Received: 5 November 2016 | Revised: 26 September 2017 | Accepted: 1 October 2017

DOI: 10.1111/mec.14383

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Genomic differentiation and patterns of gene flow between two long-tailed tit species (*Aegithalos*)

Dezhi Zhang[1,2*] | Gang Song[1*] | Bin Gao[1] | Yalin Cheng[1,2] | Yanhua Qu[1] | Shaoyuan Wu[3] | Shimiao Shao[1,2] | Yongjie Wu[1,4] | Per Alström[1,5,6] | Fumin Lei[1,2]

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

[2]College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China

[3]School of Life Sciences, Jiangsu Normal University, Xuzhou, China

[4]Key Laboratory of Bio-resources and Eco-environment of Ministry of education, College of Life Sciences, Sichuan University, Chengdu, Sichuan, China

[5]Department of Ecology and Genetics, Animal Ecology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

[6]Swedish Species Information Centre, Swedish University of Agricultural Sciences, Uppsala, Sweden

Correspondence
Fumin Lei, Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China.
Email: leifm@ioz.ac.cn

Funding information
National Science Foundation of China, Grant/Award Number: 31572249, 31630069; Strategic Priority Research Program of the Chinese Academy of Sciences, Grant/Award Number: XDB13020300; Ministry of Science and Technology of China, Grant/Award Number: 2014FY210200; Jornvall Foundation; Swedish Research Council

Abstract

Patterns of heterogeneous genomic differentiation have been well documented between closely related species, with some highly differentiated genomic regions ("genomic differentiation islands") spread throughout the genome. Differential levels of gene flow are proposed to account for this pattern, as genomic differentiation islands are suggested to be resistant to gene flow. Recent studies have also suggested that genomic differentiation islands could be explained by linked selection acting on genomic regions with low recombination rates. Here, we investigate genomic differentiation and gene-flow patterns for autosomes using RAD-seq data between two closely related species of long-tailed tits (*Aegithalos bonvaloti* and *A. fuliginosus*) in both allopatric and contact zone populations. The results confirm recent or ongoing gene flow between these two species. However, there is little evidence that the genomic regions that were found to be highly differentiated between the contact zone populations are resistant to gene flow, suggesting that differential levels of gene flow is not the cause of the heterogeneous genomic differentiation. Linked selection may be the cause of genomic differentiation islands between the allopatric populations with no or very limited gene flow, but this could not account for the heterogeneous genomic differentiation between the contact zone populations, which show evidence of recent or ongoing gene flow.

KEYWORDS
Aegithalos, gene flow, genomic differentiation, population genomics, speciation

1 | INTRODUCTION

The evolution of genetic barriers to gene flow plays a critical role during the speciation process (Coyne & Orr, 2004). Genes involved in reproductive isolation have been suggested to be resistant to gene flow and are thus important in the process of divergence between incipient species (Wu, 2001). The advent of next-generation sequencing technologies makes it possible to characterize patterns of genetic differentiation across genomes between closely related species that are still at the early stages of the speciation process. The landscape of genomic differentiation has been investigated between closely related species, subspecies or even ecomorphs. These studies have discovered highly differentiated genomic regions ("genomic differentiation islands") that are spread across the genome (Carneiro et al., 2014;

*These authors contributed equally.

Ellegren et al., 2012; Harr, 2006; Jones et al., 2012; Lawniczak et al., 2010; Malinsky et al., 2015; Martin et al., 2013; Poelstra et al., 2014; Renaut et al., 2013; Turner, Hahn, & Nuzhdin, 2005). To date, most studies have found that genomic differentiation islands are highly heterogeneous across chromosomes and are relatively concentrated in regions with low recombination rates and reduced levels of intraspecific diversity (Nachman & Payseur, 2012; Roesti, Hendry, Salzburger, & Berner, 2012; Tine et al., 2014). Differential levels of gene flow have been proposed to play a significant role in shaping the landscape of heterogeneous genomic differentiation (Nosil, Funk, & Ortiz-Barrientos, 2009; Turner, Hahn, & Nuzhdin, 2005; Wu, 2001). Genomic differentiation islands have been suggested to be shielded from gene flow, thereby contributing to the evolution of reproductive isolation in the face of gene flow (Feder, Egan, & Nosil, 2012; Feder & Nosil, 2010; Flaxman, Feder, & Nosil, 2013; Nosil, 2008; Via, 2012).

The cause of the formation of heterogeneous genomic differentiation and the role of genomic differentiation islands in driving speciation have been debated (Cruickshank & Hahn, 2014; Noor & Bennett, 2009; Payseur & Rieseberg, 2016; Pennisi, 2014; Turner & Hahn, 2010). Although genomic differentiation islands were originally regarded as "speciation islands" (Turner et al., 2005), this interpretation has been contradicted by some recent studies. For example, Clarkson et al. (2014) found that even an extremely prominent genomic differentiation island in *Anopheles* was independent of reproductive isolation. Moreover, migration-linked genes that are considered to be involved in reproductive isolation are mainly distributed outside of genomic differentiation islands in two subspecies of Swainson's Thrush (*Catharus ustulatus*) with different migration routes (Ruegg, Anderson, Boone, Pouls, & Smith, 2014). Instead of genomic differentiation islands being caused by shielding from gene flow, there is amounting evidence that heterogeneous genomic differentiation landscapes may be inherently driven by linked selection, that is background selection and selective sweeps affecting linked neutral loci at genomic regions with reduced recombination rates (Burri et al., 2015; Delmore et al., 2015; Vijay et al., 2016), rather than by reduced gene flow in these highly differentiated genomic regions (Cruickshank & Hahn, 2014).

The Black-browed Tit (*Aegithalos bonvaloti*) and Sooty Tit (*A. fuliginosus*) are two closely related long-tailed tit species occurring in East Asia. Their distribution ranges are mostly allopatric, although they meet in a rather narrow contact zone in central China (Figure 1). As resident forest birds, *A. bonvaloti* is distributed in high altitude habitats ranging from 1,500 to 4,400 m, whereas *A. fuliginosus* inhabits relatively lower altitudes ranging from 1,000 to 2,600 m (del Hoyo, Elliott, & Christie, 2008). Although these two species are believed to have diverged from each other as recently as ~0.2 million years ago (Päckert, Martens, & Sun, 2010; Päckert et al., 2012), the colours of their plumages are quite different (Figure 1). A recent study, based on two mitochondrial and six nuclear markers, found virtually no mitochondrial divergence between the two species, and ascribed this to past hybridization in combination with a selective sweep and/or genetic drift (Wang et al., 2014). The same study also found evidence of unidirectional mitochondrial introgression, from *A. fuliginosus* to *A. bonvaloti*, but much lower or negligible nuclear gene flow.

In this study, we examined genomic differentiation and gene-flow patterns between these two species based on RAD-seq data and a larger number of samples from the contact zone compared to Wang et al. (2014). Primarily, we asked the following questions: (i) Is there any evidence of nuclear gene flow between the two species, in contrast to the previous suggestions of low or negligible nuclear gene flow? (ii) If so, are highly differentiated genomic regions resistant to gene flow? (iii) Is linked selection the main cause of heterogeneous genomic differentiation?

2 | METHODS

2.1 | Sampling and DNA extraction

We sampled a total of 36 individuals from eight localities, including 21 individuals of *A. bonvaloti* (10 from the allopatric zones and 11

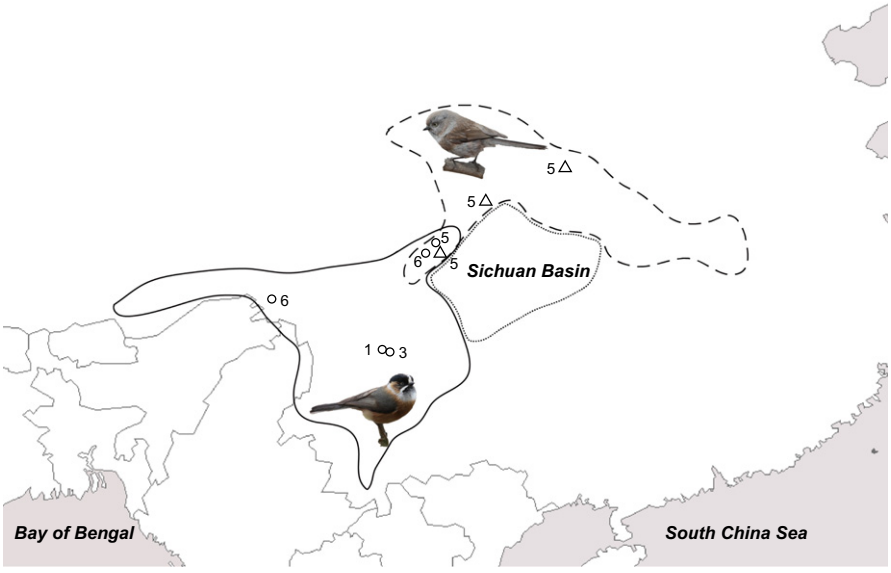

FIGURE 1 Geographical distribution of *Aegithalos bonvaloti* and *A. fuliginosus* with DNA sampling localities. The solid and dashed lines indicate the distribution ranges of *A. bonvaloti* and *A. fuliginosus*, respectively. The overlapping region represents the contact zone. The circles and triangles indicate the sampling localities of *A. bonvaloti* and *A. fuliginosus*, respectively, with the numbers representing sample sizes [Colour figure can be viewed at wileyonlinelibrary.com]

from the contact zone) and 15 individuals of *A. fuliginosus* (10 from the allopatric zones and five from the contact zone) (Figure 1; Table S1). We sampled one individual of *A. caudatus* as an out-group (Table S1). Muscle tissue was preserved in 100% ethanol and then stored at −80°C. Muscle tissue samples were collected along with specimens for the National Zoological Museum, Institute of Zoology, Chinese Academy of Sciences, Beijing. The collection work conformed to the National Wildlife Conservation Law with permission from the Forestry Department of China. No experiments using live animals were carried out. Total genomic DNA was extracted from muscle tissue using the Tissue/Cell Genomic DNA Extraction Kit (Aidlab Biotechnologies Co., Ltd., Beijing, China) according to the manufacturer's protocol.

2.2 | RAD-seq procedures and bioinformatics analysis

RAD libraries were prepared according to the protocol described by Baird et al. (2008). We used the restriction enzyme *Eco*RI to digest DNA samples and selected DNA fragments with a sequence length ranging from 200 to 400 bp for library construction. All prepared libraries were sequenced using Illumina HiSeq2500 with a paired-end read length of 125 bp on two lanes. Raw reads were first processed to remove adapter sequences, low-quality reads (those with over 50% of bases having Phred quality scores less than 5) and poly-N reads (those with ≥10% unidentified nucleotides) using in-house scripts. Library preparation, NGS sequencing and processing of raw reads were conducted by the *Novogene Bioinformatics Institute* (Beijing, China).

Because reference sequences were not available for *A. bonvaloti* and *A. fuliginosus*, we assembled the genome sequences of the *A. bonvaloti* sample with the greatest number of reads as the reference genome (pseudo-reference genome, PRG, Rheindt, Fujita, Wilton, & Edwards, 2014) using VELVET 1.2.10 (Zerbino & Birney, 2008). We also assembled a PRG of an *A. fuliginosus* sample using the same standard to reduce the putative bias if only one focal species was used as a reference. All reads were aligned to the PRG using BWA 0.7.12 (Li & Durbin, 2009), and variants were called using SAMtools 0.1.19 (Li et al., 2009) using the "mpileup" module for all 37 samples with the settings "mpileup -g -u -S –D." PCR duplicates were removed before variant calling using SAMtools 0.1.19 with the "rmdup" command. SNPs were filtered using VCFtools 0.1.12b (Danecek et al., 2011) and bcftools (Li et al., 2009) according to the following criteria: (i) quality value ≥30; (ii) only bi-allelic SNPs were retained; (iii) genotype depth ranged from 2 to 1,000; (iv) SNPs with ≥7 genotypes out of the 10 or 11 samples from the allopatric and contact zone populations of *A. bonvaloti* and the allopatric population of *A. fuliginosus*, and SNPs with ≥4 genotypes of the five samples from the contact zone population of *A. fuliginosus*; (v) homogeneous SNPs throughout all *A. bonvaloti* and *A. fuliginosus* samples were filtered out to remove SNPs between the out-group species and the two focal species.

Except for local inversions and rearrangements, the conserved chromosomal synteny of birds (Backström et al., 2008; Ellegren, 2010) allows the mapping of sequences to the genomes of other avian species (Rheindt et al., 2014; Ruegg et al., 2014). We mapped contigs from the PRG to the genome sequences of the Great Tit (*Parus major*) (Laine et al., 2016) using BLAST+ 2.2.26. As four of five individuals of *A. fuliginosus* in the contact zone were females (Table S1), which have heterozygous sex chromosomes (ZW), we only consider autosomes in this study. To reduce possible mismatches, we used only contigs that had a single hit with an *e*-value $<10^{-40}$.

2.3 | Phylogenetic and population structure analysis

Concatenated SNPs were used to infer maximum-likelihood (ML) trees in RAxML 8.1 (Stamatakis, 2014). The ASC_GTRGAMMA model (rate heterogeneity with ascertainment bias correction for SNP data), as recommended by the manual, was used to estimate the ML tree for all 36 *A. bonvaloti* and *A. fuliginosus* and separately for the 20 allopatric individuals; 100-bootstrap replicates were used, and *A. caudatus* was set as the out-group taxon. Because ascertainment bias correction of RAxML treated some SNPs as invariant (see the RAxML v8.2.X manual, pages 27–28), two subsets were generated for constructing the two trees after discarding unqualified SNPs.

We used FRAPPE 1.1 (Tang, Peng, Wang, & Risch, 2005) to analyse the population structure of these two species with a *K* value setting ranging from 2 to 4, and a maximum of 10,000 iterations. We only retained one SNP per contig for FRAPPE analysis. Principal component analysis (PCA) was performed using GCTA 1.24 (Yang, Lee, Goddard, & Visscher, 2011), and we also performed multidimensional scaling (MDS) as implemented in PLINK 1.9 (Purcell et al., 2007) using the same data set.

2.4 | Estimation of population genomic parameters

We calculated F_{ST} (fixation index) and Tajima's *D* using VCFtools 0.1.12b (Danecek et al., 2011), and D_{XY} (interpopulation/-species nucleotide divergence) and π (intrapopulation/species nucleotide diversity) were estimated using the Python script egglib_sliding_windows.py (https://github.com/johnomics; Martin, Davey, & Jiggins, 2015). All parameters were calculated using a 100 kb nonoverlapping sliding-window approach along each of the chromosomes. We also performed a 200 kb nonoverlapping sliding-window analysis to avoid the putative bias introduced by a single window size. Windows of 100 kb with fewer than 10 variants were removed, as were windows of 200 kb with fewer than 20 variants. We calculated F_{ST} and D_{XY} between the allopatric populations of *A. bonvaloti* and *A. fuliginosus*, as well as between the contact zone populations of these two species. The mean F_{ST} values produced by VCFtools were used in this analysis, and negative F_{ST} values were converted to zero. Tajima's *D* and π were calculated for each of the four populations. We defined highly differentiated genomic regions (high F_{ST} regions, HFRs) as those where the absolute *Z*-score (standard score) of F_{ST} values in a window was ≥3 (representing approximately the top 2%;

cf. similar to Carneiro et al., 2014) based on the allopatric populations, while the remaining genomic regions represented genomic background (low F_{ST} regions, LFRs). We also evaluated how alternative definitions of HFRs, as the top 5% and 10% of F_{ST} regions, respectively, would affect the results.

Gene flow acts to decrease both F_{ST} and D_{XY}, but unlike the absolute measure of divergence D_{XY}, the relative measure of divergence F_{ST} has an inherent bias with respect to π and is higher in genomic regions with lower π (Charlesworth, 1998; Charlesworth, Nordborg, & Charlesworth, 1997; Cruickshank & Hahn, 2014). Therefore, we only focused on D_{XY}. We compared D_{XY} in the contact zone populations to D_{XY} in the allopatric populations, and subtracted the former from the latter to obtain the decreased values of D_{XY} that would indicate the degree of gene flow, with larger decreases in values of D_{XY} indicating higher rates of gene flow. Here, we only compared the decreases of D_{XY} of HFRs to LFRs between the contact zone populations with potential gene flow.

2.5 | Linkage disequilibrium analysis

We estimated linkage disequilibrium (LD) patterns for each of the four populations. BEAGLE 4.1 (Browning & Browning, 2007) was used for genotype phasing. The phased genotypes were used to calculate the correlation coefficient (r^2) between any two SNPs using VCFtools 0.1.12b (Danecek et al., 2011) with the "–hap-r2" option. Average r^2 was plotted against physical distance in base pairs with R 3.3.3 (R Development Core Team, 2008). The SNPs generated from the PRG of *A. fuliginosus* that had successfully mapped to the Great Tit's genome with minor allele frequency ≥0.1 were used for LD analysis.

2.6 | Demographic inference

∂a∂i 1.6.3 (Gutenkunst, Hernandez, Williamson, & Bustamante, 2009) was used to infer demographic histories between both the allopatric and the contact zone populations. We compared three different divergence models for the allopatric populations: no migration model (NM), isolation with asymmetric migration model (IaM) and secondary contact with asymmetric migration model (SCaM). For the contact zone populations, we only compared the latter two models with asymmetric migration: IaM and SCaM models. The NM and IaM models indicate divergence between the two populations without or with asymmetric gene flow, respectively, while the SCaM model indicates the two populations first diverged in allopatry without gene flow, followed by a period of asymmetric gene flow.

2.7 | *D*- and f_d-statistic

We used a *D*-statistic approach ("ABBA-BABA test") to distinguish gene flow from incomplete lineage sorting and to estimate gene flow in a four-taxon framework (Durand, Patterson, Reich, & Slatkin, 2011; Green et al., 2010). *D*-statistic is used to detect gene flow between an inner-group P1/P2 and a third inner-group P3. We set the allopatric *A. fuliginosus* as P1, the contact zone *A. fuliginosus* as P2, the contact zone *A. bonvaloti* as P3 and *A. caudatus* as the outgroup. We also used the f_d-statistic derived from the *D*-statistic, as the latter is not reliable when applied in small genomic regions (Martin et al., 2015). We then compared both *D*- and f_d-statistic values of HFRs to LFRs between the contact zone populations. Both *D*- and f_d-statistics were computed using the Python script egglib_sliding_windows.py (https://github.com/johnomics; Martin et al., 2015).

3 | RESULTS

3.1 | RAD-seq data output

We obtained high-quality sequences from RAD-sequencing that averaged Q20 > 95% and Q30 > 89%, with an average mapping rate >85%, an average breadth of coverage >71% and an average sequencing depth >4.8 for both species used as PRG (Table S2). A total of 672,228 and 625,868 contigs made up the PRG for *A. bonvaloti* and *A. fuliginosus*, respectively, of which 54% (363,760 contigs) and 47% (294,337 contigs), respectively, successfully mapped to the Great Tit genome. The average contig lengths were 290 and 263 bp for the PRG of *A. bonvaloti* and *A. fuliginosus*, respectively. Most contig lengths were ≤500 bp (Fig. S1). A total of 372,187 and 381,386 SNPs were obtained after the filtering steps described above using *A. bonvaloti* and *A. fuliginosus* as the PRG, respectively, of which 275,846 (*bonvaloti* data set) and 296,051 (*fuliginosus* data set) successfully mapped to the Great Tit autosomes.

3.2 | Phylogeny and population structure

In the *bonvaloti* data set, 82,677 and 73,128 SNPs were retained to construct ML trees for all and only allopatric samples, respectively. For the *fuliginosus* data set, 88,562 and 78,710 SNPs were retained for the two ML trees. When only samples from the allopatric zones were included, both species formed monophyletic groups with 100% bootstrap support (Figures 2a and S2a). However, when samples from both the allopatric and contact zones were included, *A. fuliginosus* was no longer supported as monophyletic, and the monophyly of *A. bonvaloti* was not strongly supported (Figures 2b and S2b). Instead, all *A. bonvaloti* and two contact zone *A. fuliginosus* formed a clade with moderate bootstrap support (76% and 80%) (Figures 2b and S2b).

In total, 66,558 and 69,967 SNPs from the *bonvaloti* and *fuliginosus* data sets, respectively, were retained for FRAPPE, PCA and MDS analyses. According to FRAPPE, eight of 11 *A. bonvaloti* and four of five *A. fuliginosus* from the contact zone have a mixed origin with $K = 2$, but the extent of mixing in *A. fuliginosus* was much higher than in *A. bonvaloti* (Figures 2c and S2c). In addition, two *A. bonvaloti* individuals from the allopatric zone also showed evidence of mixed origin. With $K = 3$ and $K = 4$, some *A. bonvaloti* from the allopatric zone showed intermediate status between individuals from the remainder of the allopatric zone and the contact zone, but no *A. fuliginosus* from the allopatric zone displayed this pattern (Figures 2c and S2c). The PCA and MDS confirmed the results of

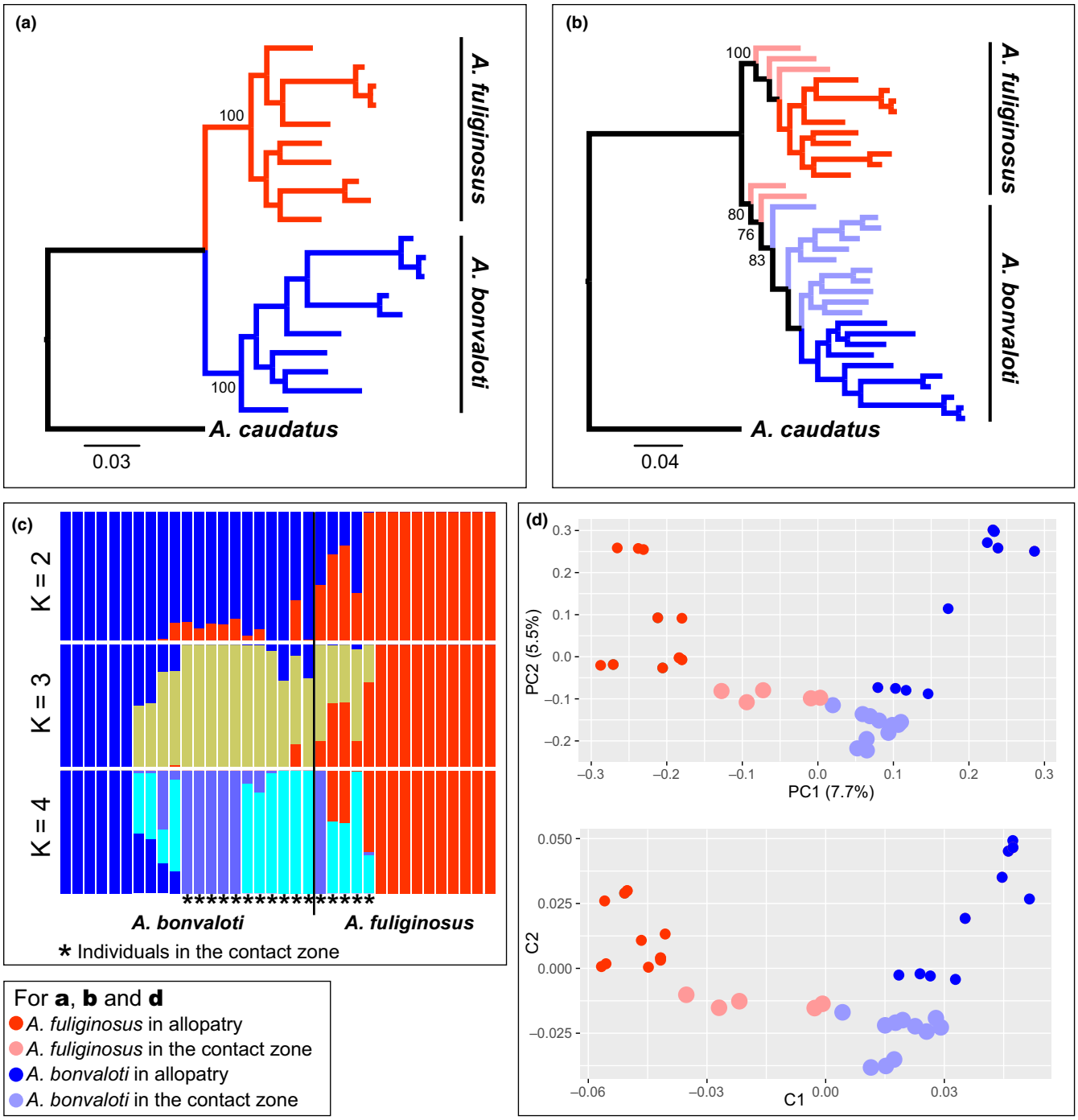

FIGURE 2 Population genetic structure of *Aegithalos bonvaloti* and *A. fuliginosus* using *A. fuliginosus* as pseudo-reference genome. (a) Maximum-likelihood tree based on samples from only the allopatric populations. (b) Maximum-likelihood tree based on samples from both the allopatric and contact zone populations. (c) Population genetic structure of all *A. bonvaloti* and *A. fuliginosus* samples when $K = 2$, 3 and 4, respectively. Asterisks indicate individuals from the contact zone. (d) Principal component analysis (top) and multidimensional scaling (bottom) plot of all *A. bonvaloti* and *A. fuliginosus* samples. Colour codes in a, b and d explained in inset [Colour figure can be viewed at wileyonlinelibrary.com]

the phylogeny and FRAPPE analysis, in which neither species formed a clear cluster, and individuals from the contact zone occupied an intermediate position between the two allopatric populations along PC1 and C1, respectively, and were generally closer to each other than to individuals from the allopatric zones along PC2 and C2, respectively (Figures 2d and S2d).

3.3 | Population genomic parameters

Genomic differentiation between *A. bonvaloti* and *A. fuliginosus* based on F_{ST} was found to be highly heterogeneous for both the allopatric and contact zone populations, and their main HFRs were generally coincident (Figure 3). The mean F_{ST} and D_{XY} values were

0.05 and 0.00061, respectively, for the allopatric populations, and 0.02 and 0.00042 for the contact zone populations, respectively (Table S3, here we only show the 100 kb windowed analysis based on the *fuliginosus* data set). Both F_{ST} and D_{XY} were significantly decreased for the contact zone populations compared to the allopatric populations (Fig. S3).

Although D_{XY} and Tajima's D were not significantly lower for HFRs than for LFRs for the allopatric populations (Figures 4 and S4), mean D_{XY} and Tajima's D showed consistently lower mean values for HFRs than for LFRs across different data sets and analyses using different window sizes (Table S3). For the contact zone populations, D_{XY} was significantly lower for HFRs than for LFRs, whereas the opposite was true for Tajima's D (Figures 4 and S4). π was significant lower for HFRs than for LFRs in both species and both the allopatric and contact zone populations (Figures 4 and S4).

When HFRs were defined as the top 5% and top 10% F_{ST} genomic regions, respectively, we observed the opposite results for Tajima's D for the allopatric populations. Although this was not significant, it still suggested higher mean values for HFRs than for LFRs across different data sets and different window sizes for both species (Figure S5, Table S3). Along with the reduction of F_{ST} in HFRs when the range of HFRs expanded from top ~2% (Z [F_{ST}] ≥ 3) to top 10% genomic regions, Tajima's D showed an increasing trend for the allopatric populations, but a decreasing trend for the contact zone populations; π showed an increasing trend for both species and both the allopatric and contact zone populations; D_{XY} showed an increasing trend for the contact zone populations, but no obvious increasing or decreasing trend for the allopatric populations (Table S3). For the decreases of D_{XY}, there was no significant difference between HFRs and LFRs across different data sets and different window sizes (Fig. S6). The slope plots of D_{XY} in the allopatric populations against the contact zone populations also suggested that there was no obvious difference in the slopes between HFRs and LFRs (Fig. S7).

3.4 | LD and ∂a∂i estimation

Linkage disequilibrium as measured by the correlation coefficient (r^2) decreased below 0.2 within less than 1,000 bp for all the four populations, and the patterns of LD decay between all four populations were highly consistent (Fig. S8). And this rapid decay of LD would exclude the autocorrelation between adjacent windows, as the window sizes used in this study were much broader than LD decay distance. We tested several divergence models between both the allopatric and the contact zone populations implemented in ∂a∂i. The SCaM model had the best fit for both the allopatric and the contact zone populations (Fig. S9, Table S4).

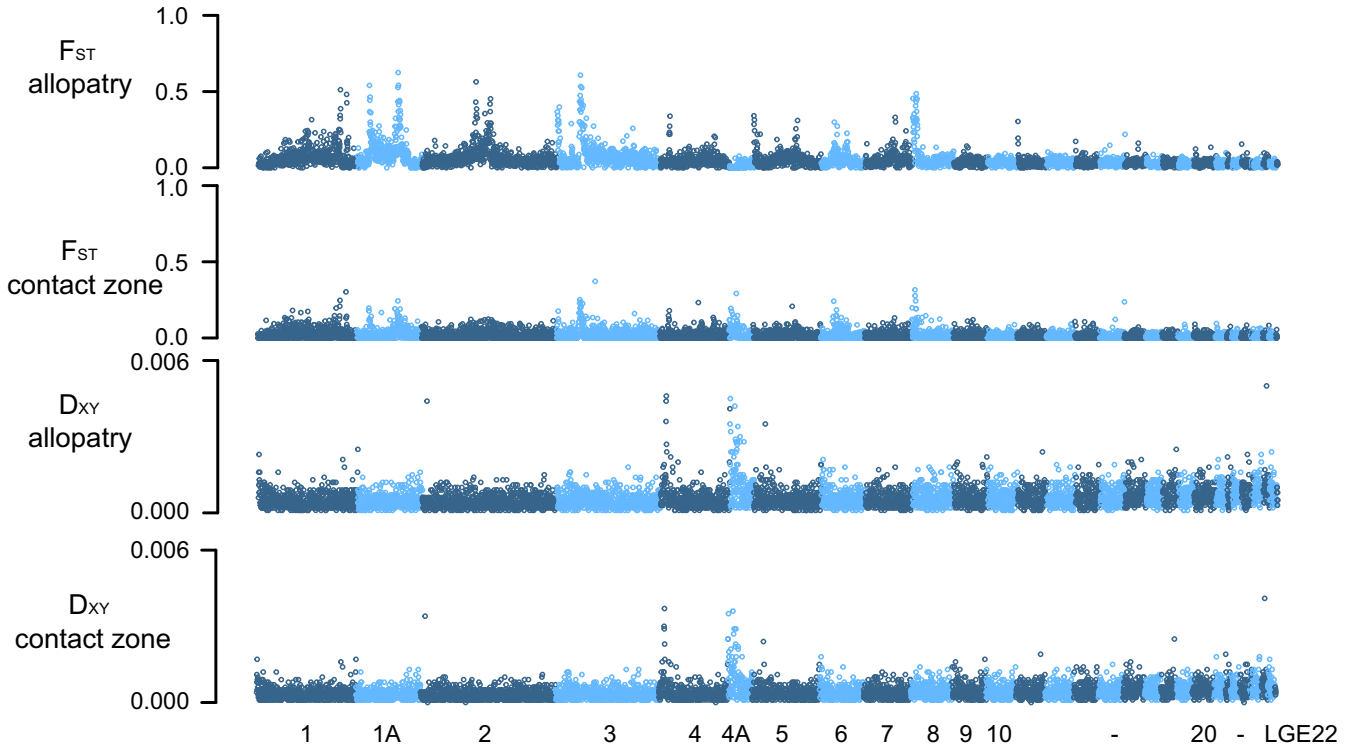

FIGURE 3 Genomic differentiation patterns based on F_{ST} and D_{XY} in 100 kb sliding windows along each of the chromosomes between *Aegithalos bonvaloti* and *A. fuliginosus* from both the allopatric and contact zone populations. Each dot denotes a value for a window. From top to bottom, the panels represent F_{ST} for the allopatric populations, F_{ST} for the contact zone populations, D_{XY} for the allopatric populations and D_{XY} for the contact zone populations. Alternating colours indicate the different chromosomes. The chromosome numbers are indicated at the bottom, and their arrangement is identical for each panel [Colour figure can be viewed at wileyonlinelibrary.com]

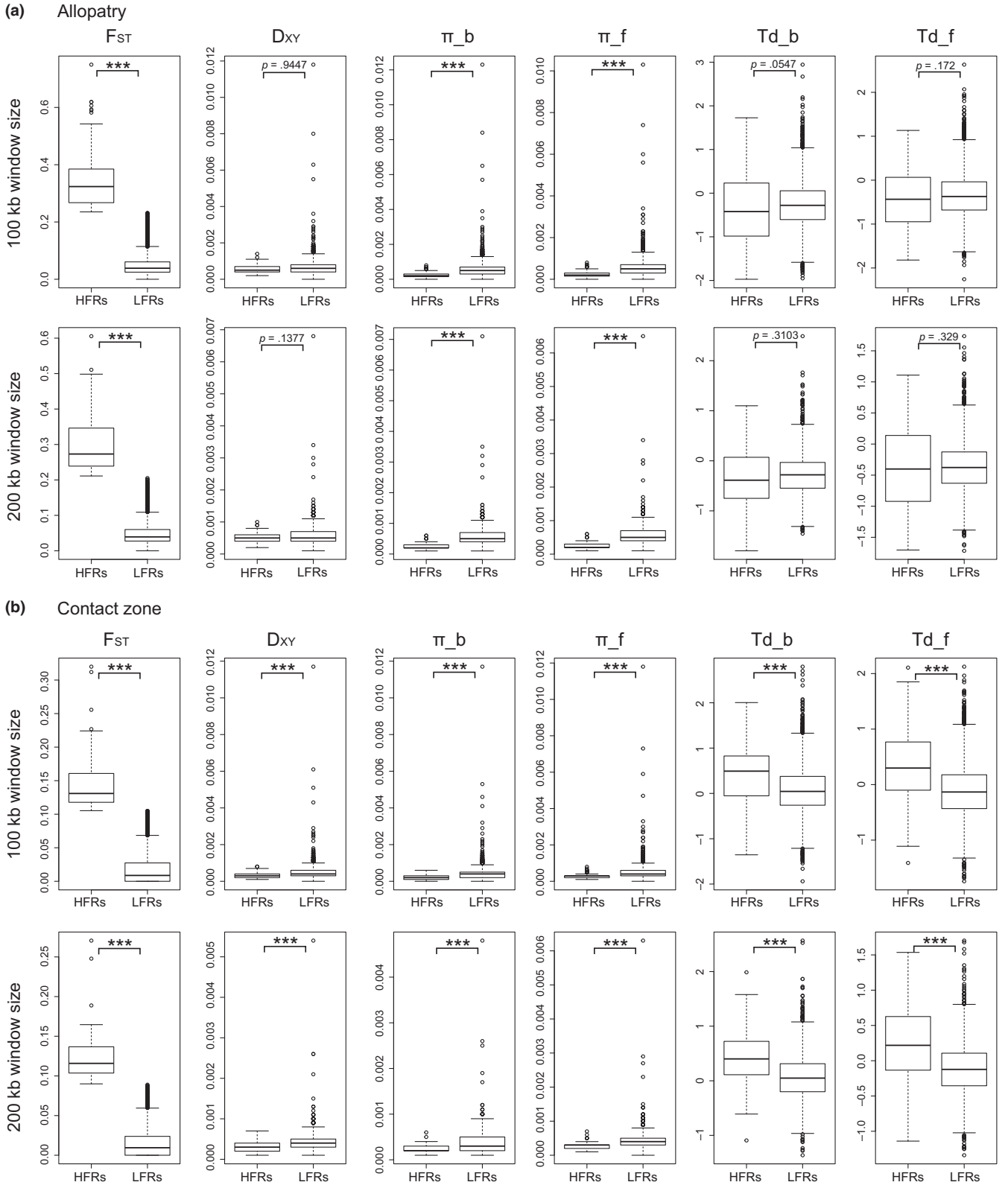

FIGURE 4 Population genomic parameter comparison between HFRs and LFRs when HFRs were defined as a *Z*-score of $F_{ST} \geq 3$ in (a) the allopatric populations and (b) the contact zone populations using *Aegithalos fuliginosus* as PRG. π_b and π_f represent π of *A. bonvaloti* and *A. fuliginosus*, respectively; Td_b and Td_f represent Tajima's *D* of *A. bonvaloti* and *A. fuliginosus*, respectively. ***Indicates $p < 0.001$, and the nonsignificant *p* values are shown with numerical values (Wilcoxon rank sum test)

3.5 | *D*- and f_d-statistic

We conducted *D*- and f_d-statistic analyses for both HFRs and LFRs. *D* values were much greater for HFRs than for LFRs across different data sets and different window sizes when HFRs were defined as *Z* (F_{ST}) ≥3 or the top 5% F_{ST} regions (Table S5), and f_d values were much higher for HFRs than for LFRs only when HFRs were defined as Z (F_{ST}) ≥3 across different data sets and window sizes (Table S5).

4 | DISCUSSION

4.1 | Nuclear gene flow between *A. bonvaloti* and *A. fuliginosus* in the contact zone

Although substantial unidirectional mitochondrial introgression was found from *A. fuliginosus* to *A. bonvaloti* in a previous study, nuclear gene flow was found to be limited (Wang et al., 2014). This limited nuclear gene flow might have been biased by insufficient sampling from the contact zone or the small number of loci or both. This was indeed suggested to be the case, as in this study, we found recent or ongoing nuclear gene flow between the contact zone populations of these two species. Although both incomplete lineage sorting and gene flow can contribute to paraphyly between closely related species (Degnan & Rosenberg, 2009; Funk & Omland, 2003; Petit & Excoffier, 2009), the observed pattern of nonmonophyly exclusively when the contact zone populations were included can only be explained by introgression. The observed spatial distribution of the samples in the PCA and MDS also indicated gene flow across the contact zone. In addition, the FRAPPE results suggested a mixed origin for most of the individuals from the contact zone, with a higher level of admixture for *A. fuliginosus* than for *A. bonvaloti*. This indicates bidirectional gene flow between the two focal species, with stronger gene flow from *A. bonvaloti* to *A. fuliginosus* than in the opposite direction. Our ∂a∂i analysis also indicated that secondary contact with asymmetric gene flow was the most likely demographic model between both the allopatric and the contact zone populations of the two species. Of course, other evolutionary histories of the two species are possible, for example that these species arose sympatrically in the contact zone, with the allopatric populations emerging later through range expansion. However, given that Wang et al. (2014) found that most contact zone individuals of *A. bonvaloti* possessed *A. fuliginosus* mitochondrial haplotypes, but none of the allopatric individuals of *A. bonvaloti* had *A. fuliginosus* mitochondrial genes, previous data more strongly support our proposed model of allopatric speciation followed by secondary contact. Further, our analyses demonstrate that the contact zone populations have lower π values than the allopatric population (Table S3), whereas the opposite result would be predicted under a model of sympatric speciation followed by range expansion. For the allopatric populations, the model of secondary contact with asymmetric gene flow indicates that gene flow happened in the past, but does not imply recent or ongoing gene flow between them.

The pure parental status of *A. fuliginosus* from the allopatric zone demonstrates the limited gene flow between the allopatric and contact zone populations of this species. In contrast, the intermediate status of *A. bonvaloti* from the allopatric zone suggests higher genetic exchange between the contact zone and the allopatric zone populations. Both F_{ST} and D_{XY} values would be reduced by gene flow (Cruickshank & Hahn, 2014), and we observed a marked decrease in both these parameters between the contact zone populations compared to the allopatric populations. In addition, the *D*- and f_d-statistic values, as well as the D_{XY} decreases for both HFRs and LFRs, indicated that gene flow has occurred between the two focal species at both these regions.

Wang et al. (2014) suggested that the observed unidirectional mitochondrial gene flow from *A. fuliginosus* to *A. bonvaloti* might be explained by the brighter plumage of *A. bonvaloti* compared to that of *A. fuliginosus* or the possible female-biased southward dispersal of *A. fuliginosus*, which might make hybridization between male *A. bonvaloti* and female *A. fuliginosus* more likely than between female *A. bonvaloti* and male *A. fuliginosus*. They also suggested that both species underwent population size growth and expanded their contact zones to the Sichuan basin during the last glacial maximum, which may have facilitated hybridization. Our results indicated stronger nuclear gene flow from *A. bonvaloti* to *A. fuliginosus* than in the opposite direction in contrast with the reverse pattern for mitochondrial DNA. In general, in shifting hybrid zones, gene flow occurs from the invaded population to the invading one, as a result of the smaller number of individuals of the invading species than of the invaded in the contact zone (Currat, Ruedi, Petit, & Excoffier, 2008; Rheindt & Edwards, 2011). However, in the case of the two long-tailed tits, nothing is known about the temporal and spatial dynamics of the contact zone, or whether the population sizes differ between the two species in the contact zone. If Wang et al.'s (2014) hypothesis that dispersal into the contact zone has been mainly by female *A. fuliginosus* is true, this would be consistent with the observed different directions of gene flow between mitochondrial and nuclear loci.

4.2 | Little evidence of highly differentiated genomic regions exhibiting resistance to gene flow

If the HFRs would be shielded from gene flow, the differences in D_{XY} values between the contact zone and allopatric populations should be smaller in the HFR regions than in the genomic background. However, we did not observe this pattern. Although lower D_{XY} values of specific genomic regions compared to genomic background have been used as an indicator of gene flow (Cruickshank & Hahn, 2014; Smith & Kronforst, 2013; Zhang, Dasmahapatra, Mallet, Moreira, & Kronforst, 2016), this pattern can also be caused by reduction of ancestral population size in HFRs before lineage splitting (Burri et al., 2015; Nachman & Payseur, 2012). We could not exclude this possibility and suggest that the lower D_{XY} in HFRs than in LFRs between the contact zone populations might have been caused by both gene flow and reduced ancestral population size. We also found that these highly differentiated genomic regions showed larger *D*- and f_d-statistic values than the remaining genomic regions,

which has also been noted in *Heliconius* butterflies (Kronforst et al., 2013). In conclusion, our analyses found little evidence that the HFRs were shielded from gene flow. However, it should be noted that this was based on an analysis of all HFRs, and therefore, some narrow genomic regions in HFRs related to reproductive isolation may have been overlooked. Using a sliding-window approach, it is possible that some narrow genomic regions with high F_{ST} may have been missed in LFRs.

4.3 | Linked selection may not be the main cause of the observed heterogeneous genomic differentiation between contact zone populations

As described above, the HFRs showed no resistance to gene flow, indicating that differential levels of gene flow across the genome are not the main cause of heterogeneous genomic differentiation. Moreover, all of the peaks in these genomic differentiation islands were much lower or even absent in contact zone populations, and some of these differentiation islands even disappeared, indicating that gene flow could erode or even erase genomic differentiation islands. For the allopatric populations, we observed lower intraspecific diversity (π) and Tajima's *D* in HFRs than in the genomic background. The reduced Tajima's *D* indicates that allele frequency spectra were skewed towards rare alleles, which is a strong signal of linked selection (Burri et al., 2015; Delmore et al., 2015; Ellegren et al., 2012). Burri et al. (2015), Delmore et al. (2015) and Ellegren et al. (2012) suggested that linked selection is the main cause of genomic differentiation islands and the lower intraspecific diversity within them. In the present study, we found a similar pattern for the allopatric populations. However, we observed lower intraspecific diversity, π, but higher Tajima's *D*, in HFRs than genomic background for the contact zone populations. Recent or ongoing gene flow between the target populations in previous studies is either absent (Burri et al., 2015; Ellegren et al., 2012) or undetermined (Delmore et al., 2015), and therefore, it is possible that linked selection may only account for heterogeneous genomic differentiation in the absence of or with limited gene flow. Here, we found that linked selection could not explain the observed heterogeneous genomic differentiation when recent or ongoing gene flow is considered.

4.4 | RAD-seq reliability in population genomics

Because of the low cost, independence of a genomic references and the high throughput, RAD-seq has been widely used in the past decade in population genomics in evolutionary and ecological studies, especially for nonmodel organisms (Shafer et al., 2016). Although RAD-seq has many advantages, it still exhibits inherent bias. This bias may be introduced by allele dropout and null alleles, PCR duplicates and genotyping errors, and by variance in the depth of coverage among loci (Andrews, Good, Miller, Luikart, & Hohenlohe, 2016). Allelic dropout and null alleles can cause heterozygotes to be mistaken as homozygotes (Andrews, Good, Miller, Luikart, & Hohenlohe, 2016), which would cause an overestimation of F_{ST} and an underestimation of genetic diversity (Arnold, Corbett-Detig, Hartl, & Bomblies, 2013; Gautier et al., 2013). The low depth of coverage as used in the present study could also introduce the same bias. That would influence the overall levels of polymorphism, meaning that the overall F_{ST} and π in this study may have been overestimated and underestimated, respectively. However, this should not cause a bias in the comparisons between HFRs and LFRs. Although much concern has been focused on the biases associated with different laboratory protocols, little is known about the downstream analysis (Shafer et al., 2016). Large differences exist between reference-based and de novo approaches, and reference-based approaches to a closely related genome were recommended by Shafer et al. (2016). The pseudo-reference genome approaches using RAD-seq data, as used in this study (and also Rheindt et al., 2014), might be suitable alternatives, but this has not been evaluated to date (Shafer et al., 2016).

5 | CONCLUSIONS

In this study, we systematically investigated the genomic differentiation and gene-flow patterns between *A. bonvaloti* and *A. fuliginosus* using RAD-seq data. We confirm the presence of recent or ongoing gene flow between these two species in the contact zone. Gene flow may have occurred in both HFRs and genomic background, but we find little evidence that HFRs may have been shielded from gene flow. We propose that linked selection may account for the heterogeneous genomic differentiation observed between the allopatric populations, but that it may not be the cause of the heterogeneous genomic differentiation observed between the contact zone populations. Overall, our study provides new insights into the correlation between selection, differentiation and gene flow between closely related species in evolutionary biology.

ACKNOWLEDGEMENTS

We thank Matthew W. Hahn, James B. Pease and Simon H. Martin for assistance, and Scott V. Edwards for suggestions. We also thank editor Anna Santure and four anonymous reviewers for helpful comments and suggestions in developing the manuscript. We thank Zuohua Yin, Wenjuan Wang, Chuanyin Dai and Xiaoyang Wang for specimen collections, and Chenxi Jia and Wenjuan Wang for providing images of the two species. This research was funded by the National Science Foundation of China (no. 31572249, 31630069), the Strategic Priority Research Program of the Chinese Academy of Sciences (grant no. XDB13020300), and a grant from the Ministry of Science and Technology of China (no. 2014FY210200) to F.L. and was further supported by the Jornvall Foundation and the Swedish Research Council (to P.A.).

DATA ACCESSIBILITY

NGS sequences produced in this study were deposited at the NCBI Sequence Read Archive (SRA) under Bioproject PRJNA352289.

AUTHOR CONTRIBUTIONS

The study program was conceived by F.L, and the experiment was designed by D.Z. and F.L. The sample collection was carried out by D.Z., S.S. and Y.W. The data were analysed by D.Z., B.G. and G.S., and data analysis was assisted by Y.C. and S.S. The manuscript was written by D.Z., G.S. and F.L. The manuscript was revised and commented by Y.Q., S.W. and P.A.

ORCID

Dezhi Zhang http://orcid.org/0000-0003-1523-3168
Per Alström https://orcid.org/0000-0001-7182-2763

REFERENCES

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*, 81–92. https://doi.org/10.1038/nrg.2015.28

Arnold, B., Corbett-Detig, R. B., Hartl, D., & Bomblies, K. (2013). RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, *22*, 3179–3190. https://doi.org/10.1111/mec.12276

Backström, N., Karaiskou, N., Leder, E. H., Gustafsson, L., Primmer, C. R., Qvarnstrom, A., & Ellegren, H. (2008). A gene-based genetic linkage map of the collared flycatcher (*Ficedula albicollis*) reveals extensive synteny and gene-order conservation during 100 million years of avian evolution. *Genetics*, *179*, 1479–1495. https://doi.org/10.1534/genetics.108.088195

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., … Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, *3*, e3376. https://doi.org/10.1371/journal.pone.0003376

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. *The American Journal of Human Genetics*, *81*, 1084–1097. https://doi.org/10.1086/521987

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., … 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*, 1656–1665. https://doi.org/10.1101/gr.196485.115

Carneiro, M., Albert, F. W., Afonso, S., Pereira, R. J., Burbano, H., Campos, R., … Ferrand, N. (2014). The genomic architecture of population divergence between subspecies of the European rabbit. *PLoS Genetics*, *10*, e1003519. https://doi.org/10.1371/journal.pgen.1003519

Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, *15*, 538–543. https://doi.org/10.1093/oxfordjournals.molbev.a025953

Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, *70*, 155–174. https://doi.org/10.1017/S0016672397002954

Clarkson, C. S., Weetman, D., Essandoh, J., Yawson, A. E., Maslen, G., Manske, M., … Donnelly, M. J. (2014). Adaptive introgression between *Anopheles* sibling species eliminates a major genomic island but not reproductive isolation. *Nature Communications*, *5*, 4248. https://doi.org/10.1038/ncomms5248

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sunderland, MA: Sinauer Associates.

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*, 3133–3157. https://doi.org/10.1111/mec.12796

Currat, M., Ruedi, M., Petit, R. J., & Excoffier, L. (2008). The hidden side of invasions: Massive introgression by local genes. *Evolution*, *62*, 1908–1920. https://doi.org/10.1111/j.1558-5646.2008.00413.x

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution*, *24*, 332–340. https://doi.org/10.1016/j.tree.2009.01.009

del Hoyo, J., Elliott, A., & Christie, D. (2008). *Handbook of the Birds of the World. Vol. 13: Penduline Tits to Shrikes*. Barcelona, Spain: Lynx Edicions.

Delmore, K. E., Hubner, S., Kane, N. C., Schuster, R., Andrew, R. L., Câmara, F., … 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*, 1873–1888. https://doi.org/10.1111/mec.2015.24.issue-8

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, *28*, 2239–2252. https://doi.org/10.1093/molbev/msr048

Ellegren, H. (2010). Evolutionary stasis: The stable chromosomes of birds. *Trends in Ecology and Evolution*, *25*, 283–291. https://doi.org/10.1016/j.tree.2009.12.004

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., … Wolf, J. B. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, *491*, 756–760. https://doi.org/10.1038/nature11584

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, *28*, 342–350. https://doi.org/10.1016/j.tig.2012.03.009

Feder, J. L., & Nosil, P. (2010). The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*, *64*, 1729–1747. https://doi.org/10.1111/j.1558-5646.2009.00943.x

Flaxman, S. M., Feder, J. L., & Nosil, P. (2013). Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. *Evolution*, *67*, 2577–2591. https://doi.org/10.1111/evo.2013.67.issue-9

Funk, D. J., & Omland, K. E. (2003). Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, *34*, 397–423. https://doi.org/10.1146/annurev.ecolsys.34.011802.132421

Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdelhué, C., Pudlo, P., … Estoup, A. (2013). The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology*, *22*, 3165–3178. https://doi.org/10.1111/mec.12089

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., … Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science*, *328*, 710–722. https://doi.org/10.1126/science.1188021

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, *5*, e1000695. https://doi.org/10.1371/journal.pgen.1000695

Harr, B. (2006). Genomic islands of differentiation between house mouse subspecies. *Genome Research*, *16*, 730–737. https://doi.org/10.1101/gr.5045006

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., … Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*, 55–61. https://doi.org/10.1038/nature10944

Kronforst, M. R., Hansen, M. E., Crawford, N. G., Gallant, J. R., Zhang, W., Kulathinal, R. J., … Mullen, S. P. (2013). Hybridization reveals the evolving genomic architecture of speciation. *Cell Reports*, *5*, 666–677. https://doi.org/10.1016/j.celrep.2013.09.042

Laine, V. N., Gossmann, T. I., Schachtschneider, K. M., Garroway, C. J., Madsen, O., Verhoeven, K. J., … Groenen, M. A. (2016). Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nature Communications*, *7*, 10474. https://doi.org/10.1038/ncomms10474

Lawniczak, M. K. N., Emrich, S. J., Holloway, A. K., Regier, A. P., Olson, M., White, B., … Besansky, N. J. (2010). Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science*, *330*, 512–514. https://doi.org/10.1126/science.1195755

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., … Turner, G. F. (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*, *350*, 1493–1498. https://doi.org/10.1126/science.aac9927

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., … Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, *23*, 1817–1828. https://doi.org/10.1101/gr.159426.113

Martin, S. H., Davey, J. W., & Jiggins, C. D. (2015). Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Molecular Biology and Evolution*, *32*, 244–257. https://doi.org/10.1093/molbev/msu269

Nachman, M. W., & Payseur, B. A. (2012). Recombination rate variation and speciation: Theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*, 409–421. https://doi.org/10.1098/rstb.2011.0249

Noor, M. A., & Bennett, S. M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, *103*, 439–444. https://doi.org/10.1038/hdy.2009.151

Nosil, P. (2008). Speciation with gene flow could be common. *Molecular Ecology*, *17*, 2103–2106. https://doi.org/10.1111/mec.2008.17.issue-9

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, *18*, 375–402. https://doi.org/10.1111/mec.2009.18.issue-3

Päckert, M., Martens, J., & Sun, Y. H. (2010). Phylogeny of long-tailed tits and allies inferred from mitochondrial and nuclear markers (Aves: Passeriformes, Aegithalidae). *Molecular Phylogenetics and Evolution*, *55*, 952–967. https://doi.org/10.1016/j.ympev.2010.01.024

Päckert, M., Martens, J., Sun, Y.-H., Severinghaus, L. L., Nazarenko, A. A., Ting, J., … Tietze, D. T. (2012). Horizontal and elevational phylogeographic patterns of Himalayan and Southeast Asian forest passerines (Aves: Passeriformes). *Journal of Biogeography*, *39*, 556–573. https://doi.org/10.1111/jbi.2012.39.issue-3

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, *25*, 2337–2360. https://doi.org/10.1111/mec.13557

Pennisi, E. (2014). Disputed islands. *Science*, *345*, 611–613. https://doi.org/10.1126/science.345.6197.611

Petit, R. J., & Excoffier, L. (2009). Gene flow and species delimitation. *Trends in Ecology and Evolution*, *24*, 386–393. https://doi.org/10.1016/j.tree.2009.02.011

Poelstra, J. W., Vijay, N., Bossu, C. M., Lantz, H., Ryll, B., Müller, I., … Wolf, J. B. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*, *344*, 1410–1414. https://doi.org/10.1126/science.1253226

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., … Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*, 559–575. https://doi.org/10.1086/519795

R Development Core Team (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0. Retrieved from http://www.R-project.org

Renaut, S., Grassa, C. J., Yeaman, S., Moyers, B. T., Lai, Z., Kane, N. C., … Rieseberg, L. H. (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications*, *4*, 1827. https://doi.org/10.1038/ncomms2833

Rheindt, F. E., & Edwards, S. V. (2011). Genetic introgression: An integral but neglected component of speciation in birds. *The Auk*, *128*, 620–632. https://doi.org/10.1525/auk.2011.128.4.620

Rheindt, F. E., Fujita, M. K., Wilton, P. R., & Edwards, S. V. (2014). Introgression and phenotypic assimilation in *Zimmerius* flycatchers (Tyrannidae): Population genetic and phylogenetic inferences from genome-wide SNPs. *Systematic Biology*, *63*, 134–152. https://doi.org/10.1093/sysbio/syt070

Roesti, M., Hendry, A. P., Salzburger, W., & Berner, D. (2012). Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology*, *21*, 2852–2862. https://doi.org/10.1111/j.1365-294X.2012.05509.x

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*, 4757–4769. https://doi.org/10.1111/mec.12842

Shafer, A. B. A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., & Wolf, J. B. W. (2016). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, *8*, 907–917. https://doi.org/10.1111/2041-210X.12700

Smith, J., & Kronforst, M. R. (2013). Do *Heliconius* butterfly species exchange mimicry alleles? *Biology Letters*, *9*, 20130503. https://doi.org/10.1098/rsbl.2013.0503

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Tang, H., Peng, J., Wang, P., & Risch, N. J. (2005). Estimation of individual admixture: Analytical and study design considerations. *Genetic Epidemiology*, *28*, 289–301. https://doi.org/10.1002/(ISSN)1098-2272

Tine, M., Kuhl, H., Gagnaire, P. A., Louro, B., Desmarais, E., Martins, R. S., … Reinhardt, R. (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications*, *5*, 5770. https://doi.org/10.1038/ncomms6770

Turner, T. L., & Hahn, M. W. (2010). Genomic islands of speciation or genomic islands and speciation? *Molecular Ecology*, *19*, 848–850. https://doi.org/10.1111/mec.2010.19.issue-5

Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, *3*, e285. https://doi.org/10.1371/journal.pbio.0030285

Via, S. (2012). Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical*

Transactions of the Royal Society of London. Series B, Biological Sciences, *367*, 451–460. https://doi.org/10.1098/rstb.2011.0260

Vijay, N., Bossu, C. M., Poelstra, J. W., Weissensteiner, M. H., Suh, A., Kryukov, A. P., & Wolf, J. B. (2016). Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature Communications*, *7*, 13195. https://doi.org/10.1038/ncomms13195

Wang, W., Dai, C., Alström, P., Zhang, C., Qu, Y., Li, S. H., ... Lei, F. (2014). Past hybridization between two East Asian long-tailed tits (*Aegithalos bonvaloti* and *A. fuliginosus*). *Frontiers in Zoology*, *11*, 40. https://doi.org/10.1186/1742-9994-11-40

Wu, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, *14*, 851–865. https://doi.org/10.1046/j.1420-9101.2001.00335.x

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*, 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, *18*, 821–829. https://doi.org/10.1101/gr.074492.107

Zhang, W., Dasmahapatra, K. K., Mallet, J., Moreira, G. R., & Kronforst, M. R. (2016). Genome-wide introgression among distantly related *Heliconius* butterfly species. *Genome Biology*, *17*, 25. https://doi.org/10.1186/s13059-016-0889-0

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Zhang D, Song G, Gao B, et al. Genomic differentiation and patterns of gene flow between two long-tailed tit species (*Aegithalos*). *Mol Ecol*. 2017;26:6654–6665. https://doi.org/10.1111/mec.14383