DR. DREW ROBERTS SCHIELD (Orcid ID : 0000-0001-7854-9480)

MR. CHRIS CR SMITH (Orcid ID : 0000-0002-6470-3413)

DR. WEI LIANG (Orcid ID : 0000-0002-0004-9707)

DR. YANG LIU (Orcid ID : 0000-0003-4580-5518)

MISS SHEELA PHANSALKAR TURBEK (Orcid ID : 0000-0003-1080-9375)

DR. MATTHEW REED WILKINS (Orcid ID : 0000-0002-3162-6749)

Article type : Original Article

Sex-linked genetic diversity and differentiation in a globally distributed avian species complex

Drew R. Schield[1, §], Elizabeth S. C. Scordato[1,2], Chris C. R. Smith[1], Javan K. Carter[1], Sidi Imad Cherkaoui[3], Sundev Gombobaatar[4], Said Hajib[5,†], Saad Hanane[5], Amanda K. Hund[1,6], Kazuo Koyama[7], Wei Liang[8], Yang Liu[9], Najib Magri[5], Alexander Rubtsov[10], Basma Sheta[11], Sheela P. Turbek[1], Matthew R. Wilkins[1,12], Liu Yu[13], and Rebecca J. Safran[1]

Affiliations:

1. Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, USA

2. Department of Biological Sciences, California State Polytechnic University, Pomona, CA USA

3. Ecole Supérieure de Technologie de Khénifra, Sultan Moulay Slimane University, Béni-Mellal, Morocco

4. National University of Mongolia and Mongolian Ornithological Society, Ulaanbaatar, Mongolia

5. Water and Forests Department, Forest Research Center, Rabat-Agdal, Morocco

6. Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN USA

7. Japan Bird Research Association, Tokyo, Japan

8. Ministry of Education Key Laboratory for Ecology of Tropical Islands, Key Laboratory of Tropical Animal and Plant Ecology of Hainan Province, College of Life Sciences, Hainan Normal University, Haikou, China

9. State Key Laboratory of Biocontrol, School of Ecology, Sun Yat-sen University, Guangzhou, Guangdong, China

10. State Darwin Museum, Moscow, Russia

11. Zoology Department Faculty of Science, Damietta University, New Damietta City, Egypt

13. Collaborative for STEM Education and Outreach, Vanderbilt University, Nashville, TN USA

14. Key Laboratory for Biodiversity Sciences and Ecological Engineering, Ministry of Education, College of Life Sciences, Beijing Normal University, Beijing, China

[†]Deceased

[§]To whom correspondence should be addressed: Drew Schield, Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO 80309 USA. *Email:* drew.schield@colorado.edu *phone:* 763-498-4271

Running title: Z-linked genetic variation in barn swallows

Keywords: barn swallows, *Hirundo*, population genomics, selection, sex chromosomes, speciation

Abstract

Sex chromosomes often bear distinct patterns of genetic variation due to unique patterns of inheritance and demography. The processes of mutation, recombination, genetic drift, and selection also influence rates of evolution on sex chromosomes differently than autosomes. Measuring such differences provides information about how these processes shape genomic variation and their roles in the origin of species. To test hypotheses and predictions about patterns of autosomal and sex-linked genomic diversity and differentiation, we measured population genetic statistics within and between populations and subspecies of the barn swallow (*Hirundo rustica*) and performed explicit comparisons between autosomal and Z-linked genomic regions. We first tested for evidence of low Z-linked genetic diversity and high Z-linked population differentiation relative to autosomes, then for evidence that the Z chromosome bears greater ancestry information due to faster lineage sorting. Finally, we investigated geographic clines across hybrid zones for evidence that the Z chromosome is resistant to introgression due to selection against hybrids. We found evidence that the barn swallow mating system, demographic history, and linked selection each contribute to low Z-linked diversity and high Z-linked differentiation. While incomplete lineage sorting is rampant across the genome, our results indicate faster sorting of ancestral polymorphism on the Z. Finally, hybrid zone analyses indicate barriers to introgression on the Z chromosome suggesting that sex-linked traits are important in reproductive isolation, especially in migratory divide regions. Our study highlights how selection, gene flow, and demography shape sex-linked genetic diversity and underscores the relevance of the Z chromosome in speciation.

1. Introduction

Regions of the genome diverge at different rates during speciation, leading to the commonly observed pattern of heterogeneity in genomic differentiation between populations and species (Wu 2001; Hohenlohe et al. 2010; Ellegren et al. 2012; Martin et al. 2013; Poelstra et al. 2014; Wolf and Ellegren 2017). Indeed, the genomic landscape of population differentiation is often strikingly non-uniform, with ‘peaks’ and ‘valleys’ of divergence across the genome, suggesting that a subset of genomic regions are involved in reproductive isolation between populations. Sex chromosomes are important in this regard because differences in effective population size, recombination rate, gene flow, and the efficacy of selection on sex chromosomes relative to autosomes may culminate in a more rapid accumulation of sex-linked divergence during speciation (Qvarnström and Bailey 2009). Comparisons between autosomes and sex chromosomes therefore provide a natural experiment to examine how evolutionary processes shape genetic diversity across the genome

(Bachtrog et al. 2011; Ellegren 2011). Further, sex chromosomes are often assumed to play a prominent role in the formation and maintenance of species, and empirical investigation into sex-linked reproductive isolation is necessary to evaluate this assumption.

Sex chromosomes often exhibit low within-population genetic diversity and high between-population differentiation relative to autosomes driven by demographic effects, the influence of low recombination rate and linked selection, reduced gene flow, or a combination of these factors (Aquadro et al. 1994; Charlesworth 1996; Betancourt et al. 2004; Mank et al. 2009; Burri et al. 2015). In some cases, sex chromosomes themselves are considered genomic 'islands' of speciation (e.g., Ellegren et al. 2012; Martin et al. 2013), or are enriched for peaks of population differentiation (e.g., Poelstra et al. 2014; Toews et al. 2016). This trend is consistent with the general expectation that sex chromosomes play a disproportionate role in speciation for several reasons. First, sex chromosomes have distinct patterns of inheritance from autosomes. While each individual in a diploid sexual population inherits two copies of each autosome, the inheritance of sex chromosomes varies depending on sex determination system. In species with female heterogamety, males (ZZ) inherit a copy of the Z chromosome from both parents, while females (ZW) inherit only the patrilineal Z. Further, the female-specific W chromosome is only inherited matrilineally. These inheritance patterns result in distinct demography compared to autosomes. Under equilibrium conditions, Z and W chromosomes will be present in a population at 3/4 and 1/4 frequency relative to autosomes (Caballero 1995; Charlesworth 2001; Pool and Nielsen 2007; Mank et al. 2010), and are therefore expected to harbor proportionately lower genetic diversity. However, variance in reproductive success between the sexes, sex-biased mutation rates, and variation in the strength of selection on sex chromosomes may also impact standing genetic variation. Finally, sex chromosomes encode genes and regulators underlying sexually dimorphic traits (Rice 1984), traits involved in species recognition (Sæther et al. 2007), traits subject to sexual selection (Corl and Ellegren 2012), and traits that confer a fitness advantage to one sex over the other (e.g., sexually antagonistic alleles; Charlesworth and Charlesworth 1978). Accordingly, beyond their role in sex determination, sex chromosomes are predicted to have special relevance in the origins of assortative mating and reproductive isolation in a wide array of taxa (Charlesworth et al. 1987; Albert and Otto 2005; Sæther et al. 2007; Presgraves 2008).

Explicit comparisons between sex-linked and autosomal genetic diversity can reveal features of sex chromosome evolution relevant to reproductive isolation, and can also provide information regarding the relative strengths of different evolutionary forces in shaping genetic diversity across the genome (Hellborg and Ellegren 2004). Bird species have ZW sex determination and have become important model systems for

empirical studies of autosomal and sex-linked genetic diversity. These studies collectively support that low Z-linked genetic diversity and high relative differentiation on the Z are ubiquitous features of avian speciation (e.g., Ellegren et al. 2012; Burri et al. 2015; Lavretsky et al. 2015; Irwin et al. 2016; Battey 2020; Wang et al. 2020; reviewed in Irwin 2018). However, the precise mechanisms driving this pattern and resultant ratios of Z chromosome/autosome (Z/A) differentiation differ widely among species (reviewed in Irwin 2018). Theory predicts that alleles associated with genetic sex, reproduction, and mate selection will accumulate on the sex chromosomes (Rice 1984; Kirkpatrick and Hall 2004b). Consistent with this prediction, Z-linked alleles underlying sexually selected plumage coloration have been identified in several bird species (Sæther et al. 2007; Toews et al. 2016; Campagna et al. 2017). Birds also exhibit a diversity of mating systems with variation in male or female reproductive success; biases in sex-specific reproductive success can further skew relative sex-linked genetic diversity (Charlesworth 2001). Male-biased mutation rates and associated rates of Z chromosome evolution (i.e., 'fast-Z' or 'large-Z' effects resulting from a faster rate of adaptive evolution or greater genetic drift on the Z chromosome than autosomes) have also been identified in diverse bird species (Ellegren 2009; Mank et al. 2009; Elgvin et al. 2011; Dean and Mank 2014; Wang et al. 2014; Lavretsky et al. 2015; Battey 2020). These factors, among others, may be major determinants of Z-linked population genetic diversity and differentiation.

Among bird species, groups of closely related lineages with geographic variation in sexually selected traits may be particularly valuable, providing information on sex-linked and autosomal genetic diversity at the earliest stages of speciation. An example of such a system is the barn swallow (*Hirundo rustica*) species complex. Barn swallows are nearly globally distributed and comprise six subspecies (*Hirundo rustica rustica*, *H. r. savignii*, *H. r. transitiva*, *H. r. gutturalis*, *H. r. tytleri*, and *H. r. erythrogaster*; for brevity we will refer to populations using subspecies names only). The six subspecies occupy breeding ranges throughout the Northern Hemisphere (Fig. 1A; del Hoyo and Elliott 2014) and are estimated to have diverged from a common ancestor less than one million years ago (Zink et al. 2006; Dor et al. 2010). Subspecies differ in combinations of traits involved in sexual signaling (i.e., plumage coloration and tail streamer lengths; Møller 1994; Safran et al. 2005, 2016; Vortman et al. 2013; Liu et al. 2018), and there is evidence of geographic variation in the strength of sexual selection on these traits (Romano et al. 2017a,b; reviewed in Scordato and Safran 2014). There is also variation in migratory behavior among barn swallows, with two non-migratory subspecies (*savignii* in Egypt and *transitiva* in Israel) and other subspecies that migrate over long distances to overwinter in lower latitude regions of Sub-Saharan Africa, Asia, and South America (del Hoyo and Elliott 2014). Barn swallows are socially monogamous, yet extra-pair copulation is common (Smith et al. 1991; Møller and Tegelström 1997;

Kleven et al. 2006). This mating system may lead to excess variance in male reproductive success, and therefore reduce relative Z-linked genetic diversity below theoretically predicted equilibrium ratios (i.e., Z/A < 3/4; Fig. 1B-C). Barn swallow populations have also experienced very recent bottlenecks and subsequent expansions during the Holocene (Smith et al. 2018). Because sex chromosomes are hypothesized to equilibrate to demographic conditions more rapidly than autosomes (Pool and Nielsen 2007), recent bottlenecks may be a source of further decreases in relative genetic diversity on the Z chromosome. These combined aspects of barn swallow ecology and demographic history provide a foundation for hypothesis testing related to sex-linked and autosomal genetic diversity.

If rapid differentiation of sex chromosomes is relevant to reproductive isolation between lineages, it follows that sex-linked genomic regions will also be more resistant to introgression in hybrid zones (Payseur and Rieseberg 2016). Previous studies have characterized hybrid zone transects between the Asian barn swallow subspecies *rustica, tytleri,* and *gutturalis* in regions of Russia, Mongolia, and China (Scordato et al. 2017, 2020). These hybrid zones vary both in geographic width and apparent degree of reproductive isolation between parental lineages. Scordato et al. (2020) also found variation in geographic clines of ventral plumage color (a trait associated with mate choice in certain subspecies) and carbon isotope values (associated with non-breeding location and the potential presence of migratory divides) across transects, further supporting variation in the degree of assortative mating and reproductive isolation in barn swallow secondary contact zones. Sampling of individuals from these transects affords an opportunity to test the hypothesis that the Z chromosome is resistant to introgression due to selection against hybrids with mixtures of parental traits.

In this study, we estimate genomic diversity and differentiation in the barn swallow species complex using a whole genome resequencing dataset, including samples of all subspecies and hybrid zone transects in Asia. We make explicit comparisons between patterns on the Z chromosome and autosomes to test theoretical and empirical predictions for Z-linked genetic diversity and differentiation and to understand the role of the Z chromosome in speciation (Fig. 1D; Table 1). In the absence of reproductive skew, demographic changes, or selection, genetic diversity on the Z should be roughly 3/4 of autosomal diversity due to genetic drift, after accounting for sex-biased mutation rates (black lines in Fig. 1B-C; Charlesworth 2001). Excess variance in reproductive success in either sex can cause deviations from this expectation (Irwin 2018), and species with greater variance in male reproductive success (e.g., barn swallows; Møller 1994) may experience reductions in the effective population size of the Z chromosome, specifically. That is, if a relatively small number of males reproduce, this will necessarily reduce population-level variation on the Z chromosome (of which males have

two copies), and to a lesser degree for autosomes (i.e., two copies are in both sexes). The theoretical minimum ratio for Z-to-autosomal diversity under extreme cases of variance in male reproductive success is equal to 0.56 (orange shaded areas in Fig. 1B-C; Table 1; Charlesworth 2001). Such values might be observed if, for example, a single male is responsible for all offspring as a consequence of mating with many females in the absence of other factors that would influence the transmission of the Z chromosome (Vicoso and Charlesworth 2009). We consider genomic patterns in barn swallows in light of these predictions, and how sex-linked diversity may be shaped by a combination of factors related to reproductive skew, demography, selection, and gene flow. We then test the hypothesis that the Z chromosome represents a genomic 'differentiation island' in barn swallows. The combined effects of faster lineage sorting and reduced gene flow on the Z chromosome due to selection may drive an increased signal of genetic structure between lineages, and we test the hypothesis that Z-linked diversity and differentiation are products of rapid lineage sorting relative to rates on autosomes. Finally, we test the hypothesis that reduced introgression on the Z chromosome due to selection against hybrids has produced excess population differentiation between barn swallow lineages, indicating a role for Z-linked regions in speciation.

2. Materials and Methods

2.1 Sampling design and genome resequencing

We obtained blood samples from each of the six barn swallow subspecies (n = 160, range = 8–34 individuals per subspecies; Supplementary Table S1). Multiple localities were sampled for wide-ranging subspecies (e.g., *rustica*, *tytleri*, and *gutturalis*). Our sampling also included individuals from previously identified hybrid zones in Asia (Scordato et al. 2017, 2020), and a single Wire-tailed Swallow (*H. smithii*) for use as an outgroup, based on the swallow phylogeny estimated in Sheldon et al. (2005). Blood was obtained using medial metatarsal venipuncture and stored in DNA lysis buffer. DNA was extracted using Qiagen DNeasy kits following the manufacturer's protocol. Purified DNA products were quantitated using a Qubit fluorometer and genomic libraries were prepared for each sample using Illumina Nextera XT kits at the University of Colorado Boulder BioFrontiers Institute using sample-specific barcodes. Prepared libraries were sequenced on two Illumina NovaSeq 6000 lanes using 150 bp paired end reads. The raw sequencing data used in this study are available on the NCBI short-read archive (accession PRJNA323498).

2.2 Variant calling and filtering

Paired read data for each individual from the two sequencing lanes were concatenated, then quality filtered using Trimmomatic v.0.39 (Bolger et al. 2014). Bases at the 5' and 3' ends of reads with a quality score below 20 were removed using the LEADING:20 and TRAILING:20 options. We also removed reads shorter than 32 bp (MINLEN:32) or with an average quality score below 30 (AVGQUAL:30). We mapped the filtered read data to the European barn swallow (*Hirundo rustica rustica*) reference genome (Formenti et al. 2019) using BWA 'mem' v.0.7.17 (Li and Durbin 2009) with default settings. We then sorted output BAM files and quantified mapping statistics using the Samtools v.1.10 (Li et al. 2009) 'sort' and 'stat' tools, respectively.

After mapping read data to the reference genome, we called genomic variants using the GATK v.4.0.8.1 best-practices workflow (McKenna et al. 2010; Van der Auwera et al. 2013). To facilitate analyses requiring an outgroup, we called variants among ingroup samples plus *H. smithii*. We first used the 'HaplotypeCaller' tool to genotype individuals, and specified '--ERC GVCF ' to generate a genomic VCF per sample, then generated an 'all-sites' VCF, calling variant sites across individuals using the 'GenotypeGVCFs' tool. We filtered raw variant calls using the GATK 'VariantFiltration' tool, imposing the following filters on SNPs using information in the site format fields: variant confidence by depth (QD < 2.0), strand-bias (FS > 60.0), among sample mapping quality (MQ < 40.0), mapping quality of heterozygous sites (MQRankSum < -12.5), and distance of variant sites from ends of reads (ReadPosRankSum < -8.0), following the GATK tutorial on hard-filtering variants and settings used in Hooper et al. (2019). We also used 'VariantFiltration' to mask SNPs in repeat regions from the barn swallow genome after converting the NCBI GFF version 'GCA_003692655.1_Chelidonia_genomic.gff' to BED format and sorting using BEDtools (Quinlan and Hall 2010). We further identified any spurious female heterozygous sites on Z-linked scaffolds, and conservatively recoded them as missing genotypes in all individuals. We removed all indels and any SNP variants that did not pass these filter settings using BCFtools v1.10.2 (Li et al. 2009). We used VCFtools v.0.1.17 (Danecek et al. 2011) to remove singleton variants (--mac 2), non-biallelic variants (--min-alleles 2, --max-alleles 2), SNPs with a minor allele frequency lower than 0.05 (--maf 0.05), sites not meeting various missing data thresholds after hard filtering (--max-missing 0.2, 0.4, and 0.6), and any individuals missing greater than 75% of genotypes after filtering steps. Singleton and minor-allele frequency filters were applied after removing sites genotyped in fewer than 60% of individuals (i.e., --max-missing 0.4).

The barn swallow reference genome includes large scaffolds (N50 = 26 Mbp) that are not localized or ordered according to their orientation within chromosomes. To enable comparisons between the Z chromosome and

autosomes, and to order scaffolds for genome scans of population genetic summary statistics, we aligned the barn swallow genome to chromosome-assigned scaffolds of the collared flycatcher (*Ficedula albicollis*) genome version FicAlb1.5 (Ellegren et al. 2012) using MashMap v.2.0 (Jain et al. 2017, 2018). We specified a required mapping segment length of 10 kb for barn swallow scaffolds shorter than one Mbp, and a mapping segment length of 50 kb for scaffolds greater than one Mbp. MashMap outputs many short alignment segments between query and reference sequences, and we assigned barn swallow scaffolds to chromosomes if: 1) more than half of the alignment segments per scaffold aligned to the same chromosomes, and 2) more than half of the total scaffold length aligned to the same chromosome. All other scaffolds were excluded from analysis. We approximated the start position for chromosome-assigned barn swallow scaffolds by subtracting half of the scaffold length from the median alignment position among short alignments to the assigned chromosome. We assigned ~ 1.01 Gbp (91%) of the barn swallow genome to Collared Flycatcher chromosomes using this procedure (Supplementary Fig. S1).

2.3 Genetic diversity within populations: does reproductive skew explain Z chromosome diversity?

We tested the hypothesis that Z-linked diversity in barn swallows matches theoretical predictions in the absence of reproductive skew, demographic changes, or selection by measuring nucleotide diversity, or the average pairwise differences between chromosomes in a population (π), across the genome. We then calculated the ratio of Z chromosome-to-autosome π (π_{ZA}) within lineages. We measured π in non-overlapping 100 kilobase (kb) genomic windows using pixy (Korunes and Samuk 2020), using the 'all-sites' VCF as input and applying depth filters to all sites (DP>3). We calculated the mean π value for autosomes and the Z chromosome after dividing the mean Z value by 1.1, a generalized avian Z:autosome mutation rate (μ) ratio (Axelsson et al. 2004; Wang et al. 2014; Oyler-McCance et al. 2015; Irwin 2018), to account for the effects of male mutation bias in comparisons of autosomal and Z-linked diversity. We performed these calculations in each of the six described subspecies and each of the three sampled hybrid zones. We repeated these analyses on a 'male-only' dataset to confirm that including females did not strongly influence the results. Three subspecies (*rustica*, *gutturalis*, and *tytleri*) were sampled from multiple localities (Supplementary Table S1), and we calculated π for each locality represented by at least two individuals (25 localities total). We compared distributions of π on autosomes and the Z chromosome using Welch's two-sample *t*-tests and examined relationships between Z-linked and autosomal π, and between π_{ZA} and the mean value of Z-linked and autosomal π after accounting for male-biased mutation.

To determine whether heterogeneity in genetic diversity corresponded with skews in the allele frequency spectrum and associated departures from neutral expectations (e.g., shifts towards an excess or scarcity of rare alleles), we measured Tajima's *D* (Tajima 1989) in 100 kb sliding windows using VCFtools (Danecek et al. 2011). Because minor allele frequency and singleton filters will lead to an underestimate of the number of segregating sites in a given population, we used variant calls prior to implementing these filters for Tajima's *D* analyses. Small sample sizes have also been shown to influence Tajima's *D* estimates (Tennessen et al. 2012; Subramanian 2016), so we limited these analyses to each subspecies and hybrid zone (minimum n = 8), and did not analyze specific localities. We ran analyses on the male-only dataset and on the full dataset with three missing data thresholds (--max-missing 0.2, --max-missing 0.4, and --max-missing 0.6) to confirm that results were robust to the inclusion of females and missing data.

2.4 Population differentiation: does the Z chromosome exhibit high relative differentiation?

Measures of relative population differentiation, such as *Fst*, are sensitive to population-level processes that affect allele frequencies. Accordingly, differentiation islands may evolve as a consequence of multiple diversity-reducing scenarios of genomic divergence, including positive selection, background selection, and divergence hitchhiking, where regions experiencing gene flow will evolve lower differentiation (Burri 2017). Differences in autosomal and sex-linked genetic diversity are also predicted to correspond to different rates of lineage sorting and fixation of derived alleles (e.g., Pease and Hahn 2013). We tested the hypotheses that genomic variation in within-population genetic diversity is associated with relative allelic differentiation and that the Z chromosome represents an 'island' of differentiation by measuring pairwise Weir and Cockerham's *Fst* (Weir and Cockerham 1984) between barn swallow lineages. Here, we would predict a negative relationship between genetic diversity (π) and relative population differentiation if linked selection is largely responsible for shaping variation across the genome. *Fst* was measured from biallelic SNPs with minor-allele frequencies >= 0.05 in 100 kb sliding windows across the genome using VCFtools. We performed all pairwise comparisons between subspecies in order to investigate relative differentiation across a continuum of lineage divergence. *Fst* comparisons were not phylogenetically independent, however, so in order to evaluate lineage-specific allele frequency change we calculated population branch statistics (*PBS*; Shriver et al. 2004; Yi et al. 2010) for each subspecies. We also measured relative differentiation between each hybrid zone and its respective parental subspecies (e.g., *rustica* x *gutturalis* versus *rustica* and versus *gutturalis*) using windowed *Fst* estimates. We examined relationships between differentiation statistics and π estimates using Spearman rank order correlation coefficients, and tested for differences in autosomal and Z-linked relative differentiation using two-sample Mann-Whitney $\cup$ tests.

2.5 Genetic structure: evidence for greater population structure on the Z chromosome?

To test the hypothesis that the Z chromosome bears a greater signal of genetic structure between lineages than autosomes, we examined population genetic structure among barn swallow populations and subspecies using several approaches. Because tight physical linkage of SNPs can bias estimates of population structure, we first thinned our variant dataset to remove ingroup SNPs with minor-allele frequency >= 0.05 within 100 bp of one another by sampling one SNP at random. We then performed principal component analysis (PCA) on autosomal, Z-linked, and combined SNP datasets using the SNPRelate package v.1.22.0 (Zheng et al. 2012) in R (R Core Team 2017). We then used the likelihood-model approach implemented in ADMIXTURE (Alexander et al. 2009) to estimate the number of genetic clusters in each SNP dataset. We used Plink v.1.90 (Purcell et al. 2007) to convert SNP data into ADMIXTURE input format, then ran ADMIXTURE on each dataset across a series of 1-16 *K* cluster values with 200 bootstrap replicates per run. We evaluated the amount of variation explained by various *K* values using the cross-validation procedure.

2.6 Phylogenetic analyses: faster lineage sorting and reduced gene flow on the Z chromosome?

To quantify variation in rates of lineage sorting (and potential variation in gene flow) across the genome, we examined support for alternative topologies for four sets of barn swallow subspecies triplets using the topology weighting procedure described in Martin and Van Belleghem (2017). The bifurcation of the barn swallow species complex into two groups consisting of European versus Asian and North American subspecies is well supported, while relationships within each group are unresolved (Zink et al. 2006; Dor et al. 2010). We therefore designed triplet topologies to examine support at two phylogenetic levels, first testing relationships in the two unresolved clades with shallow divergence (triplet 1: *rustica*, *savignii*, and *transitiva*; triplet 2: *erythrogaster*, *tytleri*, *gutturalis*), then testing support for relationships spanning the primary split between subspecies groups (triplet 3: *savignii*, *transitiva*, and *erythrogaster*; triplet 4: *erythrogaster*, *tytleri*, and *savignii*). The phylogenetic uncertainty in each shallow triplet provided an opportunity to test whether one topology was supported at higher frequency across the genome, or on the Z chromosome specifically. We had strong *a priori* expectations for the best-supported topologies in triplets 3 and 4, allowing for further comparisons of support for the major split between European and Asian/North American subspecies between autosomes and the Z chromosome. Signal of strong isolation by distance has been observed in barn swallows (Safran et al. 2016a). We therefore limited our sampling of wide-ranging subspecies to localities in the core of their ranges (i.e., *gutturalis* from Changsha, China and *rustica* from Moscow, Russia). Each of these localities

was represented by three individuals, so we randomly sampled three individuals of other subspecies in each triplet. For each triplet, we examined genomic support for alternative topologies by first estimating neighbor-joining gene trees in sliding windows of 1,000 SNPs using the GTR model in Phyml v.3.3.2 (Guindon et al. 2010) implemented in the Python script 'phyml_sliding_windows.py' (https://github.com/simonhmartin/genomics_general), assigning *H. smithii* as the outgroup for all analyses. We then measured relative support for alternative topologies in each triplet using TWISST (https://github.com/simonhmartin/twisst), which quantifies topology weights by sampling subtrees within the taxon tree (Martin and Van Belleghem 2017). We used the 'complete' weighting method in TWISST, which determines the exact weight of alternative topologies across all subtrees. We compared the proportions of subtrees supporting alternative topologies for autosomes and the Z chromosome and compared support for specific topologies using Mann-Whitney $\cup$ tests in R (Core Team 2017).

To further disentangle whether reduced gene flow on the Z chromosome also contributes to its relative diversity, we calculated introgression statistics from genealogies using the ABBA-BABA framework. We tested whether rates of allele sharing due to introgression, measured using the $\hat{f}_d$ statistic (Martin et al. 2015), differed for autosomes and the Z chromosome between parental and hybrid populations in Asia. Here, we would predict lower $\hat{f}_d$ values on the Z chromosome if it has reduced gene flow relative to autosomes. We performed three analyses, testing the topologies (P1: *savignii*, P2: *rustica*, P3: *rustica* x *tytleri*, O: *H. smithii*), (P1: *savignii*, P2: *rustica*, P3: *rustica* x *gutturalis*, O: *H. smithii*), and (P1: *erythrogaster*, P2: *gutturalis*, P3: *gutturalis* x *tytleri*, O: *H. smithii*). We calculated the frequency of 'ABBA' and 'BABA' patterns, Patterson's *D* statistics (Durand et al. 2011), and $\hat{f}_d$ in 2 kb sliding windows, specifying at least 50 SNPs be present within a window to be analyzed using the Python script 'ABBABABAwindows.py' (https://github.com/simonhmartin/genomics_general). Because $\hat{f}_d$ between P2 and P3 is meaningless if *D* is negative (Martin et al. 2015), we removed windows with negative *D* from analysis, then compared distributions of $\hat{f}_d$ for autosomes and the Z chromosome using Mann-Whitney $\cup$ tests.

2.7 Geographic clines: evidence for restricted gene flow on the Z chromosome?

We performed geographic cline analyses across six hybrid zone transects in Asia to further test the hypothesis that there is restricted gene flow on the Z chromosome in regions of secondary contact between *rustica*, *tytleri*, and *gutturalis* subspecies. We leveraged the reduced-representation genotype-by-sequencing (GBS) dataset analyzed in Scordato et al. (2020) to increase our sampling of individuals from parental and admixed

populations (n = 1,288 individuals; Supplementary Fig. S2). We processed the GBS data following Scordato et al. (2020), with minor adjustments. Briefly, we demultiplexed raw sequencing files using the Stacks v2.5 'process_radtags' module (Catchen et al. 2013; Rochette et al. 2019) and quality trimmed the demultiplexed data using Trimmomatic v0.36 (Bolger et al. 2014) using the same settings as above for the whole genome resequencing data. Filtered reads were mapped to the *H. rustica* reference genome using bwa 'mem' (Li and Durbin 2009) and we called variants using bcftools (Li et al. 2009). We retained high quality SNPs by filtering sites with a median depth less than 7 or with greater than 20% missing data among individuals. We then removed SNPs with a minor allele frequency < 0.05 and separately parsed variants from the Z and autosomes.

We used HZAR (Derryberry et al. 2014) to analyze the hybrid zones transects, which fell into two categories: those spanning inferred migratory divides associated with the Tibetan Plateau, and those that do not appear to be associated with known barriers to gene flow. The three migratory divide transects represent secondary contact zones between *rustica – tytleri* in Russia, *rustica – tytleri* in Mongolia, and *rustica – gutturalis* in China (Scordato et al. 2017, 2020). To obtain a continuous measure of ancestry through hybrid zones from multiple genetic backgrounds, we performed PCA analyses for the Z chromosome and autosomes using SNPRelate (Zheng et al. 2012). We used the resulting PC1 scores as input for cline-fitting after normalizing PC scores so that ancestry measures ranged between 0 and 1, then calculated means and variances among populations. All samples within 20 kilometer (km) intervals were defined as a population and distance between populations was calculated using Haversine distance in km. We fit clines to autosomal and Z chromosome ancestry values using cline models with five distinct exponential tails (none, both, right, left, and mirrored) using the MCMC procedure in HZAR. Each model estimated the cline center (*c*; km from the westernmost point in each transect) and width (*w*), equal to 1/maximum slope of the cline, while estimating the mean and variance as additional free parameters. We initialized each model using 1 x 10^5 generations of burn-in, then ran three independent MCMC chains for 1 x 10^6 generations after performing new fit request based on initial chains. We inspected each run for convergence and concatenated the chain results for each model prior to model selection. We identified the best-fit model using AICc and determined the maximum-likelihood parameters and confidence intervals associated with the best model for each transect.

We used the neutral diffusion equation $w \approx \sigma t^{0.5}$ (Barton and Gale 1993) to test the hypothesis that Z-linked ancestry clines fit expectations of no selection or reproductive isolation. This equation approximates the expected cline width under a scenario of neutrality, based on the root mean square dispersal per generation (σ) and the number of generations since secondary contact (t). Following Scordato et al. (2020), we used the values

σ = 42 km based on barn swallow dispersal distances reported in Møller (1994) and Paradis et al. (1998) and t = 20, assuming a hybrid zone age of at least 20 years and a generation time of one year (Zink et al. 2006).

3 Results

3.1 Genome resequencing and variant calling

Our whole genome sequencing and quality filtering steps produced an average of 50 $\pm$ 9.29 million reads per sample. An average of 99.8 $\pm$ 0.03% of reads mapped to the reference genome, corresponding to an estimated 5.96 $\pm$ 1.23-fold predicted coverage based on a 1.2 Gb genome size. Our variant calling and hard filtering steps using GATK produced 65,062,523 biallelic SNPs. Further filtering steps for specific downstream analyses included the removal of SNPs that did not meet missing data thresholds (max-missing 0.2 = 62,408,527 SNPs, max-missing 0.4 = 64,236,731 SNPs, max-missing 0.6 = 64,907,500 SNPs), removal of singletons (47,282,930 SNPs), and removal of ingroup SNPs with minor-allele frequencies below 0.05 (12,061,130 SNPs), and within 100 bp of another SNP (4,668,243 SNPs). One sample had nearly completely missing data after filtering steps and was removed from all analyses, resulting in a final ingroup sampling of n = 159 (Supplementary Table S1).

3.2 Genetic diversity within populations: does reproductive skew explain Z chromosome diversity?

Mean nucleotide diversity (π) within barn swallows was 0.0043 and the genomic landscape of diversity was highly conserved across subspecies (Spearman's rank order correlation coefficients, r values > 0.97, p-values < 2.2×10^{-16}; Supplementary Table S2), as predicted by evidence of very recent divergence (e.g., Dor et al. 2010; Smith et al. 2018). Estimates with and without the inclusion of females were highly correlated (r values > 0.95, p-values < 2.2×10^{-16}; see Methods), thus we report on the results from the full (i.e., female and male) dataset. Nucleotide diversity estimates were lower on the Z chromosome than autosomes for all subspecies, hybrid zones, and localities examined (Welch's two-sample t-tests, p-values < 2.2×10^{-16}; Fig. 2A; Supplementary Table S3). Mean autosome and Z chromosome π estimates were positively correlated (r = 0.84, p-value = 1.91×10^{-7}), however, the ratio of Z:autosome π (π_{ZA}) was 0.48 $\pm$ 0.063, much lower than the equilibrium population expectation of 0.75 (Fig. 2B). Empirical π_{ZA} values were also generally below or very near the theoretical minimum threshold under extreme variance in male reproductive success (orange shaded area in Fig. 2B-C).

Tajima's *D* was higher and more variable on the Z chromosome than autosomes across barn swallow subspecies (autosome mean = -0.94 $\pm$ 0.46, Z mean = -0.55 $\pm$ 0.58). Genome-wide distributions of Tajima's *D* were shifted towards values less than zero, indicative of skewed allele frequencies due to an excess of rare alleles following recent demographic changes or due to selection (Supplementary Fig. S3). Z-linked distributions were consistently right-shifted relative to autosomal distributions, indicating a higher frequency of intermediate frequency alleles. We observed this pattern regardless of whether females were included or excluded from analysis, and found similar estimates of *D* at various missing data thresholds (see Methods; Supplementary Table S4). These findings suggest that the lower effective population size of the Z chromosome and potential for transmission distortion through variance in male reproductive success may cause the Z chromosome to equilibrate to demographic changes faster than autosomes through genetic drift (Wall et al. 2002; Pool and Nielsen 2007), leading to fewer rare alleles on the Z chromosome, specifically.

3.3 Population differentiation: does the Z chromosome exhibit high relative differentiation?

Genome-wide relative population differentiation (*Fst*) was modest across pairs of barn swallow subspecies (mean *Fst* = 0.022 $\pm$ 0.016; Fig. 3A-B; Supplementary Table S5). There were, however, regions of accentuated differentiation scattered across chromosomes and concentrated on the Z chromosome and Chromosome 4A (Fig. 3C). Mean Z-linked *Fst* estimates were significantly higher (1.77 to 8.1 $\times$) than autosomes across comparisons (Mann-Whitney $\cup$ tests, *p*-values =< 0.0016; Fig. 3A; Supplementary Table S5). There was also a positive correlation between Z-linked *Fst* and the ratio of Z:autosome *Fst* (Fst_{ZA}; Pearson's product-moment correlation, $r = 0.58$, *p*-value = 0.005; Fig. 3B). We found no evidence for an association between Fst_{ZA} and autosomal *Fst* (*p*-value = 0.66), however, suggesting that Z/A ratios are largely driven by Z-linked patterns (Fig. 3B).

Relative population differentiation was also elevated on the Z chromosome in comparisons between hybrid zone regions and parental subspecies (Mann-Whitney $\cup$ tests, *p*-values $< 2.2 \times 10^{-16}$; Supplementary Table S5). In two cases, *tytleri* versus *rustica* $\times$ *tytleri* and *tytleri* versus *gutturalis* $\times$ *tytleri*, *Fst* values were too low to allow for a statistical comparison, likely due to the fact that our whole-genome sampling included admixed individuals that were substantially backcrossed to *tytleri* (see Fig. 4); these individuals were not included in any analyses including 'parental' *tytleri*. In several comparisons, however, pairwise Fst_{ZA} was higher between a given parental subspecies and hybrids than between pairwise comparisons of respective parental subspecies, potentially due to reduced gene flow on the Z chromosome in hybrid zones. For example, mean Fst_{ZA} between *gutturalis* and *gutturalis* $\times$ *rustica* hybrids was 8.93 compared to 4.65 between parental

gutturalis and *rustica*. Similarly, Fst_{ZA} between *gutturalis* and *gutturalis* × *tytleri* hybrids was 4.14, while Fst_{ZA} between *gutturalis* and *tytleri* was 3.89.

Distributions of lineage-specific allelic differentiation (*PBS*) were also higher on the Z chromosome than autosomes in each subspecies (Fig. 3C; Supplementary Table S6; Mann-Whitney ∪ tests, *p*-values < 2.2×10^{-16}). The landscape of differentiation across the Z chromosome was highly heterogeneous, with multiple differentiation islands appearing in each subspecies (Fig. 3C). There were also numerous localized peaks of differentiation on autosomes, and Chromosome 4A had consistently elevated *PBS* relative to other autosomes (Fig. 3C). Consistent with the recurrent evolution of the differentiation landscape, genome-wide *PBS* distributions were positively correlated between subspecies (Spearman rank order correlations, $r = 0.23 – 0.59$, *p*-values < 2.2×10^{-16}), with a more pronounced relationship on the Z, specifically ($r = 0.42 – 0.78$, *p*-values < 2.2×10^{-16}). Further, *PBS* and π were negatively correlated in all subspecies (Spearman rank order correlations, $r = -0.26 – -0.54$, *p*-values < 2.2×10^{-16}), indicating a general increase in relative differentiation in genomic regions with low genetic diversity.

3.4 Genetic structure: evidence for greater population structure on the Z chromosome?

PCA on all, autosomal, and Z-linked SNPs each identified three major genetic clusters consisting of *rustica* + *savignii* + *transitiva*, *gutturalis*, and *tytleri* + *erythrogaster* (Fig. 4A-D). Samples from hybrid zones fell between respective parental subspecies clusters. In the case of the *rustica* × *tytleri* hybrid zone, most samples clustered with *tytleri*, consistent with an overall greater degree of backcrossing with *tytleri* in samples from this region. Although PCA results consistently supported the presence of three genetic clusters, the percent variance explained by the first two principal components (PCs) differed. In analyses based on all SNPs and autosomal SNPs, PCs 1 and 2 explained only ~2% and ~1% of variance, respectively (Fig. 4B-C), while PCs 1 and 2 explained 8.21% and 2.84% of variance in Z-linked SNPs (Fig. 4D). Moreover, PC 1 for Z-linked SNPs strongly differentiated the cluster comprising European subspecies from the Asian and North American subspecies, with PC 2 further delineating *gutturalis* from the *tytleri* + *erythrogaster* cluster.

ADMIXTURE analyses based on Z-linked SNPs yielded support for three genetic clusters (i.e., $K = 3$) based on the cross-validation procedure (CV error = 0.626; Fig. 4E, Supplementary Fig. S4). Similar to our PCA results, Z-linked clusters corresponded to groupings of *rustica* + *savignii* + *transitiva*, *gutturalis*, and *tytleri* + *erythrogaster*. The next best-supported Z chromosome model was $K = 2$ (CV error = 0.628), which strongly distinguished between *rustica* + *savignii* + *transitiva* and *gutturalis* + *tytleri* + *erythrogaster* groups

(Supplementary Fig. S5). Under both models, samples from hybrid zones showed intermediate assignment to genetic clusters of respective parental subspecies. We did not find appreciable additional resolution with higher values of *K*, except that some admixed individuals were assigned to genetic clusters under $K = 4$ and $K = 5$ models (Supplementary Fig. S5). There was less evidence of population genetic structure from autosomal SNPs, with similar support for $K = 1$ and $K = 2$ models (CV error = 0.551 and 0.552, respectively). The results for the autosomal $K = 2$ model are shown in Fig. 4F, showing intermediate assignment of most individuals to both clusters. The autosomal $K = 2$ model also did not strongly delineate European from Asian and North American genetic clusters.

3.5 Phylogenetic analyses: faster lineage sorting and reduced gene flow on the Z chromosome?

Topology weighting analyses revealed greater phylogenetic signal on the Z chromosome than autosomes in each of the triplet experiments we tested (Fig. 5). Two 'shallow' triplets included subspecies in the European and Asian/North American groups, respectively. The best-supported topology in the Triplet 1 analysis was t3 (*rustica*, (*savignii*, *transitiva*)), with a higher proportion of subtrees supporting this topology for both autosomes and the Z chromosome (Fig. 5A; Mann-Whitney ∪ tests, p-values $< 2.2 \times 10^{-16}$). However, we also found a higher proportion of Z-linked subtrees supporting this topology than autosomal subtrees (Mann-Whitney ∪ test, p-value $= 5.7 \times 10^{-8}$). There was differential support for two topologies between autosomes and the Z in the Triplet 2 analysis (Fig. 5B), with a slightly higher proportion of autosomal subtrees supporting t3 (*erythrogaster*, (*tytleri*, *gutturalis*)). The Z chromosome more strongly supported t2 (*gutturalis*, (*erythrogaster*, *tytleri*)) over t1 and t3 (Mann-Whitney ∪ tests, p-values $< 2.2 \times 10^{-16}$).

Analyses including deeper-time splits between European and Asian/North American subspecies yielded comparatively strong support for one topology over alternatives (Fig. 5C-D; Triplet 3, t1: (*erythrogaster*, (*savignii*, *transitiva*)), Mann-Whitney ∪ tests, autosome and Z p-values $< 2.2 \times 10^{-16}$; Triplet 4, t1: (*savignii*, (*erythrogaster*, *tytleri*)), autosome and Z chromosome p-values $< 2.2 \times 10^{-16}$). These results reflect the expected greater degree of lineage sorting between major barn swallow groups than within the 'shallow' groups tested in Triplets 1 and 2. The proportion of subtrees supporting t1 was also comparatively higher for the Z chromosome than autosomes in both triplets (Mann-Whitney ∪ tests, p-values $< 2.2 \times 10^{-16}$), emphasizing faster lineage sorting and greater phylogenetic signal from the Z, specifically. Because topology weighting does not explicitly disentangle the effects of incomplete lineage sorting and gene flow on

genealogies (Martin and Van Belleghem 2017), we note it is possible that these results are in part driven by reduced gene flow on the Z chromosome.

The results of ABBA-BABA tests varied across comparisons (Supplementary Fig. S6). Mean $\hat{f}_d$ values were higher for autosomes than the Z chromosome in each analysis (*rustica* versus *rustica* x *tytleri* autosomal $\hat{f}_d$ = 0.162 ± 0.12, Z $\hat{f}_d$ = 0.133 ± 0.1; *rustica* versus *rustica* x *gutturalis* autosomal $\hat{f}_d$ = 0.174 ± 0.12, Z $\hat{f}_d$ = 0.173 ± 0.13; *gutturalis* versus *tytleri* x *gutturalis* autosomal $\hat{f}_d$ = 0.171 ± 0.12, Z $\hat{f}_d$ = 0.163 ± 0.123). This pattern matches our prediction for reduced sex-linked introgression in secondary contact zones, however distributions of $\hat{f}_d$ were only significantly lower on the Z chromosome between *rustica* and *rustica* x *tytleri* (Mann-Whitney ∪ test, p = 0.039).

3.6 Geographic clines: evidence for restricted gene flow on the Z chromosome?

The variant filtering steps for GBS samples used in geographic cline analyses yielded 52,514 autosomal and 1,838 Z-linked bi-allelic SNPs for analysis (Supplementary Fig. S7), from which we determined PC1 for autosomes and the Z independently. Cline analyses based on genomic PC1 in HZAR yielded evidence for steep Z-linked clines in each transect spanning migratory divides (Fig. 6A-C; Supplementary Fig. S2; Supplementary Table S7), qualitatively similar to the ancestry clines estimated in Scordato et al. (2020). The 95% confidence intervals for Z-linked clines in the Russia *rustica – tytleri* and China *rustica – gutturalis* analyses were consistent with narrow regions of introgression (Fig. 6A,C; Russia c = 349.53 km, w = ± 59.19 km; China c = 1,424.2 km, w = ± 87.11 km), while the Mongolia *rustica – tytleri* cline was comparatively wide (Fig. 6B; c = 435.2 km, w = ± 233.5 km). We note the presence of the Tarvagatai and Khangai mountain ranges in the center of the Mongolian transect and that uncertainty in estimates of the cline center and width is associated with the geographic break in the presence of swallows in this region. Autosomal clines in the two *rustica – tytleri* transects were shallow compared to Z-linked clines (Fig. 6A-B). We use the term 'shallow' to refer to the lower magnitude of change in genomic PC1 through space for autosomes than the Z chromosome (i.e., 0.29 – 0.31 versus 0.35 – 0.63 for Russia *rustica – tytleri*, 0.09 – 0.11 versus 0.18 – 0.48 for Mongolia *rustica - tytleri*). We interpret these patterns as evidence for more extensive autosomal admixture across these regions when compared to steep Z-linked clines. While the modest autosomal clines had narrow w (Supplementary Table S7), broad 95% confidence intervals (CIs) indicated uncertainty in their geographic distribution. In contrast to patterns in the two *rustica – tytleri* transects, autosomal ancestry in the *rustica – gutturalis* hybrid zone in China fit a narrow cline (Fig. 6C; autosomal c = 1,544.6 km, w = ± 12.27 km), similar to the Z

chromosome. Z-linked and autosomal cline centers for the *rustica – gutturalis* hybrid zone were also offset by 120.4 km, suggesting that introgression of the Z chromosome may extend further west than autosomes, however we interpret this with caution as the confidence intervals for both clines were largely overlapping.

Geographic clines for *tytleri – gutturalis* transects lacking known barriers to gene flow showed varied patterns (Fig. 6D-F). The autosomal cline between *tytleri* and *gutturalis* in Russia was narrower than the Z-linked cline with cline centers displaced by greater than 100 km (autosomal c = 685.9 km, w = ± 52 km; Z c = 515.9 km, w = ± 225 km), though there was also a broader 95% CI for autosomes and the inferred cline center fell between sampled populations (Fig. 6D). Results from the Mongolian transect showed evidence of extensive autosomal admixture with an inferred cline that was flat and a modest Z-linked cline with very wide CIs (c = 269.1, w = ± 2.7 km; Fig. 6E), suggesting that barriers to gene flow may be weak in this region. Finally, the *tytleri – gutturalis* transect in China had narrow clines for the Z and autosomes centered between 642 and 682 km (Fig. 6F). While c differed between autosomes and the Z chromosome clines, the CIs were nearly completely overlapping and there was no evidence of displacement between ancestry clines, similar to the *rustica – gutturalis* hybrid zone in China.

Z-linked clines of four hybrid zone transects had widths narrower than expectations based on neutral diffusion approximation and conservative assumptions about barn swallow dispersal and contact zone age (i.e., $w \approx \sigma t^{0.5}$ = 187.8 km): *rustica – tytleri* in Russia, *rustica – gutturalis* in China, *tytleri – gutturalis* in China, and *tytleri – gutturalis* in Mongolia, suggesting that selection on the Z chromosome results in restricted gene flow between parental barn swallow subspecies. Results from the other two transects suggest that restricted gene flow on the Z chromosome due to selection need not be invoked in all contact zones between barn swallow subspecies. As noted above, however, the inferred cline width for the *rustica – gutturalis* contact zone in Mongolia is associated with a geographic break due to the presence of central Mongolian mountain ranges, and we cannot conclusively rule out that this hybrid zone is maintained by selection. Collectively, geographic cline analyses provide evidence for barriers to gene flow on the Z chromosome across multiple hybrid zones despite idiosyncratic autosomal patterns and suggest that Z-linked reproductive isolation may be relevant to speciation in Asian barn swallow lineages.

Discussion

In this study, we examined multiple aspects of Z-linked diversity and differentiation in the barn swallow and tested for evidence of a role of the Z chromosome in generating and maintaining lineage diversity. Below we discuss how these facets of our study relate to life history, historical demography, biogeography, and speciation in a lineage with one of the most expansive geographic distributions among vertebrates.

4.1 Factors shaping Z-linked genetic diversity

We found repeated patterns of low genetic diversity on the Z chromosome in barn swallows (Fig. 2; Supplementary Table S3). Assuming equal sex ratios, the Z chromosome will be present in a population at 3/4 the frequency of autosomes and is expected to harbor a proportionately lower amount of the observed autosomal diversity. Empirical evidence from birds rarely fits this expectation, however. Ratios of Z-to-autosomal genetic diversity in barn swallows generally fall well below theoretical predictions, and are among the lower estimates from bird species studied using genomic data from autosomes and the Z (e.g., mean π_{ZA} equal to 0.61 for 33 species reviewed in Irwin 2018). On average, the barn swallow Z chromosome had more than two-fold lower nucleotide diversity relative to autosomes after correcting for male-biased mutation rates, with the highest observed ratio among subspecies equal to 0.58 (Supplementary Table S3). These results indicate that factors extending beyond differences in effective population size due to female heterogamety contribute to reductions in Z-linked diversity. It is therefore likely in barn swallows that mating strategies are a driver of low diversity on the Z chromosome, yet recent demographic history, linked selection, and divergence hitchhiking may also contribute to the recurrent evolution of this pattern.

In addition to the predictions for autosomal and sex-linked variation due to basic differences in effective population size, differences in mating strategies will skew sex-linked diversity depending on variation in male or female reproductive success. Barn swallows are a classic example of a species that is both socially monogamous and genetically promiscuous (Smith et al. 1991; Møller and Tegelström 1997), a mating system that disproportionately influences male reproductive success. Excess variance in male reproductive success will skew the transmission of Z chromosomes in a population, leading to a net reduction in Z-linked diversity (Caballero 1995; Charlesworth 2001). Empirical data support that ~30% of barn swallow offspring are the product of extra-pair paternity (EPP; Smith et al. 1991; Møller and Tegelström 1997; Saino et al. 1997; Møller et al. 2003; Kleven et al. 2006; Vortman et al. 2013). This suggests that excess variance in male reproductive success may play a prominent role in driving low Z-linked genetic diversity. Our results are certainly consistent

with this hypothesis, but also highlight the additional impacts of other factors on sex-linked diversity. Indeed, the theoretical minimum Z/A diversity ratio given reproductive skew against males is 0.56 (Charlesworth 2001); most of our empirical ratios fell below this value (Fig. 2), indicating that EPP alone cannot fully explain the observed ratios. We note also that the coarse estimate of EPP in barn swallows is insufficient for directly measuring the relative role of excess variance in male reproductive success on sex-linked diversity. Indeed, some males may have no offspring in a given breeding season while other males may have equal or greater success compared to females. A more comprehensive evaluation of EPP and male reproductive success within and between populations would therefore be necessary to quantify the extent to which Z-linked diversity is shaped by these factors.

Several processes and mechanisms that may contribute further to low Z-linked diversity in barn swallows include demographic shifts in the recent past, differences in recombination rates between sex chromosomes and autosomes and related effects of linked selection, and reduced sex-linked gene flow. We note that each of these mechanisms is not necessarily mutually exclusive and consider them in turn. First, studies have found evidence for recent fluctuations in effective population size during barn swallow diversification (Zink et al. 2006; Dor et al. 2010; Smith et al. 2018), including global bottlenecks followed by population recovery concurrent with increased human settlement. These changes in historical demography would be expected to skew the allele frequency spectrum of the Z chromosome relative to autosomes and correspond with sex-linked reductions in diversity due to smaller effective population size. Tajima's *D* estimates provide evidence of skewed Z-linked distributions indicative of fewer rare alleles relative to autosomes (Supplementary Fig. S3), a possible consequence of population size contractions during the shift from natural to human-mediated nesting sites in the recent past (Smith et al. 2018). Second, several lines of evidence suggest that the Z chromosome has lower recombination rates than autosomes. First, genetic diversity typically covaries with recombination rate across the genome (Smukowski and Noor 2011; Corbett-Detig et al. 2015), and recombination maps from birds consistently show low recombination rates on the Z (e.g., Backström et al. 2010; Burri et al. 2015; Singhal et al. 2015; Kawakami et al. 2017). Related to this point, Z-linked regions recombine only in males, requiring a potentially extreme male recombination bias for the effective recombination rate of the Z to match that of autosomes. However, a recent study found intriguing evidence for female-biased recombination rates in barn swallows (Malinovskaya et al. 2020). While we did not estimate recombination rates here, these factors together argue that low rates of recombination on the Z chromosome contribute to low genetic diversity through linked selection. Our results are also consistent with low Z-linked genetic diversity due to reduced gene flow on the Z chromosome, and we discuss in detail below these results and their relevance to speciation.

4.2 Allelic differentiation, gene flow, and the role of the Z chromosome in speciation

Our findings add to a growing body of data highlighting the concentration of allelic differentiation on the Z chromosome in birds (e.g., Ellegren et al. 2012; Burri et al. 2015; Lavretsky et al. 2015; Irwin et al. 2016; Hooper et al. 2019; Battey 2020; reviewed in Irwin 2018). This phenomenon has led to the conclusion that the Z itself represents a large genomic island of differentiation, divergence, and in some cases speciation. We observed a higher genomic background of differentiation on the Z combined with numerous 'peaks' of differentiation relative even to the high chromosomal background (Fig. 3). Genome-wide *Fst* between barn swallow subspecies is quite low, which makes sense given divergence from a very recent common ancestor (Zink et al. 2006; Dor et al. 2010; Smith et al. 2018). In this context, the barn swallow Z chromosome stands out as a conspicuous island of differentiation, where *Fst* and *PBS* values consistently surpass the genomic background. Higher background differentiation on the Z chromosome suggests that detailed examination of differentiation peaks may require conservative evaluation relative to the Z-linked background rather than the genome-wide background when inferring patterns of selection (Fig. 3, Supplementary Table S5). Our topology weighting analyses also indicate faster rates of lineage sorting on the Z chromosome and/or lower Z-linked gene flow, allowing us to examine support for phylogenetic relationships within very recently diverged clades. In the European subspecies group, we found the greatest support for a sister relationship between the two non-migratory subspecies *savignii* and *transitiva*, with *rustica* sister to the combined *savignii* + *transitiva* clade. In the Asian/North American group, *tytleri* (Asia) was supported as sister to *erythrogaster* (North America) rather than the other Asia-distributed subspecies *gutturalis*. These results argue that Z-linked regions could be useful for reconstructing phylogenies in barn swallows and other groups of recently-diverged bird taxa.

An important related question is whether the Z chromosome also represents a divergence and/or speciation island in barn swallows. Here, Z-linked regions would need to play a role in generating and maintaining species boundaries rather than evolving allelic differentiation solely as a by-product of low recombination in ancestral populations (Cruickshank and Hahn 2014; Burri 2017). The combined effects of linked selection driving reduced variation at linked sites and selection against hybrids on the Z could both contribute to the observed landscape of diversity and differentiation. Making a conclusive statement about the relative roles of these processes would require estimates of recombination rates and sequence divergence, which were outside of the scope of the present study. However, we observed a complement of patterns in barn swallows suggesting that the Z chromosome does not represent an 'incidental' genomic island (e.g., Turner and Hahn 2010), and is instead involved in reproductive isolation between lineages. We also emphasize that linked selection and

selection against gene flow are not necessarily mutually exclusive processes, each contributing to genomic divergence in barn swallows.

We will first explore patterns that fit predictions under the linked selection model. First, we observed significant genome-wide correlations in π between subspecies indicating a highly conserved genomic landscape of diversity ($r > 0.99$; Supplementary Table S2). This is expected given very recent, shared ancestry among barn swallow populations and also if a conserved genomic recombination landscape has mediated levels of diversity since splitting from their common ancestor (Burri 2017). Our topology weighting results support this conclusion. Given the remarkable degree of incomplete lineage sorting across the barn swallow genome overall (Fig. 5), we hypothesize that genomic regions with high diversity are largely those harboring shared ancestral polymorphism not yet purged due to linked selection. Alternatively, we expect greater support for specific tree topologies in genomic regions with less ancestral polymorphism (e.g., the Z chromosome), which we observed here. Second, genomic differentiation (*PBS*) was correlated across subspecies comparisons, similar to the relationship for π. There were also multiple *PBS* peaks shared among most or all subspecies (Fig. 3), which may have evolved through recurrent background selection in structural genomic features with low recombination rates, as observed in flycatchers (Burri et al. 2015), or as a consequence of recurrent hitchhiking (Matthey-Doret and Whitlock 2019). Indeed, some of these regions may house genes underlying divergent traits that have been repeatedly 'tapped' by selection (e.g., Chromosome 4A), a hypothesis that deserves more detailed investigation. Previous analyses linking divergent traits among barn swallow subspecies (e.g., plumage coloration, wing shape, migratory behavior) to genomic variation indicate that these traits have a complex genetic architecture (Scordato et al. 2017, 2020), which must be considered when exploring the possibility that a subset of differentiation peaks have been subject to recurrent hitchhiking effects. Third, there was a negative correlation between π and *PBS* indicating a general increase in allelic differentiation in regions of low diversity, a relationship expected if recombination rate variation and linked selection contribute to the genomic differentiation landscape (Maynard Smith and Haigh 1974; Noor and Bennett 2009; Cruickshank and Hahn 2014; Burri et al. 2015). Finally, while not direct evidence for linked selection in low recombination regions, analyses of population genetic structure indicate a greater degree of ancestry information on the Z chromosome relative to autosomes (Fig. 4) corresponding to the greater degree of lineage sorting inferred due to low recombination.

Some observed genomic patterns also provide evidence of reproductive isolation on the Z chromosome, resulting in selection against gene flow and a greater signal of lineage-specific ancestry. These patterns

suggests a more active role of Z-linked regions in barn swallow speciation than solely accumulating differentiation due to linked selection, and in which case disruptive selection on Z-linked variation would result in reduced hybrid fitness in secondary contact. Our analyses of barn swallow hybrid zones in Asia provide evidence to support this hypothesis (Fig. 6). Narrow Z chromosome clines were most prominent in two contact regions associated with migratory divides where *rustica* comes into contact with *tytleri* in Russia and *gutturalis* in China (Fig. 6A,C). In contrast to autosomal patterns in multiple hybrid zones indicating weak barriers to gene flow (e.g., Fig. 6A,B,E), there was fairly consistent evidence of selection against hybrids on the Z chromosome. Genetic diversity (π) was also low and allelic differentiation (*Fst*) high on the Z chromosome in hybrid populations, likely consequences of reduced gene flow due to selection on the Z in contact zones. These results may therefore indicate a role of Z-linked regions in reproductive isolation during secondary contact, leading to a recurrent pattern of elevated relative differentiation between parental and hybrid populations on the Z chromosome. This interpretation is also supported by results from tree-based tests of introgression and lineage sorting (Fig. 5; Supplementary Fig. S6), showing a lower incidence of incomplete lineage sorting and gene flow on the Z chromosome. Collectively, the observed patterns of diversity, differentiation, and gene flow on the barn swallow Z chromosome are consistent with verbal analogies of differentiation and speciation 'islands', however more detailed investigation into specific peaks and valleys of differentiation will be necessary to identify specific regions involved in reproductive isolation.

A related question is what genes and regulators are within genomic islands (Z-linked and autosomal), and whether these underlie divergent traits in barn swallows. Barn swallow populations differ in mate selection traits presumably under sexual selection (Møller 1994b; Safran et al. 2016b; Romano et al. 2017a). These include melanin-based plumage coloration and tail streamer lengths, two traits with potentially complex genetic architectures involving many genes and regulatory elements. Several studies have implicated Z-linked regions or identified Z-linked genes involved in species recognition, plumage coloration, or feather development with exceptional differentiation between bird lineages (e.g., Sæther et al. 2007; Toews et al. 2016; Campagna et al. 2017), raising the possiblity that Z-linked peaks could underlie plumage variation relevant to pre-zygotic isolation in barn swallows. Other divergent traits potentially involved in reproductive isolation relate to migration, including wing length and shape linked to migratory distance (Safran et al. 2016a; von Rönn et al. 2016; Scordato et al. 2017), timing of arrival to breeding habitat (Møller 1994a), and wintering destination (Scordato et al. 2020). Song also differs between barn swallow populations, and divergence in song is strongly correlated with pairwise genetic divergence (Wilkins et al. 2018). Migratory traits and behaviors probably also have very complex genetic architectures, leaving open the possibility that some of their component genes and

regulators are Z-linked. Another possible mechanism contributing to reproductive isolation is structural variation on the Z chromosome. Hooper et al. (2019) found evidence that Z-linked inversions favored by selection are also responsible for reduced gene flow on the Z in finches. Pericentric inversions are also common in birds and especially on the Z chromosome (Hooper and Price 2017), raising the intriguing possibility that structural variation also facilitates reproductive isolation in barn swallows. Further investigation will be necessary to determine whether Z-linked genes or structural variation indeed constitute barriers to gene flow through reproductive isolation.

Overall, our findings align with major predictions for an accumulation of loci involved in sex differences and under divergent sexual selection on sex chromosomes (Charlesworth et al. 1987; Kirkpatrick and Hall 2004a,b; Albert and Otto 2005), and support a role of the Z chromosome in barn swallow speciation. We also note that there were multiple peaks of genomic differentiation on autosomes, including a concentration of high differentiation on Chromosome 4A, meaning that loci involved in reproductive isolation could be distributed across chromosomes. Thorough examinations of specific genomic islands and their associations with divergent traits will be natural extensions to our findings to further evaluate the role of the Z chromosome in barn swallow speciation.

Acknowledgments

We are grateful to the following individuals for assistance in the field and the lab: Caroline Glidden, Rachel Lock, Brittany Jenkins, Nikolai Markov, Gennady Bachurin, Olga Zayatseva, Elena Shnayder, Georgy Semenov, Unurjargal Enkhbat, Bayanmunkh Dashnyam, Davaadori Enkhbayar, Wataru Kitamura, Takashi Tanioka, and Yuta Inaguma. This work was supported by National Science Foundation (NSF) postdoctoral research fellowship grant DBI-1906188 to DRS, NSF CAREER grant DEB-1149942 to RJS, NSF grant IOS-DEB-1856266 to RJS, and a National Geographic Society Committee on Research and Exploration award to ESCS. The University of Colorado Boulder occupies land within the territories of the Arapaho, Cheyenne, and Ute peoples.

References

Albert, A. Y. K., and S. P. Otto. 2005. Sexual selection can resolve sex-linked sexual antagonism. *Science* 310:119–121.

Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19:1655–1664.

Aquadro, C. F., D. J. Begun, and E. C. Kindahl. 1994. Selection, recombination, and DNA polymorphism in *Drosophila*. In *Non-neutral evolution* (pp. 46–56. Springer, Boston, MA.

Axelsson, E., N. G. C. Smith, H. Sundström, S. Berlin, and H. Ellegren. 2004. Male-biased mutation rate and divergence in autosomal, Z-linked and W-linked introns of chicken and turkey. *Mol Biol Evol* 21:1538–1547.

Bachtrog, D., M. Kirkpatrick, J. E. Mank, S. F. McDaniel, J. C. Pires, W. Rice, and N. Valenzuela. 2011. Are all sex chromosomes created equal? *Trends in Genetics* 27:350–357.

Backström, N., W. Forstmeier, H. Schielzeth, H. Mellenius, K. Nam, E. Bolund, M. T. Webster, T. Öst, M. Schneider, and B. Kempenaers. 2010. The recombination landscape of the zebra finch *Taeniopygia guttata* genome. *Genome Res* 20:485–495.

Barton, N. H., and K. S. Gale. 1993. Genetic analysis of hybrid zones. *Hybrid zones and the evolutionary process*, 13–45.

Battey, C. J. 2020. Evidence of linked selection on the z chromosome of hybridizing hummingbirds. *Evolution* 74:725–739.

Betancourt, A. J., Y. Kim, and H. A. Orr. 2004. A pseudohitchhiking model of X vs. autosomal diversity. *Genetics* 168:2261–2269.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.

Burri, R. 2017. Interpreting differentiation landscapes in the light of long-term linked selection. *Evolution Letters* 1:118–131.

Burri, R., A. Nater, T. Kawakami, C. F. Mugal, P. I. Olason, L. Smeds, A. Suh, L. Dutoit, S. Bureš, and L. Z. Garamszegi. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Res* 25:1656–1665.

Caballero, A. 1995. On the effective size of populations with separate sexes, with particular reference to sex-linked genes. *Genetics* 139:1007–1011.

Campagna, L., M. Repenning, L. F. Silveira, C. S. Fontana, P. L. Tubaro, and I. J. Lovette. 2017. Repeated divergent selection on pigmentation genes in a rapid finch radiation. *Science Advances* 3:e1602404.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: an analysis tool set for population genomics. *Mol Ecol* 22:3124–3140.

Charlesworth, B. 1996. Background selection and patterns of genetic diversity in *Drosophila melanogaster*. *Genet Res (Camb)* 68:131–149.

Charlesworth, B. 2001. The effect of life-history and mode of inheritance on neutral genetic variability. *Genet Res (Camb)* 77:153–166.

Charlesworth, B., and D. Charlesworth. 1978. A model for the evolution of dioecy and gynodioecy. *Am Nat* 112:975–997.

Charlesworth, B., J. A. Coyne, and N. H. Barton. 1987. The relative rates of evolution of sex chromosomes and autosomes. *Am Nat* 130:113–146.

Corbett-Detig, R. B., D. L. Hartl, and T. B. Sackton. 2015. Natural selection constrains neutral diversity across a wide range of species. *PLoS Biol* 13:e1002112.

Corl, A., and H. Ellegren. 2012. The genomic signature of sexual selection in the genetic diversity of the sex chromosomes and autosomes. *Evolution* 66:2138–2149.

Cruickshank, T. E., and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* 23:3133–3157.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and G. Genomes Project Analysis. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.

Dean, R., and J. E. Mank. 2014. The role of sex chromosomes in sexual dimorphism: discordance between molecular and phenotypic data. *J Evol Biol* 27:1443–1453.

del Hoyo, J., and A. Elliott. 2014. Barn swallow (*Hirundo rustica*). In *Handbook of the Birds of the World.*

Lynx Edicions, Spain.

Derryberry, E. P., G. E. Derryberry, J. M. Maley, and R. T. Brumfield. 2014. HZAR: hybrid zone analysis using an R software package. *Mol Ecol Resour* 14:652–663.

Dor, R., R. J. Safran, F. H. Sheldon, D. W. Winkler, and I. J. Lovette. 2010. Phylogeny of the genus *Hirundo* and the Barn Swallow subspecies complex. *Mol Phylogenet Evol* 56:409–418.

Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. Testing for ancient admixture between closely related populations. *Mol Biol Evol* 28:2239–2252.

Elgvin, T. O., J. S. Hermansen, A. Fijarczyk, T. Bonnet, T. Borge, S. A. Saether, K. L. Voje, and G. Sætre. 2011. Hybrid speciation in sparrows II: a role for sex chromosomes? *Mol Ecol* 20:3823–3837.

Ellegren, H. 2009. Genomic evidence for a large-Z effect. *Proc R Soc B Biol Sci* 276:361–366.

Ellegren, H. 2011. Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. *Nat Rev Genet* 12:157–166.

Ellegren, H., L. Smeds, R. Burri, P. I. Olason, N. Backstrom, T. Kawakami, A. Kunstner, H. Makinen, K. Nadachowska-Brzyska, A. Qvarnstrom, S. Uebbing, and J. B. Wolf. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* 491:756–760.

Formenti, G., M. Chiara, L. Poveda, K.-J. Francoijs, A. Bonisoli-Alquati, L. Canova, L. Gianfranceschi, D. S. Horner, and N. Saino. 2019. SMRT long reads and Direct Label and Stain optical maps allow the generation of a high-quality genome assembly for the European barn swallow (*Hirundo rustica rustica*). *Gigascience* 8:giy142.

Guindon, S., J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321.

Hellborg, L., and H. Ellegren. 2004. Low levels of nucleotide diversity in mammalian Y chromosomes. *Mol Biol Evol* 21:158–163.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet*

6:e1000862.

Hooper, D. M., S. C. Griffith, and T. D. Price. 2019. Sex chromosome inversions enforce reproductive isolation across an avian hybrid zone. *Mol Ecol* 28:1246–1262.

Hooper, D. M., and T. D. Price. 2017. Chromosomal inversion differences correlate with range overlap in passerine birds. *Nat Ecol Evol* 1:1526–1534.

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

Irwin, D. E., M. Alcaide, K. E. Delmore, J. H. Irwin, and G. L. Owens. 2016. Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. *Mol Ecol* 25:4488–4507.

Jain, C., A. Dilthey, S. Koren, S. Aluru, and A. M. Phillippy. 2017. A fast approximate algorithm for mapping long reads to large reference databases. In *International Conference on Research in Computational Molecular Biology* (pp. 66–81). Springer, Boston, MA.

Jain, C., S. Koren, A. Dilthey, A. M. Phillippy, and S. Aluru. 2018. A fast adaptive algorithm for computing whole-genome homology maps. *Bioinformatics* 34:i748–i756.

Kawakami, T., C. F. Mugal, A. Suh, A. Nater, R. Burri, L. Smeds, and H. Ellegren. 2017. Whole-genome patterns of linkage disequilibrium across flycatcher populations clarify the causes and consequences of fine-scale recombination rate variation in birds. *Mol Ecol* 26:4158–4172.

Kirkpatrick, M., and D. W. Hall. 2004a. Male-biased mutation, sex linkage, and the rate of adaptive evolution. *Evolution* 58:437–440.

Kirkpatrick, M., and D. W. Hall. 2004b. Sexual selection and sex linkage. *Evolution* 58:683–691.

Kleven, O., F. Jacobsen, R. Izadnegahdar, R. J. Robertson, and J. T. Lifjeld. 2006. Male tail streamer length predicts fertilization success in the North American barn swallow (*Hirundo rustica erythrogaster*). *Behav Ecol Sociobiol* 59:412–418.

Korunes, K. L., K. Samuk. 2020 Pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. bioRxiv:2020.06.27.175091.

Lavretsky, P., J. M. Dacosta, B. E. Hernández-Baños, A. Engilis Jr, M. D. Sorenson, and J. L. Peters. 2015.

Speciation genomics and a role for the Z chromosome in the early stages of divergence between Mexican ducks and mallards. *Mol Ecol* 24:5364–5378.

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

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.

Liu, Y., E. S. C. Scordato, R. Safran, and M. Evans. 2018. Ventral colour, not tail streamer length, is associated with seasonal reproductive performance in a Chinese population of Barn Swallows (*Hirundo rustica gutturalis*). *J Ornithol* 159:675–685.

Malinovskaya, L. P., K. Tishakova, E. P. Shnaider, P. M. Borodin, and A. A. Torgasheva. 2020. Heterochiasmy and Sexual Dimorphism: The Case of the Barn Swallow (*Hirundo rustica*, Hirundinidae, Aves). *Genes* 11:1119.

Mank, J. E., K. Nam, and H. Ellegren. 2009. Faster-Z evolution is predominantly due to genetic drift. *Mol Biol Evol* 27:661–670.

Mank, J. E., B. Vicoso, S. Berlin, and B. Charlesworth. 2010. Effective population size and the Faster-X effect: Empirical results and their interpretation. *Evolution* 64:663–674.

Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, C. Salazar, J. R. Walters, F. Simpson, M. Blaxter, A. Manica, J. Mallet, and C. D. Jiggins. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res*. 23:1817–1828.

Martin, S. H., J. W. Davey, and C. D. Jiggins. 2015. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol Biol Evol* 32:244–257.

Martin, S. H., and S. M. Van Belleghem. 2017. Exploring evolutionary relationships across the genome using topology weighting. *Genetics* 206:429–438.

Matthey-Doret, R., and M. C. Whitlock. 2019. Background selection and FST: consequences for detecting local adaptation. *Mol Ecol* 28:3902–3914.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S.

Gabriel, and M. Daly. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303.

Møller, A. P. 1994a. Phenotype-dependent arrival time and its consequences in a migratory bird. *Behav Ecol Sociobiol* 35:115–122.

Møller, A. P. 1994b. Sexual selection and the barn swallow. *Model systems in behavioral ecology: integrating conceptual, theoretical, and empirical approaches*, pp. 359–380.

Møller, A. P., J. Brohede, J. J. Cuervo, F. De Lope, and C. Primmer. 2003. Extrapair paternity in relation to sexual ornamentation, arrival date, and condition in a migratory bird. *Behav Ecol* 14:707–712.

Møller, A. P., and H. Tegelström. 1997. Extra-pair paternity and tail ornamentation in the barn swallow Hirundo rustica. *Behav Ecol Sociobiol* 41:353–360.

Noor, M. A., and S. M. Bennett. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity* 103:439–444.

Oyler-McCance, S. J., R. S. Cornman, K. L. Jones, and J. A. Fike. 2015. Z chromosome divergence, polymorphism and relative effective population size in a genus of lekking birds. *Heredity* 115:452–459.

Paradis, E., S. R. Baillie, W. J. Sutherland, and R. D. Gregory. 1998. Patterns of natal and breeding dispersal in birds. *J Anim Ecol* 67:518–536.

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

Pease, J. B., and M. W. Hahn. 2013. More accurate phylogenies inferred from low-recombination regions in the presence of incomplete lineage sorting. *Evolution* 67:2376–2384.

Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Müller, V. Baglione, P. Unneberg, M. Wikelski, and M. G. Grabherr. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* 344:1410–1414.

Pool, J. E., and R. Nielsen. 2007. Population size changes reshape genomic patterns of diversity. *Evolution* 61:3001–3006.

Presgraves, D. C. 2008. Sex chromosomes and speciation in *Drosophila*. *Trends Genet* 24:336–343.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* 81:559–575.

Rice, W. R. 1984. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* (1984):735–742.

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

Qvarnström, A., and R. I. Bailey. 2009. Speciation through evolution of sex-linked genes. *Heredity* 102:4–15.

R Core Team. 2017. R: A language and environment for statistical computing.

Rice, W. R. 1984. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* 38:735–742.

Rochette, N. C., A. G. Rivera-Colón, and J. M. Catchen. 2019. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol Ecol* 28:4737–4754.

Romano, A., A. Costanzo, D. Rubolini, N. Saino, and A. P. Møller. 2017a. Geographical and seasonal variation in the intensity of sexual selection in the barn swallow *Hirundo rustica*: a meta-analysis. *Biol Rev* 92:1582–1600.

Romano, A., N. Saino, and A. P. Møller. 2017b. Viability and expression of sexual ornaments in the barn swallow *Hirundo rustica*: a meta-analysis. *J Evol Biol* 30:1929–1935.

Sæther, S. A., G.-P. Sætre, T. Borge, C. Wiley, N. Svedin, G. Andersson, T. Veen, J. Haavie, M. R. Servedio, and S. Bureš. 2007. Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. *Science* 318:95–97.

Safran, R. J., C. R. Neuman, K. J. McGraw, and I. J. Lovette. 2005. Dynamic paternity allocation as a function of male plumage color in barn swallows. *Science*. 309:2210–2212.

Safran, R. J., E. S. Scordato, M. R. Wilkins, J. K. Hubbard, B. R. Jenkins, T. Albrecht, S. M. Flaxman, H. Karaardic, Y. Vortman, A. Lotem, P. Nosil, P. Pap, S. Shen, S. F. Chan, T. L. Parchman, and N. C. Kane. 2016a. Genome-wide differentiation in closely related populations: the roles of selection and geographic isolation. *Mol Ecol* 25:3865–3883.

Safran, R. J., Y. Vortman, B. R. Jenkins, J. K. Hubbard, M. R. Wilkins, R. J. Bradley, and A. Lotem. 2016b.

The maintenance of phenotypic divergence through sexual selection: An experimental study in barn swallows *Hirundo rustica*. *Evolution* 70:2074–2084.

Saino, N., C. R. Primmer, H. Ellegren, and A. P. Møller. 1997. An experimental study of paternity and tail ornamentation in the barn swallow (*Hirundo rustica*). *Evolution* 51:562–570.

[dataset] Schield, D. R., E. S. C. Scordato, C. C. R. Smith, J. K. Carter, S. I. Cherkaoui, S. Gombobaatar, S. Hajib, S. Hanane, A. K. Hund, K. Koyama, W. Liang, Y. Liu, N. Magri, A. Rubtsov, B. Sheta, S. P. Turbek, M. R. Wilkins, L. Yu, and R. J. Safran. Barn swallow whole genome resequencing data. NCBI SRA PRJNA323498.

Scordato, E. S. C., and R. J. Safran. 2014. Geographic variation in sexual selection and implications for speciation in the Barn Swallow. *Avian Res* 5:1.

Scordato, E. S. C., C. C. R. Smith, G. A. Semenov, Y. Liu, M. R. Wilkins, W. Liang, A. Rubtsov, G. Sundev, K. Koyama, and S. P. Turbek. 2020. Migratory divides coincide with reproductive barriers across replicated avian hybrid zones above the Tibetan Plateau. *Ecol Lett* 23:231–241.

Scordato, E. S. C., M. R. Wilkins, G. Semenov, A. S. Rubtsov, N. C. Kane, and R. J. Safran. 2017. Genomic variation across two barn swallow hybrid zones reveals traits associated with divergence in sympatry and allopatry. *Mol Ecol* 26:5676–5691.

Sheldon, F. H., L. A. Whittingham, R. G. Moyle, B. Slikas, and D. W. Winkler. 2005. Phylogeny of swallows (Aves: Hirundinidae) estimated from nuclear and mitochondrial DNA sequences. *Mol Phylogenet Evol* 35:254–270. Elsevier.

Shriver, M. D., G. C. Kennedy, E. J. Parra, H. A. Lawson, V. Sonpar, J. Huang, J. M. Akey, and K. W. Jones. 2004. The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Hum Genomics* 1:274.

Singhal, S., E. M. Leffler, K. Sannareddy, I. Turner, O. Venn, D. M. Hooper, A. I. Strand, Q. Li, B. Raney, C. N. Balakrishnan, S. C. Griffith, G. McVean, and M. Przeworski. 2015. Stable recombination hotspots in birds. *Science* 350:928–932.

Smith, C. C. R., S. M. Flaxman, E. S. C. Scordato, N. C. Kane, A. K. Hund, B. M. Sheta, and R. J. Safran. 2018. Demographic inference in barn swallows using whole-genome data shows signal for bottleneck and

subspecies differentiation during the Holocene. *Mol Ecol* 27:4200–4212.

Smith, H. G., R. Montgomerie, T. Põldman, B. N. White, and P. T. Boag. 1991. DNA fingerprinting reveals relation between tail ornaments and cuckoldry in barn swallows, *Hirundo rustica*. *Behav Ecol* 2:90–98.

Smith, J. M., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. *Genet Res (Camb)* 23:23–35.

Smukowski, C. S., and M. A. F. Noor. 2011. Recombination rate variation in closely related species. *Heredity* 107:496–508.

Subramanian, S. 2016. The effects of sample size on population genomic analyses - implications for the tests of neutrality. *BMC Genomics* 17:123.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.

Tennessen, J. A., A. W. Bigham, T. D. O'Connor, W. Fu, E. E. Kenny, S. Gravel, S. McGee, R. Do, X. Liu, and G. Jun. 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337:64–69.

Toews, D. P. L., S. A. Taylor, R. Vallender, A. Brelsford, B. G. Butcher, P. W. Messer, and I. J. Lovette. 2016. Plumage genes and little else distinguish the genomes of hybridizing warblers. *Curr Biol* 26:2313–2318.

Turner, T. L., and M. W. Hahn. 2010. Genomic islands of speciation or genomic islands and speciation? *Mol Ecol* 19:848–850.

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, and J. Thibault. 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinforma* 43:10–11.

Vicoso, B., and B. Charlesworth. 2009. Effective population size and the faster-X effect: an extended model. *Evolution* 63:2413–2426.

von Rönn, J. A. C., A. B. A. Shafer, and J. B. W. Wolf. 2016. Disruptive selection without genome-wide evolution across a migratory divide. *Mol Ecol* 25:2529–2541.

Vortman, Y., A. Lotem, R. Dor, I. Lovette, and R. J. Safran. 2013. Multiple sexual signals and behavioral reproductive isolation in a diverging population. *Am Nat* 182:514–523.

Wall, J. D., P. Andolfatto, and M. Przeworski. 2002. Testing models of selection and demography in *Drosophila simulans*. *Genetics* 162:203–216.

Wang, S., S. Rohwer, D. R. de Zwaan, D. P. L. Toews, I. J. Lovette, J. Mackenzie, and D. Irwin. 2020. Selection on a small genomic region underpins differentiation in multiple color traits between two warbler species. *Evol. Lett.*

Wang, Z., J. Zhang, W. Yang, N. An, P. Zhang, G. Zhang, and Q. Zhou. 2014. Temporal genomic evolution of bird sex chromosomes. *BMC Evol Biol* 14:250.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 1358–1370.

Wilkins, M. R., E. S. C. Scordato, G. A. Semenov, H. Karaardiç, D. Shizuka, A. Rubtsov, P. L. Pap, S.-F. Shen, and R. J. Safran. 2018. Global song divergence in barn swallows (*Hirundo rustica*): exploring the roles of genetic, geographical and climatic distance in sympatry and allopatry. *Biol J Linn Soc* 123:825–849.

Wolf, J. B. W., and H. Ellegren. 2017. Making sense of genomic islands of differentiation in light of speciation. *Nat Rev Genet* 18:87.

Wu, C. I. 2001. The genic view of the process of speciation. *J Evol Biol* 14:851–865.

Yi, X., Y. Liang, E. Huerta-Sanchez, X. Jin, Z. X. Cuo, J. E. Pool, X. Xu, H. Jiang, N. Vinckenbosch, T. S. Korneliussen, H. Zheng, T. Liu, W. He, K. Li, R. Luo, X. Nie, H. Wu, M. Zhao, H. Cao, J. Zou, Y. Shan, S. Li, Q. Yang, Asan, P. Ni, G. Tian, J. Xu, X. Liu, T. Jiang, R. Wu, G. Zhou, M. Tang, J. Qin, T. Wang, S. Feng, G. Li, Huasang, J. Luosang, W. Wang, F. Chen, Y. Wang, X. Zheng, Z. Li, Z. Bianba, G. Yang, X. Wang, S. Tang, G. Gao, Y. Chen, Z. Luo, L. Gusang, Z. Cao, Q. Zhang, W. Ouyang, X. Ren, H. Liang, H. Zheng, Y. Huang, J. Li, L. Bolund, K. Kristiansen, Y. Li, Y. Zhang, X. Zhang, R. Li, S. Li, H. Yang, R. Nielsen, J. Wang, and J. Wang. 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329:75–78.

Zheng, X., D. Levine, J. Shen, S. M. Gogarten, C. Laurie, and B. S. Weir. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28:3326–3328.

Zink, R. M., A. Pavlova, S. Rohwer, and S. V Drovetski. 2006. Barn swallows before barns: population histories and intercontinental colonization. *Proc R Soc B Biol Sci* 273:1245–1251.

Data Accessibility

Data supporting the conclusions of this study have been deposited to the NCBI short-read archive (accession PRJNA323498). The computational workflow and associated analysis scripts used in this work are available at https://github.com/drewschield/Z-chromosome_analysis_hirundo.

Author Contributions

DRS designed the research, performed the research, analyzed data, and wrote the paper. ESCS, CCRS, and JKC performed research, analyzed data, and edited the paper. SIC, SG, SH, SH, AKH, KK, WL, YL, NM, AR, BS, SPT, MRW, and LY performed field work and sample collection and edited the paper. RJS designed the research, performed field work and sample collection, contributed reagents, and edited the paper.

Tables

Table 1. Predictions of Z-linked diversity and differentiation based on theoretical and empirical studies, including the impacts of various evolutionary scenarios and processes on the ratio of Z chromosome-to-autosomal diversity.

Scenario	Effect on Z-linked diversity	Predictions	Citations
Neutrality; equal sex ratio	Z = 3/4A	Genetic diversity is exactly proportional to the frequency of Z chromosomes in the population due to a lack of reproductive skew, demographic changes, and selection	Charlesworth 2001; Pool and Nielsen 2007
Reproductive skew against males	Z < 3/4A	Minimum Z/A = 9/16 under extreme variance in male reproductive success	Caballero 1995; Charlesworth 2001
Reproductive skew against females	Z > 3/4A	Maximum Z/A = 9/8 under extreme variance in female reproductive success	Caballero 1995; Charlesworth 2001
Population bottleneck	Z =< 3/4A	The Z will have proportionally fewer or greater rare alleles than autosomes, depending on the stage of the bottleneck (i.e., contraction versus expansion, respectively); the allele frequency spectrum on the Z will show signatures of more recent demographic changes due to drift than autosomes	Pool and Nielsen 2007
Background selection	Z < 3/4A	Purifying selection removes variation at linked sites; low recombination results in low diversity and high relative differentiation on the Z	Aquadro et al. 1994; Noor and Bennett 2009; Cruickshank and Hahn 2014; Burri et al. 2015; Wolf and Ellegren 2017
Positive selection	Z < 3/4A	Genetic hitchhiking removes variation at linked sites; low diversity and high differentiation in regions linked to selected variants on the Z	Maynard Smith and Haigh 1974; Charlesworth 1996; Betancourt et al. 2004; Nachman and Payseur 2012
Selection against gene flow	Z < 3/4A	Selection against hybrids in Z-linked regions will reduce relative genetic diversity and increase differentiation on the Z, specifically	Via 2012; Nachman and Payseur 2012

Z = Z chromosome, A = autosome, Z/A = ratio of Z-linked versus autosomal genetic diversity.

Figure Captions

Figure 1. Study system, theoretical predictions of Z-linked and autosomal genetic diversity ratios with variance in reproductive success between sexes, and overview of the study. (A) Range map of the barn swallow species complex redrawn from Scordato et al. (2017). Illustrations highlight phenotypic variation between the six subspecies and are reprinted with permission from artist Hilary Burn. Bold arrows summarize the hypothesized biogeographical history of the species complex, including a recolonization of Northeastern Asia from North America. Red = *rustica*; dark red = *savignii*; lavender = *transitiva*; blue = *gutturalis*; gold = *tytleri*; turquoise = *erythrogaster*. **(B)** Theoretical predictions for the relationship between autosomal and Z-linked nucleotide diversity redrawn from Irwin (2018). The black diagonal line shows the expected relationship under equilibrium conditions, where $\pi_{ZA} = 3/4$. The shaded regions represent the theoretical ranges of values explained by excess variance in female (blue) and male (orange) reproductive success. **(C)** Predictions for the relationship between mean genetic diversity between autosomes and the Z chromosome and the π_{ZA} ratio, with the black line again representing the expectation under equilibrium conditions and shaded regions depicting ranges of values possible under excess variance in sex-specific reproductive success. **(D)** Overview of our approach, starting with comparisons of autosomal and Z-linked genetic diversity, then testing whether differences in diversity are related to allelic differentiation, rates of lineage sorting, and levels of introgression between lineages. Bold black arrows illustrate the analytical workflow and conceptual connections between questions. In each panel, grey and green lines and shapes represent hypothetical autosomal and Z-linked patterns, respectively. In the right two panels, black dashed lines depict gene trees that match the species tree shown in bold. Red dashed lines represent gene trees that do not match the species tree or are the result of introgression.

Figure 2. Autosomal and Z-linked nucleotide diversity (π) throughout the barn swallow species complex. **(A)** Distributions of autosomal π and Z chromosome π for subspecies and hybrid zones. Autosomal π distributions are shown as grey boxes and whiskers, and Z-linked distributions are shown in green. Bold horizontal lines represent median π values. The shaded region encompasses the three hybrid zone comparisons. **(B)** Relationship between autosomal and Z-linked π for specific localities, subspecies, and hybrid zones. **(C)** Relationship between mean π between autosomes and the Z chromosome and π_{ZA} ratio. Shaded regions in both panels correspond with theoretical predictions shown in Fig. 1. Open circles represent estimates for specific sampled localities within subspecies, closed black circles represent combined estimates for subspecies, and closed grey circles represent estimates for hybrid populations.

Figure 3. Autosomal and Z-linked population differentiation statistics (*Fst* and *PBS*). (A) Relationship between mean autosomal and Z-linked *Fst*. The bold diagonal line represents a hypothetical 1:1 relationship. **(B)** Relationship between Z-linked *Fst* and the ratio of Z:autosome *Fst* (Fst_{ZA}). The dashed green line shows the correlation between Fst_Z and Fst_{ZA} and the solid horizontal line represents the hypothetical relationship if Fst_Z and Fst_A were equal. In panels A and B, filled circles represent pairwise *Fst* between subspecies and open circles represent *Fst* between hybrid zones and respective parental populations. **(C)** Genome scans of lineage-specific relative differentiation (*PBS*) for autosomes (light and dark blue) and the Z chromosome (green) in 100 kb sliding windows in each barn swallow subspecies, according to their approximate order based on collared flycatcher chromosomes.

Figure 4. PCA and population structure analyses. (A) Map of sampled localities throughout the barn swallow breeding range. The colors of points represent subspecies and hybrid zones and correspond with the colors used in the PCA results in panels B-D. **(B)** Results of PCA analyses for all SNPs. **(C)** PCA results for autosome-only SNPs. **(D)** PCA results for Z chromosome-only SNPs. **(E)** Inferred ancestry assignments from ADMIXTURE analysis of Z-linked SNPs under a best-supported $K = 3$ model. Each vertical bar represents the probability of assignment of an individual to one or more genetic clusters. **(F)** Ancestry assignments from ADMIXTURE analysis of autosomal SNPs under a $K = 2$ model. Colored horizontal bars above the ancestry plots correspond to colors used in panels A-D for subspecies and hybrid zones.

Figure 5. Topology support for four sets of triplet phylogenies across the genome. In each panel, the subspecies analyzed are shown to the left and alternative topologies (i.e., t1-t3) are shown in the center column. The right side of each panel shows distributions of support as boxplots for topologies t1, t2, and t3 based on the proportion of subtrees across sliding windows for autosomes and the Z chromosome. The genome-wide best-supported topology for each triplet is highlighted. **(A)** Triplet 1: *rustica*, *savignii*, and *transitiva*. **(B)** Triplet 2: *erythrogaster*, *tytleri*, and *gutturalis*. **(C)** Triplet 3: *savignii*, *transitiva*, and *erythrogaster*. **(D)** Triplet 4: *erythrogaster*, *tytleri*, and *savignii*. Triplets 1 and 2 test 'shallow' topologies within European and Asian/North American subgroups, respectively. Triplets 3 and 4 test deeper topologies spanning the split between supported subgroups.

Figure 6. Geographic clines for autosomes and the Z chromosome in hybrid zone transects. The distance in km from the westernmost hybrid zone transect location is shown on each horizontal axis, and ancestry trait

values based on normalized genomic PC1 scores are shown on vertical axes. Maximum-likelihood clines and 95% confidence intervals for autosomes and the Z chromosome are shown as grey and green lines and shaded areas, respectively. Plus signs denote sampled populations across transects. Transects associated with migratory divides are shown in panels **A-C**. Transects not associated with known geographic barriers to gene flow are shown in panels **D-F**.

Supplementary Material Captions

Supplementary Table S1. Samples used in this study, locality data, and sequence data mapping statistics assuming a genome size of 1.2 Gbp.

Supplementary Table S2. Spearman's rank order correlation coefficients for pairwise comparisons of nucleotide diversity across the genome of barn swallow subspecies. All correlations were significant at $p <$ 2.2e-16.

Supplementary Table S3. Nucleotide diversity (π) statistics for autosomes and the Z chromsosome and the ratio of Z-linked to autosomal π (πZA) for barn swallow subspecies, hybrid zones, and specific localities. Values denoted with an asterisk were calculated after accounting for male-mutation bias. Results of Welch's two-sample *t*-tests comparing autosomal and Z chromosome π distributions are provided.

Supplementary Table S4. Summary of Tajima's *D* statistics for autosomes and the Z chromosome based on filtered datasets, including a male-only dataset, and the full dataset filtered to allow 'max-missing' thresholds of 0.2, 0.4, and 0.6 in VCFtools.

Supplementary Table S5. Relative population differentiaton (*Fst*) statistics for autosomes and the Z chromosome, along with results of Mann-Whitney ∪ tests comparing distributions. *Statistical analyses for the *tytleri* versus *rustica* x *tytleri* and *tytleri* versus *tytleri* x *gutturalis* comparisons were not possible due to a high degree of backcrossed individuals in the hybrid population.

Supplementary Table S6. Summary of population branch statistics (*PBS*) for autosomes and the Z chromosome and results of Mann-Whitney ∪ tests comparing autosomal and Z-linked distributions for each barn swallow subspecies.

Supplementary Table S7. Results of geographic cline analyses in HZAR for hybrid zone transects in Asia. Results based on autosomal and Z-linked genomic PC1 scores are presented separately, including the inferred cline model, cline center *c*, cline width *w*, and observed trait value range.

Supplementary Figure S1. Results of MashMap alignment of barn swallow (*Hirundo rustica*) reference genome scaffolds to the collared flycatcher (*Ficedula albicollis*) chromosomes, using the settings -s 50000 -f one-to-one and --pi 90. Purple and blue segments represent alignments in the forward and reverse orientation, respectively.

Supplementary Figure S2. Sampling localities for GBS dataset used for geographic cline analyses of hybrid zones in Asia, redrawn from Scordato et al. (2020). Colors correspond to previously described 'zones' or regions of parental or admixed ancestry throughout Asia. The white arrows point to locations of inferred migratory divides in Russia, Mongolia, and China.

Supplementary Figure S3. Density distributions of Tajima's *D* from 100 kb sliding windows of autosomes (grey) and the Z chromosome (green). Panels **A-F** show distributions for the six barn swallow subspecies. Panels **G-I** show distributions for samples from hybrid zone regions.

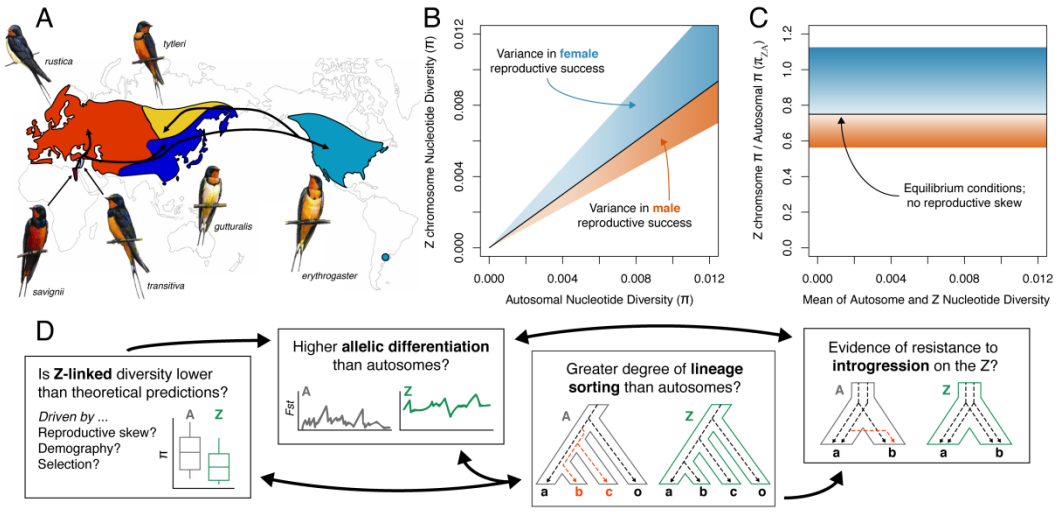
Supplementary Figure S4. Cross-validation (CV) error for increasing values of *K* genetic clusters in ADMIXTURE analyses of autosomes and the Z chromosome. Dashed lines in each panel intersect with the best-supported value of *K* (i.e., lowest CV error).

Supplementary Figure S5. ADMIXTURE results for $K = 2$, 3, 4, and 5 genetic clusters models based on Z-linked SNPs (the best-fit model based on CV error was $K = 3$, discussed in the Main Text). Vertical bars represent individuals and depict the probability of assignment to one genetic cluster or the other. Abbreviations: er = *erythrogaster*, gu = *gutturalis*, rg = *rustica* x *gutturalis*, rt = *rustica* x *tytleri* hybrids, ru = *rustica*, sa = *savignii*, tg = *tytleri* x *gutturalis* hybrids, tr = *transitiva*, ty = *tytleri*.

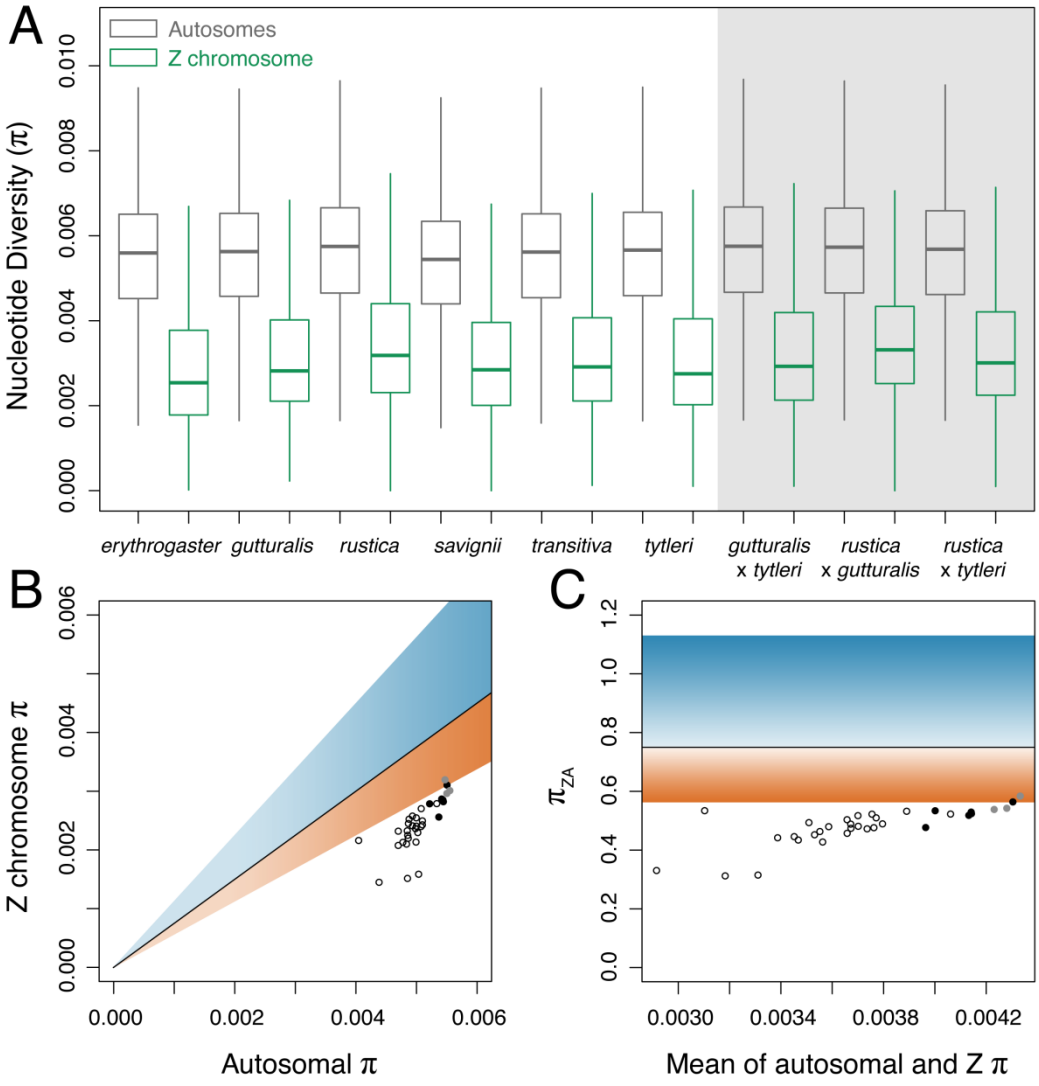
Supplementary Figure S6. Results of ABBA-BABA introgression statistics ($\hat{f}_d$) between parental subspecies and hybrid zone populations for autosomes (grey) and the Z chromosome (green). **A**. $\hat{f}_d$ between *rustica* and *rustica* x *tytleri*. The asterisk denotes that autosomal and Z-linked distributions were significantly different. **B**. $\hat{f}_d$ between *rustica* and *rustica* x *gutturalis*. **C**. $\hat{f}_d$ between *gutturalis* and *tytleri* x *gutturalis*. Abbreviations: er

= *erythrogaster,* gu = *gutturalis*, o = outgroup (*H. smithii*), rg = *rustica* x *gutturalis*, rt = *rustica* x *tytleri*, ru = *rustica*, sa = *savignii*, tg = *tytleri* x *gutturalis*.

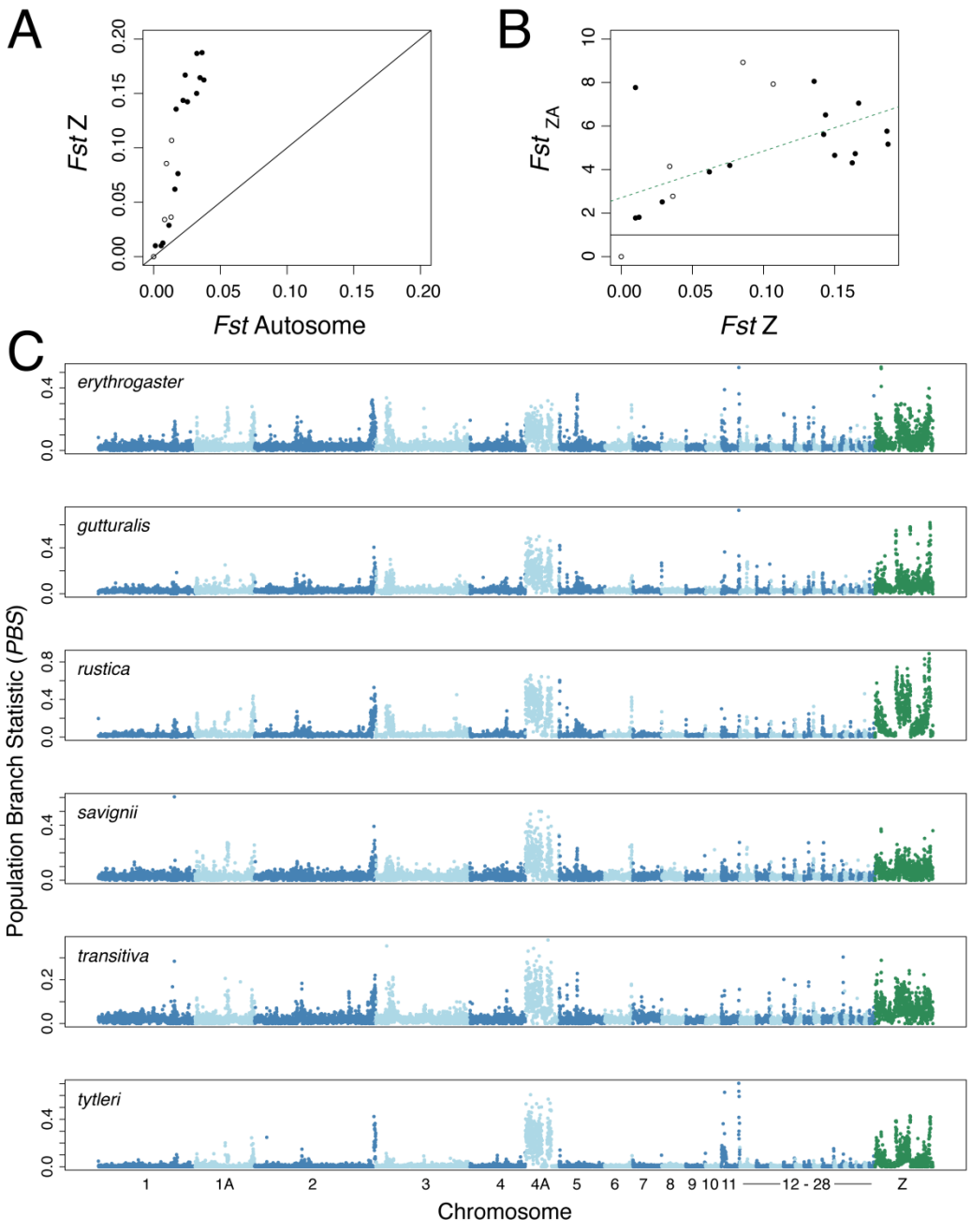
Supplementary Figure S7. Genomic locations of GBS loci analyzed in geographic cline analyses. Blue dots depict the locations of SNPs called from GBS loci and are shown over the genomic scan of *PBS* for *rustica*.



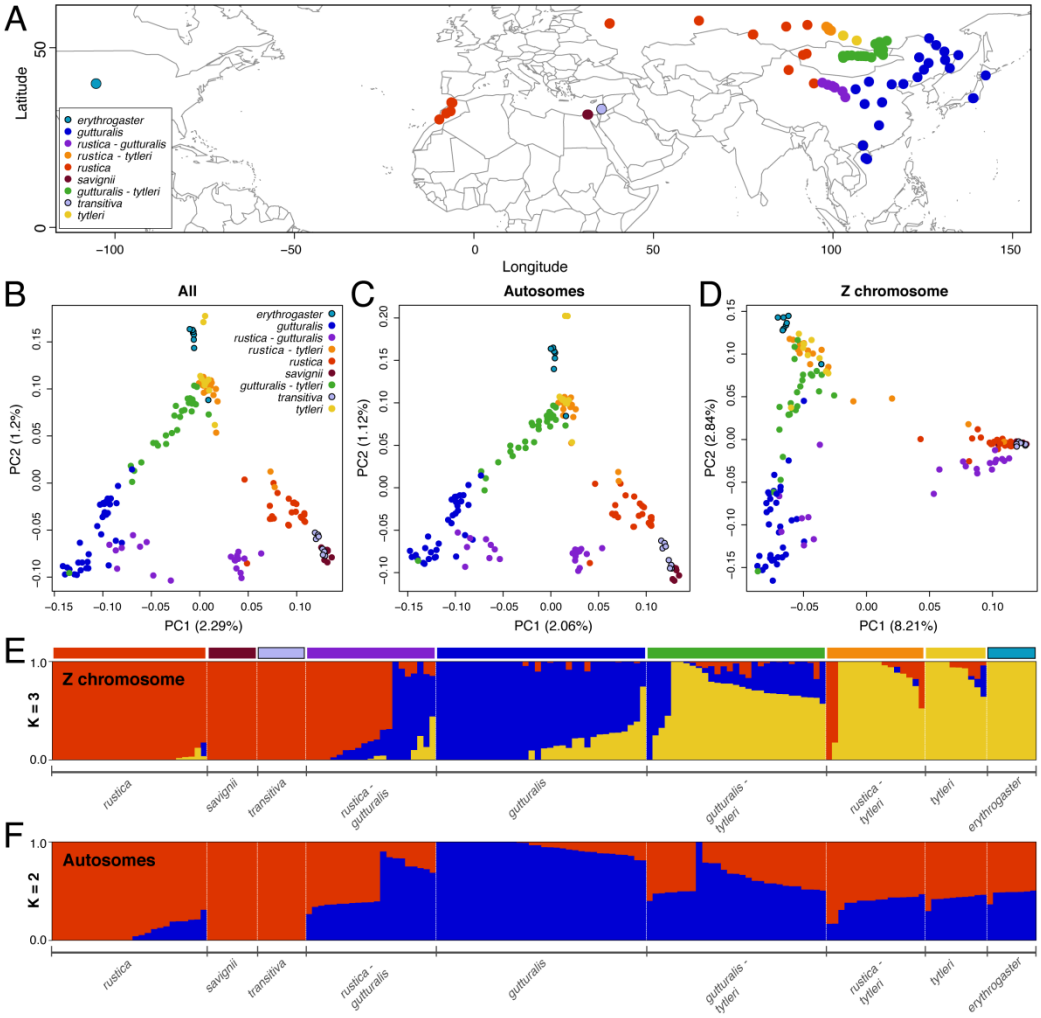
mec_15885_f1.tif



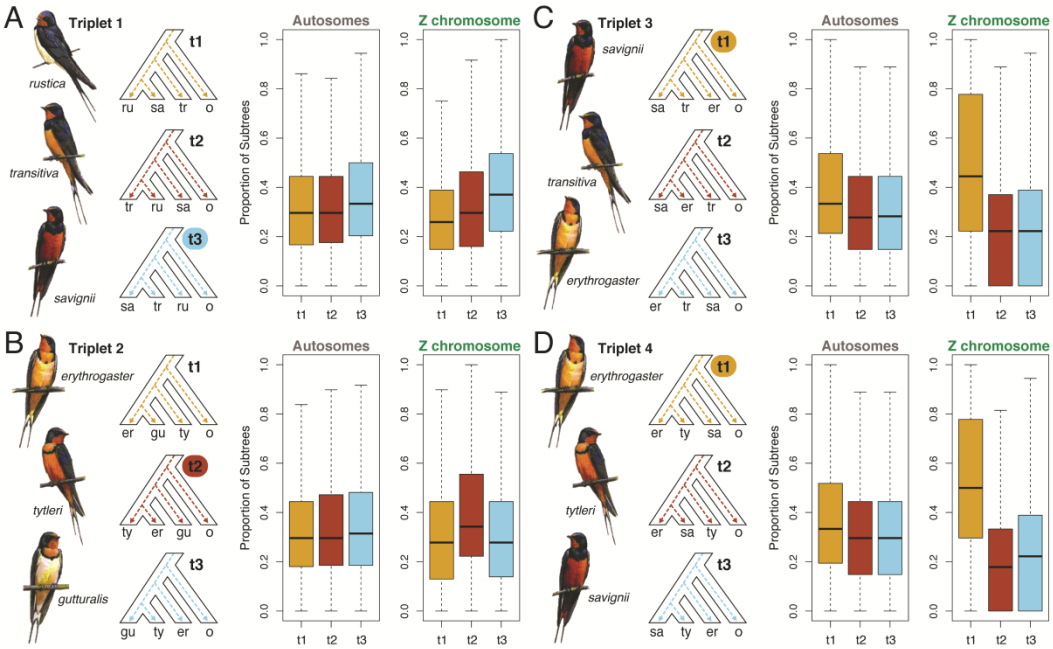
mec_15885_f2.tif



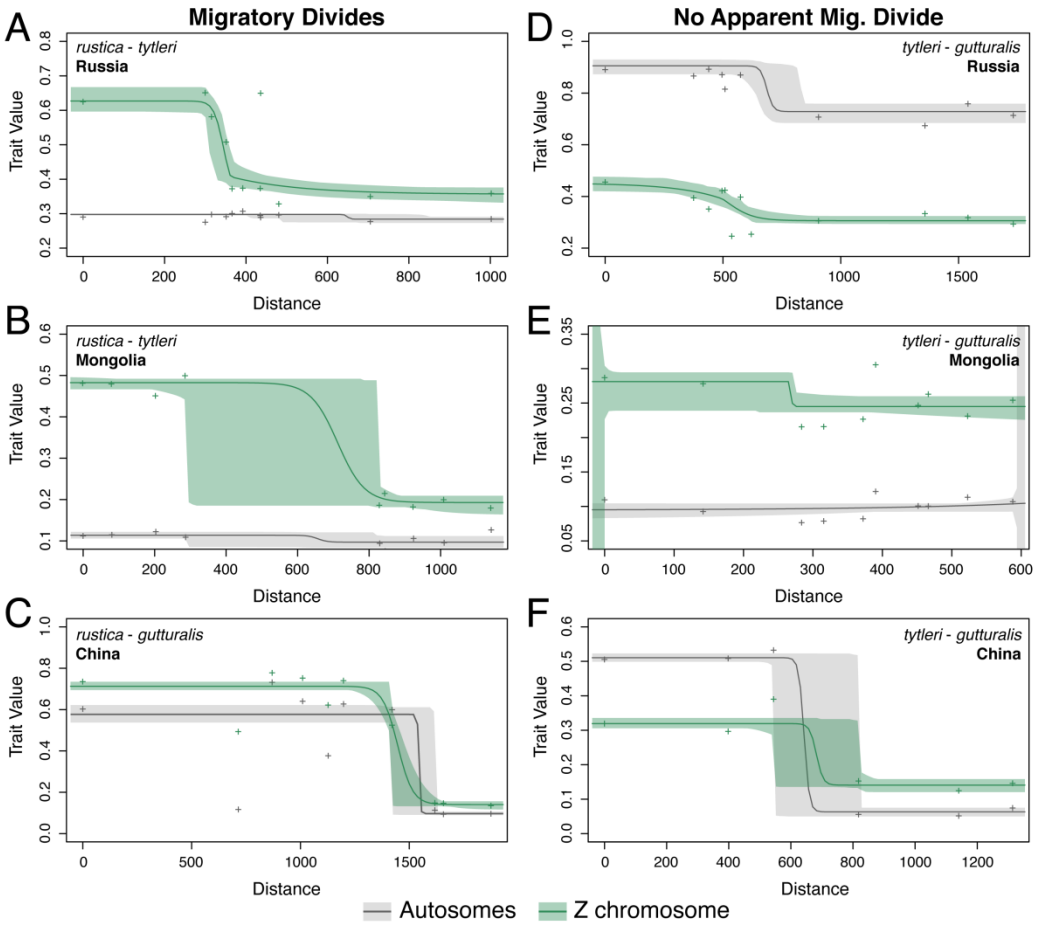
mec_15885_f3.tif



mec_15885_f4.tif



mec_15885_f5.tif



mec_15885_f6.tif