



OPEN

Genomic architecture of migration timing in a long-distance migratory songbird

Evelien de Greef¹✉, Alexander Suh^{2,3}, Matt J. Thorstensen¹, Kira E. Delmore^{4,5} & Kevin C. Fraser^{1,5}

The impact of climate change on spring phenology poses risks to migratory birds, as migration timing is controlled predominantly by endogenous mechanisms. Despite recent advances in our understanding of the underlying genetic basis of migration timing, the ways that migration timing phenotypes in wild individuals may map to specific genomic regions requires further investigation. We examined the genetic architecture of migration timing in a long-distance migratory songbird (purple martin, *Progne subis subis*) by integrating genomic data with an extensive dataset of direct migratory tracks. A moderate to large amount of variance in spring migration arrival timing was explained by genomics (proportion of phenotypic variation explained by genomics = 0.74; polygenic score $R^2 = 0.24$). On chromosome 1, a region that was differentiated between migration timing phenotypes contained genes that could facilitate nocturnal flights and act as epigenetic modifiers. Overall, these results advance our understanding of the genomic underpinnings of migration timing.

Climate change affects spring phenology in temperate zones and could have significant, negative impacts on migratory animals¹. For example, migrants must synchronize arrival at breeding grounds to coincide with seasonal resources availability². These resources are becoming available earlier and it is unclear if migrants will be able to match these advances in timing, potentially leading to substantial population declines³. Migration phenology may be influenced by many factors, including extrinsic factors such as rainfall⁴ and wintering habitat conditions⁵. However, migration timing is thought to be largely endogenously controlled⁶, and thus knowledge of its genetic architecture (e.g., the identity, number, and location of genetic loci involved) is essential for predicting how migrants will respond to phenological changes that accompany climate change.

Previous genetic studies provide important insights regarding migration timing, such as in genes associated with circadian and circannual rhythms^{7–12}; however, results vary across species¹³ and are limited to small portions of the genome. Another limitation associated with earlier studies was an inability to quantify migratory behavior in the wild—prior to 2007, it was not possible to track animals < 100 g on migration¹⁴ and most migratory avian species fall into this size class. We overcame these limitations here, combining high resolution genomic data with an extensive migration tracking dataset for purple martins (*Progne subis subis*), while specifically focusing on spring arrival timing given the importance of matching timing with resources critical for breeding. The purple martin is a Nearctic-neotropical migrant that travels over 7000 km between North America and South America¹⁵ and exhibits extensive latitudinal variation in migration timing. It is thus a powerful system to study migration genomics. For example, individuals breeding in the southern edge of the range in Florida may arrive as early as mid-January, while their northern counterparts in Alberta may arrive as late as June¹⁶.

Our objective was to examine the genomic architecture of migration timing by assembling a reference genome for the purple martin and integrating sequencing data with light-level geolocator tracks. We examined results from genome-wide association studies (GWAS), to calculate the proportion of phenotype variation explained by our genomic data (PVE) along with presence of any associated single nucleotide polymorphisms (SNPs). We also calculated polygenic scores (PGS) to assess the genetic predisposition of migratory timing. Lastly, we conducted a genomic differentiation analyses between different migratory phenotypes to identify regions in the genome that may influence this trait. This study expands our understanding of the whole-genome contribution to migration and yields insight into migration timing behavior.

¹Department of Biological Sciences, University of Manitoba, Winnipeg R3T 2N2, Canada. ²Department of Organismal Biology, Uppsala University, 752 36 Uppsala, Sweden. ³School of Biological Sciences, University of East Anglia, Norwich NR4 7TU, UK. ⁴Department of Biology, Texas A&M University, College Station, TX 77843, USA. ⁵These authors contributed equally: Kira E. Delmore, and Kevin C. Fraser. ✉email: eveliendegreef@gmail.com

Results

Reference and resequencing data. The final *P. subis* reference genome assembly based on long reads and linked reads was 1.17 Gb in length, consisted of 2896 scaffolds, had an N50 scaffold length of 6.13 Mb and an N50 contig length of 3.08 Mb. The annotation included 12,686 genes (SI Appendix, Table S1). The assembly length was similar to other avian genomes, which are typically between 1.0 and 1.2 Gb¹⁷. BUSCO analysis revealed that the *P. subis* genome was relatively complete with 91% of avian orthologs detected as complete sequences (89.1% being single-copy and 1.9% being duplicated), which was in the range of other non-model avian genomes¹⁸. We aligned resequencing data for 87 individuals to this reference resulting in 4.6 million SNPs after filtering. All these individuals were tracked on migration with light-level geolocators yielding precise estimates for migratory timing.

Genomic architecture of migration timing. Birds in this study exhibited considerable latitudinal variation in migratory timing (sampling locations in Table S2), ranging over 131 days for spring arrival dates (Fig. S1). Sampling locations in the breeding grounds are displayed in Fig. 1.

Estimates of PVE from Bayesian sparse linear mixed models (BSLMs)¹⁹ were high, with a median value of 0.74 and 89% equal-tailed interval (ETI) of 0.10–1.00. PGS estimated from linear mixed models²⁰ were strongly correlated with spring migration timing (R^2 of 0.24, $p < 0.001$, Fig. 2a). We assessed predictive power of the PGS model with 100 jackknife cross-validation partitions, and found that birds with lower PGS deciles exhibited earlier spring migration timing compared with individuals in higher PGS deciles with later spring migration timing (Fig. 2b).

BSLMs did not identify any specific genomic regions linked to migratory timing (Fig. S2). However, a survey of net genomic differentiation (ΔF_{ST}) between the earliest and latest spring migrants in our dataset did reveal a region of elevated differentiation on chromosome 1 (Fig. 3). ΔF_{ST} controls for processes unrelated to migration that could elevate F_{ST} (including population structure, see “Methods”). This elevation was additionally present in comparisons of early and late migrants within populations in Florida (southernmost colony) and Alberta (northernmost colony) (Fig. 3b) suggesting population structure did not generate this pattern. Reductions in nucleotide diversity and Tajima’s D indicative of a selective sweep are also present in this region, which covers 2 Mb region and consist of 13 genes (Table S3) including *ppfia2* and *nts*, which may be related to sleep^{21,22}, and *mettl25* and *acss3* that may serve as epigenetic modifiers by adding methyl groups to DNA²³ or remodeling chromatin and regulating gene expression²⁴.

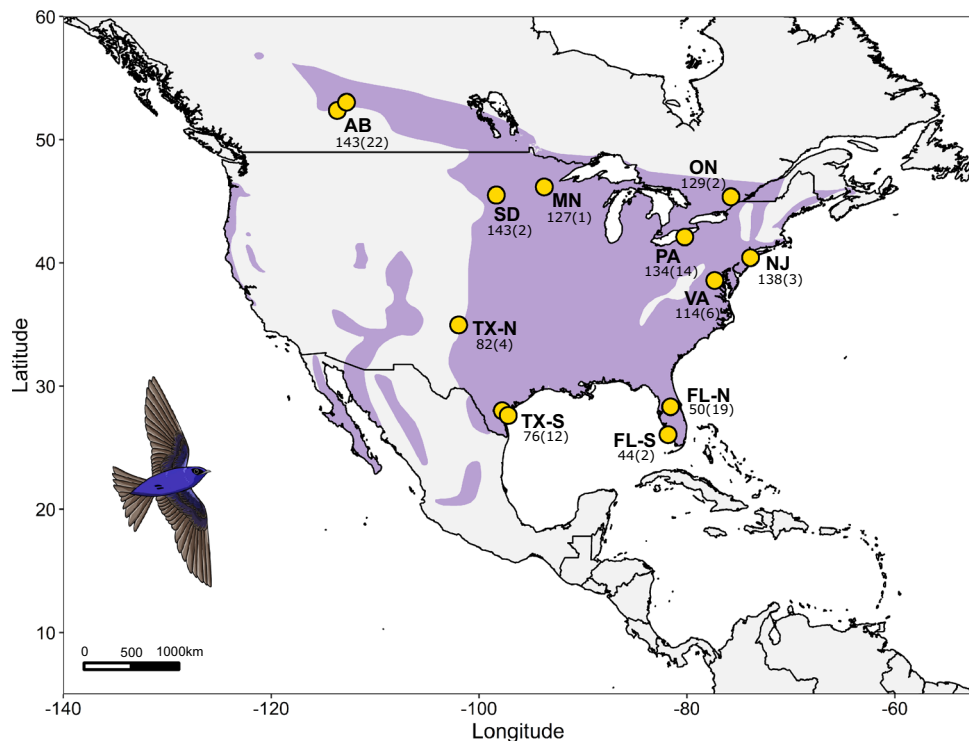


Figure 1. Purple martin breeding distribution (purple), including sampling sites for 87 individuals in their North American breeding range (yellow). The mean arrival date for each site is in ordinal date format (day of year), with the site sample size provided in parentheses. The purple martin range map layer was obtained through the IUCN red list database.

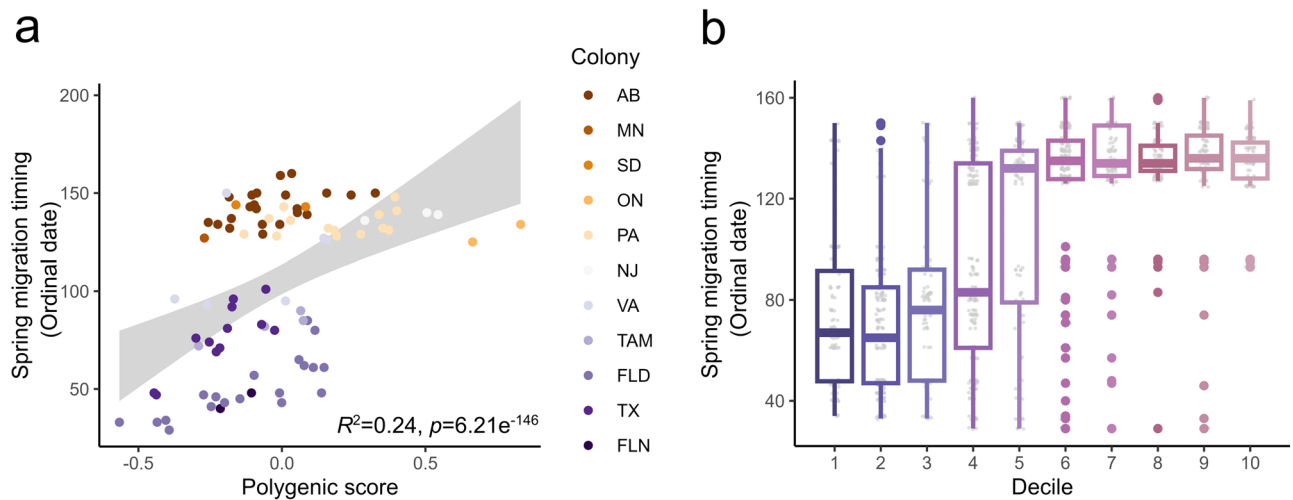


Figure 2. (a) Polygenic scores of spring migration timing for purple martins ($n = 87$) colored in order by latitude, and linear regression 95% confidence interval is colored in gray. (b) Predicted polygenic scores (PGS) partitioned by decile (lowest PGS = decile 1, highest PGS = decile 10), showing individuals (raw points in gray dots) and results summarized in box plots) in lowest deciles had earlier migration timing compared with individuals in higher deciles with later timing.

Discussion

We used one of the largest tracking datasets available for a long-distance migratory songbird and a genome-wide SNP dataset to reveal the genetic underpinnings of migration timing. The results demonstrate a genetic basis to migration timing in the purple martin. A previously unidentified 2 Mb genetically differentiated region was found on chromosome 1, illuminating a candidate genomic region underlying migration timing.

Phenotypes that vary across individuals are a result of both environmental and genetic factors, and PVE represents the proportion of variance attributed to genetic factors. The large PVE estimate (0.74) is evidence that variation in migration timing is largely influenced by genetics. The predictive utility of the polygenic model across multiple deciles showed it is possible to distinguish early and late migrants using genetic variants, which could potentially aid in estimating a birds' phenotype in the wild. With overlapping wintering grounds among purple martin colonies¹⁵, the model may also help predict an individuals' timing tendency in addition to breeding region when captured during the winter. Some species demonstrate phenotypic plasticity in migration timing²⁵, including purple martins²⁶. While the high genomic variation explaining migratory timing does not preclude phenotypic plasticity, it suggests that changes in timing would also need to occur through microevolutionary processes. It is important to understand the sources of considerable genetic variation in migratory traits⁶, and future work will inform how influences such as standing genetic variation (presence of more than one allele at a locus in a population) may play a role in facilitating rapid microevolutionary changes²⁷. While our sample size limits our estimates of PVE (the ETIs were quite wide), our PGS analysis supported the potential to predict individual migration timing with genomic data. Given the present sample size, it is remarkable that a substantial proportion of variation was explained with genomics. Whether this proportion could be even greater with larger sample sizes^{28,29} could be determined in future studies. However, collecting enough samples to capture strong statistical power in tests of genomic associations with phenotypes in field-based wildlife research will be challenging.

The lack of genomic loci in the GWAS significantly associated with spring migration timing suggests this trait is controlled by many alleles of small effect, which may have been undetected, or could have been located in assembly gaps¹⁷. However, elevated genomic differentiation on chromosome 1 provides evidence for a potential connection between migration timing in purple martins with some genes related to rest and others that could act as epigenetic modifiers. *Ppfa2* and *nts* are 2 of the 13 genes located in this region. *Ppfa2* has been linked to sleep and wakefulness in white-crowned sparrows²¹, and *nts* has been linked to sleep regulation in European mice²². While purple martins are primarily diurnal migrants, they can incorporate both day and night flights on spring migration³⁰, and the former genes could play a role in these nocturnal flights. Epigenetic modifiers may play a role in migration timing, such as methylation on the polyglutamine repeat domain of the *clock* gene affecting phenology and breeding performance³¹. Therefore, it is interesting that *mettl25* and *acss3* are in a region of high differentiation between early and late migrants. *Mettl25* may be similar to *mettl25b* and encode a methyltransferase that represses gene expression²³ and *acss3* produces acetyl-CoA which promotes gene expression by acetylating histones³². Acetyl-CoA is also important for generating, using, and storing energy²⁴. While this study suggests associations with these genes, our current data cannot pinpoint how these genes are functioning with purple martin migration timing. Further work could elucidate these mechanisms and their roles in migration timing. While estimates of F_{ST} are often considered bottom-up comparisons, we compared extreme phenotypes while controlling for population differentiation to identify genomic regions associated with migration timing. Results from this approach were further supported by a comparison of early and late migrants within populations that recovered a similar elevated pattern of genomic differentiation (Fig. 3b).

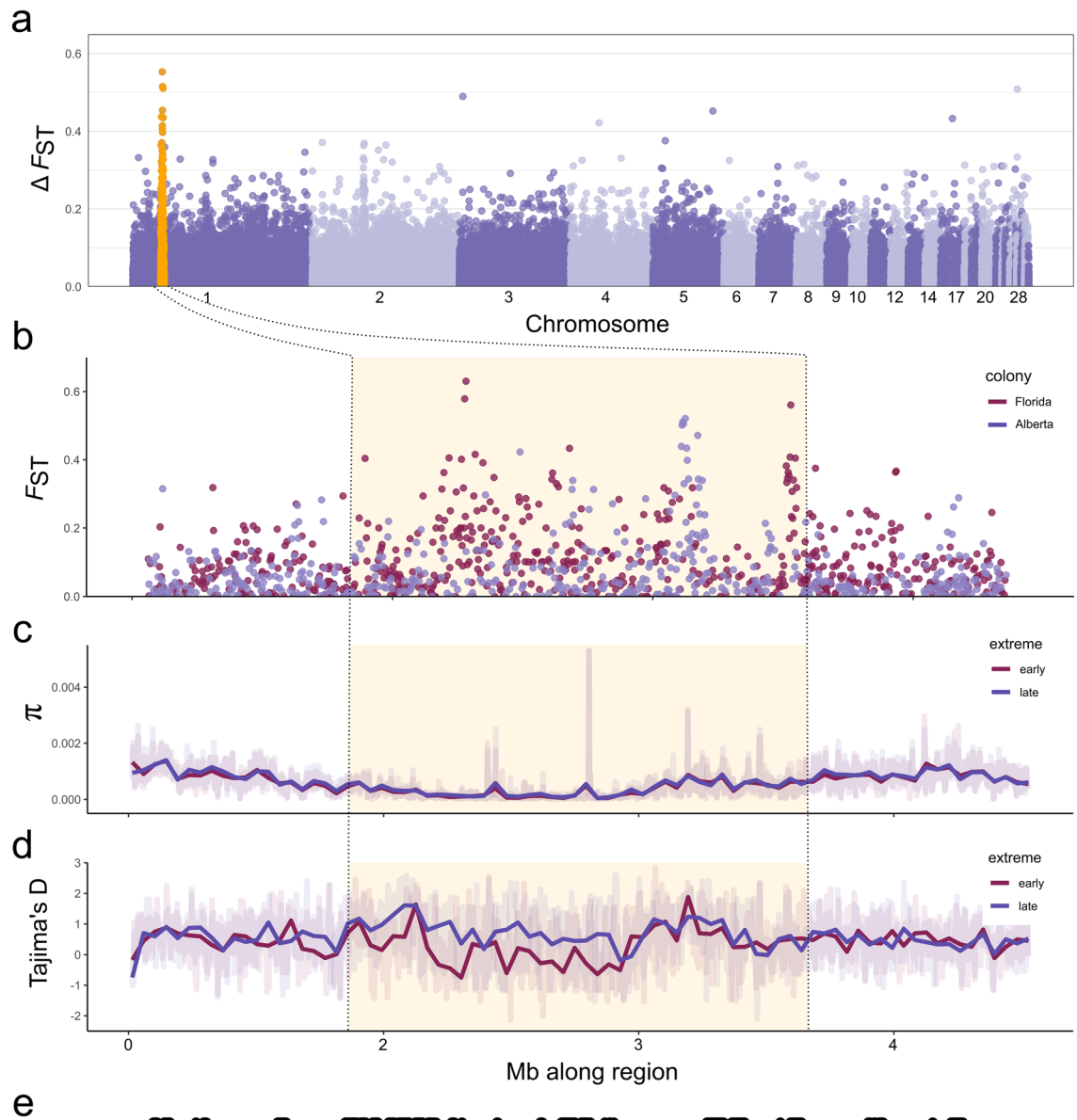


Figure 3. (a) Net genetic differentiation (ΔF_{ST}) across autosomes in 5 kb non-overlapping windows between earliest and latest spring migrants. The elevated region on chromosome 1 (defined as scaffolds with homology to chicken chromosome 1) is highlighted in orange, with plots examining this region to show (b) F_{ST} between early and late migrants within Alberta and within Florida, (c) nucleotide diversity (π) between earliest and latest migrants, (d) Tajima's D between earliest and late migrants, and (e) location of genes in this region as black dots.

This study presents novel findings on migration timing, facilitating an understanding of the components in the genomic architecture of migration timing in other long-distance migrants. The strong genomic variation and significant regions associated with purple martin migration timing could have important implications for adaptability in long-distance migrants. Many portions of the genome are conserved across other bird species and vertebrates³³ and climate change continues to affect migratory animals all over the world. Therefore, these findings bring us closer to understanding a common basis for migration, which may have broad implications for a variety of organisms.

Methods

Reference and resequencing data. The reference genome was assembled using PacBio long and 10X linked reads generated for a female martin from Manitoba, Canada. We used FALCON^{34,35} to create the initial assembly, then polished and scaffolded the genome with ArrowGrid, Pilon, and ARKS^{36–38}. The genome was annotated with MAKER³⁹. We used skimSeq (low-coverage whole-genome sequencing) to generate resequencing data⁴⁰ for an additional 87 birds (average coverage $2.7 \times$ per sample). Missing genotypes were then imputed with Beagle⁴¹, using information from the reference, surrounding genotypes, linkage disequilibrium structure, and haplotype blocks⁴². These included 45 male and 42 female blood samples collected from 13 different breed-

ing colonies across North America between 2008 and 2015 (Table S2). We filtered SNPs for quality (QUAL > 20, MQ > 20), maximum-missingness (20%), minor allele frequency (MAF > 0.05), Hardy–Weinberg equilibrium, and biallelic sites. Details on assembly, annotation, sequencing, and filtering are in supplementary information.

Light-level geolocator analysis. Light-level geolocators were mounted during the breeding season using leg-loop backpack harnesses and retrieved through recapture in the following year. Purple martin behavior of aerial foraging and use of open habitats makes light-level geolocators ideal for capturing sunrises and sunsets with minimal shading. The timing of these twilights is used to estimate the daily locations of birds over the entire year, using the midpoint of rise-set events to determine longitudes and day length for estimating latitudes⁴³. We analyzed twilight times with BASTag and GeoLight^{44,45}, producing estimated daily locations to obtain migratory departure and arrival dates. Due to the correlation of departure with arrival timing for migratory journeys (Fig. S3), we used spring arrival date as the migration timing phenotype.

Genomic architecture of migration timing. BSLMMs and LMMs were run using GEMMA¹⁹, where we included the covariates of sex, year, age, and the first principal component (PC1) from a PCA summarizing genetic variation in our dataset. We summarized results from these runs for PVE and used posterior inclusion probabilities (PIP) to identify specific SNPs with strong associations to the timing phenotypes. PIP is the probability that the SNP is associated with the phenotypic variation⁴⁶ and following⁴⁷ we considered SNPs with PIPs > 0.1 important. Polygenic models were created using the PLINK v1.9⁴⁸ and following Choi et al. (2020)'s PGS pipeline⁴⁹. Accuracy of the PGS model was assessed with jackknife cross-validation partitions using a random 85% of the dataset for the GWAS (n = 74) and the other 15% to test prediction accuracy (n = 13) over each of the 100 runs. We used VCFtools⁵⁰ to estimate F_{ST} between the 10 earliest (originating from two Florida colonies) and 10 latest (originating from two Alberta colonies, one Virginia colony, and one Pennsylvania colony) spring migrants. Since this F_{ST} could be elevated by processes unrelated to migration, including linked background selection and population structure⁵¹, we controlled for these potential effects by subtracting F_{ST} between Alberta and Florida (representing the northernmost and southernmost breeding regions) from values estimated between extreme timing phenotypes. This approach has been used in crows⁵² and blackcaps⁵³ to isolate differentiation associated with specific phenotypes. Additionally, we estimated F_{ST} between early and late migrants within Alberta and Florida populations separately to examine if we could recover the same signature of elevated F_{ST} in the same genomic region.

Data availability

Raw sequencing data is available on SRA NCBI BioProject PRJNA772931. Source code for reference genome and resequencing work is available at <https://github.com/edegreef/PUMA-reference-genome> and <https://github.com/edegreef/PUMA-resequencing-data>. Code for the polygenic analyses is available at https://github.com/BioMatt/PUMA_PGS.

Received: 22 September 2022; Accepted: 6 February 2023

Published online: 10 February 2023

References

- Davidson, S. C. *et al.* Ecological insights from three decades of animal movement tracking across a changing arctic. *Science* **370**, 712–715 (2020).
- Cohen, J. M., Lajeunesse, M. J. & Rohr, J. R. A global synthesis of animal phenological responses to climate change. *Nat. Clim. Chang.* **8**, 224–228 (2018).
- Both, C., Bouwhuis, S., Lessells, C. M. & Visser, M. E. Climate change and population declines in a long-distance migratory bird. *Nature* **441**, 81–83 (2006).
- Studds, C. E. & Marra, P. P. Rainfall-induced changes in food availability modify the spring departure programme of a migratory bird. *Proc. R. Soc. B.* **278**, 3437–3443 (2011).
- González, A. M., Bayly, N. J. & Hobson, K. A. Earlier and slower or later and faster: spring migration pace linked to departure time in a Neotropical migrant songbird. *J. Anim. Ecol.* **89**, 2840–2851 (2020).
- Liedvogel, M., Åkesson, S. & Bensch, S. The genetics of migration on the move. *Trends Ecol. Evol.* **26**, 561–569 (2011).
- Caprioli, M. *et al.* Clock gene variation is associated with breeding phenology and maybe under directional selection in the migratory barn swallow. *PLoS ONE* **7**, e35140 (2012).
- Mettler, R., Segelbacher, G. & Schaefer, M. H. Interactions between a candidate gene for migration (ADCYAP1), morphology and sex predict spring arrival in blackcap populations. *PLoS ONE* **10**, e0144587 (2015).
- Bazzi, G. *et al.* Clock gene polymorphism and scheduling of migration: a geolocator study of the barn swallow *Hirundo rustica*. *Sci. Rep.* **5**, 12443 (2015).
- Saino, N. *et al.* Polymorphism at the *Clock* gene predicts phenology of long-distance migratory birds. *Mol. Ecol.* **24**, 1758–1773 (2015).
- Bossu, C. M. *et al.* Clock-linked genes underlie seasonal migratory timing in a diurnal raptor. *Proc. R. Soc. B.* **289**, 20212507 (2022).
- O'Malley, K. G., Ford, M. J. & Hard, J. J. Clock polymorphism in Pacific salmon: evidence for variable selection along a latitudinal gradient. *Proc. R. Soc. B.* **277**, 3703–3714 (2010).
- Peterson, M. P. *et al.* Variation in candidate genes CLOCK and ADCYAP1 does not consistently predict differences in migratory behavior in the songbird genus Junco. *PLoS Research* **2** (2013).
- McKinnon, E. A. & Ten Love, O. P. years tracking the migrations of small landbirds: Lessons learned in the golden age of bio-logging. *Auk* **135**, 834–856 (2018).
- Fraser, K. C. *et al.* Continent-wide tracking to determine migratory connectivity and tropical habitat associations of a declining aerial insectivore. *Proc. R. Soc. B.* **279**, 4901–4906 (2012).
- Neufeld, L. R. *et al.* Breeding latitude is associated with the timing of nesting and migration around the annual calendar among purple martin *Progne subis* populations. *J. Ornithol.* **162**, 1009–1024 (2021).
- Peona, V. *et al.* Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. *Mol. Ecol.* **21**(1), 263–286 (2020).

18. Coelho, L. A., Musher, L. J. & Cracraft, J. A multireference-based whole genome assembly for the obligate ant-following antbird, *Rhegmatorhina melanosticta* (Thamnophilidae). *Diversity* **11**(19), 144 (2019).
19. Zhou, X., Carbonetto, P. & Stephens, M. Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genet.* **9**, e1003264 (2013).
20. Fuller, Z. L. *et al.* Population genetics of the coral *Acropora millepora*: Towards a genomic predictor of bleaching. *Science* **369**(6501) (2019).
21. Jones, S., Pfister-Genskow, M., Benca, R. M. & Cirelli, C. Molecular correlates of sleep and wakefulness in the brain of the white-crowned sparrow. *J. Neurochem.* **105**, 46–62 (2008).
22. Ma, C. *et al.* Sleep regulation by neurotensinergic neurons in a thalamo-amygdala circuit. *Neuron* **103** (2019).
23. Wong, J. M. & Eirin-Lopez, J. M. Evolution of methyltransferase-like (METTL) proteins in metazoan: a complex gene family involved in epitranscriptomic regulation and other epigenetic processes. *Mol. Biol. Evol.* **38**, 5309–5327 (2021).
24. Jia, Z. *et al.* ACS3 in brown fat drives propionate catabolism and its deficiency leads to autophagy and systemic metabolic dysfunction. *Clin. Transl. Med.* **12**, e665 (2022).
25. Muller, F. *et al.* Towards a conceptual framework for explaining variation in nocturnal departure time of songbird migrants. *Mov. Ecol.* **4**, 24 (2016).
26. Fraser, K. C. *et al.* Individual variability in migration timing can explain long-term population-level advances in a songbird. *Front. Ecol. Evol.* **7**, 324 (2019).
27. Barret, R. D. H. & Schluter, D. Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**(1), 38–44 (2008).
28. Colodro-Conde, L. *et al.* A direct test of the diathesis-stress model for depression. *Mol. Psychiatry* **23**, 1590–1596 (2017).
29. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genetics* **9**(4) (2013).
30. Lavallée, C. D. *et al.* The use of nocturnal flights for barrier crossing in a diurnally migrating songbird. *Mov. Ecol.* **9**, 21 (2021).
31. Saino, N. *et al.* Migration phenology and breeding success are predicted by methylation of a photoperiodic gene in the barn swallow. *Sci. Rep.* **7**, 45412 (2017).
32. Henry, R. A. *et al.* Changing the selectivity of p300 by acetyl-CoA modulation of histone acetylation. *ACS Chem. Biol.* **10**, 146–156 (2015).
33. Sun, H., Skorgerbø, G., Wang, Z., Liu, W. & Li, Y. Structural relationships between highly conserved elements and genes in vertebrate genomes. *PLoS ONE* **3**, e3727 (2008).
34. Chin, C. S. *et al.* Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **13**, 1050–1054 (2016).
35. Chin, C. S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569 (2013).
36. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
37. Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
38. Coombe, L. *et al.* ARKS: Chromosome-scale scaffolding of human genome drafts with linked read kmers. *BMC Bioinform.* **19**, 1–10 (2018).
39. Campbell, M. S., Holt, C., Moore, B. & Yandell, M. Genome annotation and curation using MAKER and MAKER-P. *Curr. Protocols Bioinform.* **48**, 4.11.1–4.11.39 (2014).
40. Malmberg, M. M. *et al.* Evaluation and recommendations for routine genotyping using skim whole genome re-sequencing in canola. *Front. Plant. Sci.* **9** (2018).
41. Browning, B. L. & Browning, S. R. Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* **98**, 116–126 (2016).
42. Golicz, A. A., Bayer, P. E. & Edwards, D. Skim-based genotyping by sequencing. *Methods Mol. Biol.* **1245**, 257–270 (2015).
43. Hill, R. D. Theory of geolocation by light levels. In B. J. L. Boeuf, & R. M. Laws (Ed.), *Elephant seals: Population ecology, behaviour and physiology*, pp. 227–236. Berkeley, CA: University of California Press (1994).
44. Wotherspoon, S., Summer, M. & Lisovski, S. BAStag: basic data processing for light based geolocation archival tags. Version 0.1.3. (2016).
45. Lisovski, S. & Hahn, S. GeoLight-processing and analysing light-based geolocator data in R. *Methods Ecol. Evol.* **3**, 1055–1059 (2012).
46. Gompert, Z., Lucas, L. K., Nice, C. C. & Buerkle, C. A. Genome divergence and the genetic architecture of barriers to gene flow between *Lycaeides idas* and *L. melissa*. *Evolution* **67**, 2498–2514 (2013).
47. Pfeifer, S. P. *et al.* The evolutionary history of Nebraska deer mice: local adaptation in the face of strong gene flow. *Mol. Biol. Evol.* **35**, 792–806 (2018).
48. Purcell, S. *et al.* PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
49. Choi, S. W., Mak, T. S. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analysis. *Nat Protoc* **15**, 2759–2772 (2020).
50. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
51. Cruickshank, T. E. & Hahn, M. W. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* **23**, 3133–3157 (2014).
52. Vijay, N. *et al.* Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nat. Commun.* **7**, 13195 (2016).
53. Delmore, K. *et al.* The evolutionary history and genomics of European blackcap migration. *eLife* **9**, e54462 (2020).

Acknowledgements

We thank K. Applegate, J. Barrow, L. Burgess, L. Chambers, T. Cheskey, P. Clifton, B. Dietrich, J. Fischer, G. Hvenegaard, P. Kramer, P. Mammenga, N. Mickle, M. North, M. Pearman, J. Ray, A. Ritchie, A. Savage, T. Shaheen, J. Siegrist, C. Silverio, B. Stutchbury, and J. Tautin for sample collection; M. Przeworski, Z. Fuller, and C.J. Garroway for conceptual input; F. Taborsak-Lines, M. Ormestad, P. Ewels, C. Wang, O.V. Pettersson, M.B. Mosbech, and C. Tellgren-Roth for generating data for the reference assembly; K. Jeffries, National Genomics Infrastructure (NGI)/Uppsala Genome Center, and Texas A&M AgriLife Genomics and Bioinformatics Service for laboratory support; Texas A&M High Performance Research Computing and Compute Canada for computing services. This work was supported by funding from the University of Manitoba, Texas A&M University, Natural Sciences and Engineering Research Council, Research Manitoba, the Society of Canadian Ornithologists, Sigma Xi, RFI/VR & Science for Life Laboratory, Sweden, and we further acknowledge the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure.

Author contributions

K.F. and K.D. designed and supervised the study. K.F. conducted/coordinated the collection of migration data and blood sampling. E.D. conducted laboratory work and performed genomic analyses with K.D.'s guidance. M.T. assisted with association modelling and conducted polygenic score analyses. A.S. coordinated sequencing for the reference genome and provided guidance on the assembly. E.D. analyzed light-level geolocator data K.F.'s guidance. E.D., K.F., and K.D. wrote the manuscript with feedback from M.T. and A.S.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-29470-7>.

Correspondence and requests for materials should be addressed to E.d.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023