledo, Toledo, OH 43606; and [d]Cornell Lab of Ornithology, Cornell University, Ithaca, NY 14850

Edited by Scott V. Edwards, Harvard University, Cambridge, MA, and approved August 8, 2019 (received for review June 4, 2019)

Animal migration demands an interconnected suite of adaptations for individuals to navigate over long distances. This trait complex is crucial for small birds whose migratory behaviors—such as directionality—are more likely innate, rather than being learned as in many longer-lived birds. Identifying causal genes has been a central goal of migration ecology, and this endeavor has been furthered by genome-scale comparisons. However, even the most successful studies of migration genetics have achieved low-resolution associations, identifying large chromosomal regions that encompass hundreds of genes, one or more of which might be causal. Here we leverage the genomic similarity among golden-winged (*Vermivora chrysoptera*) and blue-winged (*V. cyanoptera*) warblers to identify a single gene—*vacuolar protein sorting 13A* (*VPS13A*)—that is associated with distinct differences in migration to Central American (CA) or South American (SA) wintering areas. We find reduced sequence variation in this gene region for SA wintering birds, and show this is the likely result of natural selection on this locus. In humans, variants of *VPS13A* are linked to the neurodegenerative disorder chorea-acanthocytosis. This association provides one of the strongest gene-level associations with avian migration differences.

evolution | genomics | migration

Animal migration requires integration across a suite of traits to allow organisms to move between habitats separated by long distances (1). Evidence suggests many short-lived migratory animals navigate biannual migrations by relying on innate machinery (2). Yet, the specific genes that underlie these traits have been challenging to identify (3). The most successful studies of avian migratory directionality, for example, have identified associations with large genomic regions, representing hundreds of genes, one or more of which might be causal (4, 5).

Here we combine genome sequencing, migration tracking, and an animal system—*Vermivora* warblers—with an exceptional utility for discovering possible gene−migration associations. This utility stems from the fact that much of the genetic variation in golden-winged (*V. chrysoptera*, hereafter "*chryso*") and blue-winged (*V. cyanoptera,* hereafter "*cyano*") warblers is shared—even between individuals differing strongly in their plumage—due to a history of hybridization (6). Both breed in North America and winter in 1 of 2 neotropical regions: in South America (SA), primarily in Venezuela, or in Central America (CA), from Panama to Guatemala (7, 8). These wintering locations are predictable based on breeding locations and plumage phenotype: *chryso* in Appalachia winter in SA, whereas *chryso* breeding outside of Appalachia—as well as most *cyano*—primarily winter in CA (ref. 8 and Fig. 1).

These patterns are not simply the result of a shift in overall migration directionality. SA birds undergo a more prolonged migration, with tracks that cluster into a small geographic region as they move across the Isthmus of Panama (7). Also, SA birds tend to migrate through Florida, unlike CA wintering warblers that are more likely to circumvent the Gulf of Mexico or cross it directly.

To test for associations between single nucleotide polymorphisms (SNPs) and wintering locations, we used resequenced genomes of 70 *Vermivora* warblers classified as wintering in either CA or SA. These samples either had geolocator tracks associated with them to determine overwintering location ($n = 44$; ref. 8), or were reliably classified based on breeding location/plumage ($n = 26$). Many SNPs associated with overwintering behavior clustered within a 120-kb region of the Z chromosome (Fig. 2*C*). There is no single SNP with a highly significant association (i.e., $-\log_{10}(P) > 7$) in this small region. However, there are an exceptionally high number of moderately significant SNPs: In one 10-kb window, there are 51 SNPs with $-\log_{10}(P) > 5$; by comparison, the next highest window in the genome has only 7 such SNPs.

This region of the Z chromosome is an outlier in allele frequencies between CA and SA birds (Fig. 2*A*; region $F_{ST} = 0.22$, genome average = 0.002) and contains the top 12 most differentiated F_{ST} windows. Evidence suggests this pattern has likely been produced by natural selection in SA birds: 1) This region has reduced sequence variation in SA birds compared to those wintering in CA (region mean $\pi_{SA} = 0.0007$; mean $\pi_{CA} = 0.001$); 2) Tajima's *D*—a statistic that helps to distinguish between DNA sequences evolving neutrally versus under nonneutral processes—is much lower in SA birds compared with CA (Fig. 2*B*); and 3) this region shows elevated linkage disequilibrium compared to the surrounding chromosome. These findings are consistent with contemporary natural selection acting on SA wintering birds.

Principal components analysis (PCA) of SNPs in this region (Fig. 2*D*) clusters individuals into 3 groups. This is expected if there are 2 distinct haplotypes across this region, with some individuals heterozygous for both types (i.e., birds in the middle cluster). In this case, high values on principal component 1 (PC1) are primarily *chryso*, whereas low PC1 values comprise mostly *cyano*. SA birds have nearly uniformly high PC1 values and represent a restricted subset of variation present in CA wintering birds. We also found that individuals with intermediate PC1 values have higher heterozygosity in this region: The average percentage of heterozygous sites for birds with intermediate PC1 values (i.e., PC1 values <0.09 and > −0.09; $n = 23$) is 44.7%; for individuals with PC1 values at the ends of the spectrum (i.e., PC1 values >0.09 or < −0.09; $n = 47$) the average is 16% of sites ($P < 0.0001$, Mann−Whitney *U* test). Additional variability in PC1 scores is likely the result of recombination generating additional haplotypic variants across this region following admixture between *chryso* and *cyano*. While the positions of functionally relevant sites are not known, the clusters of points along PC1 likely provide a close approximation of the genotypes at these sites.

Author contributions: D.P.L.T. and S.A.T. designed research; D.P.L.T. performed research; D.P.L.T., H.M.S., and G.R.K. analyzed data; and D.P.L.T., S.A.T., H.M.S., G.R.K., and I.J.L. wrote the paper.

The authors declare no conflict of interest.

Data deposition: Sample information is deposited at Data Dryad, DOI: 10.5061/dryad.bs85242. Sequences are deposited at the NCBI Sequence Read Archive (SRA), http://www.ncbi.nlm.nih.gov/sra, under BioProject no. PRJNA325126.

[1]To whom correspondence may be addressed. Email: toews@psu.edu.

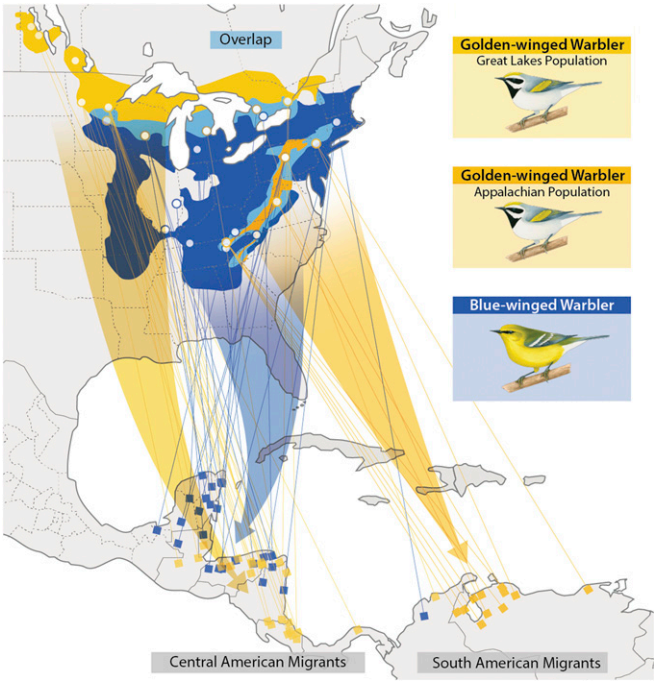

Fig. 1. Blue-winged warblers (blue) mostly migrate to CA. There are 2 breeding populations of golden-winged warblers: The Great Lakes population (light orange) also primarily migrates to CA, whereas the Appalachian population migrates to SA (darker orange). Circles show geolocator deployment; squares show predicted wintering locations. Adapted from ref. 6. Copyright (2016), with permission from Elsevier. Data from ref. 8.

Given the strong geographic connectivity between breeding, migration, and wintering locations (8), it is difficult to determine the influence of selection on individual components of the annual cycle for this chromosomal region. However, one exception within the current analysis points to differences driven by migration/wintering location: Our study included 4 *cyano* from Massachusetts (MA), all breeding east of the Appalachian Mountains. Three of these birds migrated to CA, like other *cyano*, and had low PC1 values. The fourth bird from MA, however, was the only *cyano* in our study to winter in SA. It was also one of only 2 *cyano* to have high a high PC1 value. The strength of this genomic region to predict individual wintering location will require more study, including within-site variation in migration behavior in a large number of individuals.

We cannot exclude environmental differences between CA and SA as the source of selection favoring alternative alleles in this region of the genome. However, research on the climatic and habitat associations of golden-winged warblers indicates no obvious ecological differences between the habitats occupied by CA and SA wintering birds, with both generally preferring midelevation evergreen mixed broadleaf forest (9). However, individual-level habitat data are not available from many of the nonbreeding sites, particularly in SA. Taken together, we suggest that the most obvious phenotypic difference between SA and CA birds is their distinct migration patterns, with the caveat that additional detailed studies of nonbreeding habitats could identify other correlated ecological characters in the future.

Only one gene—*VPS13A*—falls within the region associated with wintering location. *VPS13A* does not have a characterized function in birds. However, it has a known disease phenotype in humans: *VPS13A* codes for chorein, and variants of this gene are associated with the neurodegenerative disorder chorea-acanthocytosis (10). This disease is inherited as an autosomal recessive condition that includes movement disorders and dystonia. Recent work shows *VPS13A* is closely associated with mitochondria, where it may be involved in lysosomal degradation (11) and lipid transfer with the endoplasmic reticulum (12). While speculative, selection on *VPS13A* may enhance the capacity in SA wintering birds to more efficiently remove reactive oxygen species resulting from a prolonged migration (13). Determining how specific mutations in *VPS13A* might modulate variation in migratory directionality or distance, and quantifying the mechanism of selection, should be the target of future research in *Vermivora* as well as other migrant taxa.

Methods

We used blood from warblers predicted to have wintered in CA (n = 25 *chryso*, n = 23 *cyano*, and n = 2 hybrids) or SA (n = 17 *chryso*, n = 1 *cyano*, n = 2 hybrids). Most were males captured during the breeding season, although the sample also included 4 females. We outfitted 44 birds with geolocators that we retrieved during the following breeding season (i.e., one annual migration cycle). We analyzed geolocator data to derive spatially explicit probability density functions during the winter period, and presented these data previously (see refs. 7 and 8). We generated sequencing libraries with

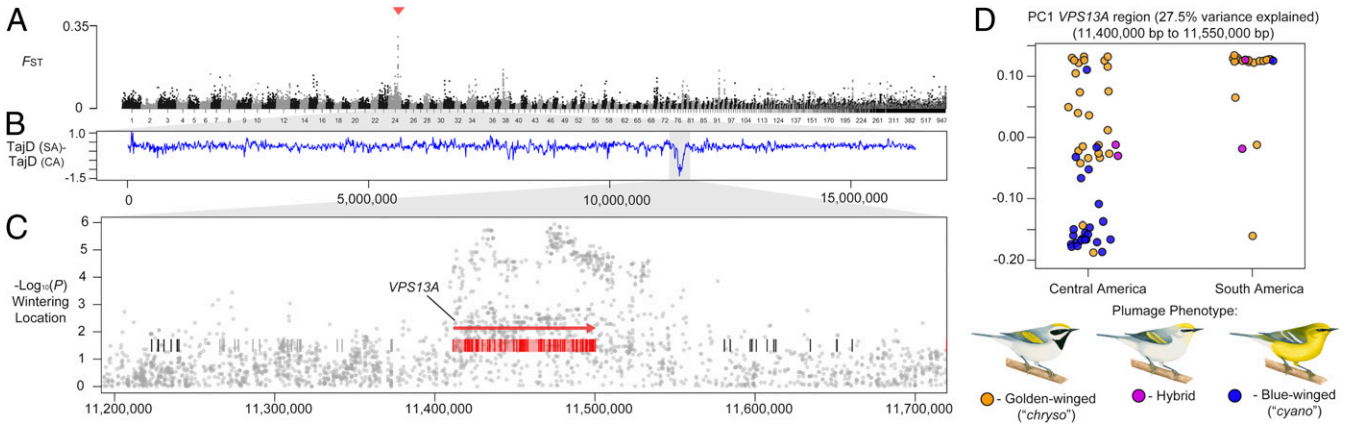

Fig. 2. Genomic variation and associations with *Vermivora* wintering locations. (*A*) F_{ST} in 10-kb nonoverlapping windows between SA and CA wintering birds, with a peak on the Z chromosome (warbler scaffold 24). (*B*) The difference in Tajima's D (TajD) between groups, with SA birds having much lower values compared to CA birds (between 11.4 Mb and 11.55 Mb). (*C*) The statistical association between SNP genotypes, wintering location, and the coding region of *VPS13A* (denoted by the red arrow and vertical lines) (*D*) PCA of the *VPS13A* region. Individuals are distinguished by wintering location and plumage phenotype. The PCA shows 3 groups; intermediate values represent more heterozygous individuals at the *VPS13A* region. SA birds have a restricted subset of variation present in other birds.

either an Illumina TruSeq PCRFree or Nano kits after ref. 6. We identified SNPs using ANGSD (14) or GATK (15). From GATK-SNPs, we estimated heterozygosity, as well as F_{ST} and π in VCFTOOLs (16) in 10-kb windows. We imputed the GATK-SNPs with BEAGLE (17) to generate input for GEMMA to compute genotype associations (18). From genotype likelihoods we estimated Tajima's D and generated input for PCAngsd (19) using ANGSD between 11.40- and 11.55-Mbp positions on warbler scaffold 24. Sample information and sequence accession numbers can be found at ref. 20.

ACKNOWLEDGMENTS. We thank D. Andersen, D. Buehler, P. Wood, S. VanWilgenburg, J. Larkin, R. Vallender, R. Ricklefs, K. Aldinger, and L. Campagna. Three anonymous reviewers provided helpful comments on previous versions of the paper. J. Ditner and L. C. Fuller produced the illustrations. This work was supported by the Natural Sciences and Engineering Council of Canada, National Science Foundation, Cornell Lab of Ornithology, US Geological Survey, US Fish and Wildlife Service, University of Minnesota, University of Tennessee, and University of Toledo.

1. T. Alerstam, *Bird Migration* (Cambridge University Press, 1990).
2. P. Berthold, Genetic control of migratory behaviour in birds. *Trends Ecol. Evol.* **6**, 254–257 (1991).
3. M. Liedvogel, S. Akesson, S. Bensch, The genetics of migration on the move. *Trends Ecol. Evol.* **26**, 561–569 (2011).
4. K. E. Delmore, D. P. L. Toews, R. R. Germain, G. L. Owens, D. E. Irwin, The genetics of seasonal migration and plumage color. *Curr. Biol.* **26**, 2167–2173 (2016).
5. M. Lundberg *et al*., Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. *Evol Lett* **1**, 155–168 (2017).
6. D. P. L. Toews *et al*., Plumage genes and little else distinguish the genomes of hybridizing warblers. *Curr. Biol.* **26**, 2313–2318 (2016).
7. G. R. Kramer *et al*., Nonbreeding isolation and population-specific migration patterns among three populations of Golden-winged Warblers. *Condor* **119**, 108–121 (2017).
8. G. R. Kramer *et al*., Population trends in *Vermivora* warblers are linked to strong migratory connectivity. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E3192–E3200 (2018).
9. K. V. Rosenberg *et al*., Dynamic distributions and population declines of golden-winged warblers. *Stud. Avian Biol.* **49**, 3–28 (2016).
10. C. Ishida, T. Makifuchi, S. Saiki, G. Hirose, M. Yamada, A neuropathological study of autosomal-dominant chorea-acanthocytosis with a mutation of *VPS13A*. *Acta Neuropathol.* **117**, 85–94 (2009).
11. S. Muñoz-Braceras, A. R. Tornero-Écija, O. Vincent, R. Escalante, VPS13A is closely associated with mitochondria and is required for efficient lysosomal degradation. *Dis. Model Mech.* **12**, dmm036681 (2019).
12. W. M. Yeshaw *et al*., Human *VPS13A* is associated with multiple organelles and influences mitochondrial morphology and lipid droplet motility. *eLife* **8**, e43561 (2019).
13. S. Jenni-Eiermann, L. Jenni, S. Smith, D. Costantini, Oxidative stress in endurance flight: An unconsidered factor in bird migration. *PLoS One* **9**, e97650 (2014).
14. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).
15. M. A. DePristo *et al*., A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
16. P. Danecek *et al*., The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
17. S. R. Browning, B. L. Browning, Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
18. X. Zhou, M. Stephens, Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* **11**, 407–409 (2014).
19. M. Fumagalli *et al*., Quantifying population genetic differentiation from next-generation sequencing data. *Genetics* **195**, 979–992 (2013).
20. D. P. L. Toews *et al*., Data from "Selection on VPS13A linked to migration in a songbird." Dryad. https://dx.doi.org/10.5061/dryad.bs85242. Accessed 16 August 2019.