Spotlight Article

Syst. Biol. XX(XX):XX–XX, 2024

https://doi.org/10.1093/sysbio/syad078
Advance Access Publication December 29, 2023

Phylogenetic Conflict Between Species Tree and Maternally Inherited Gene Trees in a Clade of *Emberiza* Buntings (Aves: Emberizidae)

Dezhi Zhang[1], Huishang She[1], Shangyu Wang[1,2], Haitao Wang[3], Shi Li[4], Yalin Cheng[1], Gang Song[1], Chenxi Jia[1], Yanhua Qu[1,*], Frank E. Rheindt[5,*], Urban Olsson[6,7], Per Alström[1,8,*], and Fumin Lei[1,2,*]

[1]*Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, 100101 Beijing, China*
[2]*College of Life Sciences, University of Chinese Academy of Sciences, 100049 Beijing, China*
[3]*School of Life Sciences, Jilin Key Laboratory of Animal Resource Conservation and Utilization, Northeast Normal University, Changchun 130024, China*
[4]*College of Animal Science and Technology, Jilin Agricultural University, Changchun 130118, China*
[5]*Department of Biological Sciences, National University of Singapore, Singapore 117543, Republic of Singapore*
[6]*Department of Biology and Environmental Science, University of Gothenburg, Box 463, SE-405 30 Gothenburg, Sweden*
[7]*Gothenburg Global Biodiversity Centre, Box 461, SE-405 30 Gothenburg, Sweden*
[8]*Animal Ecology, Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18 D, SE-752 36 Uppsala, Sweden*
[*]*Correspondence to be sent to: Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, 100101 Beijing, China; E-mail: leifm@ioz.ac.cn.*

Received 4 July 2022; reviews returned 6 December 2023; accepted 27 December 2023
Associate Editor: Bastien Boussau

Abstract.—Different genomic regions may reflect conflicting phylogenetic topologies primarily due to incomplete lineage sorting and/or gene flow. Genomic data are necessary to reconstruct the true species tree and explore potential causes of phylogenetic conflict. Here, we investigate the phylogenetic relationships of 4 *Emberiza* species (Aves: Emberizidae) and discuss the potential causes of the observed mitochondrial non-monophyly of *Emberiza godlewskii* (Godlewski's bunting) using phylogenomic analyses based on whole genome resequencing data from 41 birds. Analyses based on both the whole mitochondrial genome and ~39 kilobases from the non-recombining W chromosome reveal sister relationships between each the northern and southern populations of *E. godlewskii* with *E. cioides* and *E. cia*, respectively. In contrast, the monophyly of *E. godlewskii* is reflected by the phylogenetic signal of autosomal and Z chromosomal sequence data as well as demographic inference analyses, which—in combination—support the following tree topology: ([{*E. godlewskii*, *E. cia*}, *E. cioides*], *E. jankowskii*). Using D-statistics, we detected multiple gene flow events among different lineages, indicating pervasive introgressive hybridization within this clade. Introgression from an unsampled lineage that is sister to *E. cioides* or introgression from an unsampled mitochondrial + W chromosomal lineage of *E. cioides* into northern *E. godlewskii* may explain the phylogenetic conflict between the species tree estimated from genome-wide data versus mtDNA/W tree topologies. These results underscore the importance of using genomic data for phylogenetic reconstruction and species delimitation. [*Emberiza* buntings; gene flow; mitochondrion; phylogenetic conflict; W chromosome.]

There is a wide appreciation in the biological community for the two major non-exclusive causes of phylogenetic conflict among orthologous gene trees (hereafter gene trees is used) or between gene trees and species trees, namely incomplete lineage sorting (ILS) and interspecific gene flow (Pamilo and Nei 1988; Maddison 1997; Nichols 2001; Avise and Robinson 2008; Degnan and Rosenberg 2009). ILS is positively correlated with ancestral effective population size and negatively correlated with internode divergence time (Avise and Robinson 2008; Hibbins and Hahn 2022). Interspecific hybridization is pervasive across the tree of life and may lead to genetic introgression from one species into the other (Mallet 2005; Rheindt and Edwards 2011; Ottenburghs et al. 2015; Payseur and Rieseberg 2016). Early phylogenetic reconstructions typically used a small number of genes due to the limitations of sequencing technology (Felsenstein 2004). Phylogenetic relationships based on few genes need to be treated with caution due to the potential sampling bias caused by ILS and gene flow. Fortunately, the use of whole genome data can overcome this sampling bias effectively, except in rare cases where gene flow between non-sister lineages occurs over a large part of the genome (e.g., Fontaine et al. 2015; Small et al. 2020; Zhang et al. 2021).

Emberiza godlewskii (Godlewski's bunting) is a passerine bird species widespread in temperate East Asia (Fig. 1a). Two subspecies groups are often recognized, a northern one comprising the subspecies *godlewskii*, *decolorata*, and *omissa* (Central Asia, southern Russia, Mongolia and northern China) and a southern one comprising the subspecies *khamensis* and *yunnanensis* (south-central China) (Copete 2020). Three independent phylogenetic and population-genetic analyses, each based on 3 mitochondrial and two nuclear loci, have suggested that northern populations of *E. godlewskii* are sister to *E. cioides* (meadow bunting), while the southern group is sister to *E. cia* (rock bunting), and that the divergence time between northern *E. godlewskii* and *E. cioides* was slightly deeper than that between

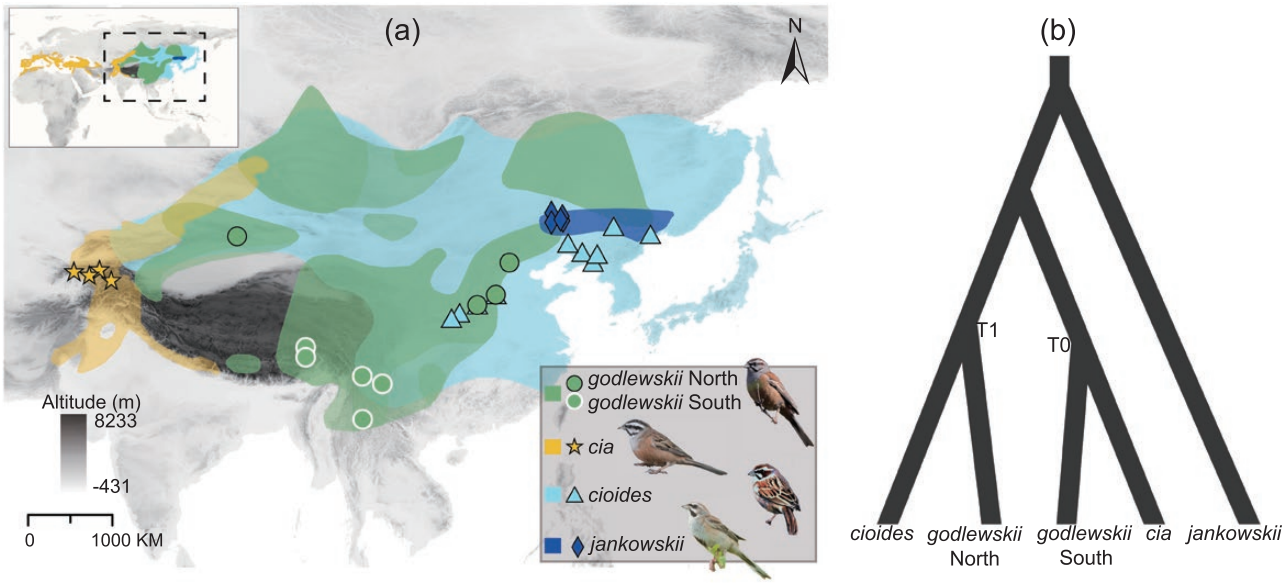

FIGURE 1. Distributions and phylogenetic relationships suggested by Päckert et al. (2015) for the 4 *Emberiza* species under study. (a) Distribution maps and DNA sampling sites; inset (top left) shows the global distribution of these 4 species, with the dashed box representing the area covered in the present study; distribution maps are based on BirdLife (https://www.birdlife.org/) and Birds of the World (https://birdsoftheworld.org). (b) Phylogenetic relationships on the basis of 3 mitochondrial and 2 nuclear loci (from Päckert et al. 2015); the divergence time between *E. cioides* and northern *E. godlewskii* (T1) is slightly deeper than that between *E. cia* and southern *E. godlewskii* (T0); this tree may primarily represent the mitochondrial phylogenetic signal because of mtDNA's faster lineage sorting; "*godlewskii* North" and "*godlewskii* South" represent the northern and southern *E. godlewskii* populations, respectively.

southern *E. godlewskii* and *E. cia* (Fig. 1b), with an original divergence between northern and southern populations at roughly 3 million years (Päckert et al. 2015; Li et al. 2019, 2023). These findings have prompted taxonomic proposals to divide *E. godlewskii* into a northern and a southern species (Li et al. 2023). However, such a treatment would only be justified as far as the non-monophyletic placement of populations is not a phylogenetic artifact.

The alleged non-monophyly of *E. godlewskii* is biologically surprising in view of the close plumage similarity between the two subspecies groups and the more divergent appearances of the two other *Emberiza* species (Copete 2020). As these results were only based on 3 mitochondrial DNA (mtDNA) and two nuclear loci, the tree is likely to be biased towards the much stronger phylogenetic signal from the mtDNA loci because of the much faster rates of lineage sorting of mtDNA in comparison to nuclear DNA (Toews and Brelsford 2012). As mtDNA is known to introgress readily (Rheindt and Edwards 2011; Toews and Brelsford 2012), the results may not conform to the evolutionary history of these taxa. Analysis of a larger nuclear dataset would be needed to evaluate if the mitochondrial phylogeny is supported by the entirety of molecular evidence.

Using 40 resequenced genomes of *E. godlewskii*, *E. cia*, *E. cioides* and *E. jankowskii* and one resequenced genome from the outgroup species *Emberiza stewarti* (white-capped bunting), we present support for the monophyly of *E. godlewskii* and find *E. godlewskii* to be sister to *E. cia*, and these two species sister to *E. cioides*, while *E. jankowskii* is basal to all of them. However, northern and southern *E. godlewskii* emerged as sister to *E. cioides* and *E. cia*, respectively, for the maternally inherited mitochondrion and W chromosome. On the basis of our genomic datasets, we infer multiple gene flow events among different lineages and discuss the potential causes of this observed phylogenetic conflict: introgression from an unsampled lineage that is sister to *E. cioides* or introgression from an unsampled mitochondrial + W chromosomal lineage of *E. cioides* into northern *E. godlewskii*. These results underscore the importance of using genomic data for phylogenetic reconstruction and species delimitation.

MATERIALS AND METHODS

Sampling, Resequencing, and Variant Calling

The monophyly of the clade comprising *E. godlewskii*, *E. cia*, *E. cioides*, and their closest relative, *E. jankowskii*, has been strongly supported by previous studies (Fig. 1b) (Alström et al. 2008; Päckert et al. 2015; Cai et al. 2021). We sampled a total of 41 individuals, including 9 *E. godlewskii*, 11 *E. cia*, 10 *E. cioides*, 10 *E. jankowskii*, and 1 female *E. stewarti* (Fig. 2; Supplementary Table S1). The target sequence data was 15 Gb for each sample. The raw sequence output was quality-controlled to remove adapter sequences, low-quality reads, and poly-N reads in fastp 0.20.0 (Chen et al. 2018). Quality-controlled reads of all 41 samples were mapped to the reference

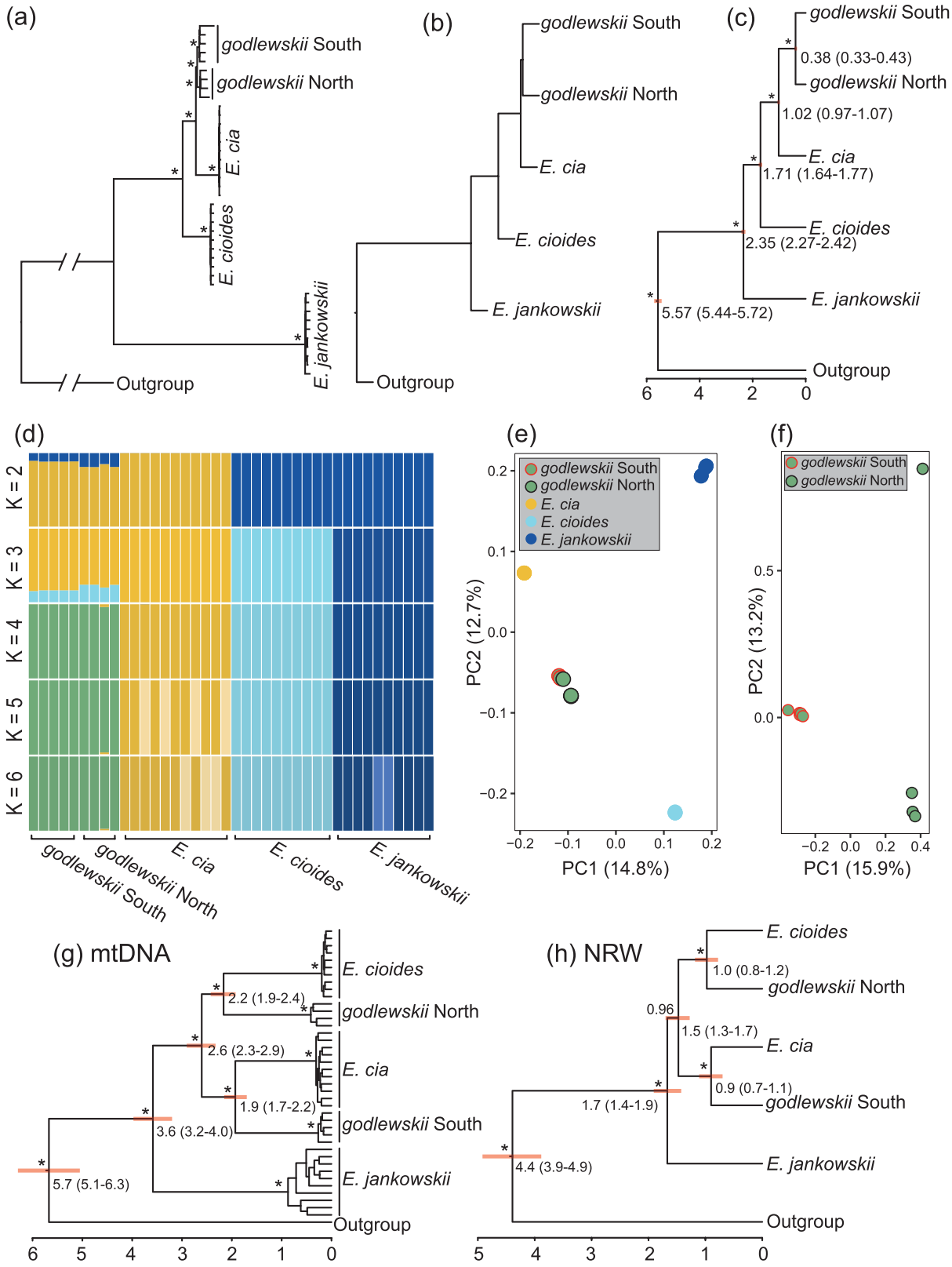

FIGURE 2. Phylogenetic and population structure analyses using genome-wide nuclear data. (a) Maximum likelihood tree estimated in IQ-TREE using 2,654,187 concatenated autosomal SNVs; the tree was rooted with the outgroup (*E. stewarti*) in FigTree; asterisks represent 100% bootstrap support. (b) Species tree estimated in MP-EST using 20,000 randomly selected 20 Kb sliding-window autosomal gene trees. (c) Species tree and divergence time estimations in StarBEAST2 using 100 randomly selected 20 Kb windows across the autosomes; divergence times with 95% highest posterior densities (in brackets) are given below the nodes and indicated by red horizontal bars; time scale in millions

genome of a congeneric species *E. elegans* (GenBank genome assembly accession: GCA_022818055.1; Zhang et al. 2023) using BWA 0.7.12 (Li and Durbin 2009). The sequencing depth for each sample was estimated using BEDtools 2.27 (Quinlan and Hall 2010). Variants were called in SAMtools 0.1.19 (Li et al. 2009). Single-nucleotide variants (SNVs) were filtered using VCFtools 0.1.12b (Danecek et al. 2011). As the sample of the outgroup species *E. stewarti* was a female with heterozygous sex chromosomes (ZW) (Supplementary Table S1), we only called autosomal SNVs in this study. A total of 20,994,339 autosomal SNVs were retained. Information on specific sampling methods, DNA extraction, library construction and sequencing, sequence read processing, variant calling, and filtering can be found in the Supplementary Materials and Methods.

Phylogenetic Analyses Using Genome-wide Data

The concatenated autosomal SNVs were used to infer a maximum likelihood (ML) tree in IQ-TREE 2.1.3 (Nguyen et al. 2015) with 1,000 bootstraps (-bb 1000) and *E. stewarti* as the outgroup. We applied the GTR+ASC substitution model suggested by the IQ-TREE manual as the concatenated alignment contained only variants. Heterozygous sites were removed from the alignment because the ascertainment bias correction (ASC) model treats heterozygous sites as invariants. Out of 20,994,339 SNVs, a total of 2,654,187 concatenated homozygous variants were used in IQ-TREE analysis (Supplementary Table S2).

In addition to the phylogenetic analyses based on the concatenated dataset, we also conducted species tree analyses using aligned consensus sequences in MP-EST 2.1 (Liu et al. 2010). MP-EST is a maximum pseudo-likelihood approach for estimating species trees under the multispecies coalescent model using a set of rooted gene trees by maximizing a pseudo-likelihood function. Using the BAM files with PCR duplicates removed generated by the previous mapping step in BWA, we employed SAMtools 0.1.19 (Li et al. 2009) and bcftools 1.3.1 (Li et al. 2009) to generate consensus sequences of 20 kilobases (Kb) sliding windows for autosomes. The minimum sequencing depth was set to 8 (vcfutils.pl vcf2fq -d 8). These consensus sequences were aligned in mafft 7.464 (Nakamura et al. 2018) with the command "--adjustdirection" and trimmed in Gblocks 0.91b (Talavera and Castresana 2007) with default parameters. Only aligned consensus sequences with a length ≥ 10 Kb were used in the downstream analyses, and a total of 47,547 autosomal consensus alignments of 20 Kb windows were retained. All the 47,547 sequences were used to reconstruct ML trees in RAxML 8.1 (Stamatakis 2014) (Supplementary Table S2). We set the substitution model as GTRGAMMA and designated *E. stewarti* as the outgroup. In consideration of computational limits, bootstrap numbers were set to only 10 with the command "-x 1234 -# 10." We randomly selected 20,000 gene trees as the input for MP-EST analysis with 10 independent runs (Supplementary Table S2).

The haploidy of the Z chromosome in female birds led us to exclude chromosome Z from the SNV-related analyses. However, the haploidy of the Z chromosome does not affect the generation of consensus sequences. Using the same criteria as described above, a total of 3583 consensus alignments of 20 Kb windows on the Z chromosome were retained. All 3583 alignments were used to reconstruct ML trees in RAxML-NG 1.2.0 (Kozlov et al. 2019) (Supplementary Table S2). We set GTR+G as the substitution model and designated *E. stewarti* as the outgroup. These 3583 gene trees were used to reconstruct the species tree for the Z chromosome in ASTRAL 5.7.8 (Zhang et al. 2018) with default parameters (Supplementary Table S2).

We estimated lineage divergence times in StarBEAST2 (Ogilvie et al. 2017) using 100 randomly selected 20 Kb sliding windows among the 47,547 autosomal consensus alignments (Supplementary Table S2). StarBEAST2 is a method in the BEAST package that infers a species tree under the multispecies coalescent (Ogilvie et al. 2017). We utilized a GTR model and applied a clock rate of 0.33% per My for all windows (Zhang et al. 2014). A total of 50 million iterations were carried out while sampling every 5000 generations. We used a birth-death prior and kept the other priors at default. The consensus trees were generated using TreeAnnotator 1.8 with a 10% burn-in and were visualized in FigTree 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/). Note that the estimated divergence time using StarBEAST2 may be biased to some extent due to the existence of gene flow among some lineages (see Results). Additionally, based on the IQ-TREE analysis (see Results), the longer branch lengths for *E. jankowskii* may suggest a higher accumulation of genetic variation compared to other lineages. Consequently, employing a strict clock for estimating divergence times in BEAST-related analyses may have introduced some bias in this analysis.

Among the 47,547 autosomal sliding window trees reconstructed above, we selected and classified trees in which all 5 clades (northern *E. godlewskii*, southern

of years; asterisks represent a posterior probability of 1.00. (d) Population structure estimated in FRAPPE based on 5,407,601 linkage-pruned autosomal SNVs at K = 2, 3, 4, 5 and 6. (e) Principal component analysis for the 4 species based on the same SNV data used in FRAPPE analysis. (f) Principal component analysis for *E. godlewskii* only. (g) Tree and divergence time estimates in BEAST based on ~16 Kb of mtDNA sequences. (h) Tree and divergence time estimates in BEAST based on ~39 Kb of concatenated non-recombining W chromosome (NRW) sequences; only one randomly selected female from each clade was used in the NRW phylogenetic analysis. For (g) and (h), divergence times with 95% highest posterior densities (in brackets) are given below the nodes and indicated by red horizontal bars; time scale in millions of years; posterior probabilities are shown above the nodes, with asterisks representing a posterior probability of 1.00. The labels *godlewskii* North and *godlewskii* South represent the northern and southern *E. godlewskii* populations, respectively.

E. godlewskii, *E. cia*, *E. cioides*, and *E. jankowskii*) were monophyletic (monophyletic tree), using the R package MonoPhy (Schwery and O'Meara 2016).

Gene trees in which all 5 clades were monophyletic accounted for only ~4% of the autosomes (see Results). Therefore, we selected one male individual with the highest sequencing depth from each clade to explore the distribution of phylogenetic topologies across both the autosomes and the Z chromosome. The female individual of *E. stewarti* was used as the outgroup species. Using the same criteria as described above, we obtained 47,562 consensus alignments of 20 Kb windows for autosomes and 3732 consensus alignments for the Z chromosome. All these consensus sequences were used to reconstruct ML trees in RAxML-NG 1.2.0 (Kozlov et al. 2019) (Supplementary Table S2).

Testing for the Most Likely Phylogeny Using fastsimcoal2

The phylogeny estimated from genome-wide data may not always accurately represent the true species tree (Fontaine et al. 2015; Small et al. 2020; Zhang et al. 2021). The sliding window tree analyses for both the autosomes and Z chromosome jointly revealed 3 antagonistic tree topologies with the highest proportions (see Results). To determine which of the 3 gene trees represents the most likely species tree, we conducted coalescent-based demographic inference in fastsimcoal2 2.6.0.3 (Excoffier et al. 2013). SNVs with high linkage disequilibrium (LD) were pruned in Plink 1.9 (www.cog-genomics.org/plink/1.9/; Chang et al. 2015) with "--indep-pairwise 50 10 0.2." A total of 5,407,601 variants were retained. We generated the joint site frequency spectrum (SFS) using these LD-pruned SNVs in the Python script easySFS.py (https://github.com/isaacovercast/easySFS) (Supplementary Table S2). A total of 12 models were tested (Supplementary Fig. S1). Details of model and parameter settings can be found in the Supplementary Materials and Methods.

Population Structure Analyses

On the basis of the 5,407,601 LD pruned SNVs, we used FRAPPE 1.1 (Tang et al. 2005) to analyze population structure for the 4 ingroup species with a K value setting ranging from 2 to 6 (Supplementary Table S2). The step number and maximum iterations were set to 200 and 10,000, respectively. Principal component analysis (PCA) was performed in Plink using the same dataset with "--pca" (Supplementary Table S2). We also conducted a PCA for *E. godlewskii* only.

Phylogenetic Analysis of Mitochondrial DNA and W Chromosomal Sequences

We assembled the mitochondrial genomes for all 41 samples in MITObim 1.8 (Hahn et al. 2013) using default parameters. We conducted phylogenetic inference and divergence time estimation in BEAST 2.4.5 (Bouckaert et al. 2014) based on the whole mtDNA sequences. Details on the mtDNA phylogenetic analysis can be found in the Supplementary Materials and Methods.

The maternally inherited W chromosome and mtDNA are known to be co-segregated in birds (Smeds et al. 2015). Therefore, they may show the same phylogenetic topology. To test this, we conducted a phylogenetic analysis of DNA from the non-recombining W chromosome (NRW). As only one female was available for the northern *E. godlewskii* population and for the outgroup species (Supplementary Table S1), in order to obtain longer homologous NRW sequences, we selected the female with the highest sequencing depth from each clade for this analysis. The female-specific NRW typically has a nearly zero coverage when mapping the male sequences to the female genome assembly (Chen et al. 2012). We first assembled the scaffold-level genome sequences using SOAPdenovo 2.04 (Luo et al. 2012) for female individuals only. We then mapped male reads to the female genome assemblies to detect potential NRW sequences. We conducted phylogenetic inference and divergence time estimation in BEAST 2.4.5 (Bouckaert et al. 2014) based on 38,994 base pair (bp) NRW sequences. Details on NRW detection and phylogenetic analysis can be found in the Supplementary Materials and Methods.

Gene Flow Test Using D-statistic

On the basis of the phylogenetic topology estimated from genome-wide data ((((southern *E. godlewskii*, northern *E. godlewskii*), *E. cia*), *E. cioides*), *E. jankowskii*) (Fig. 2), we conducted D-statistic analysis in Dsuite (Malinsky et al. 2020) using the "Dtrios" function across all possible combinations on the basis of autosomal SNVs (n = 20,994,339), with *E. stewarti* as the outgroup, and 20 Jackknife blocks for P-value estimations (Supplementary Table S2). D-statistics is also known as the ABBA-BABA test (Green et al. 2010). Under a 4-species phylogenetic framework ([{P1, P2}, P3], outgroup), a deviation from zero for a D-statistic value indicates significant gene flow between P1 and P3 or P2 and P3. However, D-statistics is unable to detect gene flow between P1 and P2, and it would fail to detect a minor gene flow event if gene flow occurred between P3 and P1 and between P3 and P2 at unequal ratios.

Demographic Inference Based on the Phylogenetic Topology Estimated from Genome-wide Data

Multiple gene flow events were revealed by D-statistics: gene flow between *E. jankowskii* and both northern and southern *E. godlewskii*; gene flow between *E. cioides* and both northern and southern *E. godlewskii*; gene flow between *E. cia* and southern *E. godlewskii*; gene flow between *E. cioides* and *E. jankowskii* (see Fig. 3 and Supplementary Table S3 in Results). Ancestral gene flow between *E. jankowskii* and the ancestral population of northern and southern *E. godlewskii*, as well as between *E. cioides* and the ancestral population of *E. godlewskii*, would also lead

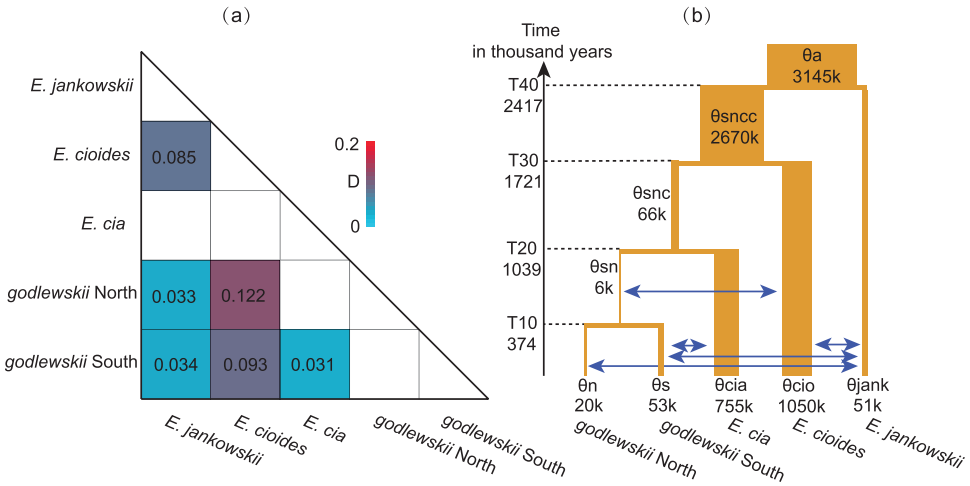

FIGURE 3. Dsuite and fastsimcoal2 analyses. (a) Heatmap of D-statistic values in Dsuite; values in the colored boxes represent average D-statistic values (see details in Supplementary Table S3); the colors of the D-statistic values are shown on the scale bar: D-statistic values increase from blue to red; a D-statistic value greater than zero indicates significant gene flow (*P*-values = 0). (b) The best-fit model (model c in Supplementary Fig. S2) among the 6 tested models in fastsimcoal2 analysis; estimated parameters included gene flow events (blue arrows), population size (θ) and divergence times (T); bidirectional arrows indicate bidirectional gene flow; θs, θn, θcia, θcio, and θjank represent population sizes of southern and northern *E. godlewskii*, *E. cia*, *E. cioides*, and *E. jankowskii*, respectively; θsn, θsnc, θsncc, and θa represent the ancestral population sizes before lineage splits; T10, T20, T30, and T40 represent divergence times between southern and northern *E. godlewskii*, between *E. godlewskii* and *E. cia*, between *E. cioides* and the former two species, and the root divergence time, respectively; values below each parameter were estimated from the best run with the highest likelihood values of model c.

to the detection of certain gene flow events between terminal lineages using D-statistics. To disentangle these confounding demographic models, we conducted demographic inference in fastsimcoal2 using the same SFS data generated above (Supplementary Table S2). Based on the phylogenetic topology estimated from genome-wide data (Fig. 2), we tested six competing models. Model (a) included all gene flow events between the terminal lineages as revealed by D-statistics (Supplementary Fig. S2a). Model (b) (Supplementary Fig. S2b) assumed gene flow between *E. jankowskii* and the ancestral lineage of *E. godlewskii* in addition to the other gene flow events between the terminal lineages as revealed by D-statistics. Model (c) (Supplementary Fig. S2c) assumed gene flow between *E. cioides* and the ancestral lineage of *E. godlewskii* in addition to the other gene flow events between the terminal lineages as revealed by D-statistics. As the D-statistics analysis had indicated the possibility of a higher level of gene flow between *E. cioides* and northern *E. godlewskii* as compared to that between *E. cioides* and southern *E. godlewskii* (Supplementary Table S3), model (d) (Supplementary Fig. S2d) incorporated an additional gene flow event between *E. cioides* and northern *E. godlewskii* in contrast to model (c). Model (e) (Supplementary Fig. S2e) assumed ancestral gene flow between *E. jankowskii* and the ancestral lineage of *E. godlewskii*, as well as ancestral gene flow between *E. cioides* and the ancestral lineage of *E. godlewskii*, in addition to the other gene flow events between the terminal lineages as revealed by D-statistics. Model (f) (Supplementary Fig. S2f) assumed no gene flow among any terminal and ancestral lineages.

We ran fastsimcoal2 with 300,000 simulations (–n = 300,000) and 20 cycles (–L = 20) and performed 20 independent runs for each model to obtain the best run with the highest likelihood values. We ran each model with the best parameter values from the best run 100 times to achieve the likelihood distributions for each model. The likelihood values will slightly differ among these 100 repeats because fastsimcoal2 approximates likelihood values with simulations. As the number of parameters differed among these models, we compared values of the Akaike information criterion (AIC) of these 100 runs of each model to determine the best-fit model. Additional parameter settings are explained in the Supplementary Materials and Methods.

RESULTS

Phylogeny and Population Structure Based on Genome-wide Nuclear Data

The ML tree using 2,654,187 concatenated autosomal SNVs estimated in IQ-TREE reconstructed *E. godlewskii* as monophyletic, with the southern and northern populations forming separate clades; *E. godlewskii* was sister to *E. cia*, with *E. cioides* sister to these, and *E. jankowskii* basal to these 3 species (Fig. 2a). Using 20,000 randomly selected 20 Kb sliding-window autosomal gene trees, the species tree analysis in MP-EST was consistent with the IQ-TREE analysis (Fig. 2b). StarBEAST2 analysis using 100 randomly selected consensus sequences of 20 Kb windows of autosomes uncovered the same tree topology as IQ-TREE and MP-EST (Fig. 2c). The Z chromosomal species tree estimated in ASTRAL, using

3583 gene trees, exhibited concordance with the aforementioned autosomal trees (Supplementary Fig. S3). We estimated a divergence time of 0.38 My (95% height posterior density (HPD): 0.33–0.43 My) between northern and southern *E. godlewskii*, 1.02 My (95% HPD: 0.97–1.07 My) between *E. cia* and *E. godlewskii*, 1.71 My (95% HPD: 1.64–1.77 My) between *E. cioides* and the former two species, and 2.35 My (95% HPD: 2.27–2.42 My) between *E. jankowskii* and the former 3 species (Fig. 2c).

Based on the sliding window phylogenetic analysis using all individuals, the 5 groups emerged as monophyletic in a total of 1881 out of 47,547 20 Kb autosomal window trees (~4%). Ten main tree topologies accounted for 94.7% of these 1881 trees (Supplementary Fig. S4). As expected, the tree in agreement with the whole genome phylogeny (Fig. 2) was the most numerous ($n = 764$; tree 1 in Supplementary Fig. S4). The tree consistent with the mitochondrial tree (as shown in Fig. 1b) ranked fifth in percentage ($n = 64$; tree 5 in Supplementary Fig. S4).

Based on the sliding window phylogenetic analysis using one single individual from each clade, we explored the distribution of topologies across the autosomes and Z chromosome. The most numerous gene tree on both the autosomes and Z chromosome, consistent with the whole genome phylogeny, accounted for 27.6% of autosomes and 48.0% of Z chromosome (Supplementary Fig. S5). The fourth-ranked tree in agreement with the mitochondrial tree accounted for 5.6% on autosomes (tree 4 in Supplementary Fig. S5). However, this same mitochondrial-consistent tree accounted for only 0.4% on the Z chromosome.

One of the gene trees with the highest proportion on both the autosomes and Z chromosome should represent the true species tree. To determine the most likely species tree, we tested 12 simplified models in fastsimcoal2 (Supplementary Fig. S1). The results showed that two models derived from tree 1, model 1-1 and model 1-2, generally had the highest likelihood and lowest AIC values (Supplementary Fig. S1; Supplementary Table S4), suggesting that tree 1 (concordance with the phylogenetic topology estimated from genome-wide data) should represent the most likely species tree.

Population structure analysis in FRAPPE using 5,407,601 LD pruned autosomal SNVs found *E. godlewskii* and *E. cia* largely forming one group and *E. cioides* and *E. jankowskii* forming another group at $K = 2$ (Fig. 2d). At $K = 3$, FRAPPE placed *E. cioides* and *E. jankowskii* in different groups, and inferred a proportion of *E. cioides* ancestry (19.1%) in *E. godlewskii* (Fig. 2d). At $K = 4$, FRAPPE divided the samples into 4 groups, adding *E. godlewskii* as a separate group, and with one sample from the northern *E. godlewskii* population showing evidence of minor shared genetic variation with *E. cia* (3.8%) (Fig. 2d). At $K = 5$ and 6, FRAPPE further revealed substructure of *E. cia* and *E. jankowskii* (Fig. 2d).

Similarly, PCA using the same dataset revealed 4 clusters corresponding to each species (Fig. 2e). A separate PCA for *E. godlewskii* only revealed a clear division between northern and southern *E. godlewskii* along PC1, while the individual of northern *E. godlewskii* that showed evidence of shared genetic variation with *E. cia* in the FRAPPE analysis clustered separately from the remaining northern *E. godlewskii* samples along PC2 (Fig. 2f).

Phylogeny and Divergence Times of mtDNA and NRW Sequences

The phylogenetic analysis in BEAST using ~16.3 Kb of mitochondrial sequences recovered northern *E. godlewskii* as sister to *E. cioides* with an estimated divergence time at 2.2 My (95% highest posterior density [HPD]: 1.9–2.4 My), and southern *E. godlewskii* as sister to *E. cia* with an estimated divergence time at 1.9 My (95% HPD: 1.7–2.2 My) (Fig. 2g). The divergence time of these 3 species was estimated at 2.6 My (95% HPD: 2.3–2.9 My), and the divergence time between these 3 species and *E. jankowskii* was estimated at 3.6 My (95% HPD: 3.2–4.0 My) (Fig. 2g). The outgroup species *E. stewarti* diverged from the ingroup species at 5.7 My (95% HPD: 5.1–6.3 My) (Fig. 2g).

The phylogenetic tree using ~39 Kb of concatenated NRW sequences was topologically identical to the mtDNA tree (Fig. 2h). Although the absolute divergence times based on NRW sequences were approximately half those of mtDNA, the divergence time between *E. cioides* and northern *E. godlewskii* (1.0 My; 95% HPD: 0.8–1.2 My) was slightly deeper than between *E. cia* and southern *E. godlewskii* (0.9 My; 95% HPD: 0.7–1.1 My) (Fig. 2h).

Dsuite Analysis and Refined Demographic Inference in fastsimcoal2

Dsuite analysis revealed significant gene flow between *E. jankowskii* and *E. cioides* (average D-statistic = 0.085 with admixture proportion between 2.84% and 4.46%), between *E. jankowskii* and southern *E. godlewskii* (average D-statistic = 0.034 with admixture proportion of 1.11%), and between *E. jankowskii* and northern *E. godlewskii* (average D-statistic = 0.033 with admixture proportion between 0.56% and 1.66%) (Fig. 3a; Supplementary Table S3). In addition, Dsuite analysis revealed significant gene flow between *E. cioides* and southern *E. godlewskii* (average D-statistic = 0.093 with admixture proportion of 7.18%), between *E. cioides* and northern *E. godlewskii* (average D-statistic = 0.122 with admixture proportion between 6.75% and 13.44%), and between *E. cia* and southern *E. godlewskii* (average D-statistic = 0.031 with admixture proportion of 2.5%) (Fig. 3a; Supplementary Table S3). Note that, due to the inherent drawback of D-statistics as described above, other undescribed gene flow events are possible. In addition, significant D-statistics values do not necessarily represent direct hybridization between certain lineages but may reflect indirect gene flow events.

Based on the most likely phylogenetic topology and the results of Dsuite analysis, a total of 6 models were

tested in fastsimcoal2 (Supplementary Fig. S2). Model (c), which assumed gene flow between *E. jankowskii* and both northern and southern *E. godlewskii*, gene flow between *E. cioides* and the ancestral population of *E. godlewskii*, gene flow between *E. cia* and southern *E. godlewskii*, as well as gene flow between *E. cioides* and *E. jankowskii*, had the highest likelihood and lowest AIC value (AIC = 273,828,954) among the 6 models (AIC of other models higher than 275×10^6) (Fig. 3b; Supplementary Table S5). Each model was compared on the basis of the best parameter values from the best run. The AIC value of model (c) was significantly lower than those of other models (all comparisons with $P < 0.001$ estimated by Wilcoxon rank sum test) (Supplementary Fig. S6). See Supplementary Table S6 for details on the estimated parameters of model (c).

Discussion

Using 3 mitochondrial and two nuclear markers, Päckert et al. (2015) recovered unexpected non-monophyly between southern and northern populations of *E. godlewskii*: southern *E. godlewskii* emerged as the sister of *E. cia* and the northern population as the sister of *E. cioides*, although the relationships were not fully supported by all loci. The non-monophyly of *E. godlewskii* was also supported by a recent study using a different set of 3 mitochondrial and two nuclear markers (Li et al. 2023). Our analysis of sequences representing the whole mitochondrion recovered the same sister relationship between southern *E. godlewskii* and *E. cia* with a divergence time of 1.9 My, and the sister relationship between northern *E. godlewskii* and *E. cioides* with a divergence time of 2.2 My (Fig. 2g). In addition, the phylogenetic analysis of the non-recombining W chromosome is topologically congruent with the mitochondrial tree (Fig. 2h), supporting co-segregation between the maternally inherited W chromosome and mtDNA in birds (Smeds et al. 2015). However, contrary to predictions of co-segregation, the divergence time estimates between mtDNA and NRW are markedly different, which might be caused by our use of a strict clock.

In stark contrast, phylogenetic analyses based on genome-wide data strongly supported southern and northern *E. godlewskii* forming a monophyletic clade (Fig. 2a–c), with an estimated divergence time between the two populations of only 0.38 My (Fig. 2c), as would be expected from traditional phenotype-based taxonomy. Furthermore, the monophyletic *E. godlewskii* emerged as sister to *E. cia*, and the clade containing these two species was sister to *E. cioides* (Fig. 2a–c). Population genomic analyses (Fig. 2d–e) based predominantly on autosomal SNVs additionally corroborated the sister relationship between northern and southern populations of *E. godlewskii*.

The analysis based on Z chromosomal data revealed a topology concordant with that of the autosomal dataset (Supplementary Fig. S3). Previous research has shown that the true species tree is disproportionately represented on the X chromosome due to reduced gene flow (Fontaine et al. 2015). The Z chromosome of birds is akin to the X chromosome in that it exhibits faster lineage sorting and lower gene flow when compared to autosomes (Irwin 2018). This characteristic may render it more suitable for accurately predicting the true species tree in the context of incomplete lineage sorting and gene flow. Although recent studies have suggested that the accuracy of phylogenetic topologies estimated from genome-wide data may be compromised by interspecific gene flow (Fontaine et al. 2015; Small et al. 2020; Zhang et al. 2021), our analyses combining autosomal and Z chromosomal data with fastsimcoal2 inference (Supplementary Fig. S1; Supplementary Table S4) invariably supports the most frequent tree topology (i.e., the topology estimated from genome-wide data) as the most likely species tree.

Our analyses indicate that the mito-nuclear discordance in our results can be explained by secondary gene flow and introgression among *Emberiza* lineages, in concordance with findings of mitochondrial introgression between *E. cioides* and *E. jankowskii* (Huang et al. 2022). We discovered multiple introgression events between different lineages (Fig. 3), indicating pervasive introgressive hybridization within this clade.

There are 4 mutually non-exclusive hypotheses that could explain the incongruence between the species tree estimated from genome-wide data and mtDNA/NRW tree topologies. First, the mtDNA and W chromosome of northern *E. godlewskii* may have been the product of genetic introgression from an unsampled lineage that is sister to *E. cioides* (Fig. 4a,b). Because our sampling does not cover the full range of *E. cioides*, we are unable to rule out the possibility that *E. cioides* comprises additional deeply diverged extant (hypothesis a) or extinct lineages (hypothesis b; Fig. 4). Second, the mtDNA and W chromosome of northern *E. godlewskii* may have arisen from introgression of an unsampled mitochondrial + W chromosomal lineage within *E. cioides* (Fig. 4c,d). ILS (deep coalescence) may theoretically lead to deep mtDNA and W chromosomal divergence within *E. cioides* (hypothesis c, Fig. 4c). Alternatively, deep divergence observed in maternally inherited markers may arise from female-specific selection (hypothesis d; Fig. 4d). For example, research on brood-parasitic birds has demonstrated deep divergence in mtDNA and the W chromosome within a single species due to differential host choice (Spottiswoode et al. 2011). These aforementioned hypotheses may also lead to certain genomic regions on the autosomes and Z chromosome exhibiting the same phylogenetic topology as the mitochondrial tree.

Future comprehensive sampling of *E. cioides* may prove or disprove hypotheses (a), (c) and (d) (Fig. 4). As for hypothesis (b), ghost introgression from an extinct lineage (Zhang et al. 2019) is challenging to demonstrate due to the general unavailability of genetic material from extinct songbirds. Considering the pervasiveness of interspecific gene flow across different stages of lineage divergence (Mallet 2005; Rheindt and Edwards

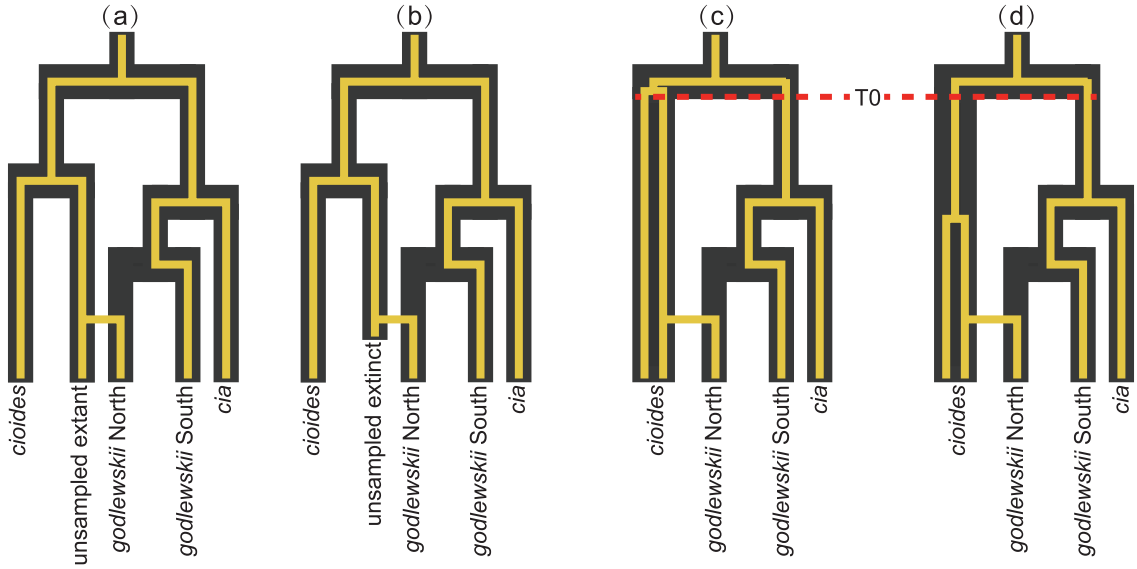

FIGURE 4. Four hypotheses potentially explaining the observed phylogenetic conflict between the species tree and mtDNA/NRW trees. The mtDNA and W chromosome of northern *E. godlewskii* may have introgressed from an unsampled (a) extant or (b) extinct lineage that is sister to *E. cioides*. The mtDNA and W chromosome of northern *E. godlewskii* could have introgressed from an unsampled mitochondrial + W chromosomal lineage within *E. cioides* (c) due to incomplete lineage sorting or (d) female-specific selection. The species tree is shown in black, with the yellow lines representing mtDNA + W chromosomal lineages. The divergence time between different lineages of mtDNA + W chromosome within *E. cioides* (T0) would be around the root divergence time (c) or shallower than the root (d). Note that the presence of gene trees on the autosomes and Z chromosome that align with the mitochondrial tree can also be accounted for by these hypotheses.

2011; Payseur and Rieseberg 2016; Ottenburghs et al. 2017) and frequent lineage extinctions (Novacek 2001; Barnosky et al. 2011), the incidence of introgression from extinct into extant species might be greatly underestimated (Zhang et al. 2019; Ottenburghs 2020).

Maternal inheritance and haploidy render lineage sorting an order of magnitude more rapid in mtDNA than in most nuclear DNA (Avise 2000; Toews and Brelsford 2012). Because of these properties, mtDNA has been the traditional workhorse of phylogenetic research on recently diverged species (Zink and Barrowclough 2008). However, phylogenetic reconstruction using mtDNA may only represent the genic evolutionary history per se, not the species tree, as indicated in this study (Fig. 2). Therefore, the use of genome-wide data is critical for phylogenetic reconstruction and species delimitation. Only when the true biological branching order is known, and we have insights into the direction and magnitude of post-speciation gene flow events, will it be possible to explore the underlying causes of phylogenetic conflicts. The results of this study challenge the previous over-reliance on datasets of limited genomic extent within taxonomy and species delimitation and call for a new research paradigm that acknowledges the importance of comprehensive chromosomal sampling in the genomic era.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.wstqjq2r1

ACKNOWLEDGMENTS

We thank the editors and 5 anonymous reviewers for the constructive comments, which improved this manuscript substantially. We thank Zuohua Yin and Xiaobing Li for sample collection. We also thank National Animal Collection Resource Center, Institute of Zoology, Chinese Academy of Sciences for assistance in specimen examination, and Loïs Rancilhac for comments.

FUNDING

This research was funded by the National Science Foundation of China (32270466 to D.Z.; 32130013 to F.L.), the Youth Innovation Promotion Association CAS (2023093 to D.Z.), the Second Tibetan Plateau Scientific Expedition and Research (STEP) program (2019QZKK0501, 2019QZKK0304-02 to F.L. and D.Z.), the National Science and Technology Basic Resources Survey Program of China (2019FY100204 to F.L.), the Swedish National Science Foundation (2019-04486), Jornvall Foundation and Mark and Mo Constantine (to P.A.), and the National Science Foundation of China (32070434) to G.S.

AUTHOR CONTRIBUTIONS

The study was designed by F.L. and D.Z. The data were analyzed by D.Z. and H.S. The manuscript was written by D.Z., H.S., P.A., U.O., and F.L. and commented on by all authors.

Conflict of Interest

The authors declare that there is no conflict of interest.

Data Availability

Whole-genome re-sequencing data were deposited at the NCBI Sequence Read Archive (SRA) under Bioproject PRJNA797667.

References

Alström P., Olsson U., Lei F., Wang H.T., Gao W., Sundberg P. 2008. Phylogeny and classification of the Old World Emberizini (Aves, Passeriformes). Mol. Phylogenet. Evol. 47:960–973.

Avise J.C. 2000. Phylogeography: the history and formation of species. Cambridge (MA) and London: Harvard University Press.

Avise J.C., Robinson T.J. 2008. Hemiplasy: a new term in the Lexicon of phylogenetics. Syst. Biol. 57:503–507.

Barnosky A.D., Matzke N., Tomiya S., Wogan G.O., Swartz B., Quental T.B., Marshall C., McGuire J.L., Lindsey E.L., Maguire K.C., Mersey B., Ferrer E.A. 2011. Has the Earth's sixth mass extinction already arrived? Nature 471:51–57.

Bouckaert R., Heled J., Kuhnert D., Vaughan T., Wu C.H., Xie D., Suchard M.A., Rambaut A., Drummond A.J. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10:e1003537.

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. J. Avian Biol. 52:6.

Chang C.C., Chow C.C., Tellier L.C.A.M., Vattikuti S., Purcell S.M., Lee J.J. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4:7.

Chen N., Bellott D.W., Page D.C., Clark A.G. 2012. Identification of avian W-linked contigs by short-read sequencing. BMC Genomics 13:183.

Chen S., Zhou Y., Chen Y., Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890.

Copete J.L. 2020. Godlewski's Bunting (Emberiza godlewskii), version 1.0. In: del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A., de Juana, E., editors. Birds of the World. Ithaca (NY): Cornell Lab of Ornithology.

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., Genomes Project Analysis G. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24:332–340.

Excoffier L., Dupanloup I., Huerta-Sanchez E., Sousa V.C., Foll M. 2013. Robust demographic inference from genomic and SNP data. PLoS Genet. 9:e1003905.

Felsenstein J. 2004. Inferring phylogenies. Sunderland (MA): Sinauer Associates.

Fontaine M.C., Pease J.B., Steele A., Waterhouse R.M., Neafsey D.E., Sharakhov I.V., Jiang X., Hall A.B., Catteruccia F., Kakani E., Mitchell S.N., Wu Y.-C., Smith H.A., Love R.R., Lawniczak M.K., Slotman M.A., Emrich S.J., Hahn M.W., Besansky N.J. 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science 347:1258524.

Green R.E., Krause J., Briggs A.W., Maricic T., Stenzel U., Kircher M., Patterson N., Li H., Zhai W., Fritz M.H., Hansen N.F., Durand E.Y., Malaspinas A.S., Jensen J.D., Marques-Bonet T., Alkan C., Prufer K., Meyer M., Burbano H.A., Good J.M., Schultz R., Aximu-Petri A., Butthof A., Hober B., Hoffner B., Siegemund M., Weihmann A., Nusbaum C., Lander E.S., Russ C., Novod N., Affourtit J., Egholm M., Verna C., Rudan P., Brajkovic D., Kucan Z., Gusic I., Doronichev V.B., Golovanova L.V., Lalueza-Fox C., de la Rasilla M., Fortea J., Rosas A., Schmitz R.W., Johnson P.L.F., Eichler E.E., Falush D., Birney E., Mullikin J.C., Slatkin M., Nielsen R., Kelso J., Lachmann M., Reich D., Paabo S. 2010. A draft sequence of the Neandertal genome. Science 328:710–722.

Hahn C., Bachmann L., Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads--a baiting and iterative mapping approach. Nucleic Acids Res. 41:e129.

Hibbins M.S., Hahn M.W. 2022. Phylogenomic approaches to detecting and characterizing introgression. Genetics 220:iyab173.

Huang L., Zhang L., Li D., Yan R., Shang W., Jiang Y., Li S. 2022. Molecular evidence of introgressive hybridization between related species Jankowski's Bunting (*Emberiza jankowskii*) and Meadow Bunting (*Emberiza cioides*) (Aves: Passeriformes). Avian Res. 13:100035.

Irwin D.E. 2018. Sex chromosomes and speciation in birds and other ZW systems. Mol. Ecol. 27:3831–3851.

Kozlov A.M., Darriba D., Flouri T., Morel B., Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35:4453–4455.

Li H., Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

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 25:2078–2079.

Li J., Song G., Chen G., Liu N., Dong L., Bao X. 2023. A new bunting species in South China revealed by an integrative taxonomic investigation of the *Emberiza godlewskii* complex (Aves, Emberidae). Mol. Phylogenet. Evol. 180:107697.

Li J., Song G., Liu N., Chang Y., Bao X. 2019. Deep south-north genetic divergence in Godlewski's bunting (*Emberiza godlewskii*) related to uplift of the Qinghai-Tibet Plateau and habitat preferences. BMC Evol. Biol. 19:161.

Liu L., Yu L., Edwards S.V. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol. Biol. 10:302.

Luo R., Liu B., Xie Y., Li Z., Huang W., Yuan J., He G., Chen Y., Pan Q., Liu Y., Tang J., Wu G., Zhang H., Shi Y., Liu Y., Yu C., Wang B., Lu Y., Han C., Cheung D.W., Yiu S.-M., Peng S., Xiaoqian Z., Liu G., Liao X., Li Y., Yang H., Wang J., Lam T.-W., Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 1:18.

Maddison W.P. 1997. Gene trees in species trees. Syst. Biol. 46:523–536.

Malinsky M., Matschiner M., Svardal H. 2020. Dsuite - Fast D-statistics and related admixture evidence from VCF files. Mol. Ecol. Resour. 21:584–595.

Mallet J. 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. 20:229–237.

Nakamura T., Yamada K.D., Tomii K., Katoh K. 2018. Parallelization of MAFFT for large-scale multiple sequence alignments. Bioinformatics 34:2490–2492.

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. Mol. Biol. Evol. 32:268–274.

Nichols R. 2001. Gene trees and species trees are not the same. Trends Ecol. Evol. 16:358–364.

Novacek M.J. 2001. The biodiversity crisis: losing what counts. New York City: The New Press.

Ogilvie H.A., Bouckaert R.R., Drummond A.J. 2017. StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. Mol. Biol. Evol. 34:2101–2114.

Ottenburghs J. 2020. Ghost introgression: spooky gene flow in the distant past. Bioessays 42:e2000012.

Ottenburghs J., Kraus R.H.S., van Hooft P., van Wieren S.E., Ydenberg R.C., Prins H.H.T. 2017. Avian introgression in the genomic era. Avian Res. 8:181–191.

Ottenburghs J., Ydenberg R.C., Van Hooft P., Van Wieren S.E., Prins H.H.T. 2015. The Avian Hybrids Project: gathering the scientific literature on avian hybridization. Ibis 157:892–894.

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. Vertebr. Zool. 65:135–150.

Pamilo P., Nei M. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568–583.

Payseur B.A., Rieseberg L.H. 2016. A genomic perspective on hybridization and speciation. Mol. Ecol. 25:2337–2360.

Quinlan A.R., Hall I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

Rheindt F.E., Edwards S.V. 2011. Genetic introgression: an integral but neglected component of speciation in birds. Auk. 128:620–632.

Schwery O., O'Meara B.C. 2016. MonoPhy: a simple R package to find and visualize monophyly issues. PeerJ Comput. Sci. 2:e56.

Small S.T., Labbe F., Lobo N.F., Koekemoer L.L., Sikaala C.H., Neafsey D.E., Hahn M.W., Fontaine M.C., Besansky N.J. 2020. Radiation with reticulation marks the origin of a major malaria vector. Proc. Natl. Acad. Sci. USA. 117:31583–31590.

Smeds L., Warmuth V., Bolivar P., Uebbing S., Burri R., Suh A., Nater A., Bures S., Garamszegi L.Z., Hogner S., Moreno J., Qvarnstrom A., Ruzic M., Saether S.A., Saetre G.P., Torok J., Ellegren H. 2015. Evolutionary analysis of the female-specific avian W chromosome. Nat. Commun. 6:7330.

Spottiswoode C.N., Stryjewski K.F., Quader S., Colebrook-Robjent J.F.R., Sorenson M.D. 2011. Ancient host specificity within a single species of brood parasitic bird. Proc. Natl. Acad. Sci. USA. 108:17738–17742.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Talavera G., Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56:564–577.

Tang H., Peng J., Wang P., Risch N.J. 2005. Estimation of individual admixture: analytical and study design considerations. Genet. Epidemiol. 28:289–301.

Toews D.P., Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol. Ecol. 21:3907–3930.

Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinf. 19:15–30.

Zhang D., Rheindt F.E., She H., Cheng Y., Song G., Jia C., Qu Y., Alström P., Lei F. 2021. Most Genomic loci misrepresent the phylogeny of an avian radiation because of ancient gene flow. Syst. Biol. 70:961–975.

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 behavior in a songbird. Mol. Ecol. 32:381–392.

Zhang D., Tang L., Cheng Y., Hao Y., Xiong Y., Song G., Qu Y., Rheindt F.E., Alström P., Jia C., Lei F. 2019. "Ghost Introgression" as a cause of deep mitochondrial divergence in a bird species complex. Mol. Biol. Evol. 36:2375–2386.

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., Yang W., Hu J., Xiao J., Yang Z., Liu Y., Xie Q., Yu H., Lian J., Wen P., Zhang F., Li H., Zeng Y., Xiong Z., Liu S., Zhou L., Huang Z., An N., Wang J., Zheng Q., Xiong Y., Wang G., Wang B., Wang J., Fan Y., da Fonseca R.R., Alfaro-Núñez A., Schubert M., Orlando L., Mourier T., Howard J.T., Ganapathy G., Pfenning A., Whitney O., Rivas M.V., Hara E., Smith J., Farré M., Narayan J., Slavov G., Romanov M.N., Borges R., Machado J.P., Khan I., Springer M.S., Gatesy J., Hoffmann F.G., Opazo J.C., Håstad O., Sawyer R.H., Kim H., Kim K.-W., Kim H.J., Cho S., Li N., Huang Y., Bruford M.W., Zhan X., Dixon A., Bertelsen M.F., Derryberry E., Warren W., Wilson R.K., Li S., Ray D.A., Green R.E., O'Brien S.J., Griffin D., Johnson W.E., Haussler D., Ryder O.A., Willerslev E., Graves G.R., Alström P., Fjeldså J., Mindell D.P., Edwards S.V., Braun E.L., Rahbek C., Burt D.W., Houde P., Zhang Y., Yang H., Wang J., Jarvis E.D., Gilbert M.T.P., Wang J.; Avian Genome Consortium. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. Science 346:1311–1320.

Zink R.M., Barrowclough G.F. 2008. Mitochondrial DNA under siege in avian phylogeography. Mol. Ecol. 17:2107–2121.