





Genomic and phenotypic changes associated with alterations of migratory behaviour in a songbird

Dezhi Zhang¹  | Huishang She¹ | Frank E. Rheindt²  | Lei Wu^{1,3} | Huan Wang^{1,3} | Kai Zhang^{1,3} | Yalin Cheng¹ | Gang Song¹ | Chenxi Jia¹ | Yanhua Qu¹  | Urban Olsson^{4,5} | Per Alström^{1,6}  | Fumin Lei^{1,3,7}

¹Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China

²Department of Biological Sciences, National University of Singapore, Singapore, Singapore

³College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China

⁴Department of Biology and Environmental Science, University of Gothenburg, Gothenburg, Sweden

⁵Gothenburg Global Biodiversity Centre, Gothenburg, Sweden

⁶Animal Ecology, Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

⁷Centre for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China

Correspondence

Fumin Lei, Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, 100101 Beijing, China.

Email: leifm@ioz.ac.cn

Funding information

National Natural Science Foundation of China, Grant/Award Number: 32270466 and 32070434; Second Tibetan Plateau Scientific Expedition and Research (STEP) programme, Grant/Award Number: 2019QZKK0304 and 2019QZKK0501; Swedish National Science Foundation, Grant/Award Number: 2019-04486; Jornvall Foundation and Mark and Mo Constantine; the National Key Research and Development Program of China, Grant/Award Number: 2022YFC2601601

Handling Editor: Paul Hohenlohe

Abstract

The seasonal migration of birds is a fascinating natural wonder. Avian migratory behaviour changes are common and are probably a polygenic process as avian migration is governed by multiple correlated components with a variable genetic basis. However, the genetic and phenotypic changes involving migration changes are poorly studied. Using one annotated near-chromosomal level de novo genome assembly, 50 resequenced genomes, hundreds of morphometric data and species distribution information, we investigated population structure and genomic and phenotypic differences associated with differences in migratory behaviour in a songbird species, Yellow-throated Bunting *Emberiza elegans* (Aves: Emberizidae). Population genomic analyses reveal extensive gene flow between the southern resident and the northern migratory populations of this species. The hand-wing index is significantly lower in the resident populations than in the migratory populations, indicating reduced flight efficiency of the resident populations. Here, we discuss the possibility that nonmigratory populations may have originated from migratory populations though migration loss. We further infer that the alterations of genes related to energy metabolism, nervous system and circadian rhythm may have played major roles in regulating migration change. Our study sheds light on phenotypic and polygenic changes involving migration change.

KEYWORDS

Emberiza elegans, gene flow, hand-wing index, migration change, polygenic

Dezhi Zhang and Huishang She contributed equally to the study.

1 | INTRODUCTION

Migratory behaviour is ubiquitous in a range of different animals, such as insects, marine invertebrates, fish, amphibians, reptiles, mammals and birds (Alerstam et al., 2003; Turbek et al., 2018). Avian migration is particularly complex, fascinating and variable in its many manifestations (Berthold, 2001; Newton, 2007; Pulido, 2007; Zink, 2011). Avian migratory behaviour (hereafter “migration”) involves multiple correlated components, such as hyperphagia, migratory restlessness, and the innate navigation system, that are variably plastic or genetically based. Differentiation in migratory behaviour is largely due to genetic differences, particularly for short-lived birds, as mean life expectancies of such species would be insufficient to allow individual gain of migratory experience (Pulido, 2007). Such differentiation is probably reflected in genetic changes related to genes responsible for energy metabolism, the circadian rhythm, nervous system and memory (Delmore et al., 2015, 2016, 2020; Dingle, 2006; Gu et al., 2021; Lundberg et al., 2017; Ruegg et al., 2014; Toews et al., 2019). In addition, migratory behaviour appears to be somewhat plastic, for example, climate change can cause rapid change in migratory behaviour (Bauer et al., 2008; Dufour et al., 2021).

Evolutionary change in migratory behaviour may result from specialization on different wintering grounds, migration routes or migratory timing regimes, as well as loss of migration (migration loss) in previously migratory populations or a corresponding gain in resident populations (migration gain). As migration involves a suite of phenotypically and genetically correlated traits (Dingle, 2006), a loss or gain of migratory behaviour is likely to have a more immediate and profound effect on genetic differentiation than other types of change in migratory behaviour. However, the underlying causes and genetic mechanisms of loss or gain of migratory behaviour remain poorly understood.

The Old World buntings (family Emberizidae, with a single genus, *Emberiza*) consist of 44 species widely distributed across Asia, Europe and Africa, with the highest abundance in the Palearctic region (Alström et al., 2008; Päckert et al., 2015; Winkler et al., 2020). Most *Emberiza* species are migratory or partially migratory (Winkler et al., 2020). The ancestor of this clade is assumed to be migratory, and multiple losses and regains of migratory behaviour have been postulated (Cai et al., 2021), indicating a high plasticity of migratory behaviour in *Emberiza* buntings. The Yellow-throated Bunting *Emberiza elegans* is of particular evolutionary interest: it has one mainly migratory population breeding in northeast China and two disjunct probably resident populations, one mainly in the Korean peninsula and the other mainly in southwest China (Figure 1a; Copete, 2020).

Using genomic, morphometric and species distribution data from both the northern migratory and the southern resident populations, we investigated population structure and genomic and phenotypic differences associated with differences in migratory behaviour. We detected significant difference in hand-wing index and extensive interbreeding between the resident and migratory populations. By

pairwise comparisons of genetic differentiation using sliding window F_{ST} among three populations, including two parental populations and one admixed population, we further investigated the potential polygenic basis for differences in migratory behaviour.

2 | MATERIALS AND METHODS

2.1 | Genome assembling and annotation

We de novo sequenced and assembled the draft genome of a male *E. elegans* (collected from the same location as sample 23570; Table S3) using a single 8 M SMRT Cell on the PacBio Sequel II platform (Pacific Biosciences; PacBio Sequel II System) and Hi-C reads generated from Illumina Novaseq 6000 platform. The final chromosome assembly ≥ 1 Mb was aligned to *Ficedula albicollis* genome (Ellegren et al., 2012) to determine the chromosome names using BLAST+ 2.2.26. We used REPEATMASKER OPEN-4.0 (<https://www.repeatmasker.org/>) to mask repetitive DNA sequences of the genome. The genome was annotated by combining homologous protein prediction and RNA-seq data. Details of genome assembly and annotation can be found in Appendix S1.

2.2 | Sampling and resequencing

A total of 50 individuals of *E. elegans* (Figure 1a) and one female of Black-faced Bunting *E. spodocephala* were sampled (Table S3). Total genomic DNA was extracted from tissue or blood using the DNeasy blood and tissue kit (Qiagen) according to the manufacturer's protocol. DNA libraries with ~350 bp inserts were constructed and sequenced using the Illumina NovaSeq 6000 platform with a PE read length of 150 bp by Berry Genomics. Raw reads were processed to remove adapter sequences, low-quality reads (those with over 50% of bases having Phred quality scores <3) and poly-N reads (those with $\geq 3\%$ unidentified nucleotides) using FASTP 0.20.0 (Chen et al., 2018).

2.3 | Variants calling and filtering

Quality controlled reads of all samples were mapped to the reference genome using BWA 0.7.12 (Li & Durbin, 2009). Polymerase chain reaction (PCR) duplicates were removed in SAMTOOLS 0.1.19 (Li et al., 2009). Depth and breadth coverage were estimated in BEDTOOLS 2.29.0 (Quinlan & Hall, 2010). Variants were called in SAMtools using the “mpileup” module for only the 50 *E. elegans* samples. We used VCFTOOLS 0.1.12b (Danecek et al., 2011) to filter SNPs according to the following criteria: (i) quality value ≥ 30 , (ii) genotype depth ≥ 8 , (iii) only biallelic SNPs were retained, (iv) SNPs with only no missing genotypes across all individuals were retained and (v) SNPs in the repetitive genomic regions were excluded. As some individuals are females (Table S3), autosomes and Z chromosome were processed

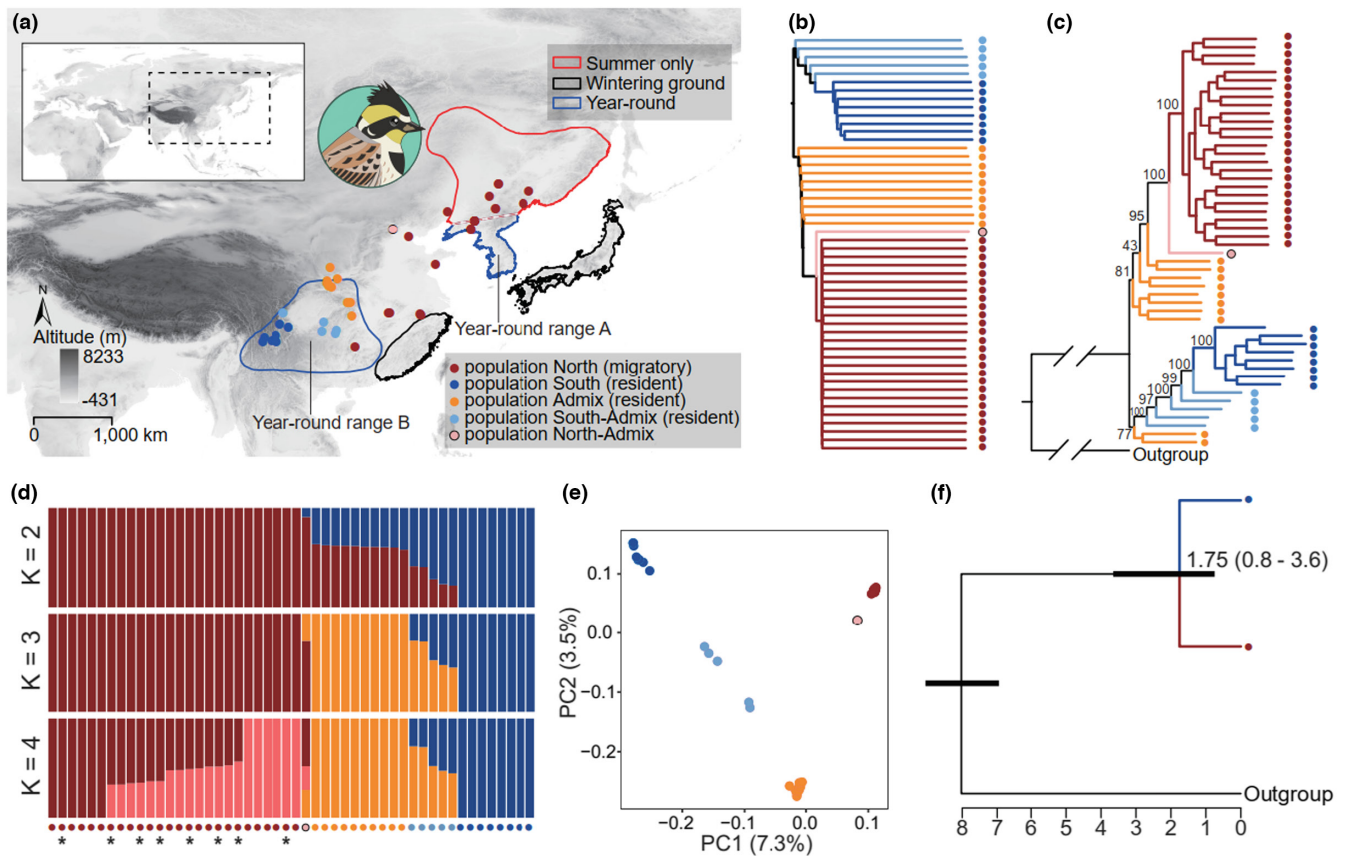


FIGURE 1 Sampling sites and population structure of *Emberiza elegans*. (a) Sampling localities and the approximate distribution range of *E. elegans*; the top left inset represents the study area; the distribution range of *E. elegans* is based on BirdLife (<https://www.birdlife.org/>); the image depicts a breeding male of population North; population North is generally migratory, populations South, Admix and South-Admix are generally resident according to the observation data in Figure S3. (b) Unrooted neighbour-joining tree based on 701,426 concatenated autosomal SNPs. (c) Rooted maximum likelihood tree using concatenated consensus sequences from 1000 randomly selected 10 kb windows; *E. spodocephala* was designated as an outgroup; the numbers represent bootstrap support values. (d) Admixture analysis using the same data set as in the NJ tree analysis at $K = 2, 3$ and 4 ; asterisks indicate samples from year-round range A. (e) Principal component analysis based on the same data set as in the NJ tree analysis; each axis displays the percentage of data explained for the first two principal components (PC). (f) Divergence time estimation in MCMCtree using the same 1000 randomly selected 10 kb consensus sequences; a root divergence time of 7–9 million years between *E. elegans* and *E. spodocephala* was used; divergence times with 95% highest posterior distribution (HPD; in brackets) is displayed above the nodes; the black horizontal bars represent 95% HPD; time scale in millions of years. [Colour figure can be viewed at wileyonlinelibrary.com]

separately. After removing SNPs from the repeated genomic regions, a total of 10,483,802 autosomal SNPs and 755,221 SNPs on Z chromosome were retained, respectively.

2.4 | Phylogenetic and population structure analyses

A total of 701,426 concatenated autosomal SNPs with a physical distance ≥ 1000 bp filtered by VCFtools were used to infer a neighbour-joining (NJ) tree using TREEBEST 1.9.2 (Vilella et al., 2009). We used SAMTOOLS 0.1.19 (Li et al., 2009) and BCFTOOLS 1.3.1 (Li et al., 2009) on the BAM files to generate 10 kilobase (kb) consensus sequences with 100 kb distance, and these consensus sequences were aligned using MAFFT 7.464 (Nakamura et al., 2018) and trimmed in GBLOCKS 0.91b (Talavera & Castresana, 2007). Due to the computational limitations, only 1000 randomly selected sequences were used to reconstruct

the phylogenetic tree in IQTREE 1.6.9 (Nguyen et al., 2015) with 1000 bootstraps. *E. spodocephala* was set as the outgroup. Using the same 1000 10 kb consensus sequences, we inferred the divergence time between population South and North using MCMCtree in PAML 4.8 (Yang, 2007). We set a root divergence time of 7–9 million years between *E. elegans* and the outgroup taxon *E. spodocephala* according to previous studies (Cai et al., 2021; Päckert et al., 2015). Admixture analysis was conducted in NGSADMIX (Skotte et al., 2013) using the same data set to TREEBEST with K values ranging from 2 to 5 and a minor allele frequency cutoff of 0.05. PCA analysis was performed in GCTA 1.24 (Yang et al., 2011) using the same data set. According to the population structure analyses, five populations were classified: population North (the genetically pure northern migratory population), population South (the genetically pure southern nonmigratory population), population Admix (intermediate between population South and North), population South-Admix (intermediate between population South and Admix), and population North-Admix (intermediate

between population North and Admix) (Figure 1; Table S3; see Section 3).

2.5 | MtDNA phylogenetic analysis

We assembled mitochondrial genomes for the 50 samples of *E. elegans* and one sample of *E. spodocephala* in MITOBIM 1.8 (Hahn et al., 2013), which relies on the sequence assembler MIRA 4.0.1 (Chevreux et al., 1999). A complete mitochondrial sequence of Chestnut-eared Bunting *Emberiza fucata* (GenBank accession: NC_033338.1) was used as the reference. All ambiguous sites of the initial assembly were converted to homozygous sites using the command "miraconvert". Mitochondrial sequences of two individuals, one from population Admix and another from population South-Admix showed extremely great differentiation compared to the other individuals of *E. elegans*, possibly due to low mitochondrion copy number in blood. These two individuals were excluded from the following mtDNA phylogenetic analysis. All mitochondrial sequences were aligned using an online web server (Madeira et al., 2019). The alignment was manually checked and trimmed. A total of 14,432 bp mitochondrial alignment was retained. A maximum likelihood mitochondrial phylogenetic tree was inferred in IQTREE 1.6.9 (Nguyen et al., 2015) with *E. spodocephala* as the outgroup taxon.

2.6 | Effective population size, linkage disequilibrium and heterozygosity estimations

Only two parental populations (South and North) were included in these analyses, while populations with hybrid origins were excluded. Autosomal data were used for effective population size (N_e), linkage disequilibrium (LD) and heterozygosity estimations. We used pairwise sequentially Markovian coalescent analyses (PSMC; Li & Durbin, 2011) for N_e estimations for population South and North. A mutation rate of 0.33% per million years (Zhang et al., 2014) and generation time of 1.84 years (Bird et al., 2020) were applied in the PSMC analysis. Although gene flow will affect effective population size estimation by affecting estimation of coalescent events in PSMC, the use of populations South and North that are relatively little affected by gene flow (see Section 3) would minimize this bias. LD was estimated for population South, North with even samples size ($n = 8$). For LD analysis, genotypes were phased in BEAGLE 4.1 (Browning & Browning, 2007) with the "gt" option; we then estimated the correlation coefficient (r^2) within 10 kb blocks in VCFTOOLS 0.1.12b (Danecek et al., 2011) with the "--hap-r2" and "--ld-window-bp 10000" options. The averaged r^2 within 100 bp was plotted against physical distance in R 3.6.3 (R Development Core Team, 2008). We used SAMTOOLS 0.1.19 (Li et al., 2009) and BCFTOOLS 1.3.1 (Li et al., 2009) on the BAM files to generate consensus fasta sequences with a minimal sequencing depth of 10 and a maximal sequencing depth of 100. We used the ratio of heterozygous sites to indicate heterozygosity, which was defined as "Heterozygosity = number of heterozygous

SNPs/(number of heterozygous SNPs + number homozygous SNPs)" in the resulting consensus sequence.

2.7 | Hand-wing index measurements of specimens

Hand-wing index (HWI) is a widely used parameter as a proxy for avian flight efficiency and dispersal ability (Sheard et al., 2020). HWI is defined as: $HWI = (100 \times \text{Kipp's distance} / \text{wing length})$, where Kipp's distance is the distance between the tip of the most distal secondary feather and the tip of the longest primary feather on the folded wing (Kipp, 1959). Here, we compared the HWI between the northern migratory and southern resident populations. We measured the HWI for a total of 164 specimens collected from 1913 to 2011, preserved at the National Zoological Museum, Institute of Zoology, Chinese Academy of Sciences, Beijing, including 100 from population North, 19 from population South and 45 from population Admix (Table S4). The wing length and Kipp's distance were measured by a single person (D.Z.). We classified these specimens to population North, South or Admix according to the sampling localities and sampling dates (Table S4). For example, a specimen was classified to a corresponding population (South, North or Admix) if it was collected from the corresponding breeding ranges of each population, and specimen collected in the nonbreeding period from southeast China, where the species does not breed, were classified as population North.

2.8 | Collection of observation data

We downloaded observation records of *E. elegans* from GBIF (the Global Biodiversity Information Facility, GBIF.org (20 June 2022) GBIF Occurrence Download: <https://doi.org/10.15468/dl.tgbacg>). Observation points far from the species' range were removed. A total of 483 and 235 observations from 2000 to 2021 generally fell in the ranges of the southern resident populations and the northern migratory population.

2.9 | Determination of migratory and resident status using observation data

Although the specific movement data of *E. elegans* are largely unknown, the migratory status of the main part of the north-eastern population and the resident status of the south-western population is well supported (Copete, 2020). By comparing the difference of the occurrence between nonwintering and wintering periods in the same distribution range, we roughly determine whether the species is migratory or nonmigratory. Wintering period is identified between November to February, and nonwintering period is identified between March to October according to other data for *E. elegans* (Copete, 2020). Significant difference between nonwintering and wintering periods of the occurrence data in the breeding

range would suggest that the population is migratory. Otherwise, it would be classified as nonmigratory. We counted the number of observations in the northern and southern breeding areas for each month, and then checked whether there was a significant difference between nonwintering and wintering periods. Note that the observation intensity between nonwintering and wintering periods might have affected the collections of observation data.

2.10 | Ecological niche modelling in Maxent

We used Maxent (Phillips et al., 2006) to predict the breeding range shifts of *E. elegans* during the Last Glacial Maximum (LGM; 0.021–0.018 million years ago) and Last Interglacial periods (LIG; 0.13–0.07 million years ago). The northern migratory and southern resident populations were processed separately. The 19 bioclimatic variables of present-day, LGM and LIG were downloaded from WORLDCLIM (<http://www.worldclim.org/>; Hijmans et al., 2005) at a resolution of 2.5 arc-min. The present-day climate data was used to predict the ecological niche of the species. Based on the premise of niche conservatism, Maxent predicts the potential breeding range during the LGM and LIG based on the historical climate. The previously used observation data downloaded from GBIF were used in Maxent analysis. Only the localities generally falling in the breeding ranges and in the potential breeding time from March to October were retained. To reduce the effect of spatial autocorrelation, localities separated from each other more than 0.1° in both longitudinal and latitudinal levels were retained. A total of 140 and 108 localities for the southern resident and northern migratory populations, respectively, were used in the final Maxent analysis. We set random test percentage to 25 and replicates to 10, and set the replicated run type to bootstrap. The maximum iterations was set to 1000. We applied a threshold rule of 10 percentile training presence and left other parameters default.

2.11 | Sliding window F_{ST} and gene annotation analyses

We calculated pairwise F_{ST} among the populations North, South and Admix in 10 kb nonoverlapping sliding window using VCFTOOLS 0.1.12b (Danecek et al., 2011). The weighted sliding window F_{ST} (estimated as a ratio of averages) was used as recommended by a previous study (Bhatia et al., 2013). Windows with fewer than five SNPs were excluded. A negative F_{ST} was converted to zero. Autosomes and the Z chromosome were processed separately. Some of the most highly differentiated genomic regions between migratory and resident populations might be causally linked to migratory behaviour change. However, genomic differentiation can also be caused by local adaptation that is unrelated to migration, and also be caused by characteristics of the genomic architecture, such as variation in recombination rate and nucleotide diversity along the genome (Burri et al., 2015; Irwin et al., 2018). To disentangle these factors, we

compared genomic differentiation (F_{ST}) between the migratory population North and the two resident populations (South and Admix). F_{ST} was calculated along 10 kb nonoverlapping sliding windows. We defined the first ΔF_{ST} ($\Delta F_{ST[NS-SA]}$) as the F_{ST} between North and South minus the F_{ST} between the two resident populations South and Admix ($\Delta F_{ST[NS-SA]} = F_{ST[North\ vs.\ South]} - F_{ST[South\ vs.\ Admix]}$) and the second ΔF_{ST} ($\Delta F_{ST[NA-SA]}$) as the F_{ST} between migratory population North and resident population Admix minus the F_{ST} between the two resident populations South and Admix ($\Delta F_{ST[NA-SA]} = F_{ST[North\ vs.\ Admix]} - F_{ST[South\ vs.\ Admix]}$). We defined genomic windows that satisfied the following criteria as the candidate genomic regions: they must exhibit a top 1% F_{ST} value in both ΔF_{ST} parameters as defined above and a lower-than-average F_{ST} between the two resident populations ($F_{ST[South\ vs.\ Admix]}$). We then calculated F_{ST} between population South and North for each SNP in these candidate windows. SNPs with the largest F_{ST} were retained and annotated using SNPeff 5.0e (Cingolani et al., 2012). Both upstream and downstream genes were retained for the intergenic SNPs. We then extracted coding sequences for all annotated genes and used KOBAS 3 (Bu et al., 2021) to conduct gene functional annotation on the basis of *Homo sapiens* database.

3 | RESULTS

3.1 | Genome assembly, annotation and whole-genome resequencing

The assembled genome is 1.22 gigabases (Gb) in length and contains 380 scaffolds with an N50 and N90 of ~74.0 megabases (Mb) and ~14.7 Mb, respectively. The average scaffold length is ~3.2 Mb, with the longest at 155.6 Mb and a total of 39 scaffolds longer than 1 Mb. The completeness of the assembly is 93.9% as estimated by BUSCO 2.0 (Simao et al., 2015). A total of 243 Mb (19.9%) of the genome was masked by REPEATMASKER OPEN-4.0 (<http://www.repeatmasker.org/>; Table S1). A total of 14,712 protein-coding genes and 1115 non-coding RNA genes were annotated. All 30 homologous autosomes and one Z chromosome were identified by comparison with the genome assembly of Collared Flycatcher *Ficedula albicollis* (Ellegren et al., 2012). Beyond these 31 known chromosomes, multiple scaffolds of a relatively large size (larger than the shortest known autosome LGE22) were also assembled (scaffold scaf_Ch08 = 68.0 Mb, scaf_Ch18 = 18.2 Mb, scaf_Ch20 = 16.8 Mb, scaf_Ch27 = 8.0 Mb and scaf_Ch34 = 5.9 Mb), and turned out to be enriched for repetitive sequences (Table S2). To explore the origin of these unknown chromosomes, we aligned them to the known chromosomes of *E. elegans*. The results showed that some of these unknown chromosomes, especially for scaf_Ch08, scaf_Ch18 and scaf_Ch20, were highly homologous with the terminal sequences of other known chromosomes (Figure S1), suggesting that they might originate from the fusion of telomeres from multiple other chromosomes.

The genomes of 50 individuals of *E. elegans* were resequenced (Figure 1a; Table S3), including the migratory population from northeast China (breeding ground, top right red line in Figure 1a),

the putatively nonmigratory population from the year-round range A (Korean peninsula and the adjacent areas, top right blue line in Figure 1a), the nonmigratory population from the year-round range B (southwest China, bottom left blue line in Figure 1a) and multiple samples from putative migratory stopover and/or wintering sites. We also resequenced one female sample of *Emberiza spodocephala* as the outgroup. The average sequencing depth and breadth of coverage for all resequenced genomes were 17.5% and 97.4% (Table S3), respectively, after removing PCR duplicates.

3.2 | Population structure

The unrooted neighbour joining (NJ) tree, based on 701,426 concatenated SNPs with a distance of ≥ 1000 base pairs, recovered two main clades (Figure 1b). One clade included the samples collected from the southwestern year-round range B (deep blue and light blue sampling sites in Figure 1a; Table S1); the other clade included the samples collected from the summer-only breeding grounds, the potential stop-over/wintering sites and the year-round range A (deep red and pale pink sampling sites in Figure 1a); samples collected from the northern part of the year-round range B were also included in this clade (orange sampling sites in Figure 1a).

We also inferred a maximum likelihood (ML) tree using concatenated consensus sequences from 1000 randomly selected 10 kb windows with *E. spodocephala* as an outgroup. The rooted ML tree based on concatenated consensus sequences is generally similar to the unrooted NJ tree based on concatenated SNPs, except that the samples from the northern part of the year-round range B (population "Admix", with orange symbols in Figure 1) were placed in different clades, with moderate bootstrap support (77% and 81%, respectively; Figure 1c). The clade with the northern samples (red and pink symbols) and the clade with the samples with blue symbols in the year-round range B received strong bootstrap support (100%), whereas the samples with orange symbols in the year-round range B were less confidently placed in the tree (Figure 1c). The nodes with the orange samples also presented low to medium bootstrap support (Figure 1c). However, both the root nodes of the deep red and deep blue samples presented 100% bootstrap support (Figure 1c).

Using the same data set as in the NJ tree analysis, admixture analysis showed that the samples from the northern part of the year-round range B (population "Admix", with orange symbols in Figure 1) were intermediate between the southwestern part of the year-round range B and the northern migratory samples (deep blue and deep red symbols in Figure 1a-c) at $K = 2$, with genetic proportions originating from the deep red and deep blue samples at 61% and 39%, respectively (Figure 1d). The other samples from the year-round range B (light blue symbols in Figure 1a-c) exhibited the hallmarks of admixture between the deep blue and the orange samples, and the single sample from the north indicated by a pale pink symbol in Figure 1a-c showed the signs of admixture between the deep red and the orange samples at $K = 3$ and 4 (Figure 1d). Similarly, in principal component analysis (PCA), the orange, pale blue and pink

samples occupied intermediate positions between the deep blue and red samples along PC1 (Figure 1e).

Combining the tree topologies (Figures 1b,c), Admixture and PCA analyses (Figure 1d,e), we therefore defined the deep red samples as population North ($n = 26$), the deep blue samples as population South ($n = 8$), the orange samples as population Admix ($n = 10$; intermediate between population South and North, generally non-migratory), the light blue samples as population South-Admix ($n = 5$; intermediate between population South and Admix), and the pale pink sample as population North-Admix ($n = 1$) (intermediate between population North and Admix) (Figure 1; Table S3). Samples matching population North collected away from the known breeding range were considered to be likely on migration or in winter quarters.

MCMCtree analysis based on the same 1000 10 kb consensus sequences revealed a divergence time of 1.75 million years between population South and North (95% highest posterior density [HPD]: 0.8–3.6 million years) (Figure 1f). The divergence time might be underestimated as MCMCtree analysis does not take into account gene flow between population South and North.

3.3 | Mitochondrial tree

Two well supported mitochondrial clades were recovered: samples from population South, Admix, South-Admix and North-Admix and four samples from population North formed one clade (bootstrap 95%) and the rest of the samples from population North formed another clade (bootstrap 98%; Figure S2). There was no obvious mitochondrial population structure within each clade (Figure S2).

3.4 | Migratory and resident status

In the northern migratory population, the number of observations on the breeding grounds (Figure 1a) was significantly lower in the putative migratory than in the nonmigratory period (Wilcoxon rank sum test, $p = .008$), with no observations in December (Figure S3a). On the contrary, in the range of the southern resident populations, the number of observations from the breeding period did not differ significantly from the nonbreeding period (Wilcoxon rank sum test, $p = .106$) (Figure S3b). The observation records corroborated that population North is generally migratory and populations South, Admix and South-Admix are generally resident.

Individuals of the putatively resident population from the year-round range A (samples with asterisks in Figure 1d; Table S3) did not exhibit any noticeable genomic differentiation from the migratory population (Figures 1c,d). However, their sampling localities were close to the very poorly known distributional boundary between these two groups (Figure 1a). Therefore, these samples were included in the migratory population North throughout the study. It should be noted that even individuals of the same population may have different migratory behaviour due to the highly plastic characteristic of migration. Therefore, the classification of migratory

behaviour in this study is relatively crude because there is no exact migration information for each sample.

3.5 | Lower genetic diversity and higher linkage disequilibrium of the resident population South compared to the migratory population North

Population South generally presented higher LD (Wilcoxon rank sum test, $p < .001$; Figure 2a), lower effective population size (N_e) estimated from PSMC analysis (Figure 2b), and significantly lower genome-wide heterozygosity (Wilcoxon rank sum test, $p < .001$; Figure 2c) than population North. Population South had lower N_e than population North only in recent times but displayed similar N_e earlier than 100,000 years ago (Figure 2b). Higher levels of genome-wide LD suggested a past bottleneck (Zhang et al., 2004).

3.6 | Lower hand-wing index of the southern resident populations compared to the northern migratory population

The HWI of the southern resident populations (population South and Admix) was significantly lower than that of the northern migratory population (population North) for both males and females (p -values $< .001$), while there were no significant differences between population South and Admix for both males and females (p -values $> .05$) (Figure 3).

3.7 | Ecological niche modelling

Despite slight over-prediction, the current breeding areas for both the northern migratory and southern resident populations were well predicted (Figure 4). We found that the breeding area of the northern migratory population was greatly reduced during the LGM period and was divided into two disconnected regions with the

southern one located in approximately the current breeding range of population South (Figure 4a). In contrast, during the LIG period, when the ice sheet was shrinking, the main breeding area was in the north, with only a small part of potential breeding range located in the current breeding range of the southern resident populations (Figure 4a). During the LIG and LGM periods, the main breeding area of the southern populations fell within the current breeding areas (Figure 4b).

3.8 | Identifying genomic regions underlying migratory behaviour change

A total of 264 windows were identified as the putative candidate genomic regions underlying the differences in migratory behaviour (red dots in Figure 5). Genes directly related to migratory change might be included in these candidate genomic regions. SNPs with the highest F_{ST} between population South and North of each window were considered as the candidate SNPs. A total of 243 genes were annotated in the *E. elegans* genome with both up- and downstream genes retained for the intergenic SNPs. According to the annotated results in KOBAS 3 (Bu et al., 2021) based on human database, 222 of these genes are known. We further analysed gene functions of these 222 known genes using gene ontology (GO) enrichment analyses in KOBAS 3 (Bu et al., 2021). As mentioned before, changes in migratory behaviour are probably reflected in changes of the genes responsible for energy metabolism, the circadian rhythm, nervous system and memory. Among 167 significantly enriched GO terms (corrected p -values $< .05$), we found that four GO terms were potentially related to energy utilization (including GO term mitochondria [GO:0005739], mitochondrial outer membrane [GO:0005741], positive regulation of fat cell differentiation [GO:0045600] and white fat cell differentiation [GO:0050872]), nine GO terms were potentially related to the nervous system (including nervous system development [GO:0007399], neuron fate commitment [GO:0048663] and neuronal cell body [GO:0043025]), and one GO term was potentially

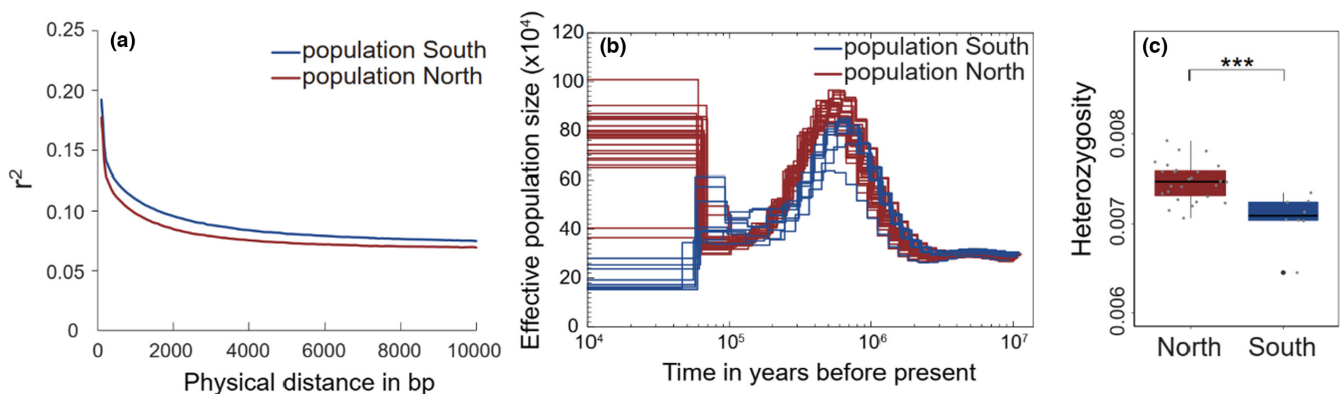


FIGURE 2 (a) Linkage disequilibrium decay for the resident population South and migratory population North; (b) PSMC estimates of change in effective population size for populations South and North; (c) Genome-wide heterozygosity is significantly lower in population South than in population North. ***Represents $p < .001$ estimated by Wilcoxon rank sum test. [Colour figure can be viewed at wileyonlinelibrary.com]

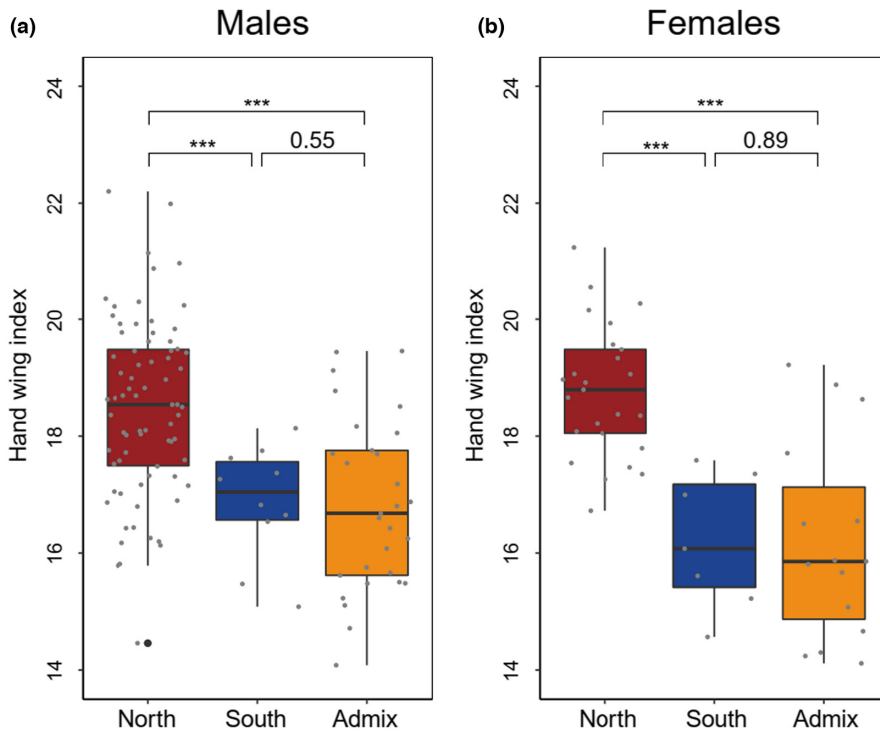


FIGURE 3 Hand-wing index (HWI) for (a) males and (b) females. HWI of the migratory population North is significantly higher than in both the resident population South and Admix for both males and females, while there is no significant differences between population South and Admix for both males and females. ***Represents $p < .001$ estimated by Wilcoxon rank sum test, and the nonsignificant p -values are displayed in numerical values. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1111/mec.16763)]

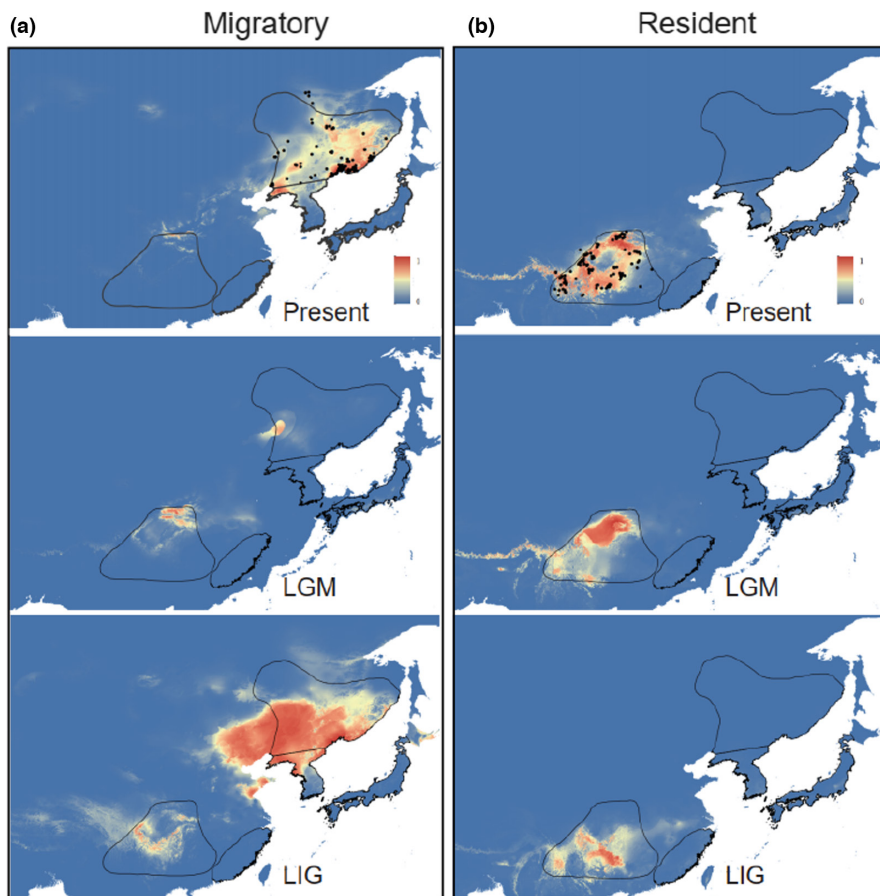


FIGURE 4 Breeding range predictions for (a) the northern migratory population and (b) the southern resident populations. From top to bottom, each panel represents the predicted results for the present, the last glacial maximum (LGM), and the last interglacial (LIG), respectively. The black lines represent the current distribution range of *Emberiza elegans* (see [Figure 1a](#)). The black dots represent the localities used in Maxent analyses. Warmer colours (with low in blue and high in red) represent higher distribution likelihoods. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/terms-and-conditions)]

related to circadian rhythm (GO:0007623) ([Table S5](#)). These 14 candidate GO terms (corrected p -values $< .05$), involving 65 candidate genes, might have played major roles in migration change ([Table S5](#)).

4 | DISCUSSION

Studies on animal migration, in particular the seasonal migration of birds, have grown explosively over the past 20 years largely due to

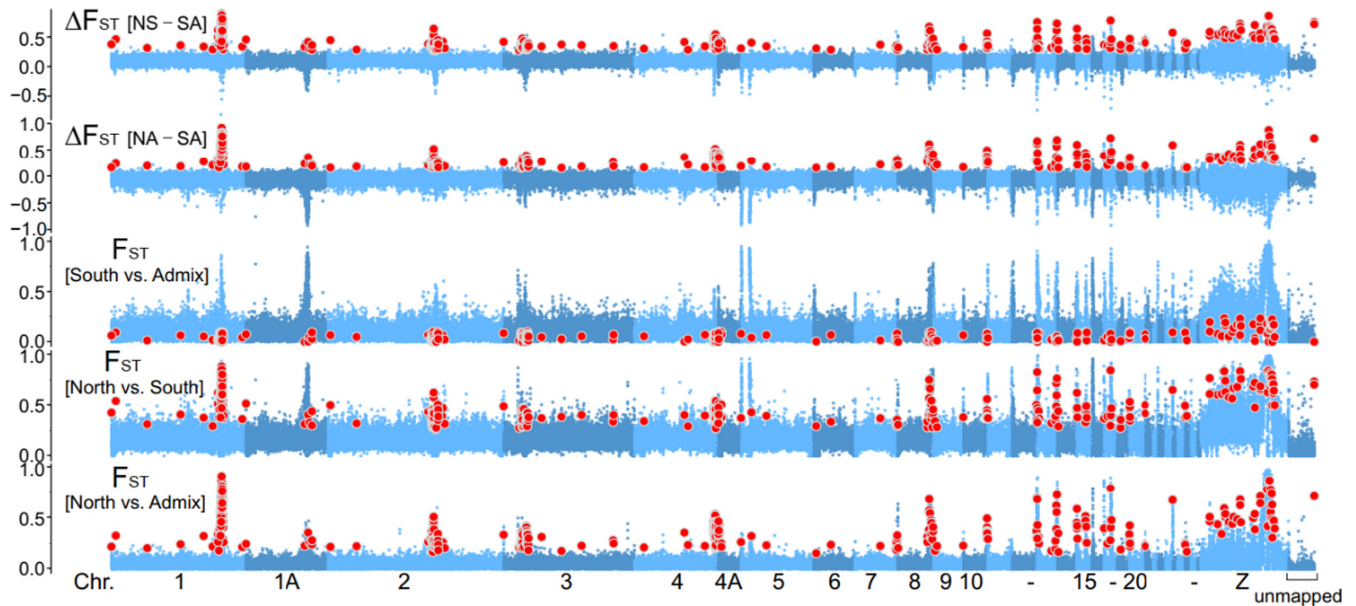


FIGURE 5 The distributions of the candidate windows potentially underlying differences in migratory behaviour. The distributions of F_{ST} and ΔF_{ST} across the genome; the red circles in each panel represent the candidate windows of excess differentiation between migratory (North) and resident populations (South and Admix); alternating dark and light blue colours indicate different chromosomes with labels underneath; note that the repeated genomic regions were removed. [Colour figure can be viewed at wileyonlinelibrary.com]

the technological revolution in tracking, remote-sensing and genome sequencing techniques (Fudickar et al., 2021). Migration gain or loss are common across different bird lineages (Cai et al., 2021; Gomez-Bahamon et al., 2020; Voelker et al., 2013; Voelker & Light, 2011; Winger et al., 2012), however, the underlying causes and genetic mechanisms remain poorly understood, possibly because avian migration involves a suite of complex phenotypically and genetically correlated traits (Dingle, 2006). By integrating genomic, morphometric and species distribution data, we investigated phenotypic and genetic differences between migratory and nonmigratory populations of *E. elegans*.

4.1 | Population structure, gene flow and phenotypic change

Population structure and phylogenetic analyses revealed a relatively deep divergence (differentiation time of 1.75 million years) between the northern migratory population (population North) and the nonmigratory population in the southernmost breeding area (population South) (Figure 1). At the northern end of the southern breeding area, a nonmigratory hybrid population (population Admix), which is intermediate between population South and North, was found, and further hybridization between population Admix and the two parental populations was also revealed (Figure 1). We also found an intermediate sample between population North and Admix (population North-Admix) (Figure 1). These results indicate extensive nuclear gene flow between migratory and nonmigratory populations. In addition, we also found mitochondrial introgression from the southern resident populations to the northern migratory population

(Figure S2). Ecological niche modelling analyses suggest that the breeding areas of the migratory population was partly overlapping with that of the southern resident populations, especially during the LGM period (Figure 4a). This may be responsible for the extensive genetic exchange between the presently geographically disjunct northern migratory and southern resident populations.

Migratory birds generally show higher HWI than nonmigratory ones (Sheard et al., 2020). A decrease of HWI is thought to be adaptive during a shift from migratory to nonmigratory behaviour (Gomez-Bahamon et al., 2020). In *E. elegans*, we also observed that the southern nonmigratory populations South and Admix displayed significantly lower HWI than the northern migratory population North (Figure 3), indicating an adaptive change of flight efficiency between migratory and nonmigratory populations.

Migration can be an ancestral or derived trait in *E. elegans*. Hence, the southern nonmigratory populations could have been derived from the northern migratory population through migration loss, or the northern migratory population may have originated from the southern nonmigratory populations through migration gain. Both migration losses and gains have been suggested to have occurred in other species of *Emberiza* (Cai et al., 2021). The southern nonmigratory population (population South) have higher LD and lower N_e and genome-wide heterozygosity than the northern migratory population (population North) (Figure 2), which may indicate signals of a possible founder effect with a reduction of nucleotide diversity (Gomez-Bahamon et al., 2020; Kondo et al., 2008) in population South, lending support to the migration loss hypothesis. However, this prediction still needs to be considered with cautions as it is based on the simplistic species evolutionary history and indirect evidence.

4.2 | Potential polygenic basis for differences in migratory behaviour

Migratory syndrome involves many phenotypically correlated traits (Dingle, 2006), and therefore, a large number of corresponding genetic differences might be involved when comparing migratory and resident populations. Using genetic differentiation scans, we identified 222 candidate genes from 264 candidate 10 kb windows that might be responsible for differences in migratory behaviour (Figure 5). We further performed gene functional annotation and found 167 significantly enriched GO terms (Table S5). Based on the assumption that the underlying genetic changes associated with migratory behaviour would most likely concern genes involved in energy metabolism, circadian rhythm, nervous system and memory (Delmore et al., 2016, 2020; Gu et al., 2021; Gwinner, 1996; Ruegg et al., 2014), we predict that 14 of these GO terms (including 65 genes) related to energy metabolism, nervous system and circadian rhythm might have played a major role in migration change of *E. elegans*. These results suggest that changes in migratory behaviour may have a polygenic basis. It should be pointed out that the classification of migratory and nonmigratory in this study is relatively crude, and future studies incorporating both individual movement data and the corresponding genomic data would greatly increase the precision of the identification of genetic basis of migratory change.

AUTHOR CONTRIBUTIONS

The study was designed by Fumin Lei and Dezhi Zhang. Sample collection was carried out by Fumin Lei, Chenxi Jia, Per Alström, Gang Song and Dezhi Zhang. The data were analysed by Dezhi Zhang, Huishang She, Lei Wu and Huan Wang. The manuscript was written by Dezhi Zhang, Huishang She, Urban Olsson, Per Alström, Frank E. Rheindt and Fumin Lei and commented on by all authors.

ACKNOWLEDGEMENTS

We thank Zuohua Yin and Xiaobing Li for sample collections. We thank the National Zoological Museum of China for the assistance of specimen examination and measurement. We also thank the editor and three anonymous reviewers for helpful comments and suggestions in improving the manuscript. This research was funded by the National Science Foundation of China (32270466 to D.Z.), the National Key Research and Development Program of China (2022YFC2601601); the Second Tibetan Plateau Scientific Expedition and Research (STEP) programme (2019QZKK0501, 2019QZKK0304 to F.L. and D.Z.); the Swedish National Science Foundation (2019-04486), Jorvall Foundation and Mark and Mo Constantine (to P.A.) and the National Science Foundation of China (32070434) to G.S.

CONFLICT OF INTEREST

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

Whole-genome resequencing data have been deposited at the NCBI Sequence Read Archive (SRA) under Bioproject PRJNA751503. The

genome assembly has been deposited at NCBI under GenBank assembly accession: GCA_022818055.1.

ORCID

Dezhi Zhang  <https://orcid.org/0000-0003-1523-3168>
 Frank E. Rheindt  <https://orcid.org/0000-0001-8946-7085>
 Yanhua Qu  <https://orcid.org/0000-0002-4590-7787>
 Per Alström  <https://orcid.org/0000-0001-7182-2763>

REFERENCES

- Alerstam, T., Hedenström, A., & Åkesson, S. (2003). Long-distance migration: Evolution and determinants. *Oikos*, 103(2), 247–260.
- Alström, P., Olsson, U., Lei, F., Wang, H. T., Gao, W., & Sundberg, P. (2008). Phylogeny and classification of the Old World Emberizini (Aves, Passeriformes). *Molecular Phylogenetics and Evolution*, 47(3), 960–973.
- Bauer, S., Van Dinther, M., Hogda, K. A., Klaassen, M., & Madsen, J. (2008). The consequences of climate-driven stop-over sites changes on migration schedules and fitness of Arctic geese. *Journal of Animal Ecology*, 77(4), 654–660.
- Berthold, P. (2001). *Bird migration: A general survey*. Oxford University Press.
- Bhatia, G., Patterson, N., Sankararaman, S., & Price, A. L. (2013). Estimating and interpreting F_{ST} : The impact of rare variants. *Genome Research*, 23, 1514–1521.
- Bird, J. P., Martin, R., Akçakaya, H. R., Gilroy, J., Burfield, I. J., Garnett, S. T., Symes, A., Taylor, J., Şekercioğlu, Ç. H., & Butchart, S. H. M. (2020). Generation lengths of the world's birds and their implications for extinction risk. *Conservation Biology*, 34(5), 1252–1261.
- Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics*, 81(5), 1084–1097.
- Bu, D., Luo, H., Huo, P., Wang, Z., Zhang, S., He, Z., Wu, Y., Zhao, L., Liu, J., Guo, J., Fang, S., Cao, W., Yi, L., Zhao, Y., & Kong, L. (2021). KOBAS-i: Intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Research*, 49(W1), W317–W325.
- Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., Suh, A., Dutoit, L., Bureš, S., Garamszegi, L. Z., Hogner, S., Moreno, J., Qvarnström, A., Ružić, M., Sæther, S. A., Sætre, G. P., Török, J., & Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Research*, 25(11), 1656–1665.
- Cai, T., Wu, G., Sun, L., Zhang, Y., Peng, Z., Guo, Y., Liu, X., Pan, T., Chang, J., Sun, Z., & Zhang, B. (2021). Biogeography and diversification of Old World buntings (Aves: Emberizidae): Radiation in open habitats. *Journal of Avian Biology*, 52(6), e02672.
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890.
- Chevreaux, B., Wetter, T., & Suhai, S. (1999). Genome sequence assembly using trace signals and additional sequence information. *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)*, 99, 45–56.
- Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6(2), 80–92.
- Copete, J. L. (2020). Yellow-throated bunting (*Emberiza elegans*), version 1.0. In J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie, & E.

- de Juana (Eds.), *Birds of the world*. Cornell Lab of Ornithology and Lynx Editions.
- 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.
- Delmore, K., Illera, J. C., Perez-Tris, J., Segelbacher, G., Lugo Ramos, J. S., Durieux, G., Ishigohoka, J., & Liedvogel, M. (2020). The evolutionary history and genomics of European blackcap migration. *eLife*, 9, e54462.
- 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.
- Delmore, K. E., Toews, D. P., Germain, R. R., Owens, G. L., & Irwin, D. E. (2016). The genetics of seasonal migration and plumage color. *Current Biology*, 26(16), 2167–2173.
- Dingle, H. (2006). Animal migration: Is there a common migratory syndrome? *Journal of Ornithology*, 147, 212–220.
- Dufour, P., de Franceschi, C., Doniol-Valcroze, P., Jiguet, F., Guéguen, M., Renaud, J., Lavergne, S., & Crochet, P. A. (2021). A new westward migration route in an Asian passerine bird. *Current Biology*, 31(24), 5590–5596.
- Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A., Uebbing, S., & Wolf, J. B. W. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, 491(7426), 756–760.
- Fudickar, A. M., Jahn, A. E., & Ketterson, E. D. (2021). Animal migration: An overview of one of Nature's great spectacles. *Annual Review of Ecology, Evolution, and Systematics*, 52(1), 479–497.
- Gomez-Bahamon, V., Marquez, R., Jahn, A. E., Miyaki, C. Y., Tuero, D. T., Laverde, R. O., Restrepo, S., & Cadena, C. D. (2020). Speciation associated with shifts in migratory behavior in an avian radiation. *Current Biology*, 30(7), 1312–1321.
- Gu, Z., Pan, S., Lin, Z., Hu, L., Dai, X., Chang, J., Xue, Y., Su, H., Long, J., Sun, M., Ganusevich, S., Sokolov, V., Sokolov, A., Pokrovsky, I., Ji, F., Bruford, M. W., Dixon, A., & Zhan, X. (2021). Climate-driven flyway changes and memory-based long-distance migration. *Nature*, 591(7849), 259–264.
- Gwinner, E. (1996). Circannual clocks in avian reproduction and migration. *Ibis*, 138(1), 47–63.
- Hahn, C., Bachmann, L., & Chevreaux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—A baiting and iterative mapping approach. *Nucleic Acids Research*, 41(13), e129.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15), 1965–1978.
- Irwin, D. E., Mila, B., Toews, D. P. L., Brelsford, A., Kenyon, H. L., Porter, A. N., Grossen, C., Delmore, K. E., Alcaide, M., & Irwin, J. H. (2018). A comparison of genomic islands of differentiation across three young avian species pairs. *Molecular Ecology*, 27(23), 4839–4855.
- Kipp, F. A. (1959). Der Handflügel-index als flugbiologisches Maß. *Die Vogelwarte*, 20, 77086.
- Kondo, B., Peters, J. L., Rosensteel, B. B., & Omland, K. E. (2008). Coalescent analyses of multiple loci support a new route to speciation in birds. *Evolution*, 62(5), 1182–1191.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760.
- Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493–496.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079.
- 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.
- Madeira, F., Park, Y., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., & Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, 47(W1), W636–W641.
- Nakamura, T., Yamada, K. D., Tomii, K., & Katoh, K. (2018). Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics*, 34(14), 2490–2492.
- Newton, I. (2007). *The migration ecology of birds* (1st ed.). Elsevier.
- Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274.
- Päckert, M., Sun, Y.-H., Strutzenberger, P., Valchuk, O., Tietze, D. T., & Martens, J. (2015). Phylogenetic relationships of endemic bunting species (Aves, Passeriformes, Emberizidae, *Emberiza koslowi*) from the eastern Qinghai-Tibet Plateau. *Vertebrate Zoology*, 65(1), 135–150.
- Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3), 231–259.
- Pulido, F. (2007). The genetics and evolution of avian migration. *Bioscience*, 57(2), 165–174.
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842.
- R Development Core Team. (2008). *R: A language and environment for statistical computing*. Vienna, Austria.
- Ruegg, K., Anderson, E. C., Boone, J., Pouls, J., & Smith, T. B. (2014). A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology*, 23(19), 4757–4769.
- Sheard, C., Neate-Clegg, M. H. C., Alioravainen, N., Jones, S. E. I., Vincent, C., MacGregor, H. E. A., Bregman, T. P., Claramunt, S., & Tobias, J. A. (2020). Ecological drivers of global gradients in avian dispersal inferred from wing morphology. *Nature Communications*, 11(1), 2463.
- Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212.
- Skotte, L., Korneliusson, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, 195(3), 693–702.
- Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, 56(4), 564–577.
- Toews, D. P. L., Taylor, S. A., Streby, H. M., Kramer, G. R., & Lockett, I. J. (2019). Selection on VPS13A linked to migration in a songbird. *Proceedings of the National Academy of Sciences of the United States of America*, 116(37), 18272–18274.
- Turbek, S. P., Scordato, E. S. C., & Safran, R. J. (2018). The role of seasonal migration in population divergence and reproductive isolation. *Trends in Ecology & Evolution*, 33(3), 164–175.
- Vilella, A. J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., & Birney, E. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Research*, 19(2), 327–335.
- Voelker, G., Bowie, R. C. K., & Klicka, J. (2013). Gene trees, species trees and earth history combine to shed light on the evolution of migration in a model avian system. *Molecular Ecology*, 22(12), 3333–3344.
- Voelker, G., & Light, J. E. (2011). Palaeoclimatic events, dispersal and migratory losses along the Afro-European axis as drivers of

- biogeographic distribution in *Sylvia* warblers. *BMC Evolutionary Biology*, 11(1), 163.
- Winger, B. M., Lovette, I. J., & Winkler, D. W. (2012). Ancestry and evolution of seasonal migration in the Parulidae. *Proceedings of the Royal Society B: Biological Sciences*, 279(1728), 610–618.
- Winkler, D. W., Billerman, S. M., & Lovette, I. J. (2020). Old World buntings (Emberizidae), version 1.0. In S. M. Billerman, B. K. Keeney, P. G. Rodewald, & T. S. Schulenberg (Eds.), *Birds of the world*. Cornell Lab of Ornithology.
- Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, 88(1), 76–82.
- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8), 1586–1591.
- Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., Storz, J. F., Antunes, A., Greenwold, M. J., Meredith, R. W., Ödeen, A., Cui, J., Zhou, Q., Xu, L., Pan, H., Wang, Z., Jin, L., Zhang, P., Hu, H., ... Wang, J. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, 346(6215), 1311–1320.
- Zhang, W., Collins, A., Gibson, J., Tapper, W. J., Hunt, S., Deloukas, P., Bentley, D. R., & Morton Newton, E. (2004). Impact of population structure, effective bottleneck time, and allele frequency on linkage disequilibrium maps. *Proceedings of the National Academy of Sciences of the United States of America*, 101(52), 18075–18080.
- Zink, R. M. (2011). The evolution of avian migration. *Biological Journal of the Linnean Society*, 104(2), 237–250.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Zhang, D., She, H., Rheindt, F. E., Wu, L., Wang, H., Zhang, K., Cheng, Y., Song, G., Jia, C., Qu, Y., Olsson, U., Alström, P., & Lei, F. (2023). Genomic and phenotypic changes associated with alterations of migratory behaviour in a songbird. *Molecular Ecology*, 32, 381–392. <https://doi.org/10.1111/mec.16763>