

# Structural genomic variation and migratory behavior in a wild songbird

Kira E. Delmore<sup>1,2</sup>, Benjamin M. Van Doren<sup>1,3,4</sup>, Kristian Ullrich<sup>1</sup>, Teja Curk<sup>5,6</sup>, Henk P. van der Jeugd<sup>5</sup> , Miriam Liedvogel<sup>1,7</sup>

<sup>1</sup>MPRG Behavioural Genomics, Max Planck Institute for Evolutionary Biology, Plön, Germany

<sup>2</sup>Department of Biology, Texas A&M University, 3528 TAMU, College Station, TX, United States

<sup>3</sup>Department of Zoology, Edward Grey Institute, University of Oxford, Oxford, United Kingdom

<sup>4</sup>Cornell Lab of Ornithology, Cornell University, Ithaca, NY, United States

<sup>5</sup>Vogeltrekstation—Dutch Centre for Avian Migration and Demography, Netherlands Institute of Ecology (NIOO-KNAW), 6700 AB Wageningen, The Netherlands

<sup>6</sup>Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany

<sup>7</sup>Institute of Avian Research “Vogelwarte Helgoland,” Wilhelmshaven, Germany

Corresponding authors: Kira E. Delmore. Email: [kdelmore@bio.tamu.edu](mailto:kdelmore@bio.tamu.edu); Miriam Liedvogel. Email: [liedvogel@evolbio.mpg.de](mailto:liedvogel@evolbio.mpg.de)

## Abstract

Structural variants (SVs) are a major source of genetic variation; and descriptions in natural populations and connections with phenotypic traits are beginning to accumulate in the literature. We integrated advances in genomic sequencing and animal tracking to begin filling this knowledge gap in the Eurasian blackcap. Specifically, we (a) characterized the genome-wide distribution, frequency, and overall fitness effects of SVs using haplotype-resolved assemblies for 79 birds, and (b) used these SVs to study the genetics of seasonal migration. We detected >15 K SVs. Many SVs overlapped repetitive regions and exhibited evidence of purifying selection suggesting they have overall deleterious effects on fitness. We used estimates of genomic differentiation to identify SVs exhibiting evidence of selection in blackcaps with different migratory strategies. Insertions and deletions dominated the SVs we identified and were associated with genes that are either directly (e.g., regulatory motifs that maintain circadian rhythms) or indirectly (e.g., through immune response) related to migration. We also broke migration down into individual traits (direction, distance, and timing) using existing tracking data and tested if genetic variation at the SVs we identified could account for phenotypic variation at these traits. This was only the case for 1 trait—direction—and 1 specific SV (a deletion on chromosome 27) accounted for much of this variation. Our results highlight the evolutionary importance of SVs in natural populations and provide insight into the genetic basis of seasonal migration.

**Keywords:** structural variants, haplotype-resolved *de novo* assembly, purifying selection, migration genetics, bird

## Lay Summary

Structural variants (SVs) represent a major source of genetic variation that is important for both adaptation and speciation. However, detailed information on how they are distributed in the genome, their frequency, and effects on the fitness of individuals in natural populations (e.g., associations with adaptive traits) is lacking. These knowledge gaps derive in large part from difficulties associated with identifying SVs in the genome (e.g., because they are highly repetitive). We filled this knowledge gap here, using new sequencing technologies to assemble *de novo* genomes for 79 Eurasian blackcaps. These blackcaps had been tracked on migration in a previous study, allowing us to expand our work to the genetics of migration as well. Migration is a complex behavioral trait that has fascinated researchers and lay people alike for decades as well. The SVs we identified had overall negative effects on fitness but a subset were linked to variation in the migratory behavior of blackcaps. These SVs were linked to genes with functions that may be directly related to migration (e.g., that help maintain circadian rhythms) or are experiencing selection that is secondary to migration (e.g., that help with immune responses).

## Introduction

Structural variants (SVs) include duplications, deletions, transpositions, and inversions. Existing research suggests that these variants represent a major source of genetic variation and could have important fitness consequences (Collins et al., 2020; Cridland et al., 2013; Qin et al., 2021). For example, SVs can have deleterious effects on fitness, disrupting functional features of the genome (e.g., exons) and/or suppressing recombination (Alonge et al., 2020; Härmälä et al., 2021; Hirsch & Springer, 2017). Suppressed

recombination can lower effective population size at the local genomic level, reducing the efficiency of purifying selection (Charlesworth et al., 2009). There is growing evidence SVs could also have the opposite effect, facilitating adaptation and speciation (Chakraborty et al., 2019; Kirkpatrick & Barton 2006; Todesco et al., 2020; Weischenfeldt et al., 2013). For example, reductions in recombination can allow co-adapted alleles at separate loci to segregate together, sheltering alleles that underlie adaptive phenotypic traits. SV breakpoints can also serve as adaptive

Received May 8, 2023; revisions received July 31, 2023; accepted September 1, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of The Society for the Study of Evolution (SSE) and European Society for Evolutionary Biology (ESEN).

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

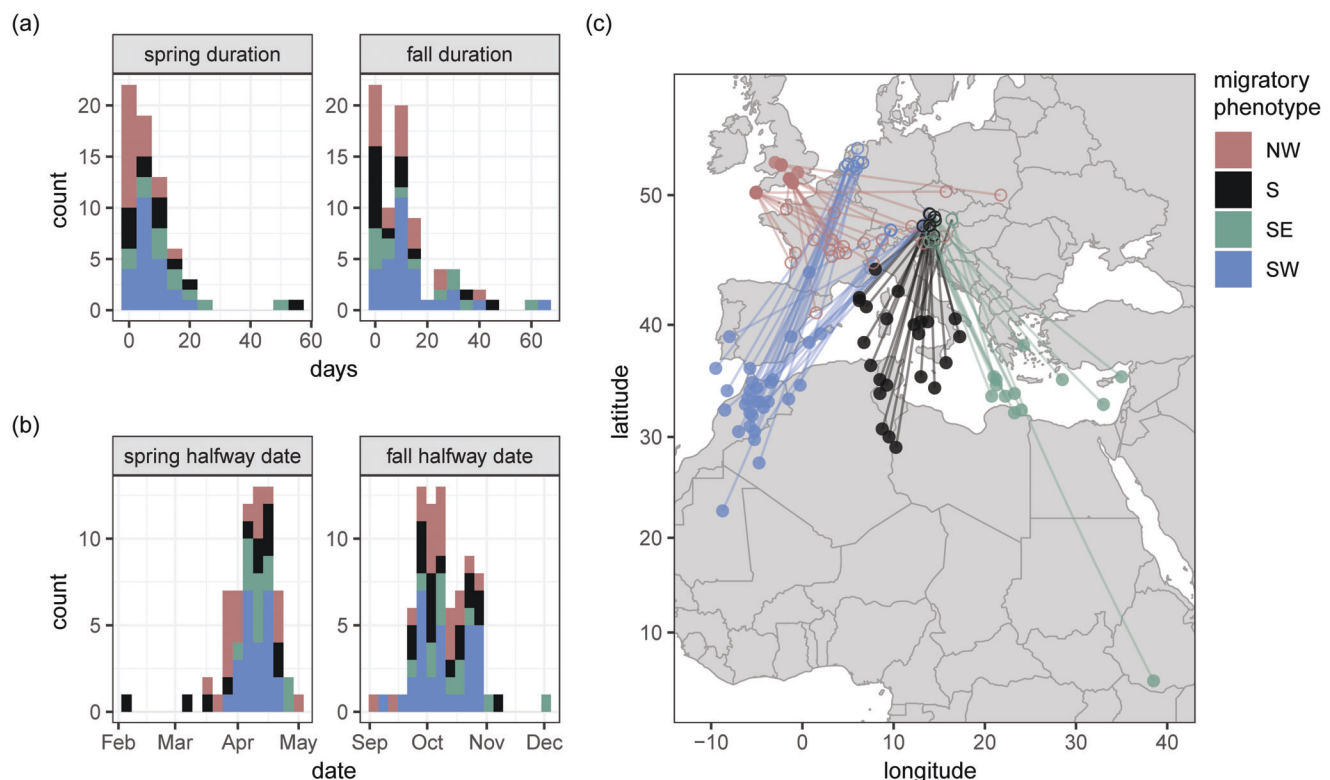
mutations themselves (Villoutreix et al., 2021). If these traits are important for maintaining reproductive isolation and gene flow with other populations is occurring, these SVs could also facilitate speciation (Kirkpatrick & Barton, 2006; Noor et al., 2001; Schwander et al., 2014).

Despite their potential fitness effects, data on the genome-wide distribution of SVs and their frequency within populations is largely lacking. This dearth of knowledge is especially true in natural populations of nonmodel organisms and is related in large part to technological limitations. Specifically, advances in sequencing technology have made it possible to obtain genome-wide data from nonmodel organisms, but existing work is often limited to short (~150 bp) sequencing reads. SVs are often larger than these reads and can be highly repetitive, making them difficult to assemble (Chakraborty et al., 2018; Huddleston & Eichler, 2016; Peona et al., 2021; Weissensteiner et al., 2020). A complete understanding of SVs will require contiguous genomes from multiple individuals where repetitive regions of the genome have been assembled accurately (Alkan et al., 2011; Huddleston et al., 2017; Lutgen et al., 2020). Linked reads are one technology that can help meet this need. They use molecular barcoding to preserve long-range sequencing information. Here we used this technology to identify SVs in a natural population of European blackcap. We have matching data on the migratory behavior of each bird, allowing us to gain inference into the genetics of seasonal migration as well.

Seasonal migration is the yearly long-distance movement of individuals between their breeding and wintering grounds. Successful migration requires the integration of several behavioral, physiological, and morphological traits (Dingle, 2006; Piersma, 2011). Decades of research has shown that there is a genetic basis to many of these traits, but the actual identity of genes underlying

them remains largely unknown. Genes controlling the circadian clock have been linked to some traits, but unbiased, genome-wide studies have only recently been applied to this question (Delmore & Liedvogel, 2016; Justen & Delmore, 2022; Liedvogel et al., 2011; Merlin & Liedvogel, 2019). Most genome-wide studies are limited to population-level comparisons, estimating genomic differentiation between populations that differ in one or more migratory traits (Delmore et al., 2015; Delmore et al., 2020a; Lundberg et al., 2017; Toews et al., 2019; von Rönne et al., 2016). These comparisons are valuable for identifying genomic regions under positive selection (i.e., areas of elevated differentiation), but caution is needed when interpreting their results as other processes can also elevate differentiation, including background selection and selection unrelated to the trait of interest (Barrett & Hoekstra, 2011; Cruickshank & Hahn, 2014; Noor & Bennett, 2009). A complete understanding of migration genetics will require complementary work at the individual level, including genome-wide association studies (GWAS) connecting specific genomic regions to individual migratory traits. GWAS will not only tell us about the genetics of individual migratory traits, but will also help us understand how these traits are modulated and integrated at the molecular level. Given their potential to shelter co-adapted alleles, SVs may be especially important for this integration. Indeed, there is already evidence that SVs underlie migratory traits in two avian systems; separate inversions on chromosome 1 underlie migratory orientation in willow warblers and wing shape in common quails (Caballero-López et al., 2022; Lundberg et al., 2017, 2023; Sanchez-Donoso et al., 2022).

The Eurasian blackcap is found throughout much of Europe, northern Africa and central Asia (Figure 1). This species exhibits considerable variation in migratory behavior—resident and migratory populations exist and among migrants, three main



**Figure 1.** Migratory behavior of Eurasian blackcaps. Timing of (A) spring and (B) fall migration and (C) map showing wintering and breeding locations ( $n = 19$  northwest, 20 south, 12 southeast, and 28 southwest).

orientations have been described (northwest [NW], southwest [SW], and southeast [SE] on fall migration) (Cramp & Brooks, 1992; Delmore et al., 2020b) (Figure 1). SE and SW migrants form a migratory divide in central Europe. Birds at the center of this divide orient in intermediate, southern (S) directions (Delmore et al., 2020b). Additional differences in the distance, timing, speed, and duration of both fall and spring migration have also been documented in migrants (Delmore et al., 2020b) (Figure 1). Researchers have capitalized on variation in the migratory behavior of blackcaps to study the genetics of migration for decades, including experimental and quantitative genetics approaches showing there is a strong genetic basis to migratory traits (Berthold et al., 1992; Helbig, 1991; Pulido & Berthold, 2010). Genetic surveys indicate that this variation arose recently and has not resulted in substantial, genome-wide differentiation (Delmore et al., 2020a; Mettler et al., 2013; Pérez-Tris et al., 2004). These genetic surveys include a recent study that used whole genome resequencing data to identify eight small genomic regions under positive selection in migrants that orient in different directions (four, three, and one in the NW, SE, and SW groups, respectively). The former study was limited to single nucleotide polymorphisms (SNPs) and population-level comparisons using individually resequenced birds distant from the contact zone with population-averaged phenotypes (Delmore et al., 2020a).

Here, we used linked reads to generate haplotype-resolved *de novo* assemblies for 79 individual blackcaps with individual-based phenotype characterization, including NW, SW, and SE migrants and individuals from the migratory divide in central Europe (Figure 1). We called SVs using these *de novo* assemblies and had three main objectives: (a) characterize SVs in natural populations, including their genome-wide distribution and overall fitness effects, (b) test for genome-wide population differentiation using these variants, and (c) use complementary population- and individual-level analyses to study the genetics of seasonal migration, including (a) local estimates of genomic differentiation to identify regions under selection and (b) GWAS to test if these regions are linked to specific migratory traits. All of the individuals used in the present study were tracked with light-level geolocators (Delmore et al., 2020b). Accordingly, we have individual-level phenotype data to run GWAS on multiple migratory traits, including direction, distance, the location of wintering grounds, and both the duration and timing of fall and spring migration.

## Results and discussion

We constructed *de novo* genome assemblies for 79 blackcaps using 10X Genomics linked-read technology (Supplementary Table S1). Final assemblies averaged 999.3 Mb in size and included an average of 1,710 scaffolds. Average scaffold and contig N50 sizes were 10.2 Mb and 106.2 kb, respectively, and >91% of the universally conserved single-copy benchmark (BUSCO) genes were present in these assemblies (Supplementary Table S1). Given considerable uncertainty associated with calling SVs (Chander et al., 2019), we used these assemblies and three separate pipelines to genotype birds, limiting our analysis to SVs called in two or more pipelines. Following these criteria, we identified between 9,246 and 12,585 SVs per individual. Note, additional information on pipelines used to genotype birds can be found in the Methods, but briefly, two pipelines started with the assembly of *de novo* genomes (two pseudohaplotypes) that were subsequently aligned to the blackcap reference genome, and one pipeline aligned reads directly to the reference. All three pipelines included heterozygous and

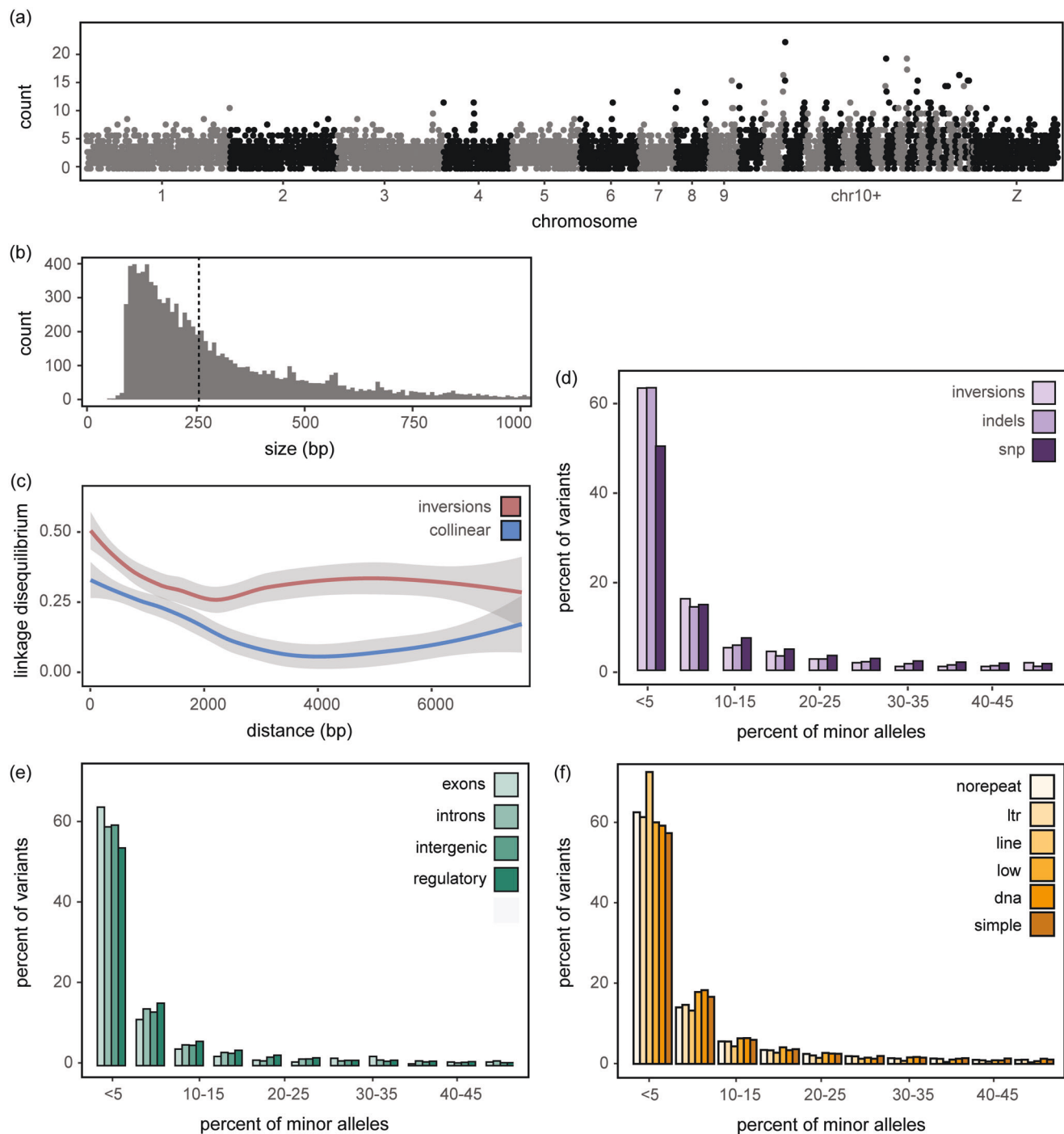
alternate homozygous genotype calls. After merging data from all individuals and filtering out variants with minor allele frequencies <0.05, our final dataset comprised 15,764 SVs (Supplementary Table S1 includes data on how many genotypes were called for each pipeline individually).

## Characterizing SVs and examining their overall fitness effects

We started our analysis by examining the distribution of SVs across the genome, counting the number of SVs in nonoverlapping windows of 200 kb. We found an average of three variants per window, with some windows harboring much larger numbers of SVs, especially on microchromosomes, where densities exceeded 15 SVs/window (Figure 2A). Similar enrichment of SVs on microchromosomes has been reported in blackcaps (on an independent Illumina WGS dataset, Bascón-Cardozo et al. 2022) and other birds and may be related to increased rates of recombination and/or enrichment of TEs on these chromosomes (e.g., Peona et al., 2022). Guanine Cytosine (GC) content is also higher on microchromosomes (Spearman's rank correlation between GC content in 200 kb windows and chromosome size =  $-0.67$  [ $p < .0001$ ]) and there is an overall positive relationship between GC content and the density of SVs independent of micro versus macrochromosomes (Spearman's rank correlation between GC content and SV density in 200 kb windows =  $0.10$  [ $p < .0001$ ]).

Deletions were the most numerous type of SV ( $n = 9341$ ), followed by insertions ( $n = 6393$ ), inversions ( $n = 24$ ), tandem duplications ( $n = 4$ ), and translocations ( $n = 1$ ). There was a bias towards smaller SVs among deletions and insertions, with median sizes of 245 and 265 bp in each class, respectively (size range of 58–65,285 bp for deletions and 52–15,209 bp for insertions; Figure 2B). The tandem duplications were a similar size (median 272 bp; range 144–357 bp), but the median size of inversions was approximately 10 times larger (2,485 bp; range 365–7,774 bp). The size of the single translocation was 10,456 bp.

We used minor allele frequency spectra (AFS) to examine the overall fitness effects of SVs using all individuals, starting with a comparison of variant types and including an AFS for SNPs called using the same sequencing data for comparison. AFS of all SV types were skewed towards rare alleles when compared to the AFS for SNPs (Figure 2D), indicating that alleles at SVs segregate at lower frequencies than alleles at SNPs and are under stronger purifying selection (i.e., are more deleterious than alleles at SNPs). This finding is supported by AFS constructed for SVs that overlap functional and nonfunctional elements in the genome, with SVs overlapping exons being more skewed towards rare alleles than variants overlapping any other feature (Figure 2E). This pattern is not as striking as the difference between SV types and SNPs, but both results are in line with theoretical predictions that SVs are often deleterious. This is also empirically supported by results from other species using a similar approach to quantify overall fitness effects associated with SVs (e.g., *Drosophila*, Zichner et al., 2013; cacao trees, Hamala et al. 2021; European crows, Weissensteiner et al., 2020; *Lycaeides* butterflies, Zhang et al. 2023). Population structure can also skew AFS, but previous genomic work on the blackcap and results below reveal limited structure in this species (Delmore et al., 2020a; Mettler et al., 2013; Pérez-Tris et al., 2004). Note, the presence of SVs can affect the quality of SNP calls and subsequent AFS but AFS built using SNPs proximate to SVs in our dataset are similar to the AFS obtained using genome-wide SNPs (Supplementary Figure S1).



**Figure 2.** Structural variants and their fitness effects. (A) Density (number/200 kb window) along the genome and (B) size distributions for indels (insertions and deletions) including dotted line for median size. (C) Average decay of linkage disequilibrium as a function of physical distance in major and minor inversion homozygotes and collinear regions. Loess curves and their standard errors are shown. Allele frequency spectrums for (D) each type of structural variant and (E) those overlapping different repeat elements in the genome.

SVs frequently occurred in repetitive regions of the genome. Specifically, we annotated SVs using a repeat library manually curated for the Eurasian blackcap (Bascón-Cardozo et al., 2022; Bours et al., 2022). Thirty-seven percent of the variants overlapped one or more repetitive elements in this library. Close to half of the repeats that overlapped these variants were simple repeats (47.8%); one quarter overlapped long terminal repeats (LTR) retrotransposons (25.5%). The rest were long interspersed nuclear elements/chicken repeat 1 (LINE/CR1) retrotransposons (17.5%),

low complexity repeats (8.2%), and to a much smaller degree DNA transposons (0.79%) and short interspersed nuclear elements (SINE) retrotransposons (0.24%; Figure 2F; similar overlap proportions were noted for each SV type alone as well [e.g., focusing on just inversions the division was 41.6%, 23.2%, 27.2%, 6.4%, and 1.5%, respectively]). The AFS for LINE/CR1 retrotransposons was skewed towards rare alleles suggesting they have the strongest deleterious effects and/or have recently been expanding. Comparable proportions of repeat elements were reported in other songbird



genomes, suggesting that transposable elements (especially LTR and LINE/CR1) are highly active in this group of organisms (Ficedula flycatchers, Suh et al., 2018; European crows, Weissensteiner et al., 2020). This overlap between TEs and SVs could also account for the skew towards rare alleles noted in our study (Figure 2D; Bergman & Bensasson 2007; Horvath et al., 2022). Note, summary statistics for AFS confirmed the patterns we documented as well. For example, Tajima's D was lowest for LINE/CR1 compared to other repeat classes (−1.04 vs. −0.81 to −0.61 for the remaining classes).

Only a small number of inversions were identified in our dataset, but they exhibited molecular signatures expected of this variant type. For example, we estimated linkage disequilibrium (LD) in inversions and a control set of colinear regions (same sizes and number). As expected, LD dropped off rapidly in colinear regions (at ~2,000 bp). However, this was not the case in inversions, with SVs continuing to exhibit elevated levels of LD as distances between variants increased beyond 2,000 bp (4,000 bp and beyond) (Figure 2c).

### Genome-wide levels of population differentiation at SVs

Consistent with previous genetic surveys using molecular tools to characterize population structure of European blackcaps, we found little evidence for population structure using SVs. Previous studies using marker-based approaches, such as mitochondrial haplotypes, microsatellites, but also genome-wide SNP-based approaches clearly show that population structure among medium-distance migrants with distinct migratory orientation is very low (Delmore et al., 2020a; Mettler et al., 2013; Pérez-Tris et al., 2004). We assigned birds from the present study to NW, SW, SE, or intermediate (S) groups using their vector between breeding and wintering locations to characterize autumn migratory direction (Supplementary Table S1) and tested if we could recover population structure using SVs. Specifically, we summarized genetic variation at SVs using a principal component analysis (PCA). We limited this analysis to autosomal chromosomes because our dataset comprises both males and females and the sex chromosomes accounted for a large amount of documented variation (Supplementary Figure S2A). Once the sex chromosomes were excluded, only the first principal component (PC) explained a significant proportion of the variation present in the dataset ( $p = .0033$ , eigenvalue = 1.17). Birds did not cluster based on breeding or wintering location (Supplementary Figure S2B). This lack of structure was true even when contrasting birds with breeding locations furthest away from the contact zone (e.g., the Netherlands vs. Austria) (Supplementary Figure S2C). And genome-wide estimates of  $F_{ST}$  support this lack of genetic variation (average weighted values of  $F_{ST}$  ranged from 0.021 between SW and NW and 0.031 between SE and NW). Combined with previous genetic surveys, these results indicate that differences in migration do not generate strong genome-wide differentiation at SVs or any other genetic marker examined in the blackcap thus far (Delmore et al., 2020b).

### Local genomic patterns of differentiation

Low levels of genetic differentiation between populations that exhibit distinct differences in phenotypic traits are ideal for work on the genetic basis of phenotypic traits, as genomic regions that underlie these traits should stand out against the backdrop of limited differentiation. Accordingly, we used a series of analyses aimed at identifying genomic regions that underlie migratory traits in blackcaps. We started with population branch statistics (PBS) (Yi et al., 2010), an  $F_{ST}$ -based statistic that can be applied to comparisons with more than two groups and identifies allele

frequency differences specific to each group. We limited our analysis to SVs in NW, SW, and SE migrants, excluding birds exhibiting intermediate (S) orientations to contrast the most extreme phenotypes. Among SW and SE migrant, we limited our analysis to birds that bred in a geographically confined area across the migratory divide in Austria (i.e., excluded birds breeding in the Netherlands) to minimize any potential confounding effects of even small amounts of population structure.

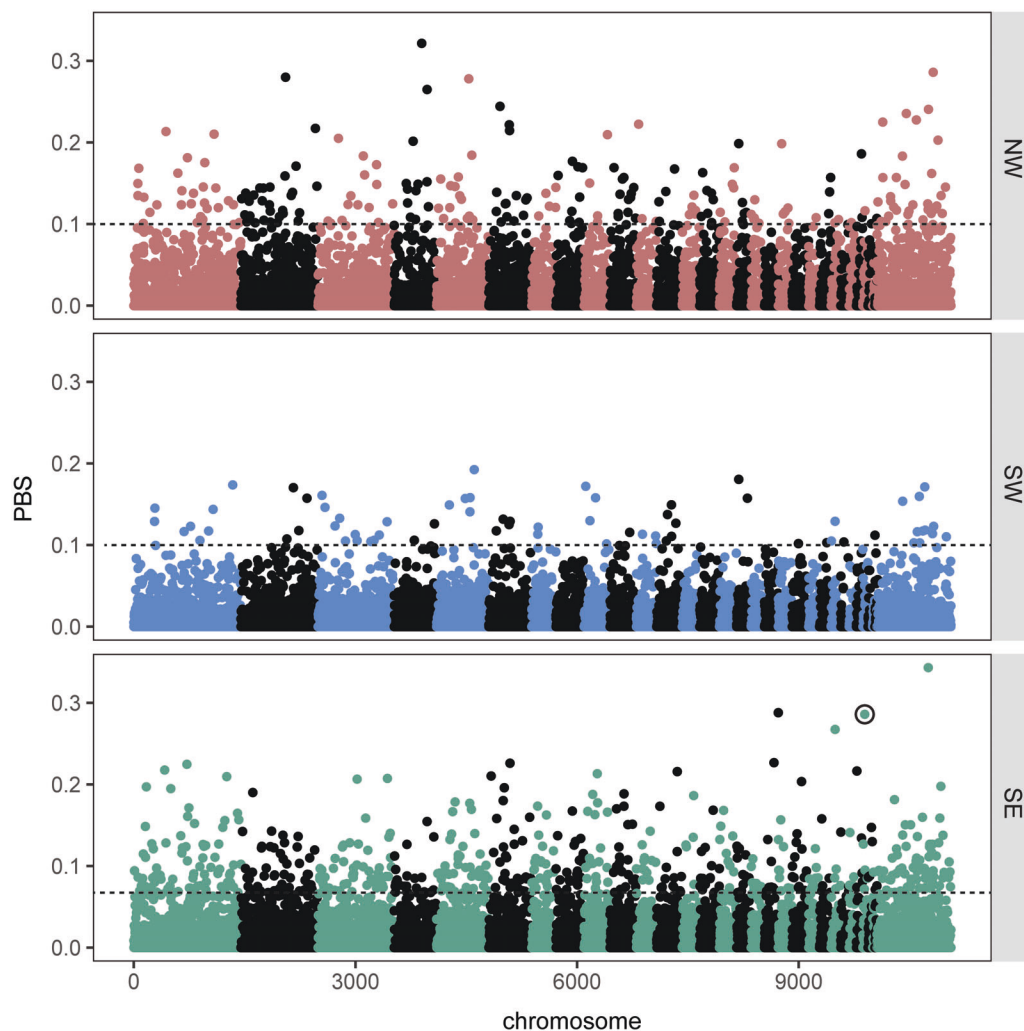
PBS was lowest for SW birds, and very few variants stood out against baseline levels in this group (average PBS in SW = 0.009; 0.013 in both NW and SE) (Figure 3). Both NW and SE migrating birds had several SVs that stood out against baseline levels of PBS (Figure 3). Variants exhibiting extreme values of PBS may be under positive selection and important for encoding variation in migratory behavior of these birds. Accordingly, we extracted genes that overlapped SVs in the top 5% of the PBS distribution for each phenotype (PBS > 0.06 [SW], 0.10 [NW], and 0.10 [SE]; 172 [SW], 155 [NW], and 162 [SE] genes, respectively; only 1/24 inversions fell in the top 5% of any PBS distributions with the remaining loci being deletions or insertions) and ran a functional enrichment analysis, looking for gene ontology (GO) terms, biological pathways and regulatory motifs that were overrepresented in these genes. Note, caution should be used when interpreting these results as many additional processes beyond selection can also elevate genomic differentiation (Barrett & Hoekstra, 2011; Cruickshank & Hahn, 2014; Noor & Bennett, 2009).

Regulatory motifs are sequences of DNA that are bound by transcription factors. Ebox is a regulatory motif that was enriched in NW migrants. The Ebox motif is often bound by basic helix-loop-helix (bHLH) transcription factors and is of significance for migration as several genes that regulate the circadian clock are bHLH transcription factors that bind Ebox motifs (Cassone, 2014). In a response to changes in photoperiod, the circadian clock entrains circadian (possibly also circannual) rhythms as well as initiates migratory behavior (Gwinner, 1996; Visser et al., 2010). Interestingly, we documented similar enrichment of Ebox motifs in our previous genomic survey of blackcaps, using SNPs and PBS to identify genomic regions under selection in the same migratory phenotypes (Delmore et al., 2020a). No functional enrichment was found in the SW migrants.

One GO term was enriched in the SE migrants: "MAPK cascade" (mitogen-activated protein kinase cascade). This cascade is highly conserved across vertebrates and important for translating extracellular signals to intracellular responses. Of particular relevance to seasonal migration, MAPK cascades facilitate learning and memory, consolidating learning following specific behaviors and eliciting memory formation (Atkins et al., 1998; Day & Sweatt, 2011; Thomas & Haganir, 2004). MAPK cascades are also important for mounting immune responses in many vertebrates. Considerable research has focused on the relationship between immunity and migration, with several studies suggesting that migrants suppress their immune system on migration, allowing them to allocate more resources to this costly life history event. It has also been noted that migrants with different routes and wintering sites likely encounter different parasites throughout their annual cycle; local adaptation associated with this variation may also drive changes in the immune system (Altizer et al., 2011; Eikenaar et al., 2018; Møller & Erritzøe, 1998).

### Genome-wide association analyses and individual migratory traits

The former analyses used local estimates of differentiation to identify genomic regions that distinguish the main migratory



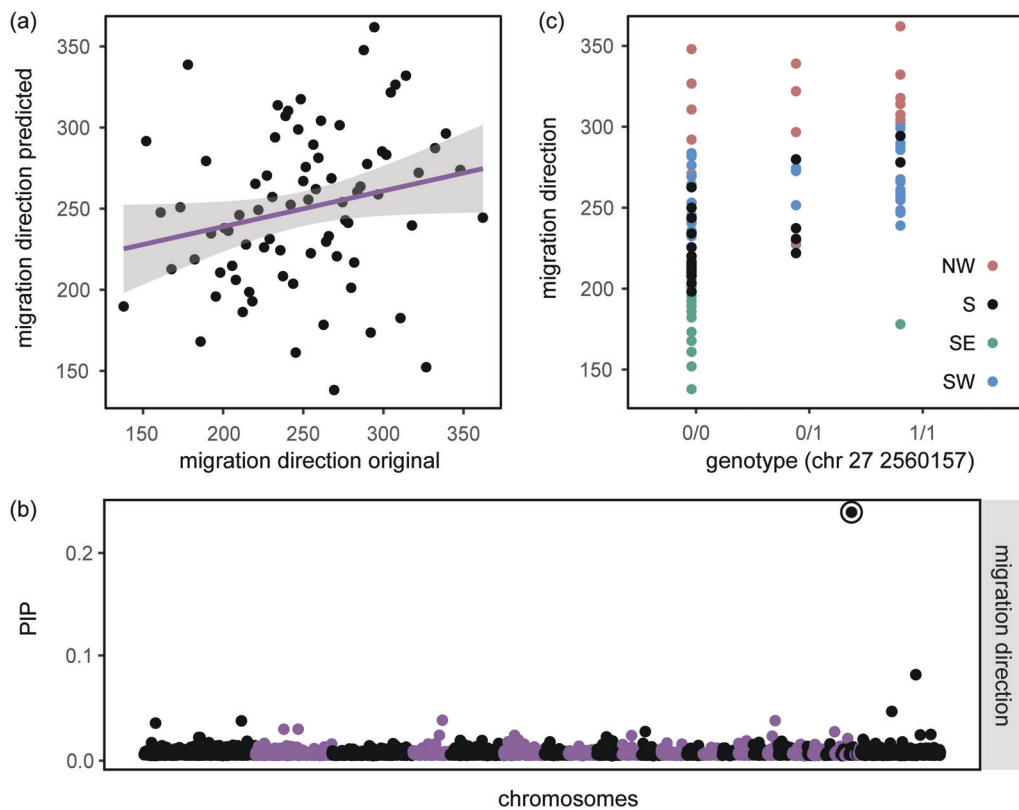
**Figure 3.** Signatures of selection. Population branch statistics estimated for birds that migrated along northwest, southwest, or southeast (SE) routes from their breeding grounds. Structural variant circled in black in the SE panel was also identified in our genome-wide association analyses (Fig. 4). The 5% cutoff used for enrichment analyses is shown with a dotted line for each phenotype.

phenotypes present in blackcaps. In this last set of analyses, we broke migration behavior down into distinct traits and examined the genetic basis of each one with GWAS. We used all birds in these analyses (i.e., added birds exhibiting intermediate [S] orientations back in to the analysis along with those breeding in the Netherlands) and focused on seven traits: direction (the vector orientation between breeding and wintering sites), distance (direct connection between breeding and wintering sites, in km), wintering location (longitude), and both the duration (days) and timing (date when birds reached the halfway point between breeding and wintering sites) of fall and spring migration.

We started our analyses by estimating PVE (the proportion of phenotype variation explained by genetic variation in our SV set) for each trait. We used Bayesian sparse linear mixed models (BSLMMs) for these analyses (Zhou et al., 2013). BSLMMs use a markov chain monte carlo (MCMC) algorithm to fit all variants to the phenotype simultaneously and control for population structure with a kinship matrix. We ran 20 million MCMC steps, extracting parameter values every 10,000 steps. Mean values of PVE across these steps ranged from 0.45 to 0.82 (direction =  $0.73 \pm 0.27$  [SD], distance =  $0.45 \pm 0.29$ , winter longitude =  $0.82 \pm 0.28$ , fall timing =  $0.47 \pm 0.28$ , fall duration =  $0.50 \pm 0.30$ , spring timing =  $0.77 \pm 0.29$ , and spring duration  $0.46 \pm 0.29$ ). These values are relatively

high, suggesting our SVs capture a good amount of the variation present in the migratory traits we measured. Standard deviations around these means, however, are quite wide and indicate we have limited precision in these estimates. Accordingly, we ran a complementary set of analyses obtaining polygenic scores (PGS) for each individual and trait. Specifically, we randomly masked phenotypic values for a subset of individuals and tested if we could predict these missing values with the remaining dataset. Migratory orientation was the only trait where predicted and actual phenotypic values were correlated (Figure 4A;  $r = 0.22$ ,  $F_{1,77} = 3.92$ ,  $p = .05$ , all  $p$ -values for remaining traits  $> .11$ ). Combined, results from PVE and PGS analyses suggest that even though PVE values were relatively high for all traits, we can only be confident that our SVs capture sufficient variation in migratory direction. This does not necessarily mean there is no genetic basis to the remaining traits; rather, future work using larger sample sizes and additional variants (e.g., SNPs) may be needed to explain variation in these traits. For now, we focus our remaining analyses on migratory direction.

Beyond PVEs, BSLMMs also estimate posterior inclusion probabilities (PIPs) for each variant. PIPs represent the proportion of MCMC iterations where a variant has a nonzero effect size. One variant stood out against the rest in our BSLMM for migration direction—a 710 bp deletion on chromosome 27 that had a PIP of 0.28 (Figure 4B). Birds



**Figure 4.** Results from genome-wide association analysis using migration direction during autumn. (A) Relationship between original values of migratory orientation and those predicted by cross-validation procedure ( $p = .05$ ). (B) Posterior inclusion probabilities for all variants, highlighting the deletion on chromosome 27 that exhibited elevated population branch statistics in southeast migrating birds (also highlighted in Fig. 3). (C) Relationship between genotypes at the deletion on chromosome 27 and migration direction ( $p < .0001$ ). Individuals are colored by their migratory phenotype.

homozygous for the reference allele oriented in directions that were further east (Figure 4C). This variant also stood out in our analysis of genomic differentiation, exhibiting an elevated estimate of PBS in SE migrants (Figure 3). SE migrants were nearly fixed for the reference allele at this locus. The relationship between migratory orientation and genotypes at this locus remained even when we limited the dataset to S migrants (the group with the most variation in migratory direction) ( $F_{2,16} = 8.84$ ,  $p = .003$ ) suggesting variation at this SV is related to orientation, not just fixation in SE migrants. Combined, these findings suggest that this variant is under selection in SE migrants and helps control migratory orientation across blackcaps. Concerning the actual identity/effect of this variant, it occurs in an intron of T cell receptor alpha chain (TRA). TRA helps T cells respond to specific antigens in their cellular environment (Göbel et al., 1994) and this finding continues to support a connection between immune response and migration in blackcaps; recall, the functional category “MAPK cascade” was enriched in SE migrating birds. This pathway plays a role in memory and learning and is also important for mounting immune responses in many vertebrates. Together, these findings add support for an important role of immunity in the context of migration behavior, for example, related to the fact that birds using different migratory routes are challenged with different environments throughout their annual cycle, and local adaptation associated with this variation may facilitate changes in the immune system.

## Conclusion

We conducted an extensive study of SVs in natural population of migratory songbirds, using *de novo* assembled genomes to genotype 79 individually phenotyped birds at thousands of SVs. We

found evidence for purifying selection on SVs, suggesting they have an overall deleterious effect on fitness also supporting previous work on SVs. We also documented considerable overlap between SVs and transposable elements, suggesting transposable elements comprise a large proportion of SVs and genetic variation in the genome. We did not find evidence for genome-wide population differentiation between blackcaps with different migratory strategies, but our individual-based phenotypic characterization indicates local genetic variation at SVs does account for a large proportion of the phenotypic variation observed in specific migratory traits.

Seasonal migration is a complex behavior that comprises many traits. SVs like inversions are strong candidates for capturing loci that underlie complex behaviors and evidence from other systems has connected inversions with migration (e.g., an inversion on chromosome 1 underlies migratory direction in willow warblers) (Caballero-López et al., 2022; Lundberg et al., 2017). We did not make such a connection here; we only identified a small number of small inversions suggesting there are several paths for the evolution of complex traits like migration. LD was reduced in these inversions suggesting they are suppressing recombination, but they did not exhibit signatures of selection and were not linked to any of our focal migratory traits. Blackcaps only began to diverge recently (30,000 years ago (Delmore et al., 2020a)) and seasonal migration is highly dynamic in this species (e.g., the NW population only recently [in the last 70 years] started to grow in size) (Cramp & Brooks, 1992). Accordingly, it is possible that inversions will capture genetic variation underlying migratory behavior in the future, but that is not currently the case. Inversions are

only one type of SV; deletions, insertions, as well as translocations or duplications can also drastically alter phenotypes with important evolutionary consequences, such as the text book example of industrial melanism that is the darker morph of the peppered moth (Hof et al., 2016), or as an example within the songbirds, plumage color divergence between hooded and carrion crows (*Corvus corone cornix*, *C. c. corone*, respectively) has been linked to a LTR retrotransposon insertions in crows, where hooded crows are homozygous for the insertion (Weissensteiner et al., 2020).

Future work using additional long-read technologies (e.g., PacBio HiFi and HiC) may uncover additional variants that could be connected to migration in the system but for now, we conclude that seasonal migration has a highly polygenic basis in blackcaps. Beyond the deletion on chromosome 27, none of the SVs identified here stood out in our analyses of selection or GWAS. We reported similar findings in a previous study using an independent dataset of genome-wide SNP data (Delmore et al., 2020b). Interestingly, the Ebox regulatory motif identified in the NW migrants here was also identified in enrichment analyses with population-averaged phenotypes based on SNP data and could reveal a mechanism through which multiple migratory traits could be controlled by a similar mechanism. The Ebox motif is bound by transcription factors that regulate circadian rhythms in birds. Circadian rhythms are important for migration (e.g., songbirds like blackcaps switch from diurnal to nocturnal behavior on migration and circadian rhythms likely entrain circannual rhythms which are important for migratory timing). Perhaps seasonal migration in blackcaps is regulated by a small number of transcription factors that affect expression at multiple genes. In a general context, we see recurrent pathways and functional categories connected to migration in many systems, including immunity, circadian rhythm regulation, learning, and memory. Although the actual genes under selection do not seem to match in an across species comparison, we would assume that the adaptation of the central pathways needs to be optimized to and constrained by the migratory niche of each species or population, that is, each species adapt their phenotype in a specific way, fitting to its ecological demands, which could explain consistency in general regulatory pathways despite the apparent lack of commonly identified genes.

## Materials and methods

### Sampling and phenotypic analysis

We included data from 79 blackcaps in the present study. A subset of these birds were captured using mist nets on the breeding grounds in Austria ( $n = 45$ ) and the Netherlands ( $n = 16$ ); the remaining birds were captured on the wintering grounds in the UK ( $n = 18$ ; Supplementary Table S1). We obtained blood samples from each bird and fitted them with light-level geolocators using leg-loop backpack harnesses. Light-level geolocators record light intensity data at specific time intervals. These data are stored until the devices are retrieved at which point light intensity data are converted to day length and time of local midday and used to estimate daily longitude and latitude (McKinnon & Love, 2018).

We describe methods used to analyze light-level geolocator data in full in Delmore et al. (2020b). Of relevance for the present study, we categorized birds into four broad phenotypic classes (migrating NW, SW, SE, and S in fall) using their wintering locations. For birds wintering north of 37.5° N, we considered those west of 5° E to be SW migrants, those east of 20° E to be SE migrants, and those between 5 and 20° E to have intermediate southerly (S) routes. For birds wintering south of 37.5° N, we used a cut-off of 0° instead of 5° E to distinguish SW from S because

these longer routes require less of a westerly component to reach the same longitude.

We estimated migration direction and distance by fitting a rhumb line between their breeding and wintering sites. We estimated timing by identifying the shortest distance route (i.e., a great circle routes) between their breeding and wintering sites and determining the date when birds reached 50% of the way between these sites. Duration was estimated as the number of days it took each bird to travel from early (30%) to late (70%) migration stages and speed as migration distance divided by duration.

### Assemblies and variant calling

High molecular weight DNA was extracted from blood samples, 10X Chromium libraries were constructed and sequenced using Illumina technology (150 bp, paired end) by Novogene (Hongkong). The mean molecule length of resulting libraries was 32,657 bp (range 10,062–54,206 bp) and sequencing reached a mean coverage of 55X (range 7–74X; based on reads aligned to reference by LongRanger [see below]; Supplementary Table S1).

We called SVs using three pipelines. The first two pipelines relied on *de novo* assemblies of reads. Specifically, we assembled reads into two parallel pseudohaplotypes (phased contigs and scaffolds) with Supernova2 (Weisenfeld et al., 2017) and used two different approaches to align these pseudohaplotypes to the blackcap reference genome (Ishigohoka et al., 2021) and call genotypes in relation to the reference genome: (a) MUMmer4 (Marçais et al., 2018) for alignment and MUM&Co (-g 1080000000 -b) (Hämälä et al., 2021) to call genotypes and (b) Minimap2 (Li, 2018) for alignment and SVIM-Asm (diploid, tandem duplications as insertions and interspersed duplications as insertions) (Heller & Vingron, 2021) for genotype calling. We aligned 10X reads directly to the blackcap reference for the third pipeline. We used LongRanger wgs for alignment (average mapping rate of 87%) and the same program to genotype SVs (--vcfmode gatk) (Marks et al., 2019).

Once SVs were genotyped, we used a series of filters to identify high-quality variants. Starting at the level of individuals, we limited the dataset to variants identified by at least two callers and had matching genotypes. We used SURVIVOR (settings 1000 2 0 0 0 50) (Jaffares et al., 2017) and a custom R script to conduct this filtering. Variants with strings of >10Ns were also removed to reduce potential errors caused by contig scaffolding. We generated a multi-individual vcf (i.e., merged variants across individuals) using SURVIVOR (settings 1000 4 0 0 0 50) and limited our analyses to SVs with maf > 0.05 using vcftools (Danecek et al., 2011). We focused on five types of SVs: insertions (sequence inserted into query), deletions (sequence deleted from the query), tandem duplications (sequence duplicated in the query), inversions (sequence with reversed orientation), and translocations (sequence moved between chromosomes).

Following Hamala et al. (2021), we chose a random set of 50 SVs for visual validation in IGV. We used alignments from LongRanger and confirmed the presence of all but one of these SVs (see examples in Supplementary Figure S3, including one of the main variants identified in our subsequent analyses), suggesting false positives are rare in our dataset and likely related to the stringent filtering we applied.

### Population genetics and GWAS

We used AFS to examine the overall fitness effects of SVs. We constructed these AFS using vcf2sfs in R (Liu et al., 2018) and included all SVs (i.e., not excluding those with MAF < 0.05 at this stage). Following Hämälä et al. (2021), we combined insertions



and deletions into the same category (INDEL) for this analysis because these variant types cannot be defined based on the reference and this could affect inferences related to fitness. We used scripts from Hämälä et al. (2021) to estimate LD (squared Pearson's correlation coefficients) between arrangements at inversions and a random set of colinear regions with the same distribution of sizes as inversions.

We used a PCA to examine genome-wide patterns of genomic differentiation. This analysis was conducted using smartpca (EIGENSOFT version 5.0) after standardizing loci to have equal variance.

We used estimates of PBS to identify SVs exhibiting signatures of selection. PBS is similar to  $F_{ST}$  but can be used with more than two populations and identifies selection specific to one population. We estimated this parameter in two steps, calculating  $F_{ST}$  between NW, SW, and SE migrants using the estimator derived by Hudson (Hudson et al., 1992) and scripts from Hämälä et al. (2021). These estimates of  $F_{ST}$  were then converted to PBS following (Zhan et al., 2014) ( $T = \log$  transformed estimates of  $F_{ST}$ , example is for SE population):  $(T^{SE-NW} + T^{SE-SW} - T^{NW-SW})/2$ . Raw estimates of  $F_{ST}$  can be found in Supplementary Figure S4.

We used two different programs to look for functional enrichment at genes overlapping SVs showing evidence for selection (i.e., with PBS values in the top 5% of the distribution): (a) BINGO (Maere et al., 2005) to look for enrichment of specific GO categories and (b) go:Profiler (Raudvere et al., 2019) to look for enrichment in additional functional databases, including biological pathways, regulatory motifs of transcription factors and microRNAs and protein–protein interactions. We used a custom annotation for the blackcap in BINGO and an annotation for the chicken in go:Profiler.

GWAS were run as BSLMMs in Genome-wide efficient mixed model association algorithm (GEMMA) (Zhou et al., 2013). BSLMMs are adaptive models that include linear mixed models (LMM) and Bayesian variable selection regression as special cases and that learn the genetic architecture from the data. These models are run separately for each phenotype but allele frequencies at all variants are considered together and included as the predictor variable. A kinship matrix is also included to control for factors that influence the phenotype and are correlated with genotypes (e.g., population structure). We ran four independent chains for each BSLMM, with a burn-in of 5 million steps and a subsequent 20 million MCMC steps (sampling every 1,000 steps). We report one hyperparameter from this model (PVE: the proportion of variance in phenotypes explained by all SVs, also called chip heritability) and focus on two variant-specific parameters: PIPs and ( $\beta$ , variant effects). We calculated genetic correlations between traits by identifying SNPs with  $PIP > 0.01$  and correlated model-averaged estimates of  $\beta$  ( $\beta$  weighted by their PIPs) (Comeault et al., 2014; Gompert et al., 2019). In order to facilitate comparisons across traits and limit the effects of outliers, we normal quantile-transformed all of our phenotypic traits before running these analyses. We also regressed each trait against sex to remove the effects of this variable on migratory traits.

We used a cross-validation procedure to obtain polygenic scores for each individual (and trait) (Gompert et al., 2019; Villoutreix et al., 2022). Specifically, we masked the phenotype of 25% of the sampled individuals and reran BSLMMs using the remaining individuals and same parameters as the original BSLMM (with only one MCMC chain and the “predict Suh-1” plugin in GEMMA). We repeated this procedure four times for each trait, obtaining predicted values (polygenic scores) for each individual and used a linear model to estimate correlation between predicted values and the original phenotype of each bird.

## Supplementary material

Supplementary material is available online at *Evolution Letters*.

## Data availability

Phenotypic data used in the present study are already available at <https://doi.org/10.5061/dryad.2280gb5qc>. Raw sequence data (from 10X Chromium libraries) generated for the present study can be found under ENA project PRJEB65115.

## Author contributions

Conceptualization: K.E.D. and M.L.; sampling: K.E.D., B.M.V.D., T.C., H.J., and M.L.; formal analysis: K.E.D. and K.U.; writing: K.E.D. with input from B.M.V.D., K.U., H.J., and M.L.

## Funding

This project was supported by funding from the Max Planck Society (MFFALIMN0001 grant to ML), the German Science Foundation (project Nav05 within SFB 1372—Magnetoreception and Navigation in Vertebrates to ML) and the National Science Foundation (National Science Foundation Grant IOS-2143004 to KED).

*Conflict of interest:* The authors declare no conflict of interest.

## Acknowledgments

We thank Matthias Weissensteiner and Claire Mérot for advice early in our analysis; Tuomas Hämälä and Hannah Justen for help implementing scripts for a subset of our population genetic and GWAS analyses, respectively; and Corinna Langebrake, Georg Manthey, and Joe Wynn for general discussion. This work would not be possible without the enthusiasm and support received in the field to collect both accurate phenotype data as well as genetic material. We are particularly grateful to Ben Sheldon, Robbie Phillips, Greg Conway, Graham Roberts, Tania Garrido-Garduño, Britta Meyer, Timo Hasselmann, Hannah Justen, Juan Sebastian Lugo Ramos, Ivan Maggini, Wolfgang Vogl, and Leonida Fusani next to many more keen fieldworkers and house owners for assistance generating the phenotypic dataset. All work was carried out under approval of the respective institutional ethics and animal welfare committee and national authorities. Specifically, permit numbers for work in Austria: GZ BMWFW-68.205/0048-WF/V/3b/2016 and BMWFW-68.205/0139-WF/V/3b/2016 according to §§ 26ff. of Animal Experiments Act, TVG 2012. In the UK, geolocator deployment was approved by the University of Oxford Animal Welfare Ethical Review Body, and fieldwork was conducted under licenses from the British Trust for Ornithology, approved by the Special Methods Technical Panel. Permit number for work in the Netherlands: AVD801002016519 valid 27-6-2016 through 31-5-2021, issued by the Centrale Commissie Dierproeven.

## References

- Alkan, C., Coe, B. P., & Eichler, E. E. (2011). Genome structural variation discovery and genotyping. *Nature Reviews. Genetics*, **12**(5), 363–376. <https://doi.org/10.1038/nrg2958>
- Alonge, M., Wang, X., Benoit, M., Soyk, S., Pereira, L., Zhang, L., Suresh, H., Ramakrishnan, S., Maumus, F., Ciren, D., Levy, Y.,

- Harel, T. H., Shalev-Schlosser, G., Amsellem, Z., Razifard, H., Caicedo, A. L., Tieman, D. M., Klee, H., Kirsche, M., ... Lippman, Z. B. (2020). Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell*, **182**, 145–161.e23. <https://doi.org/10.1016/j.cell.2020.05.021>
- Altizer, S., Bartel, R., & Han, B. A. (2011). Animal migration and infectious disease risk. *Science*, **331**(6015), 296–302. <https://doi.org/10.1126/science.1194694>
- Atkins, C. M., Selcher, J. C., Petraitis, J. J., Trzaskos, J. M., & Sweatt, J. D. (1998). The MAPK cascade is required for mammalian associative learning. *Nature Neuroscience*, **1**(7), 602–609. <https://doi.org/10.1038/2836>
- Barrett, R. D. H., & Hoekstra, H. E. (2011). Molecular spandrels: Tests of adaptation at the genetic level. *Nature Reviews. Genetics*, **12**(11), 767–780. <https://doi.org/10.1038/nrg3015>
- Bascón-Cardozo, K., Bours, A., Manthey, G., Pruischer, P., Durieux, G., Duthel, J., Odenthal-Hesse, L., & Liedvogel, M. (2022). Fine-scale map reveals highly variable recombination rates associated with genomic features in the European blackcap. *Authorea Preprints*, <https://doi.org/10.22541/au.165423614.49331155/v1>
- Bergman, C. M., & Bensasson, D. (2007). Recent LTR retrotransposon insertion contrasts with waves of non-LTR insertion since speciation in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, **104**(27), 11340–11345. <https://doi.org/10.1073/pnas.0702552104>
- Berthold, P., Helbig, A. J., Mohr, G., & Querner, U. (1992). Rapid micro-evolution of migratory behaviour in a wild bird species. *Nature*, **360**(6405), 668–670. <https://doi.org/10.1038/360668a0>
- Bours, A., Pruischer, P., Bascón-Cardozo, K., Odenthal-Hesse, L., & Liedvogel, M. (2022). The blackcap (*Sylvia atricapilla*) genome reveals a species-specific accumulation of LTR retrotransposons. *Sci Rep*, **13**, 16471. <https://doi.org/10.1038/s41598-023-43090-1>
- Caballero-López, V., Lundberg, M., Sokolovskis, K., & Bensch, S. (2022). Transposable elements mark a repeat-rich region associated with migratory phenotypes of willow warblers (*Phylloscopus trochilus*). *Molecular Ecology*, **31**(4), 1128–1141. <https://doi.org/10.1111/mec.16292>
- Cassone, V. M. (2014). Avian circadian organization: A chorus of clocks. *Frontiers in Neuroendocrinology*, **35**(1), 76–88. <https://doi.org/10.1016/j.yfrne.2013.10.002>
- Chakraborty, M., Emerson, J. J., Macdonald, S. J., & Long, A. D. (2019). Structural variants exhibit widespread allelic heterogeneity and shape variation in complex traits. *Nature Communications*, **10**(1), 1–11.
- Chakraborty, M., VanKuren, N. W., Zhao, R., Zhang, X., Kalsow, S., & Emerson, J. J. (2018). Hidden genetic variation shapes the structure of functional elements in *Drosophila*. *Nature Genetics*, **50**(1), 20–25. <https://doi.org/10.1038/s41588-017-0010-y>
- Chander, V., Gibbs, R. A., & Sedlazeck, F. J. (2019). Evaluation of computational genotyping of structural variation for clinical diagnoses. *GigaScience*, **8**(9), giz110. <https://doi.org/10.1093/gigascience/giz110>
- Charlesworth, B., Betancourt, A. J., Kaiser, V. B., & Gordo, I. (2009). Genetic recombination and molecular evolution. *Cold Spring Harbor Symposia on Quantitative Biology*, **74**, 177–186. <https://doi.org/10.1101/sqb.2009.74.015>
- Collins, R. L., Brand, H., Karczewski, K. J., Zhao, X., Alföldi, J., Francioli, L. C., Khera, A. V., Lowther, C., Gauthier, L. D., Wang, H., Watts, N. A., Solomonson, M., O'Donnell-Luria, A., Baumann, A., Munshi, R., Walker, M., Whelan, C. W., Huang, Y., Brookings, T., ... Talkowski, M. E.; Genome Aggregation Database Production Team. (2020). A structural variation reference for medical and population genetics. *Nature*, **581**(7809), 444–451. <https://doi.org/10.1038/s41586-020-2287-8>
- Comeault, A. A., Soria-Carrasco, V., Gompert, Z., Farkas, T. E., Buerkle, C. A., Parchman, T. L., & Nosil, P. (2014). Genome-wide association mapping of phenotypic traits subject to a range of intensities of natural selection in *Timema cristinae*. *The American Naturalist*, **183**(5), 711–727. <https://doi.org/10.1086/675497>
- Cramp, S., & Brooks, D. J. (1992). *Handbook of the birds of Europe, the Middle East and North Africa. The birds of the western Palearctic*, vol. VI. Warblers. Oxford University Press.
- Cridland, J. M., Macdonald, S. J., Long, A. D., & Thornton, K. R. (2013). Abundance and distribution of transposable elements in two drosophila QTL mapping resources. *Molecular Biology and Evolution*, **30**(10), 2311–2327. <https://doi.org/10.1093/molbev/mst129>
- Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**(13), 3133–3157. <https://doi.org/10.1111/mec.12796>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R.; 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, **27**(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Day, J. J., & Sweatt, J. D. (2011). Epigenetic mechanisms in cognition. *Neuron*, **70**(5), 813–829. <https://doi.org/10.1016/j.neuron.2011.05.019>
- Delmore, K. E., Hübner, S., Kane, N. C., Schuster, R., Andrew, R. L., Câmara, F., Guigó, R., & Irwin, D. E. (2015). Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. *Molecular Ecology*, **24**(8), 1873–1888. <https://doi.org/10.1111/mec.13150>
- Delmore, K. E., Illera, J. C., Pérez-Tris, J., Segelbacher, G., Lugo Ramos, J. S., Durieux, G., Ishigohoka, J., & Liedvogel, M. (2020a). The evolutionary history and genomics of European blackcap migration. *eLife*, **9**, e54462. <https://doi.org/10.7554/eLife.54462>
- Delmore, K. E., & Liedvogel, M. (2016). Investigating factors that generate and maintain variation in migratory orientation: A primer for recent and future work. *Frontiers in Behavioral Neuroscience*, **10**, 3. <https://doi.org/10.3389/fnbeh.2016.00003>
- Delmore, K. E., Van Doren, B. M., Conway, G. J., Curk, T., Garrido-Garduño, T., Germain, R. R., Hasselmann, T., Hiemer, D., van der Jeugd, H. P., Justen, H., Lugo Ramos, J. S., Maggini, I., Meyer, B. S., Phillips, R. J., Remisiewicz, M., Roberts, G. C. M., Sheldon, B. C., Vogl, W., & Liedvogel, M. (2020b). Individual variability and versatility in an eco-evolutionary model of avian migration. *Proceedings of the Royal Society B*, **287**(1938), 20201339. <https://doi.org/10.1098/rspb.2020.1339>
- Dingle, H. (2006). Animal migration: Is there a common migratory syndrome? *Journal of Ornithology*, **147**(2), 212–220. <https://doi.org/10.1007/s10336-005-0052-2>
- Eikenaar, C., Isaksson, C., & Hegemann, A. (2018). A hidden cost of migration? Innate immune function versus antioxidant defense. *Ecology and Evolution*, **8**(5), 2721–2728. <https://doi.org/10.1002/ece3.3756>
- Göbel, T. W., Chen, C. L., Lahti, J., Kubota, T., Kuo, C. -L., Aebbersold, R., Hood, L., & Cooper, M. D. (1994). Identification of T-cell receptor alpha-chain genes in the chicken. *Proceedings of the National Academy of Sciences*, **91**(3), 1094–1098. <https://doi.org/10.1073/pnas.91.3.1094>
- Gompert, Z., Brady, M., Chalyavi, F., Saley, T. C., Philbin, C. S., Tucker, M. J., Forister, M. L., & Lucas, L. K. (2019). Genomic evidence of genetic variation with pleiotropic effects on caterpillar fitness and plant traits in a model legume. *Molecular Ecology*, **28**(12), 2967–2985. <https://doi.org/10.1111/mec.15113>

- Gwinner, E. (1996). Circadian and circannual programmes in avian migration. *Journal of Experimental Biology*, **199**(Pt 1), 39–48. <https://doi.org/10.1242/jeb.199.1.39>
- Hämälä, T., Wafula, E. K., Guiltinan, M. J., Ralph, P. E., dePamphilis, C. W., & Tiffin, P. (2021). Genomic structural variants constrain and facilitate adaptation in natural populations of *Theobroma cacao*, the chocolate tree. *Proceedings of the National Academy of Sciences*, **118**(35), e2102914118. <https://doi.org/10.1073/pnas.2102914118>
- Helbig, A. J. (1991). Inheritance of migratory direction in a bird species: A cross-breeding experiment with SE- and SW-migrating blackcaps (*Sylvia atricapilla*). *Behavioral Ecology and Sociobiology*, **28**(1), 9–12.
- Heller, D., & Vingron, M. (2021). SVIM-asm: Structural variant detection from haploid and diploid genome assemblies. *Bioinformatics*, **36**(22–23), 5519–5521. <https://doi.org/10.1093/bioinformatics/btaa1034>
- Hirsch, C. D., & Springer, N. M. (2017). Transposable element influences on gene expression in plants. *Biochimica et Biophysica Acta (BBA)—Gene Regulatory Mechanisms, Plant Gene Regulatory Mechanisms and Networks*, **1860**(1), 157–165. <https://doi.org/10.1016/j.bbagr.2016.05.010>
- Hof, A., Campagne, P., Rigden, D. J., Yung, C. J., Lingley, J., Quail, M. A., Hall, N., Darby, A. C., Saccheri, I. J. (2016). The industrial melanism mutation in British peppered moths is a transposable element. *Nature*, **534**, 102–105.
- Horvath, R., Menon, M., Stitzer, M., & Ross-Ibarra, J. (2022). Controlling for variable transposition rate with an age-adjusted site frequency spectrum. *Genome Biology and Evolution*, **14**(2), evac016. <https://doi.org/10.1093/gbe/evac016>
- Huddleston, J., Chaisson, M. J. P., Steinberg, K. M., Warren, W., Hoekzema, K., Gordon, D., Graves-Lindsay, T. A., Munson, K. M., Kronenberg, Z. N., Vives, L., Peluso, P., Boitano, M., Chin, C.-S., Korlach, J., Wilson, R. K., & Eichler, E. E. (2017). Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome Research*, **27**(5), 677–685. <https://doi.org/10.1101/gr.214007.116>
- Huddleston, J., & Eichler, E. E. (2016). An incomplete understanding of human genetic variation. *Genetics*, **202**(4), 1251–1254. <https://doi.org/10.1534/genetics.115.180539>
- Hudson, R. R., Slatkin, M., & Maddison, W. P. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics*, **132**(2), 583–589. <https://doi.org/10.1093/genetics/132.2.583>
- Ishigohoka, J., Bascón-Cardozo, K., Bours, A., Fuß, J., Rhie, A., Mountcastle, J., Haase, B., Chow, W., Collins, J., Howe, K., Uliano-Silva, M., Fedrigo, O., Jarvis, E. D., Pérez-Tris, J., Illera, J. C., Liedvogel, M. (2021). Distinct patterns of genetic variation at low-recombining genomic regions represent haplotype structure. *BioRxiv preprint*, <https://doi.org/10.1101/2021.12.22.473882>.
- Jeffares, D. C., Jolly, C., Hoti, M., Speed, D., Shaw, L., Rallis, C., Balloux, F., Dessimoz, C., Bähler, J., & Sedlazeck, F. J. (2017). Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nature Communications*, **8**, 14061. <https://doi.org/10.1038/ncomms14061>
- Justen, H., & Delmore, K. E. (2022). The genetics of bird migration. *Current Biology*, **32**(20), R1144–R1149. <https://doi.org/10.1016/j.cub.2022.07.008>
- Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, **173**(1), 419–434. <https://doi.org/10.1534/genetics.105.047985>
- Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**(18), 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Liedvogel, M., Åkesson, S., & Bensch, S. (2011). The genetics of migration on the move. *Trends in Ecology & Evolution*, **26**(11), 561–569. <https://doi.org/10.1016/j.tree.2011.07.009>
- Liu, S., Ferchaud, A. -L., Grønkaer, P., Nygaard, R., & Hansen, M. M. (2018). Genomic parallelism and lack thereof in contrasting systems of three-spined sticklebacks. *Molecular Ecology*, **27**(23), 4725–4743. <https://doi.org/10.1111/mec.14782>
- Lundberg, M., Liedvogel, M., Larson, K., Sigeman, H., Grahn, M., Wright, A., Åkesson, S., & Bensch, S. (2017). Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. *Evolution Letters*, **1**(3), 155–168. <https://doi.org/10.1002/evl3.15>
- Lundberg, M., Mackintosh, A., Petri, A., & Bensch, S. (2023). Inversions maintain differences between migratory phenotypes of a songbird. *Nature Communications*, **14**(1), 452. <https://doi.org/10.1038/s41467-023-36167-y>
- Lutgen, D., Ritter, R., Olsen, R. A., Schielzeth, H., Gruselius, J., Ewels, P., García, J. T., Shirihai, H., Schweizer, M., Suh, A., & Burri, R. (2020). Linked-read sequencing enables haplotype-resolved resequencing at population scale. *Molecular Ecology Resources*, **20**(5), 1311–1322. <https://doi.org/10.1111/1755-0998.13192>
- Maere, S., Heymans, K., & Kuiper, M. (2005). BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, **21**(16), 3448–3449. <https://doi.org/10.1093/bioinformatics/bti551>
- Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., & Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *PLoS Computational Biology*, **14**(1), e1005944. <https://doi.org/10.1371/journal.pcbi.1005944>
- Marks, P., Garcia, S., Barrio, A. M., Belhocine, K., Bernate, J., Bharadwaj, R., Bjornson, K., Catalanotti, C., Delaney, J., Fehr, A., Fiddes, I. T., Galvin, B., Heaton, H., Herschleb, J., Hindson, C., Holt, E., Jabara, C. B., Jett, S., Keivanfar, N., ... Church, D. M. (2019). Resolving the full spectrum of human genome variation using Linked-Reads. *Genome Research*, **29**(4), 635–645. <https://doi.org/10.1101/gr.234443.118>
- McKinnon, E. A., & Love, O. P. (2018). Ten years tracking the migrations of small landbirds: Lessons learned in the golden age of bio-logging. *The Auk*, **135**(4), 834–856. <https://doi.org/10.1642/auk-17-202.1>
- Merlin, C., & Liedvogel, M. (2019). The genetics and epigenetics of animal migration and orientation: Birds, butterflies and beyond. *Journal of Experimental Biology*, **222**(Pt Suppl 1), jeb191890. <https://doi.org/10.1242/jeb.191890>
- Mettler, R., Schaefer, H. M., Chernetsov, N., Fiedler, W., Hobson, K. A., Ilieva, M., Imhof, E., Johnsen, A., Renner, S. C., Rolshausen, G., Serrano, D., Wesolowski, T., & Segelbacher, G. (2013). Contrasting patterns of genetic differentiation among blackcaps (*Sylvia atricapilla*) with divergent migratory orientations in Europe. *PLoS One*, **8**(11), e81365. <https://doi.org/10.1371/journal.pone.0081365>
- Møller, A. P., & Erritzøe, J. (1998). Host immune defence and migration in birds. *Evolutionary Ecology*, **12**(8), 945–953. <https://doi.org/10.1023/a:1006516222343>
- Noor, M. A., & Bennett, S. M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, **103**(6), 439–444. <https://doi.org/10.1038/hdy.2009.151>
- Noor, M. A. F., Grams, K. L., Bertucci, L. A., & Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences*, **98**(21), 12084–12088. <https://doi.org/10.1073/pnas.221274498>
- Peona, V., Blom, M. P., Frankl-Vilches, C., Milá, B., Ashari, H., Thébaud, C., Benz, B. W., Christidis, L., Gahr, M., Irestedt, M., & Suh, A.



- (2022). The hidden structural variability in avian genomes. *BioRxiv preprint*, <https://doi.org/10.1101/2021.12.31.473444>.
- Peona, V., Blom, M. P., Xu, L., Burri, R., Sullivan, S., Bunikis, I., Liachko, I., Haryoko, T., Jönsson, K. A., Zhou, Q., & Irestedt, M. (2021). Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. *Molecular Ecology Resources*, **21**, 263–286.
- Pérez-Tris, J., Bensch, S., Carbonell, R., Helbig, A., & Tellería, J. L. (2004). Historical diversification of migration patterns in a passerine bird. *Evolution*, **58**, 1819–1832.
- Piersma, T. (2011). Flyway evolution is too fast to be explained by the modern synthesis: proposals for an “extended” evolutionary research agenda. *Journal of Ornithology*, **152**(S1), 151–159. <https://doi.org/10.1007/s10336-011-0716-z>
- Pulido, F., & Berthold, P. (2010). Current selection for lower migratory activity will drive the evolution of residency in a migratory bird population. *Proceedings of the National Academy of Sciences*, **107**(16), 7341–7346. <https://doi.org/10.1073/pnas.0910361107>
- Qin, P., Lu, H., Du, H., Wang, H., Chen, W., Chen, Z., He, Q., Ou, S., Zhang, H., Li, X., Li, X., Li, Y., Liao, Y., Gao, Q., Tu, B., Yuan, H., Ma, B., Wang, Y., Qian, Y., ... Li, S. (2021). Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell*, **184**(13), 3542–3558.e16. <https://doi.org/10.1016/j.cell.2021.04.046>
- Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, **47**(W1), W191–W198. <https://doi.org/10.1093/nar/gkz369>
- Rönn, J.A.C. von, Shafer, A. B. A., & Wolf, J. B. W. (2016). Disruptive selection without genome-wide evolution across a migratory divide. *Molecular Ecology*, **25**, 2529–2541.
- Sanchez-Donoso, I., Ravagni, S., Rodríguez-Teijeiro, J. D., Christmas, M. J., Huang, Y., Maldonado-Linares, A., Puigcerver, M., Jiménez-Blasco, I., Andrade, P., Gonçalves, D., Friis, G., Roig, I., Webster, M. T., Leonard, J. A., & Vilà, C. (2022). Massive genome inversion drives coexistence of divergent morphs in common quails. *Current Biology*, **32**, 462–469.e6. <https://doi.org/10.1016/j.cub.2021.11.019>
- Schwander, T., Libbrecht, R., & Keller, L. (2014). Supergenes and complex phenotypes. *Current Biology*, **24**(7), R288–R294. <https://doi.org/10.1016/j.cub.2014.01.056>
- Suh, A., Smeds, L., & Ellegren, H. (2018). Abundant recent activity of retrovirus-like retrotransposons within and among flycatcher species implies a rich source of structural variation in songbird genomes. *Molecular Ecology*, **27**(1), 99–111. <https://doi.org/10.1111/mec.14439>
- Thomas, G. M., & Huganir, R. L. (2004). MAPK cascade signalling and synaptic plasticity. *Nature Reviews Neuroscience*, **5**(3), 173–183. <https://doi.org/10.1038/nrn1346>
- Todesco, M., Owens, G. L., Bercovich, N., Légaré, J. -S., Soudi, S., Burge, D. O., Huang, K., Ostevik, K. L., Drummond, E. B. M., Imerovski, I., Lande, K., Pascual-Robles, M. A., Nanavati, M., Jahani, M., Cheung, W., Staton, S. E., Muñoz, S., Nielsen, R., Donovan, L. A., ... Rieseberg, L. H. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*, **584**(7822), 602–607. <https://doi.org/10.1038/s41586-020-2467-6>
- Toews, D. P., Taylor, S. A., Streby, H. M., Kramer, G. R., & Lovette, I. J. (2019). Selection on VPS13A linked to migration in a songbird. *Proceedings of the National Academy of Sciences*, **116**, 18272–18274.
- Villoutreix, R., Ayala, D., Joron, M., Gompert, Z., Feder, J. L., & Nosil, P. (2021). Inversion breakpoints and the evolution of supergenes. *Molecular Ecology*, **30**(12), 2738–2755. <https://doi.org/10.1111/mec.15907>
- Villoutreix, R., Carvalho, C.F. de, Gompert, Z., Parchman, T. L., Feder, J. L., & Nosil, P. (2022). Testing for fitness epistasis in a transplant experiment identifies a candidate adaptive locus in *Timema* stick insects. *Philosophical Transactions of the Royal Society B*, **377**, 20200508.
- Visser, M. E., Caro, S. P., Oers, K. van, Schaper, S. V., & Helm, B. (2010). Phenology, seasonal timing and circannual rhythms: Towards a unified framework. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 3113–3127.
- Weischenfeldt, J., Symmons, O., Spitz, F., & Korbel, J. O. (2013). Phenotypic impact of genomic structural variation: Insights from and for human disease. *Nature Reviews Genetics*, **14**(2), 125–138. <https://doi.org/10.1038/nrg3373>
- Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., & Jaffe, D. B. (2017). Direct determination of diploid genome sequences. *Genome Research*, **27**(5), 757–767. <https://doi.org/10.1101/gr.214874.116>
- Weissensteiner, M. H., Bunikis, I., Catalán, A., Francoijs, K. -J., Knief, U., Heim, W., Peona, V., Pophaly, S. D., Sedlazeck, F. J., Suh, A., Warmuth, V. M. (2020). Discovery and population genomics of structural variation in a songbird genus. *Nature Communications*, **11**, 1–11.
- Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X. P., Pool, J. E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliusen, T. S., Zheng, H., Liu, T., He, W., Li, K., Luo, R., Nie, X., Wu, H., Zhao, M., Cao, H., ... Zou, J. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, **329**(5987), 75–78. <https://doi.org/10.1126/science.1190371>
- Zhan, S., Zhang, W., Niitepöld, K., Hsu, J., Haeger, J. F., Zalucki, M. P., Altizer, S., de Roode, J. C., Reppert, S. M., & Kronforst, M. R. (2014). The genetics of monarch butterfly migration and warning colouration. *Nature*, **514**(7522), 317–321. <https://doi.org/10.1038/nature13812>
- Zhang, L., Chaturvedi, S., Nice, C. C., Lucas, L. K., & Gompert, Z. (2023). Population genomic evidence of selection on structural variants in a natural hybrid zone. *Molecular Ecology*, **32**(6), 1497–1514. <https://doi.org/10.1111/mec.16469>
- Zhou, X., Carbonetto, P., & Stephens, M. (2013). Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genetics*, **9**(2), e1003264. <https://doi.org/10.1371/journal.pgen.1003264>
- Zichner, T., Garfield, D. A., Rausch, T., Stütz, A. M., Cannavó, E., Braun, M., Furlong, E. E. M., & Korbel, J. O. (2013). Impact of genomic structural variation in *Drosophila melanogaster* based on population-scale sequencing. *Genome Research*, **23**(3), 568–579. <https://doi.org/10.1101/gr.142646.112>