

## ORIGINAL ARTICLE OPEN ACCESS

# Recombination Rate and Recurrent Linked Selection Shape Correlated Genomic Landscapes Across a Continuum of Divergence in Swallows

Drew R. Schield<sup>1,2</sup>  | Javan K. Carter<sup>1,3</sup>  | Megan G. Alderman<sup>2</sup> | Keaka Farleigh<sup>2</sup> | Dylan K. Highland<sup>2</sup> | Rebecca J. Safran<sup>1</sup>

<sup>1</sup>Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado, USA | <sup>2</sup>Department of Biology, University of Virginia, Charlottesville, Virginia, USA | <sup>3</sup>RTI International, Research Triangle Park, North Carolina, USA

**Correspondence:** Drew R. Schield ([drew.schild@virginia.edu](mailto:drew.schild@virginia.edu))

**Received:** 26 December 2024 | **Revised:** 10 July 2025 | **Accepted:** 29 July 2025

**Handling Editor:** Yanhua Qu

**Funding:** This work was supported by Division of Biological Infrastructure (1906188, 2409958). Division of Integrative Organismal Systems (1856266). Division of Environmental Biology (1149942).

**Keywords:** divergent selection | genetic hitchhiking | passerines | population genomics | recombination | speciation

## ABSTRACT

Disentangling the drivers of genomic divergence during speciation is essential to our broader understanding of the generation of biological diversity. Genetic changes accumulate at variable rates across the genome as populations diverge, leading to heterogeneous landscapes of genetic differentiation. The ‘islands of differentiation’ that characterise these landscapes harbour genetic signatures of the evolutionary processes that led to their formation, providing insight into the roles of these processes in adaptation and speciation. Here, we study swallows in the genus *Hirundo* to investigate genomic landscapes of differentiation between species spanning a continuum of evolutionary divergence. Genomic differentiation spans a wide range of values ( $F_{ST} = 0.01–0.8$ ) between species, with substantial heterogeneity in genome-wide patterns. Genomic landscapes are strongly correlated among species ( $\rho = 0.46–0.99$ ), both at shallow and deep evolutionary timescales, with broad evidence for the role of linked selection together with recombination rate in shaping genomic differentiation. Further dissection of genomic islands reveals patterns consistent with a model of ‘recurrent selection’, wherein differentiation increases due to selection in the same genomic regions in ancestral and descendant populations. Finally, we use measures of the site frequency spectrum to differentiate between alternative forms of selection, providing evidence that genetic hitchhiking due to positive selection has contributed substantially to genomic divergence. Our results demonstrate the pervasive role of recurrent linked selection in shaping genomic divergence despite a history of gene flow and underscore the importance of non-neutral evolutionary processes in predictive frameworks for genomic divergence in speciation genomics studies.

## 1 | Introduction

Reproductive isolation evolves as populations shift away from one another in geographic space, ecological niche, species interactions, mating preferences and other key aspects

of their biology. Evidence from a wide variety of organisms (e.g., stickleback fish, Hohenlohe et al. 2010; *Ficedula* flycatchers, Ellegren et al. 2012; *Heliconius* butterflies, Nadeau et al. 2012 and Martin et al. 2013; *Populus* trees, Wang et al. 2016; *Timema* stick insects, Riesch et al. 2017; *Mimulus*

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *Molecular Ecology* published by John Wiley & Sons Ltd.

monkeyflowers, Stankowski et al. 2019; *Neodiprion* sawflies, Bendall et al. 2022) indicates that genetic divergence associated with these shifts accumulates at different rates across the genome due to the variable effects of mutation, recombination, natural and sexual selection, gene flow and genetic drift. Uncovering how these evolutionary processes influence genomic divergence is therefore paramount to our understanding of the formation of species. There is particular interest in the roles of different forms of selection in speciation, with a substantial amount of research seeking to reveal how divergent selection and barriers to gene flow promote reproductive isolation between populations; and in turn, how these manifest as genomic changes that define species (Nosil et al. 2009; Seehausen et al. 2014; Burri 2017; Wolf and Ellegren 2017). Fittingly, the analogy of genomic ‘landscapes’ has been adopted to describe the striking variation in genetic diversity and divergence across the genome, often resembling a rugged landscape with ‘islands’ or ‘peaks’ shaped by interacting evolutionary forces during speciation (Turner and Hahn 2007; Noor and Bennett 2009; Nosil et al. 2009; Nosil and Feder 2012; Martin et al. 2013; Cruickshank and Hahn 2014; Burri et al. 2015; Burri 2017). Along with the potential to reveal how specific processes shape genomic divergence between incipient species, comparing genomic landscapes among groups of related species may reveal transitions from a concentration of reproductive isolation in a small number of loci of large effect (i.e., the ‘genic’ phase of speciation, Wu 2001; Via 2009; Feder et al. 2012) to broader genome-wide differentiation, providing empirical perspectives on the timescale to a ‘genomic’ phase of speciation (Flaxman et al. 2013; Feder et al. 2014).

A growing body of evidence supports that genomic regions of elevated differentiation can arise as a consequence of alternative processes either in the presence or absence of gene flow. Some interpretations suggest that these regions harbour high differentiation (typically measured using relative measures of between-population differentiation such as  $F_{ST}$ ) because they contain loci that are causal for reproductive isolation and experience reduced gene flow relative to the rest of the genome, thus being referred to as ‘islands of speciation’ (Turner et al. 2005). However, variable rates of gene flow are not the only explanation for the formation of such regions of punctuated differentiation, as genome-wide heterogeneity can arise even between allopatric populations. Indeed, elevated differentiation may evolve after the onset of reproductive isolation as a result of locally accelerated lineage sorting due to the combined effects of recombination rate variation, linked selection and genetic drift in the absence of gene flow (Noor and Bennett 2009; Cruickshank and Hahn 2014; Burri et al. 2015; Han et al. 2017). Regions of high differentiation may be indirectly related to speciation in this case, being sometimes referred to as ‘incidental islands’ or, more generally, ‘divergence islands’ or ‘differentiation islands’ (Harr 2006; Ellegren et al. 2012; Nadeau et al. 2012; Renaut et al. 2013; reviewed in Burri 2017). Hereafter, we use the analogy of islands of differentiation to describe features of the genomic landscape and to investigate their evolutionary causes.

Alternative models have been developed to describe the evolution of islands of differentiation during speciation due to different sources of selection with and without gene flow (figure 1A,

redrawn from Cruickshank and Hahn 2014 and Irwin et al. 2016, 2018). The framework for interpreting patterns of differentiation under these models examines relationships in commonly used population genetic summary statistics—including between-population relative differentiation,  $F_{ST}$  (Weir and Cockerham 1984); between-population nucleotide divergence,  $d_{xy}$  (Nei and Li 1979); and within-population nucleotide diversity,  $\pi$  (Nei and Li 1979). These measures are related, with  $d_{xy}$  and  $\pi$  being calculated as the average proportion of nucleotides that differ between homologous sequences between and within populations, respectively.  $F_{ST}$  summarises a composition of these two measures (i.e.,  $F_{ST} = \frac{d_{xy} - \pi}{d_{xy}}$ ) and is considered a relative measure of divergence, or rather ‘differentiation’, because it is strongly influenced by within-population genetic diversity (i.e.,  $F_{ST}$  can increase due to a high proportion of allelic differences between populations but also due to low  $\pi$  within populations; Slatkin 1991; Hudson et al. 1992; Charlesworth 1998; Noor and Bennett 2009). As summary measures of nucleotide differences within and between populations,  $\pi$  and  $d_{xy}$  are also expected to be proportional to the coalescent time between pairs of homologous sequences, assuming a constant mutation rate since divergence from a common ancestor (Charlesworth 1998). Comparing these statistics across genomic regions can thus reveal information about the evolutionary processes that shape heterogeneity in the genomic landscape, which are described in three models for the formation of genomic islands of differentiation considered in this study (Figure 1). These models differ in the presence or absence of gene flow, the sources of selection driving differentiation between populations during speciation and predicted relationships between summary measures of genetic diversity and divergence.

Under the first model, ‘divergence with gene flow’ (Figure 1A, left), the geographic ranges of two diverging populations remain connected and the populations exchange alleles at some frequency. Loci contributing to reproductive isolation act as barriers to gene flow while neutrally evolving regions of the genome are comparatively porous to gene flow and allele frequencies are homogenised between populations (Wu 2001; Via 2009; Feder and Nosil 2010; Poelstra et al. 2014). As a result, we expect reproductive isolation loci (and nearby linked regions) to exhibit elevated  $F_{ST}$  with locally reduced  $\pi$  due to selection against introgression and elevated  $d_{xy}$  due to higher average coalescent times than regions under selective neutrality. Consequently,  $F_{ST}$  is expected to be negatively correlated with  $\pi$  and positively correlated with  $d_{xy}$  under the ‘divergence with gene flow’ model. Under the second model, ‘selection in allopatry’ (Figure 1A, middle), increases in  $F_{ST}$  are explained by selection within populations, wherein an ancestral population separates, and descendant populations diverge in the absence of gene flow and experience selection in distinct or shared regions of the genome (Noor and Bennett 2009; Nachman and Payseur 2012; Vijay et al. 2017). This process leads to reduced  $\pi$  within descendant populations in regions under selection whereas  $d_{xy}$  is similar for selected and neutral loci because within-population selection does not influence ancestral variation or the time to coalescence (Charlesworth 1998; Cruickshank and Hahn 2014). Under this model,  $F_{ST}$  is expected to be negatively correlated with  $\pi$  and to show no strong relationship with  $d_{xy}$ .

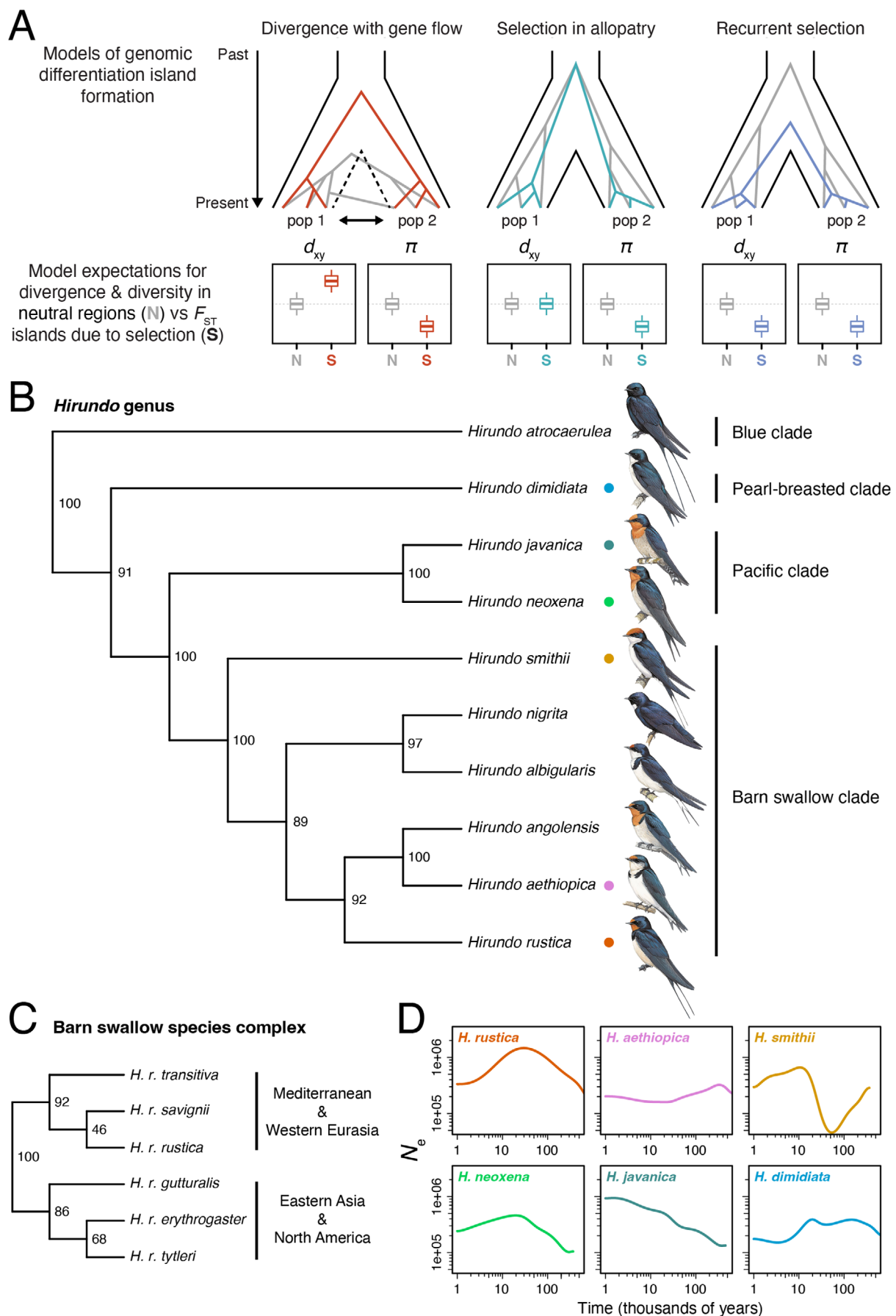


FIGURE 1 | Legend on next page.

**FIGURE 1** | Models of genomic islands of differentiation and the phylogeny and demographic history of the study system. (A) Schematic representations of models of genomic differentiation island formation, redrawn from Cruickshank and Hahn (2014) and Irwin et al. (2016, 2018). Top panels: Illustrations depict populations diverging from a common ancestor over time, with individual genealogies shown for loci in genomic regions evolving under neutrality (grey) and selection (colours representing loci in differentiation islands). Lower panels: Because the sources of selection differ among models, so do predicted patterns for between-population genetic divergence ( $d_{xy}$ ) and within-population genetic diversity ( $\pi$ ) in islands of differentiation ( $F_{ST}$ ) relative to neutral regions. (B) Phylogenetic relationships among *Hirundo* species sampled in this study, estimated using coalescent-based species tree inference on genome-wide SNPs. Nodal values indicate bootstrap support. Coloured circles highlight species represented by  $n > 1$  samples in our study. (C) Inferred relationships for the six barn swallow (*Hirundo rustica*) subspecies that fall into two major clades, one composed of subspecies in the Mediterranean and western Eurasia and the other composed of subspecies in eastern Asia and North America. (D) Estimates of effective population size ( $N_e$ ) change for six *Hirundo* species over the last 500 ka. Both x- and y-axes are shown on a  $\log_{10}$  scale. Estimates of population history were restricted to species with  $n > 1$  and colours correspond to panel (B). Swallow illustrations by Hilary Burn © Lynx Edicions.

The third model, ‘recurrent selection’ (Figure 1A, right), differs from the ‘selection in allopatry’ model as the same genomic regions experience selection in both descendant populations and their common ancestor (Cruickshank and Hahn 2014). Here, differentiation islands are expected to have low  $d_{xy}$  relative to regions of the genome under selective neutrality due to reduced ancestral variation prior to divergence (Cruickshank and Hahn 2014; Irwin et al. 2016, 2018). These same regions experience further reductions in genetic diversity due to selection in daughter populations, with associated decreases in  $\pi$  and increases in  $F_{ST}$ . Accordingly,  $F_{ST}$  islands can evolve under recurrent selection even in the absence of gene flow, where  $F_{ST}$  is expected to be negatively correlated with both  $\pi$  and  $d_{xy}$ . This explanation for the formation of differentiation islands may be especially relevant when selection occurs in shared structural features of the genome (e.g., centromeres), when the effects of linked selection are pronounced in low recombination regions (Noor and Bennett 2009; Burri et al. 2015; Han et al. 2017) or when loci are repeatedly involved in adaptation and reproductive isolation (Via and West 2008; Feder and Nosil 2010; Via 2012; Vijay et al. 2016). Because signatures of recurrent selection could be produced by either background purifying selection (Charlesworth et al. 1993) or divergent positive selection (i.e., ‘genetic hitchhiking’; Maynard Smith and Haigh 1974), additional population genetic tests are often required to distinguish between these alternative sources of selection and their effects on the genomic landscape. Importantly, genomic divergence may be underlain by multiple processes and these models are not mutually exclusive mechanisms to explain islands of differentiation. Furthermore, speciation events occurring with sudden and complete geographic, ecological or reproductive isolation are probably very rare, meaning that empirical patterns matching predictions of the ‘selection in allopatry’ and ‘recurrent selection’ models have likely also been influenced by gene flow at earlier phases of the speciation process. Nonetheless, interpreting empirical patterns in the context of this model framework enables tests of the relative importance of evolutionary processes in genomic divergence, yielding informative insights about the speciation process.

This framework and related approaches have been fruitfully applied to a wide diversity of organisms to reveal how alternative sources of selection shape genome-wide heterogeneity in differentiation at various stages of the speciation process. Of note is widespread evidence for differentiation landscapes with patterns predominately explained by recurrent selection rather than divergence with gene flow (e.g., Burri et al. 2015;

Delmore et al. 2015; Irwin et al. 2016; Stankowski et al. 2019; Jiang et al. 2023; Glover et al. 2024), including in some cases between populations with a known occurrence of historical or contemporary gene flow. An emergent property of these case studies is the ‘repeatability’ or ‘conservation’ of the genomic landscape, in which the patterns of differentiation are tightly correlated even between independent pairs of populations and species. These findings raise the possibility that the signal of divergence with gene flow has been overwritten by long-term background selection or recurrent hitchhiking in genomic regions related to reproductive isolation. Alternatively, genomic signatures of reproductive isolation in regions independent of those under recurrent selection may become swamped out as differentiation becomes more genomically widespread at intermediate and later stages of speciation. Here, studies comparing genomic landscapes in groups of species spanning a continuum of divergence and with variation in the potential for gene flow (i.e., allopatry vs. partial sympatry) hold promise for clarifying the pervasiveness of recurrent selection and its roles with respect to reproductive isolation and correlations among genomic landscapes during diversification.

Here, we study the evolution of the genomic divergence landscape across the swallow genus *Hirundo*. Our sampling of whole genomes includes 10 *Hirundo* species (the majority of species in the genus; Table S1) and spans multiple scales of divergence, from very recent divergence between subspecies of barn swallow (*H. rustica*) ~11,000 years ago (Smith et al. 2018; Schield, Carter, et al. 2024) to the divergence from the *Hirundo* common ancestor ~5 million years ago (Schield, Brown, et al. 2024), enabling us to study how genomic differentiation has been shaped at various stages of the speciation process. Species of *Hirundo* have an inferred African origin (Zink et al. 2006; Dor et al. 2010; Schield, Brown, et al. 2024), with a number of species occupying contemporary geographic distributions within sub-Saharan Africa (Figure S1; del Hoyo et al. 2004; Winkler et al. 2020) while others have expanded into Eurasia, Southeast Asia, Oceania and North America. Barn swallows, in particular, are among the most widespread songbirds in the world, having expanded to occupy a breeding distribution across the Holarctic (Zink et al. 2006). *Hirundo* species also vary in their degree of social versus solitary nesting, sedentary versus migratory life histories and plumage features, including variation in dorsal and ventral melanism, presence or absence of colourful throat patches and the length of forked tail feathers (Turner and Rose 1989; Turner 2018). These plumage traits are used to signal to potential mates in barn swallows, which have become a model for studying the role of



divergent sexual selection in phenotype evolution and speciation (Møller 1988; Saino et al. 1997; Safran et al. 2005; Scordato and Safran 2014; Wilkins et al. 2015; Safran, Vortman, et al. 2016; Romano et al. 2017; Hund et al. 2020; Lotem et al. 2022).

Barn swallows have also been the focus of genomic studies to examine comparative genomic structure (Formenti et al. 2019; Secomandi et al. 2023), population genetic structure (Safran, Scordato, et al. 2016), sex chromosome evolution (Schield et al. 2021), the interplay between selection and gene flow in hybrid zones (Scordato et al. 2017, 2020; Turbek et al. 2022; Schield, Carter, et al. 2024), historical demography (Smith et al. 2018; Lombardo et al. 2022) and the genetic basis of reproductive isolation (Schield, Carter, et al. 2024). A recent study also examined population genetic structure, phylogeography and demography in the Pacific swallow clade (*H. neoxena*, *H. javanica* and relatives) using genomic data (Broyles et al. 2023). These studies provide rich information about the evolutionary processes shaping genomic variation within and between closely related populations, yet these perspectives are limited to the early stages of the speciation process. Here, we combine genomic resources for barn swallows with whole genomes from a majority of other *Hirundo* species to gain broader perspectives on drivers of genomic landscapes at both early and later stages of evolutionary divergence. Using this combined data set, we address the following questions: (i) how correlated are genomic landscapes of divergence among *Hirundo* species; (ii) are genomic islands of differentiation explained by one or more linked selection models with associated sources of selection (e.g., Figure 1A); (iii) does the potential for gene flow influence support for alternative models of differentiation; and (iv) how prevalent is divergent positive selection in the formation of islands of differentiation during speciation? By answering these questions through investigation of an entire radiation of species, we may gain a more complete understanding of how selection interacts with other evolutionary forces to shape patterns of genomic divergence during speciation.

## 2 | Materials and Methods

### 2.1 | Sampling, Genome Sequencing and Variant Calling

We obtained tissue samples for 47 individuals representing nine species of *Hirundo* and outgroup *Petrochelidon pyrrhonota* and generated whole genome sequencing data to complement genomic data for barn swallows (*H. rustica*) and an individual wire-tailed swallow (*H. smithii*) generated previously (Smith et al. 2018; Schield et al. 2021; Schield, Carter, et al. 2024). For new samples, we extracted genomic DNA using Qiagen DNeasy kits following the manufacturer's protocol, quantified purified DNA concentrations using a Qubit fluorometer and constructed genome sequencing libraries using Illumina Nextera Flex kits at the University of Colorado BioFrontiers Institute. We then sequenced the genome libraries at Novogene on Illumina NovaSeq 6000 lanes using 150bp paired end reads, targeting a fold-coverage of 10× per individual. We combined the newly sequenced data with previously generated data for a total sampling of 196 individuals (Data S1), including *H. atrocerulea* (*n* = 1), *H. dimidiata* (*n* = 8), *H. javanica* (*n* = 5), *H. neoxena* (*n* = 10), *H.*

*smithii* (*n* = 10), *H. nigrita* (*n* = 1), *H. albigularis* (*n* = 1), *H. angolensis* (*n* = 1), *H. aethiopica* (*n* = 5), *H. rustica rustica* (*n* = 43), *H. r. savignii* (*n* = 12), *H. r. transitiva* (*n* = 13), *H. r. gutturalis* (*n* = 46), *H. r. tytleri* (*n* = 21), *H. r. erythrogaster* (*n* = 16) and outgroup *P. pyrrhonota* (*n* = 3). All sequencing data used in this study are available on the NCBI short-read archive (accession PRJNA323498).

We used Trimmomatic v0.39 (Bolger et al. 2014) to quality trim and filter raw reads using the settings LEADING:20 TRAILING:20 MINLEN:32 AVGQUAL:30. We then used BWA 'mem' v0.7.17 (Li and Durbin 2009) with default settings to map filtered reads to the North American barn swallow (*H. r. erythrogaster*) reference genome assembly bHirRus1 (Secomandi et al. 2023), with scaffold-to-chromosome assignments based on Schield, Carter, et al. (2024). We used Samtools v1.10 (Li et al. 2009) to sort mapped reads and quantify coverage statistics. We called variants using the GATK v4.0.8.1 best-practice workflow (McKenna et al. 2010; Van der Auwera et al. 2013). We first ran 'HaplotypeCaller' to call individual variants using the '--ERC GVCF' option, then ran 'GenotypeGVCFs' to call variants among the cohort of total samples and generate an 'all-sites' VCF consisting of both variant and invariant genotypes. We used GATK 'VariantFiltration' to flag genotypes failing the following filtering thresholds: 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). We identified heterozygous genotypes in females on the Z chromosome and conservatively masked these sites in all individuals. We also masked genotypes in repetitive regions annotated in Schield, Carter, et al. (2024). Finally, we used BCFtools v1.10.2 (Li et al. 2009) to recode indels, masked sites and sites flagged using the filters above as missing genotypes. We applied additional filters in BCFtools and VCFtools v0.1.17 (Danecek et al. 2011) to extract biallelic SNPs based on the proportion of missing genotypes and/or minor allele frequency for specific analyses (see below).

### 2.2 | Phylogeny, Demographic History and Introgression

We estimated phylogenetic relationships within *Hirundo* using concatenated maximum likelihood and coalescent species tree approaches. We selected a random individual for each taxon with *n* > 1 (including outgroup *P. pyrrhonota*), then filtered to retain biallelic autosomal SNPs with no missing genotypes, which we further thinned to retain a single SNP per 10kb. We converted the SNP data set into alignments using 'vcf2phyliip.py' (<https://github.com/edgardomortiz/vcf2phyliip>) and pruned invariant sites among the focal samples using 'ascbias.py' ([https://github.com/btmartin721/raxml\\_ascbias](https://github.com/btmartin721/raxml_ascbias)). We first performed maximum likelihood analysis using RAXML-NG v0.7.0 (Kozlov et al. 2019) with the concatenated SNP alignment as a single partition, specifying the GTGTR4+G+ASC\_LEWIS substitution model to account for SNP ascertainment bias (Lewis 2001). We performed 10 initial parsimony tree searches and assessed nodal support for the best tree using 100 bootstrap replicates. We then used SVDquartets (Chifman and Kubatko 2014, 2015), implemented in PAUP\* v4.0 (Swofford 2003), to estimate the species tree in a coalescent

framework based on support for taxon quartets from site patterns in the alignment. We specified *P. pyrrhonota* as a monophyletic outgroup in the analysis and assessed support for the inferred species tree topology using 100 bootstrap replicates.

We inferred demographic histories of *Hirundo* species using the sequentially Markov coalescent model implemented in 'SMC++' v1.15.2 (Terhorst et al. 2017). We focused these analyses on species with  $n > 1$  in our data set to enable estimates of effective population sizes at both deep and shallow timescales (sampling multiple individuals provides greater resolution of coalescent events in the recent past; Schiffels and Durbin 2014). For *H. rustica*, we focused analysis on a representative population of *H. r. rustica* sampled in Karasuk, Russia ( $n = 10$ ), though a previous analysis demonstrated that choice of focal population should not strongly influence inference of historical demography (Smith et al. 2018). We chose diploid genotypes from five random individuals of each species to represent 'distinguished' lineages and converted autosomal SNPs to SMC input format using the 'vcf2smc' function, masking long runs of homozygosity 50 kb using the option '-c 50000'. We then ran 'estimate' to fit cubic spline models of population size between 1000 and 500,000 generations using a composite likelihood based on the sum of log-likelihoods for each pair of distinguished lineages, assuming a per-generation mutation rate of  $2.3 \times 10^{-9}$  (Smeds et al. 2016) and a generation time of 1 year (Zink et al. 2006).

To test for evidence of introgression between species, we calculated Patterson's *D* statistics (Durand et al. 2011) based on ABBA-BABA tests of derived allele patterns in 'Dsuite' v0.5 (Malinsky et al. 2021). ABBA-BABA tests use a four-taxon topology (((P1, P2), P3), O), with an expected ancestral (A) and derived (B) allele pattern of BBAA (Martin et al. 2013). However, alternative ABBA or BABA patterns can be produced through incomplete lineage sorting (ILS), introgression or a combination of these processes (Malinsky et al. 2021). Equal frequencies of ABBA and BABA patterns are expected under ILS ( $D = 0$ ; Malinsky et al. 2021), whereas introgression between P3 and either P1 or P2 will result in an excess of ABBA or BABA patterns and a corresponding *D* statistic that deviates significantly from 0 (assessed using a *Z*-score). We compared the frequency of ABBA and BABA site patterns and calculated Patterson's *D* using the 'Dtrios' function with default settings, which tests all possible trios of populations in our data (i.e., P1, P2 and P3), with *Petrochelidon pyrrhonota* as the outgroup taxon. We corrected for multiple testing using the false discovery rate (FDR; Benjamini and Hochberg 1995) and considered any *D* statistic with an FDR-corrected *p*-value  $\leq 0.05$  as evidence of introgression. We also ran 'Dtrios' using 30 jack-knife blocks to assess significance instead of the default value of 20 to determine if different parameter settings influenced inferences of introgression. These analyses yielded identical results; therefore, we report only results from analysis using default settings. For a detailed comparison of introgression with the genomic landscape of differentiation, we also measured  $f_d$  (Martin et al. 2015), which is proportional to the effective migration rate. We performed  $f_d$  analyses using 'ABBABABAwindows.py' ([https://github.com/simonmartin/genomics\\_general](https://github.com/simonmartin/genomics_general)) between *H. smithii* and *H. aethiopica* and *H. dimidiata*, respectively, as these species are represented by  $n > 1$  samples in our dataset and also occur in parapatry or partial sympatry with the potential for recent or contemporary gene flow. We calculated  $f_d$  in sliding window resolutions matching other statistics (see 'Population

genetic summary statistics' section below) and required at least 100 biallelic SNPs be present in each window to be included in analysis.

## 2.3 | Recombination Rate and Exon Density

We estimated recombination rates in *Hirundo* species using 'pyrho' v0.1.0 (Kamm et al. 2016; Spence and Song 2019), which uses composite likelihood to infer the per-generation recombination rate from population genomic data by explicitly incorporating an estimate of demographic history. We retained biallelic SNPs that were polymorphic within each species (but which could be either sorting or fixed between species) and only sampled males on the Z chromosome. We then ran the 'lookup' function to generate a likelihood lookup table based on the sample size and population size history inferred using 'SMC++' for each species. After generating lookup tables, we ran 'hyperparam' to assess the fit of block penalty and window size hyperparameters to the data. We then ran 'optimize' to estimate recombination rates under a block penalty of 10 and window size of 50, scaled by the assumed per-generation mutation rate used for demographic inference (Smeds et al. 2016). To examine genome-wide variation in recombination rate, we used 'bedtools' v2.31.0 (Quinlan and Hall 2010) to calculate the mean recombination rate in non-overlapping windows of various resolutions (e.g., 1 Mb, 100 kb, 50 kb). We also calculated the mean recombination rate in sliding windows with intermediate step sizes (e.g., 1 Mb windows with 100 kb step) to visualise measures of between-species divergence and within-species genetic diversity in the context of recombination rate variation. We calculated Spearman's rank correlation coefficients to assess the conservation of the genomic recombination landscape among *Hirundo* species; all statistical analyses were performed using values calculated in non-overlapping windows. To obtain a measure of the density of targets of selection across the genome, we measured exon density as the proportion of sites per genomic window annotated as exons in the barn swallow genome annotation (Secomandi et al. 2023).

## 2.4 | Population Genetic Summary Statistics and Summaries of the Site Frequency Spectrum

We used 'pixy' v1.2.7beta1 (Korunes and Samuk 2021) to calculate between-population genetic differentiation (Weir and Cockerham's  $F_{ST}$ ; Weir and Cockerham 1984), between-population nucleotide divergence ( $d_{xy}$ ; Nei and Li 1979) and within-species nucleotide diversity ( $\pi$ ; Nei and Li 1979) across the genome, using both variant and invariant site information in the 'all-sites' VCF describe above. We performed analysis in non-overlapping 1-Mb, 100-kb and 50-kb windows as well as 1-Mb sliding windows with a 100-kb step size, as done for recombination rate. Because  $F_{ST}$  could not be determined between pairs of species with  $n = 1$  (e.g., *H. angolensis* vs. *H. atrocaerulea*), we removed these comparisons from further analysis. We otherwise performed pairwise analysis between all species, and all subspecies within *H. rustica*.

To detect signatures of selection in genomic islands of differentiation, we measured two summaries of the site frequency

spectrum, Tajima's  $D$  (Tajima 1989) and Fay & Wu's  $H$  (Fay and Wu 2000) for *Hirundo* species with  $n > 1$ . Tajima's  $D$  compares the mean number of pairwise genetic differences ( $\theta_\pi$ ) to the number of segregating sites ( $\theta_s$ ) in a sequence to detect departures from selective neutrality. Divergent positive selection is expected to produce an excess of rare alleles as new mutations arise following a selective sweep (Hermisson and Pennings 2017), producing a skew in the site frequency spectrum and lower values of Tajima's  $D$  due to larger  $\theta_s$  within selected loci and linked neutral variation. However, long-term purifying selection against deleterious mutations within the same genomic regions may also produce low values of  $D$ , making it potentially difficult to disentangle the effects of divergent positive selection in the presence of background selection (Enard et al. 2014). To address this limitation, we also measured Fay & Wu's  $H$ , which leverages an outgroup to characterise the frequency of derived variants in the ingroup. Divergent selection is expected to produce an excess of high-frequency derived alleles at selected loci, yielding low values of Fay & Wu's  $H$  (like Tajima's  $D$ ). Background selection, by contrast, does not generate excess derived alleles and therefore does not skew  $H$  in the same way that it might  $D$  (Fay and Wu 2000). These measures of the site frequency spectrum can therefore be combined to distinguish the effects of selection more precisely in differentiation islands (especially in regions of low recombination; Frankham 2012). We measured Tajima's  $D$  using 'VCF-kit' v0.3.0 (Cook and Andersen 2017). We measured Fay & Wu's  $H$  using the 'calcs\_sfs\_tests' function in the R package 'rehh' v3.2.2 (Gautier et al. 2017) after polarising ancestral versus derived variants using 'polarizeVCFbyOutgroup.py' (<https://github.com/kullrich/bio-scripts/blob/master/vcf/polarizeVCFbyOutgroup.py>) with *H. atrocaerulea* as the outgroup.

## 2.5 | Comparative Analyses

We tested for broad evidence of alternative models of genomic divergence by examining relationships between landscapes of differentiation ( $F_{ST}$ ), nucleotide divergence ( $d_{xy}$ ) and nucleotide diversity ( $\pi$ ) using Spearman correlation coefficients ( $\rho$ ), calculated based on summary statistics in non-overlapping 1-Mb windows. We further quantified the degree of conservation in genomic landscapes of divergence and diversity by comparing correlations between  $\pi \sim \pi$ ,  $d_{xy} \sim d_{xy}$  and  $d_{xy} \sim \pi$  as a function of the total phylogenetic distance between landscapes, which we determined based on branch lengths between representative descendant tips ( $\pi$ ) and ancestral nodes ( $d_{xy}$ ) in our concatenated maximum likelihood phylogeny. This approach was used recently to examine genomic landscape correlation in Great Apes (Rodrigues et al. 2024) and follows from the expectation that correlations between divergence and diversity will decrease rapidly with increasing genetic distance in the absence of shared evolutionary processes. By contrast, conservation of landscapes may be stronger when shared processes shape genome-wide heterogeneity in divergence between populations and diversity within populations (e.g., through conservation of recombination rate and shared effects of long-term linked selection; Burri 2017).

Preliminary analyses revealed a positive genome-wide relationship between exon density and recombination rate (see Results). To test for associations between these genomic features and population genetic summary statistics, we performed multiple

linear regression to model the explanatory effects of both recombination rate and exon density on  $F_{ST}$ ,  $d_{xy}$  and  $\pi$ . Linear models followed the general form  $y \sim \text{rate} + \text{density}$ , where  $y = F_{ST}$ ,  $d_{xy}$  or  $\pi$ . We dissected the relationship between exon density and recombination rate further by calculating the Spearman correlation coefficient between variables measured in 50 kb windows per chromosome and compared correlations to chromosome length.

To further investigate how alternative forms of selection have shaped *Hirundo* genomic landscapes, we defined differentiation islands as outlier windows with  $F_{ST}$  above null distributions generated using a permutation approach. For a given sliding window resolution, we randomly sampled smaller windows from across the genome with a total length equal to the focal window size and calculated mean  $F_{ST}$  across the random sample of windows, repeating this process a number of times equal to the number of focal windows in the genome. For example, to generate a null distribution of  $F_{ST}$  in 1 Mb windows, we sampled 10 random 100-kb windows per 1-Mb window. This approach enabled us to detect if windows represent outlier  $F_{ST}$  islands that are physically clustered in the genome beyond what would be expected by chance. We defined outlier windows as those with  $F_{ST}$  above the maximum value in the null distribution for a given pair of species. We omitted the Z chromosome from these analyses due to the overall higher degree of Z-linked differentiation, especially between more recently diverged species (see Results). We extracted the associated  $d_{xy}$  and  $\pi$  values from these windows to compare the distributions of divergence and diversity in islands versus the genomic background. For additional comparison, we examined distributions of summary statistics in islands overlapping centromere regions versus those outside of centromeres. We performed one-way analysis of variance (ANOVA) to test for significant differences in the means of  $d_{xy}$  and  $\pi$  in genomic backgrounds, all islands and islands outside of centromeres. If ANOVA was significant, we then performed Tukey *post hoc* tests to test pairwise differences between distributions while correcting for multiple testing. Finally, we tested whether differentiation islands were enriched for signatures of divergent positive selection based on skewed summaries of the site frequency spectrum. We identified Tajima's  $D$  and Fay & Wu's  $H$  outliers using permutations as described above for  $F_{ST}$  islands, defining outliers consistent with divergent positive selection as values of  $D$  or  $H$  falling below the minimum value in the null distribution of each species. We then examined proportions of  $F_{ST}$  islands overlapping both  $D$  and  $H$  outliers compared to the genomic background and used Fisher's exact tests to test for enrichment of islands for signatures of positive selection. All statistical analyses were performed in R v1.4.2 (R Core Team 2023).

## 3 | Results

### 3.1 | Genome Sequencing and Variant Calling

Our whole genome sequencing procedure yielded a mean  $\pm$  standard deviation of  $99,375,637 \pm 40,954,922$  150bp paired end reads per sample after quality trimming. An average of  $98,004,447 \pm 40,462,032$  reads mapped to the reference genome, corresponding to  $98.6\% \pm 0.006\%$  mapped reads and  $12 \pm 4.9$  read depth per sample, assuming a genome size of 1.2



Gbp (Table S1). Variant calling and filtering steps produced 109,757,230 genome-wide SNPs (mean  $\pm$  standard deviation genotype quality =  $47.6 \pm 21.6$ ) among the *Hirundo* ingroup and the outgroup *Petrochelidon pyrrhonota* and 88,985,001 SNPs within ingroup *Hirundo* for analysis.

### 3.2 | Phylogeny, Demographic History and Introgression

Phylogenetic inference using concatenated maximum likelihood and coalescent species tree approaches based on an alignment of 138,156 autosomal SNPs (see Materials and Methods) produced well supported and consistent relationships among *Hirundo* species (Figures 1B,C, S2 and S3). Our inferred topologies align with previous phylogenetic hypotheses supporting four major clades within *Hirundo* (Figure 1B): the blue clade, pearl-breasted clade, Pacific clade and barn swallow clade (Dor et al. 2010; Carter et al. 2020; Schield, Brown, et al. 2024). We find two topological differences between our current analysis and previous hypotheses from Dor et al. (2010) based on sequencing a nuclear gene and six mitochondrial genes, and Carter et al. (2020) based on mitochondrial genomes. First, our results support the sister relationship between *H. albigularis* and *H. nigrita*, whereas *H. albigularis* was inferred to be the outgroup species to the remaining barn swallow clade in both Dor et al. (2010) and Carter et al. (2020). Second, our tree provides strong support that *H. smithii* is sister to the remaining barn swallow clade, within which *H. albigularis* + *H. nigrita* are nested (Figures 1B and S2). Inferred relationships within the barn swallow clade are in agreement with the more recent phylogenetic hypothesis for the swallow family Hirundinidae based on genome-wide ultraconserved elements (Schield, Brown, et al. 2024).

Relationships within the barn swallow (*H. rustica*) strongly support two major clades: a ‘western’ clade comprised of subspecies in the Mediterranean and western Eurasia (*H. r. transitiva*, *H. r. savignii* and *H. r. rustica*) and an ‘eastern’ clade comprised of subspecies in eastern Asia and North America (*H. r. gutturalis*, *H. r. erythrogaster* and *H. r. tyleri*). The presence of these two clades is also consistent with strong evidence from prior studies for more substantial genetic structure between ‘western’ and ‘eastern’ subspecies groups (Dor et al. 2010; Safran, Scordato, et al. 2016; Carter et al. 2020; Schield et al. 2021). We find minor disagreement between maximum likelihood and coalescent-based inferences for the western clade, with *H. r. rustica* being alternatively supported as sister to *H. r. savignii* + *H. r. transitiva* or to *H. r. savignii* (Figures 1C and S2), albeit with low bootstrap support values in both analyses. Low nodal support values and discordance between approaches are unsurprising for barn swallow subspecies, given their very recent divergence from a common ancestor and thus the prevalence of shared ancestral variation (Zink et al. 2006; Safran, Scordato, et al. 2016; Smith et al. 2018; Schield et al. 2021; Schield, Carter, et al. 2024).

We used coalescent models to infer the effective population size ( $N_e$ ) history for six *Hirundo* species represented by  $n > 1$  in our sampling design. Our results indicate that *Hirundo* species have experienced substantial fluctuations in  $N_e$  over the course of diversification (Figure 1D), with highly idiosyncratic trajectories among species through time. For instance, several

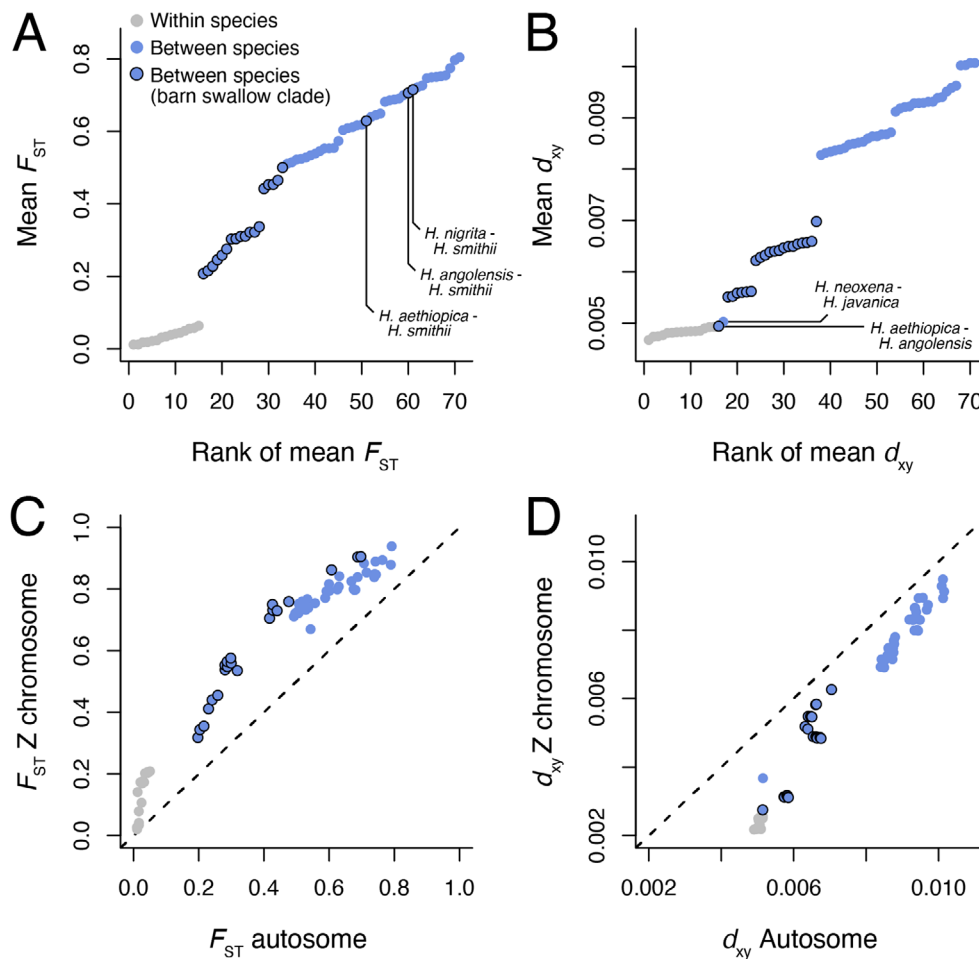
species have experienced population bottlenecks in the recent past (~10,000 years) following substantial population expansions during the Pleistocene (e.g., *H. rustica*, *H. smithii* and *H. neoxena*; see also Smith et al. 2018). The population trajectory of *H. smithii* is particularly notable due to massive population growth after near extinction roughly 50,000 years ago, followed again by recent decrease from  $N_e \sim 7 \times 10^5$  to  $N_e \sim 2 \times 10^5$ . *H. neoxena* appears to have experienced recent contraction in  $N_e$ , though to a lesser degree than *H. smithii* and *H. rustica*. Other species have experienced population expansion in the recent past, with modest increases in  $N_e$  in *H. aethiopica* and *H. dimidiata* (following an inferred bottleneck ~10,000–20,000 years ago) and a large-scale expansion in *H. javanica* during the Holocene.

Tests of introgression using the ABBA-BABA framework further support that diversification within *Hirundo* has occurred in the presence of gene flow, with introgression detected among multiple ancestral branches based on patterns of derived allele inheritance among trios of *Hirundo* species (Figure S4; Data S1). Importantly, we conservatively interpret evidence for introgression between one taxon and all taxa sharing a more recent common ancestor as evidence of introgression between ancestral branches (e.g., positive  $D$  values between *H. atrocaerulea* and each species within the barn swallow clade with  $p$ -values  $< 0.05$  after FDR correction). Using this interpretation, we detect evidence for introgression between six branches, including substantial historical introgression among species within the barn swallow clade (Figure S4). In descending order of time since divergence (labelled A–F in Figure S4), these include introgression between the ancestors of (A) *H. atrocaerulea* and the barn swallow clade (*H. smithii*, *H. nigrita*, *H. albigularis*, *H. angolensis*, *H. aethiopica*, *H. rustica* and presumably *H. lucida*, which was not sampled in this study), (B) *H. dimidiata* and the barn swallow clade, (C) the Pacific clade (*H. javanica*, *H. neoxena* and relatives) and *H. smithii*, (D) the Pacific clade and the (*H. rustica*, (*H. aethiopica*, *H. angolensis*)) subclade within the barn swallow clade, (E) *H. smithii* and the (*H. rustica*, (*H. aethiopica*, *H. angolensis*)) subclade, and (F) the (*H. nigrita*, *H. albigularis*) and (*H. rustica*, (*H. aethiopica*, *H. angolensis*)) subclades. These results provide key context for our interpretations of genomic divergence landscapes and underscore the need to consider the possibility that genomic islands of differentiation have been shaped at least in part by differential gene flow during speciation within *Hirundo*.

### 3.3 | Genomic Landscapes of Differentiation, Divergence and Diversity

To investigate how genomic divergence has evolved during speciation in *Hirundo*, we first calculated genome-wide relative population differentiation ( $F_{ST}$ ) and nucleotide divergence ( $d_{xy}$ ) between pairs of species. Although  $F_{ST}$  is a relative measure of divergence because it is influenced by levels of within-population nucleotide diversity ( $\pi$ ), whereas  $d_{xy}$  is not (Charlesworth 1998; Noor and Bennett 2009; Cruickshank and Hahn 2014), genome-wide  $F_{ST}$  should nonetheless provide a representative picture of the overall degree of divergence between species and subspecies; indeed, average genome-wide  $F_{ST}$  and  $d_{xy}$  are positively correlated (Spearman rank correlation,  $\rho = 0.88$ ,  $p < 2.2 \times 10^{-16}$ ; Figure S5). We find a large range of genome-wide



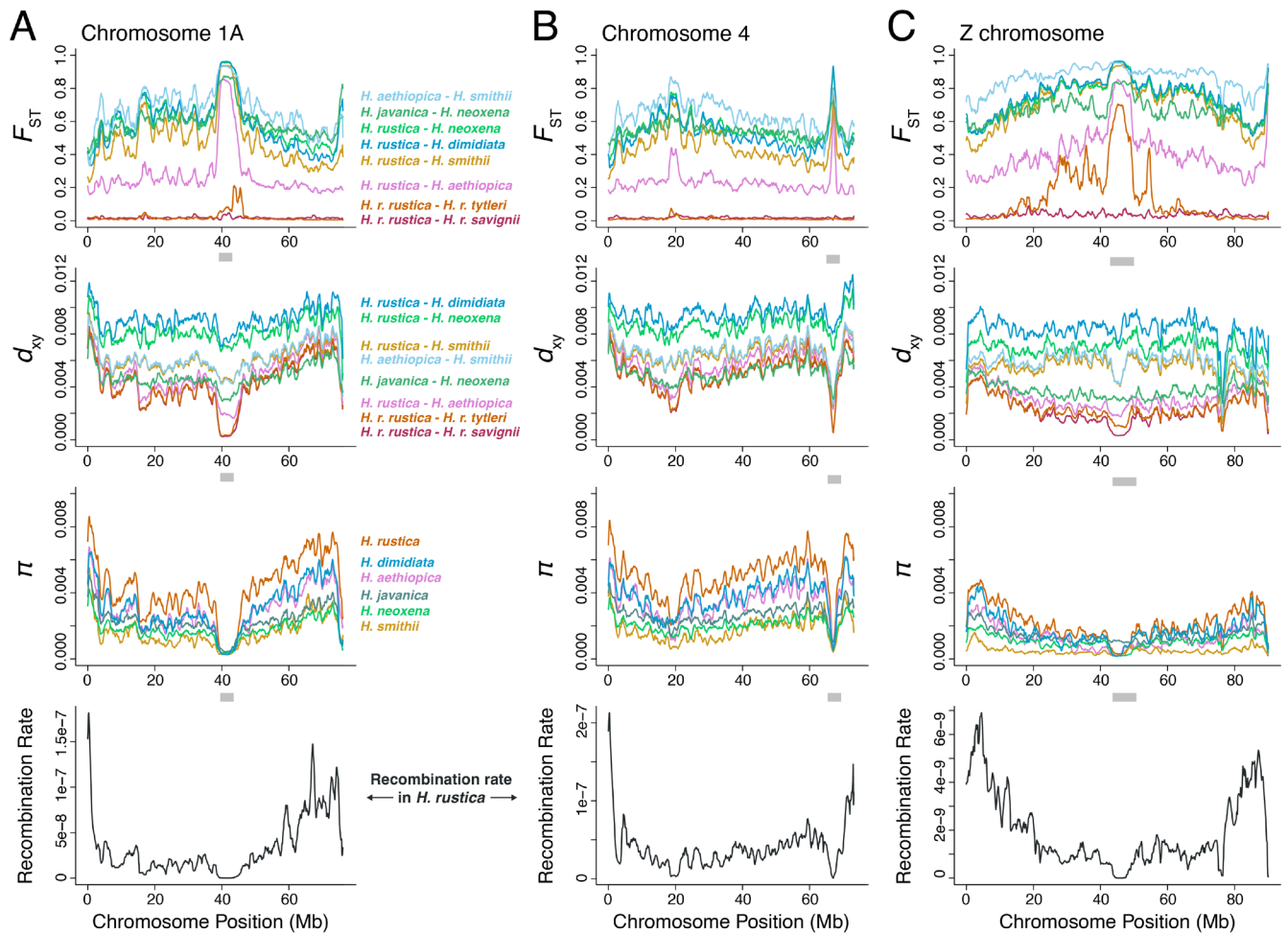


**FIGURE 2** | Genome-wide population differentiation ( $F_{ST}$ ) and nucleotide divergence ( $d_{xy}$ ). (A) Points represent rank order mean genome-wide  $F_{ST}$  between pairs of *Hirundo* species and subspecies. Grey points are  $F_{ST}$  values for pairs of barn swallow subspecies (i.e., within-species comparisons). Closed blue points are  $F_{ST}$  values between species within the barn swallow clade (*H. rustica*, *H. aethiopica*, *H. angolensis*, *H. albigularis*, *H. nigrita* and *H. smithii*). Open blue points are values for all other between-species comparisons (e.g., *H. rustica* vs. *H. atrocaerulea*). High  $F_{ST}$  values for several pairs of species within the barn swallow clade are labelled. (B) Points show rank order mean genome-wide  $d_{xy}$  between pairs of subspecies and species. Low  $d_{xy}$  values for pairs of species within clades are labelled. (C) Mean autosomal and Z-linked  $F_{ST}$ . (D) Mean autosomal and Z-linked  $d_{xy}$ . Dashed diagonal lines in (C) and (D) represent hypothetical 1:1 relationships between population differentiation and divergence on autosomes and the Z chromosome.

mean  $F_{ST}$  among lineages spanning nearly an order of magnitude (e.g.,  $F_{ST}=0.012$  between the barn swallow subspecies *H. rustica erythrogaster* and *H. r. tytleri* and  $F_{ST}=0.804$  between *H. atrocaerulea* and *H. smithii*; Figure 2; Data S2). These results highlight a rapid accumulation of genome-wide differentiation within *Hirundo* since splitting from a common ancestor ~5 million years ago (Schield, Brown, et al. 2024). We also find two major transitions in genome-wide mean  $F_{ST}$  among *Hirundo* lineages. First, there is a rapid transition from very low  $F_{ST}$  between *H. rustica* subspecies (Figure 2A; ‘within species’) to substantial differentiation among species within the barn swallow clade (e.g.,  $F_{ST,H. rustica-H. aethiopica}=0.25$ ). This is consistent with *H. rustica* subspecies being at an early stage of the speciation process with ongoing gene flow and few genomic regions of accentuated divergence (Safran, Scordato, et al. 2016; Scordato et al. 2017, 2020; Schield et al. 2021; Schield, Carter, et al. 2024). By contrast, genome-wide  $F_{ST}$  between most species within the barn swallow clade ranges between 0.2 and 0.4 (Figure 2A; ‘between species (barn swallow clade)’), although some show very high differentiation (e.g., *H. aethiopica* - *H. smithii*, *H. angolensis* - *H.*

*smithii* and *H. nigrita* - *H. smithii*), likely due to the combined effects of historical demography in *H. smithii* (Figure 1D) and more ancient divergence between it and other species within the barn swallow clade (Figures 1B and S3). We also observe a second transition towards much higher  $F_{ST}$  for between-species comparisons outside of the barn swallow clade (Figure 2A; ‘between species’) indicating the evolution of strong genome-wide differentiation between *Hirundo* species.

Genome-wide mean  $d_{xy}$  also shows expected increases between progressively divergent lineages, ranging between 0.0047 and 0.0101 (Figure 2B; Data S2). However,  $d_{xy}$  does not exhibit the same gap in values between *H. rustica* subspecies versus between species within the barn swallow clade as we observe for  $F_{ST}$ . Rather, we find one major transition towards higher  $d_{xy}$  for between-species comparisons outside of the barn swallow clade. This is unsurprising given the much greater phylogenetic distance between these species (Figure S3) and illustrates the combined effects of coalescent time and  $N_e$  on genome-wide  $F_{ST}$ . As for  $F_{ST}$ ,  $d_{xy}$  is lowest between *H. rustica* subspecies, though other



**FIGURE 3** | Landscapes of between-species population differentiation ( $F_{ST}$ ), between-species nucleotide divergence ( $d_{xy}$ ), within-species nucleotide diversity ( $\pi$ ) and recombination rate across Chromosome 1A (A), Chromosome 4 (B) and the Z chromosome (C). Summary statistics and recombination rate are shown as genome scans in 1 Mb sliding windows with a 100-kb step size. Lines are coloured according to specific between-species or within-species values with matching colour labels. The approximate locations of centromeres are shown as grey boxes.

closely related pairs of species also exhibit low genome-wide  $d_{xy}$  (e.g., *H. aethiopica* – *H. angolensis* and *H. neoxena* – *H. javanica*). Together, genome-wide  $F_{ST}$  and  $d_{xy}$  indicate that genomic divergence ranges widely among *Hirundo* species and the comparison of these estimates indicates that highly elevated  $F_{ST}$  between certain species is due to low within-species genetic diversity (i.e., higher  $F_{ST}$  than would be expected a priori based on divergence time) while these pairs show predictable  $d_{xy}$  values relative to others based on their divergence history.

We also find that  $F_{ST}$  is higher on the Z chromosome than autosomes between all species and subspecies (Figure 2C; Data S2). This commonly observed pattern can be explained at least in part by the lower effective population size of the Z chromosome relative to autosomes but likely also reflects reductions in genetic diversity due to more pronounced linked selection on the Z chromosome (see detailed interpretations of linked selection below). Consistent with this explanation,  $d_{xy}$  is uniformly lower on the Z chromosome than autosomes across species (Figure 2D; Data S2). Notably, the difference between Z-linked and autosomal  $F_{ST}$  is more accentuated between more recently diverged taxa (e.g., 11.4× higher Z chromosome  $F_{ST}$  between *H. r. rustica* and *H. r. tyleri*) than those with more ancient divergence (e.g.,

1.11× higher Z chromosome  $F_{ST}$  between *H. atrocaerulea* and *H. neoxena*).

Next, we performed genome scans of population genetic summary statistics to characterise genomic divergence in greater detail. The genomic landscape is highly heterogeneous, with numerous peaks and valleys of  $F_{ST}$  between species characterising genetic variation across chromosomes (Figures 3 and S6–S8). Genome scans of  $F_{ST}$  reflect our findings from genome-wide summaries (Figure 2), with overall levels of differentiation that increase with divergence time between lineages. To illustrate this phenomenon, in Figure 3, we show multiple scans of  $F_{ST}$  between *H. rustica* and other *Hirundo* species on two autosomes (Chromosome 1A and Chromosome 4) and the Z chromosome (Figure 3A–C; whole genome details are shown in Figures S6–S8). These results emphasise the major transition towards elevated genome-wide differentiation between species from extremely low  $F_{ST}$  between *H. rustica* subspecies (though note much higher Z-linked  $F_{ST}$  between *H. r. rustica* – *H. r. tyleri* than between *H. r. rustica* – *H. r. savignii*; Figure 3C). All comparisons show a concentration of high  $F_{ST}$  values in approximate centromere regions, which have low recombination rates (Figure S9), whereas  $F_{ST}$  tends to be lower in chromosomal

regions with higher recombination rates—we explore this relationship in greater detail below. Genomic regions with elevated  $F_{ST}$  also tend to have low  $d_{xy}$  between species and low  $\pi$  within species, consistent with reduced ancestral genetic diversity and the diversity-reducing effects of selection within populations.

Genome scans of summary statistics also reveal remarkable consistency in regional patterns among lineages (Figure 3), true even for independent pairs of species. For example, peaks and valleys of  $F_{ST}$  between *H. rustica* and *H. aethiopica* show collinear peaks and valleys between *H. neoxena* and *H. javanica*. More broadly,  $F_{ST}$  landscapes are strongly correlated for all pairs of species and subspecies (Data S3;  $\rho$  range = 0.42–0.99,  $p$ -values  $< 2.2 \times 10^{-16}$ ). Similarly, regions with low  $d_{xy}$  between one pair of species are likely to exhibit low  $d_{xy}$  between other species and the same is true for  $\pi$  within species (i.e., valleys of nucleotide diversity in *H. dimidiata* are valleys of diversity in *H. javanica*, *H. smithii* and so on). To further explore the conservation of genomic landscapes of diversity and divergence, we calculated correlations between  $\pi$  and  $d_{xy}$  measured in sliding windows and investigated the strength of these correlations with increasing phylogenetic distance between respective landscapes (see Figure S10 for examples of phylogenetic distance calculations). These comparisons indicate a high degree of conservation in genomic landscapes across species (Figure 4; Data S4). This is especially the case for  $\pi \sim \pi$  correlation coefficients, which are very strong at both shallow and deep phylogenetic distances between landscapes ( $\rho$  range = 0.84–0.99; Figure 4A), though correlations are also strong for  $d_{xy} \sim d_{xy}$  ( $\rho$  range = 0.56–0.97) and  $d_{xy} \sim \pi$  ( $\rho$  range = 0.46–0.96) landscapes across evolutionary distances (Data S4). In Figure 4B, we show the relationships for  $F_{ST}$ ,  $d_{xy}$  and mean  $\pi$  for three exemplar pairs of species (*H. rustica* – *H. aethiopica*, *H. neoxena* – *H. javanica* and *H. smithii* – *H. dimidiata*), which illustrate at a finer resolution the strong correlations between these landscapes across scales of divergence. These exemplars span several scales of divergence (Figure 5), vary in geographic distribution (i.e., allopatry or partial sympatry; Figure S1; Table S2) and are representative of patterns across *Hirundo*, broadly. We focus on these species pairs for downstream interpretations of models of linked selection and the formation of differentiation islands. Details for correlations between landscapes of differentiation, divergence and diversity for all species can be found in Data S3 and S4.

### 3.4 | Recombination Rate and Linked Selection Shape Genomic Landscapes

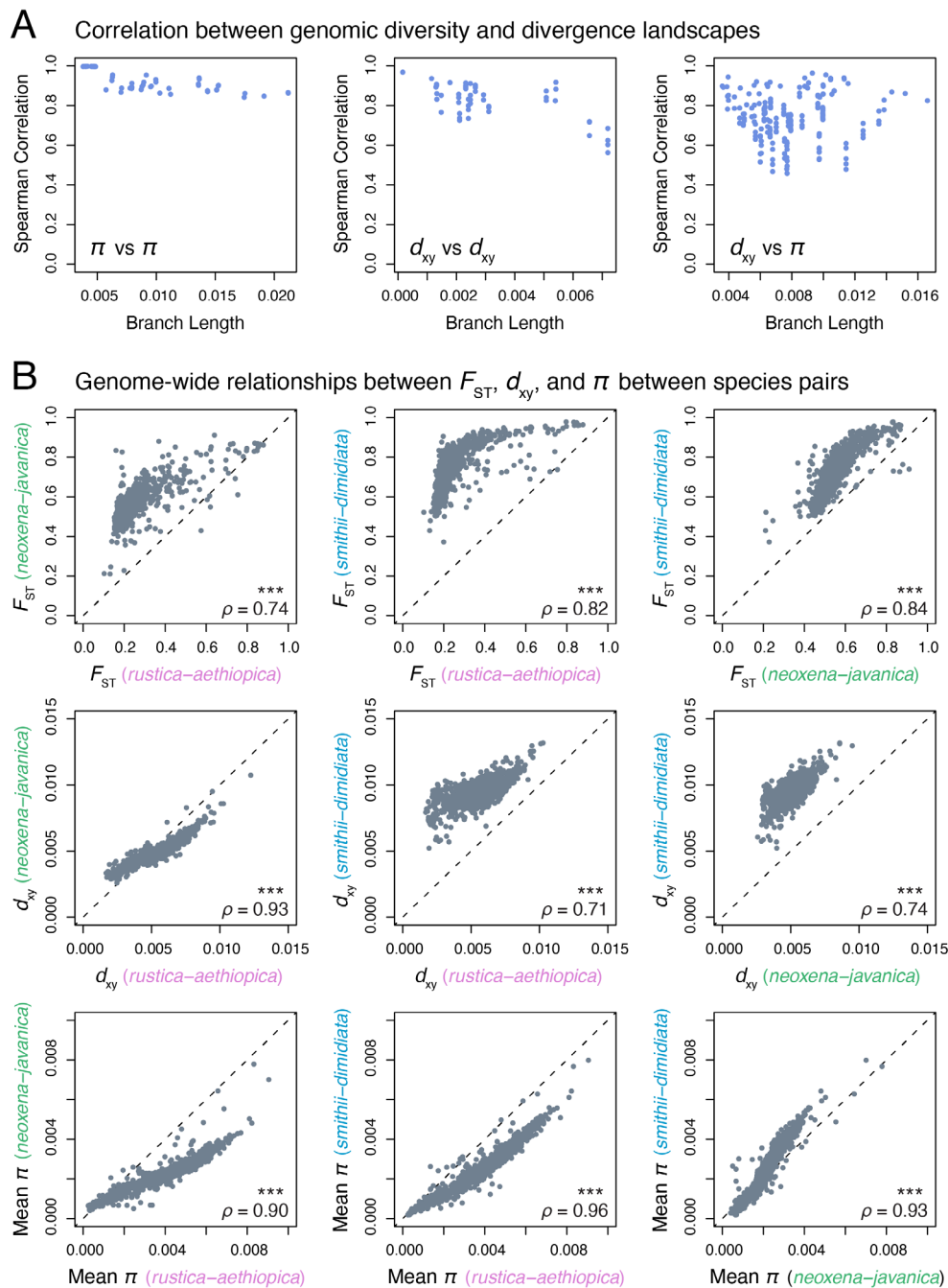
The degree of conservation in genomic landscapes suggests that shared evolutionary processes (e.g., recombination and selection) have shaped divergence among *Hirundo* species. We therefore tested whether the patterns of variation are predicted by models of linked selection by examining the relationships between population genetic summary statistics, recombination rate variation and the density of targets of selection in sliding windows (Figures 5 and S11–S13).

Whereas genome-wide average  $F_{ST}$  and  $d_{xy}$  are positively correlated (Figure S5; see ‘Genomic Landscapes’ section above), we find consistent negative correlations between  $F_{ST}$  and  $d_{xy}$  in sliding windows (Figure 5A–E; Data S5;  $\rho = -0.47$  to  $-0.69$ ,

$p$ -values  $< 2.2 \times 10^{-16}$ ) and between  $F_{ST}$  and  $\pi$  ( $\rho = -0.76$  to  $-0.96$ ,  $p$ -values  $< 2.2 \times 10^{-16}$ ), indicating that regions with high differentiation harbour lower levels of ancestral variation and reduced variation in descendent populations. Accordingly,  $\pi$  and  $d_{xy}$  are positively correlated, as described above (Figures 4 and S12; Data S4 and S5;  $\rho = 0.68$ – $0.98$ ,  $p$ -values  $< 2.2 \times 10^{-16}$ ). We used the *H. rustica* recombination map to examine the relationships between recombination rate and summary statistics, as recombination rate variation is strongly correlated among *Hirundo* species (Table S3;  $\rho = 0.81$ – $0.94$ ,  $p$ -values  $< 2.2 \times 10^{-16}$ ). Recombination rate and exon density are positively correlated across sliding windows (Figure S11A;  $\rho = 0.52$ ,  $p < 2.2 \times 10^{-16}$ ); thus, we investigated their relationships with summary statistics together using multiple linear regression (MLR; see Materials and Methods).  $F_{ST}$  is significantly negatively correlated with recombination rate (Figure 5; Data S5; MLR,  $t = -15.1$  to  $-31.6$ ,  $p$ -values  $< 2.2 \times 10^{-16}$ ;  $r^2 = 0.23$ – $0.56$ ) but correlations between  $F_{ST}$  and exon density are weak or absent (Data S5). For example, MLR reveals no effect of exon density on  $F_{ST}$  *H. rustica*–*H. aethiopica* ( $t = -1.36$ ,  $p = 0.172$ ) and *F\_{ST}* *H. neoxena*–*H. javanica* ( $t = -0.85$ ,  $p = 0.393$ ) while there is a weak effect of exon density on  $F_{ST}$  *H. smithii*–*H. dimidiata* ( $t = -2.04$ ,  $p = 0.041$ ). We find positive correlations between recombination rate and  $\pi$  (Figure S12; MLR,  $t = 23.7$ – $30.6$ ,  $p$ -values  $< 2.2 \times 10^{-16}$ ;  $r^2 = 0.39$ – $0.53$ ) and  $d_{xy}$  (Figure S13; MLR,  $t = 14$ – $29.4$ ,  $p$ -values  $< 2.2 \times 10^{-16}$ ;  $r^2 = 0.16$ – $0.51$ ). Correlations between exon density and both  $\pi$  and  $d_{xy}$  are largely weak or absent, similar to  $F_{ST}$  (see Data S5 for full results for all species). The lack of strong genome-wide effects of exon density on summary statistics may be explained by variable relationships between exon density and recombination rate among chromosomes. Indeed, many smaller chromosomes have the expected negative correlation between recombination rate and exon density while several of the largest chromosomes have positive correlations (Figure S11B). Nonetheless, genomic relationships between recombination rate and summary statistics are consistent with models of linked selection, supporting that recombination rate variation has profound impacts on rates of fixation in genomic regions experiencing genetic hitchhiking due to positive selection and background purifying selection.

### 3.5 | Signatures of Selection in Islands of Differentiation

To further discern how regions of elevated genomic differentiation have evolved between *Hirundo* species, we examined the distributions of  $d_{xy}$  and  $\pi$  in differentiation islands to test predictions under three models of linked selection that can explain island formation, described in the Introduction and depicted in Figure 1A. As expected under each model and predicted based on genome-wide patterns,  $F_{ST}$  islands consistently have lower within-population nucleotide diversity than the genomic background, measured as mean  $\pi$  between each pair of species (Figure 6; Tables S4 and S5; ANOVA,  $p$ -values  $< 2.2 \times 10^{-16}$ ). This is the case for all islands as well as islands outside of centromere regions, specifically (Table S5; Tukey *post hoc* tests;  $p$ -values  $< 1 \times 10^{-4}$ ), and distributions of mean  $\pi$  in all islands versus non-centromere islands are not significantly different (Tukey *post hoc* tests;  $p$ -values  $\geq 0.12$ ).  $F_{ST}$  islands also have lower  $d_{xy}$  between species relative to the genomic background (Figure 6; Tables S4 and S6; ANOVA,  $p$ -values  $< 2.2 \times 10^{-16}$ ;

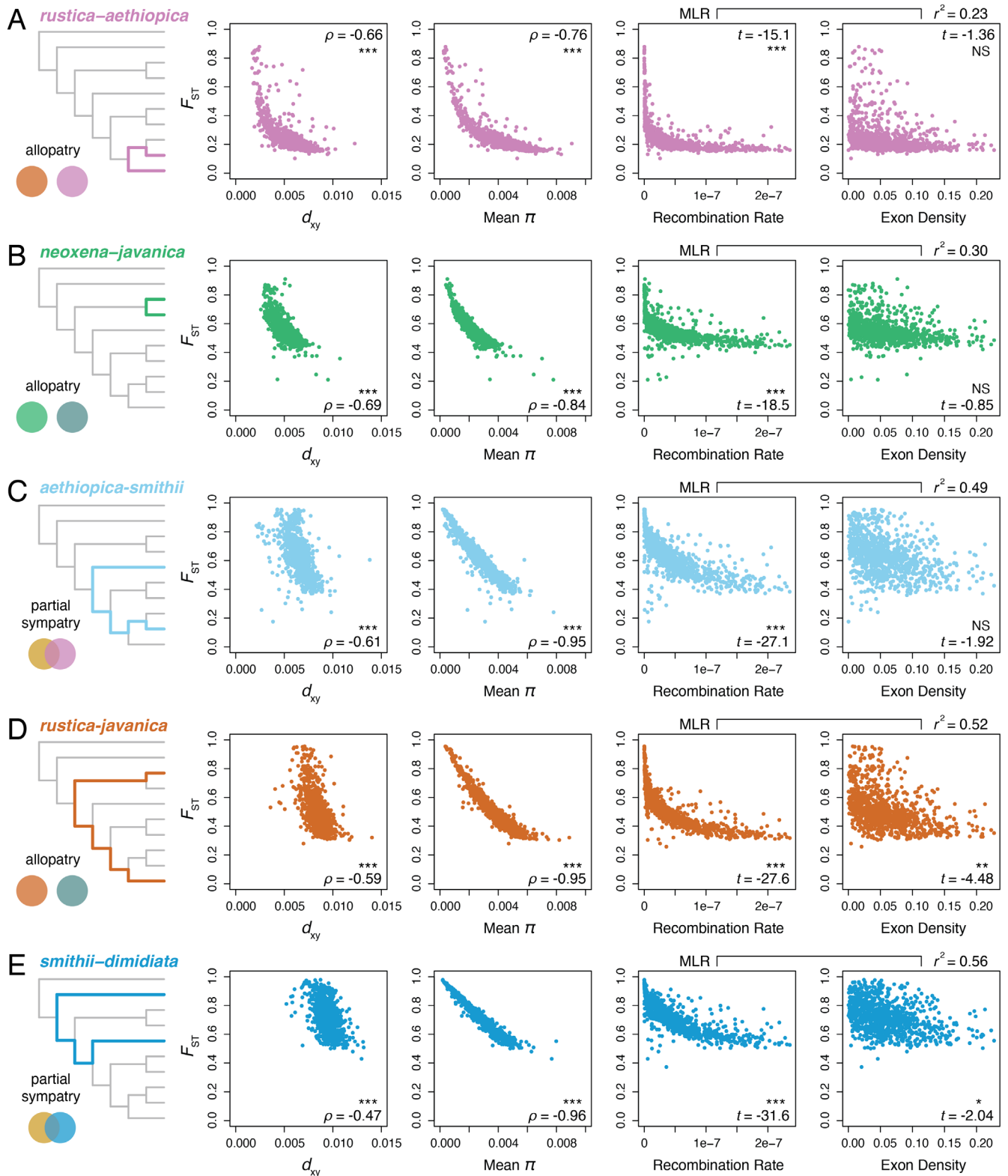


**FIGURE 4** | Correlation between genomic landscapes of diversity and divergence across *Hirundo*. (A) Spearman rank correlation coefficients (points) between genome-wide  $\pi \sim \pi$  (left),  $d_{xy} \sim d_{xy}$  (centre), and  $d_{xy} \sim \pi$  (right), shown as a function of the phylogenetic distance between pairs of landscapes and their common ancestor, measured as the total branch length. Correlations were calculated for each pair of diversity and divergence landscapes that do not share branches (see Materials and Methods) based on mean values in non-overlapping 1 Mb windows; representative correlations for  $\pi$  and  $d_{xy}$  shown in panel (B). (B) Correlations between landscapes of  $F_{ST}$  (top),  $d_{xy}$  (centre) and mean  $\pi$  (bottom) for three independent species pairs: *H. rustica* – *H. aethiopica*, *H. neoxena* – *H. javanica* and *H. smithii* – *H. dimidiata*. Points represent mean values of each summary statistic in non-overlapping 1 Mb windows. Spearman correlations ( $\rho$ ) are labelled, \*\*\* $p < 2.2 \times 10^{-16}$ .

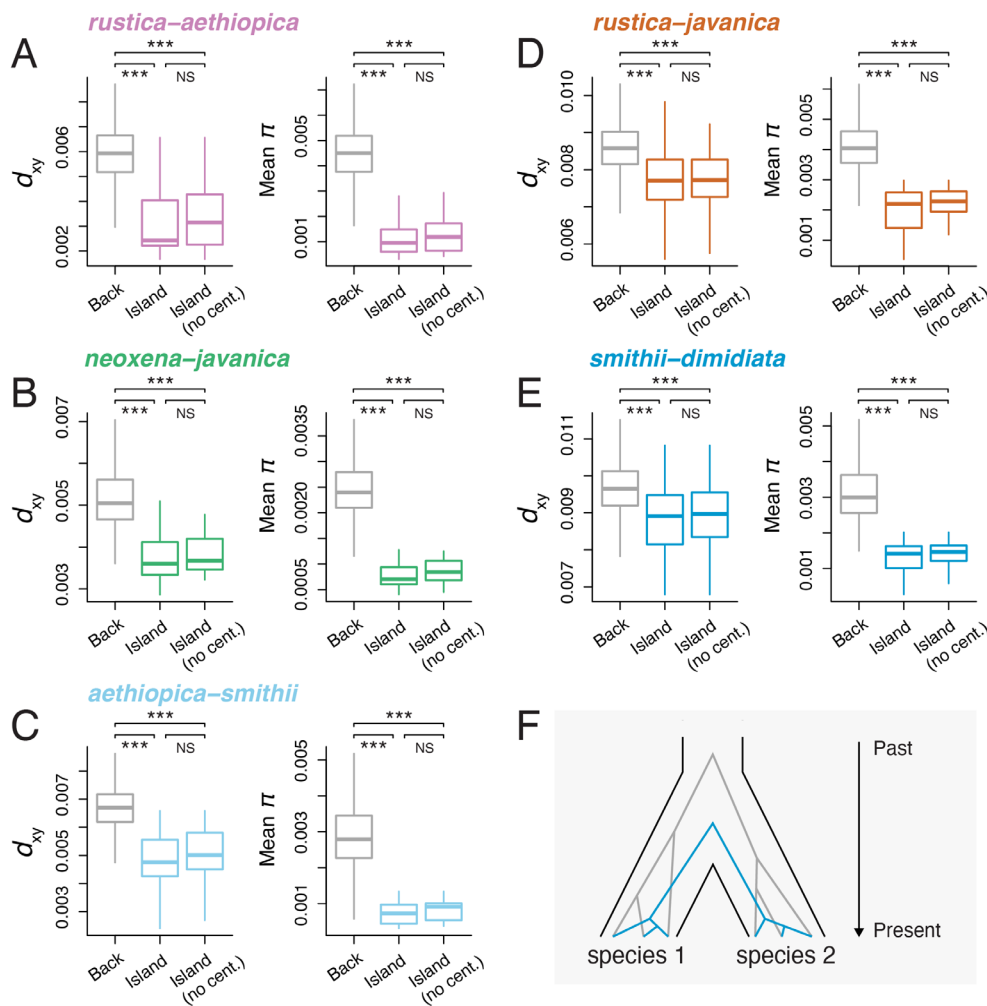
Tukey *post hoc* tests,  $p$ -values  $< 1 \times 10^{-4}$ ), consistent with reduced ancestral variation and thus more rapid coalescence in  $F_{ST}$  island regions. Together, these patterns match the predictions of the recurrent selection model (Cruickshank and Hahn 2014; Burri et al. 2015; Irwin et al. 2018), supporting that islands of differentiation have largely evolved through selection in the same genomic regions in both ancestral and descendant populations.

The causes of differentiation island formation are not necessarily mutually exclusive; thus, we were curious whether the potential for gene flow between pairs of species influences the relationships between genetic diversity and divergence in  $F_{ST}$  islands. We divided species pairs into groups based on their contemporary geographic distributions (i.e., strict allopatry vs. parapatry/partial sympatry; Figure S1; Table S2), and compared Spearman correlation coefficients between  $d_{xy}$  and  $\pi$  in differentiation





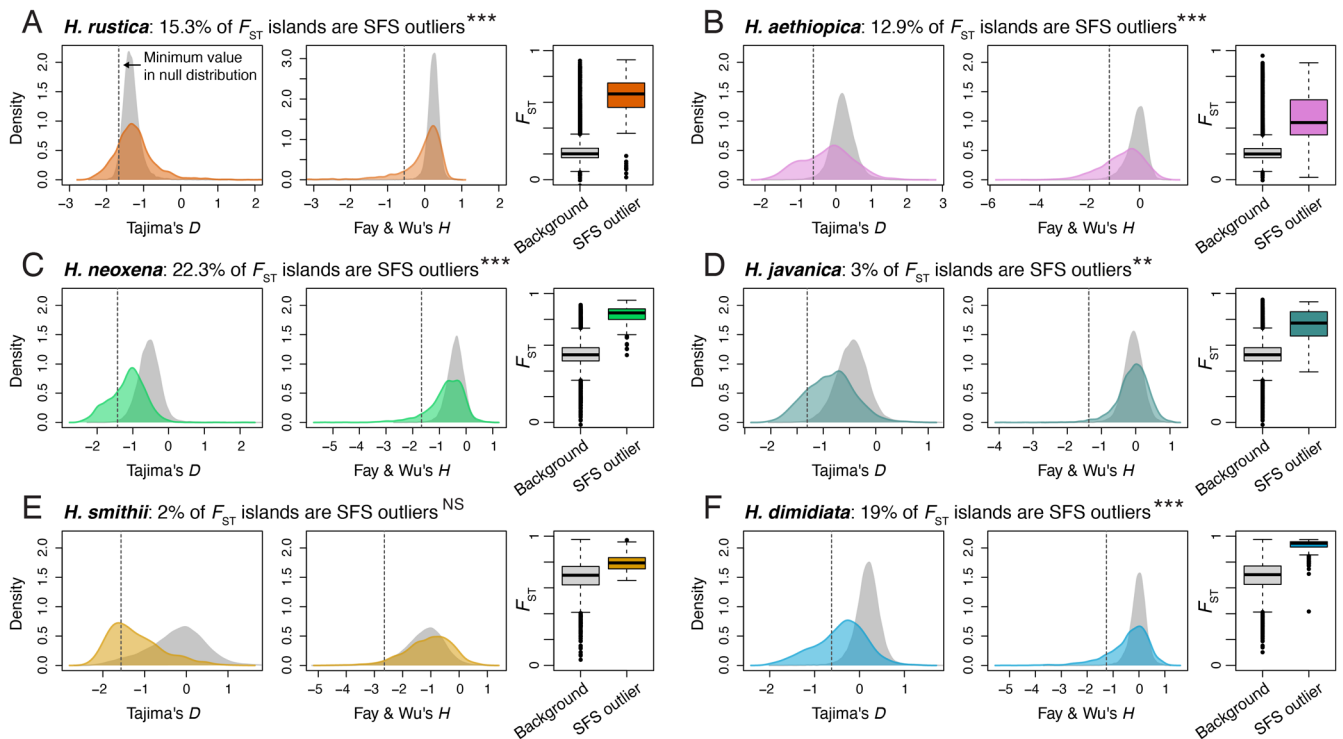
**FIGURE 5** | Genome-wide relationships between landscapes of genetic differentiation and divergence, diversity, recombination rate and exon density illustrated by *H. rustica* – *H. aethiopica* (A), *H. neoxena* – *H. javanica* (B), *H. aethiopica* – *H. smithii* (C), *H. rustica* – *H. javanica* (D) and *H. smithii* – *H. dimidiata* (E) species pairs. Panels to the left in (A)–(E) show phylogenetic distance between pairs of species as shaded branches and summarise the present-day geographic arrangement of species pairs (allopatric versus partial sympatry). The centre two panels show correlations between  $F_{ST}$  and  $d_{xy}$  and mean  $\pi$ , respectively, with labels for Spearman correlation coefficients ( $\rho$ ). The right two panels show correlations between  $F_{ST}$  and recombination rate and exon density, respectively, with summaries of multiple linear regression (MLR) to test the effects ( $r^2$  and  $t$ ) of genomic features on summary statistics. All statistical comparisons are based on mean values in non-overlapping 1 Mb windows. In all panels, \*\*\* $p < 2.2 \times 10^{-16}$ ; \* $p = 0.05$ ; NS, not significant.



**FIGURE 6** | Distributions of nucleotide divergence ( $d_{xy}$ ) and nucleotide diversity ( $\pi$ ) in differentiation ( $F_{ST}$ ) islands between species are consistent with the effects of recurrent selection, illustrated by patterns in *H. rustica* – *H. aethiopica* (A), *H. neoxena* – *H. javanica* (B), *H. aethiopica* – *H. smithii* (C), *H. rustica* – *H. javanica* (D) and *H. smithii* – *H. dimidiata* (E). Panels to the left show distributions of  $d_{xy}$  in the genomic background ('Back'),  $F_{ST}$  island regions ('Island') and  $F_{ST}$  islands outside of centromere regions ('Island no cent.'). Right panels show distributions of mean  $\pi$  in background and  $F_{ST}$  island regions. Bold centre lines in boxplots represent the median of each distribution, boxes summarise the interquartile range (IQR) and whiskers represent 1.5 IQR. Statistical summaries of ANOVA and Tukey *post hoc* tests are labelled; \*\*\* $p < 2.2 \times 10^{-16}$ ; NS, not significant. (F) Schematic representation of the recurrent selection model supported by summary statistics, wherein  $F_{ST}$  islands exhibit both reduced nucleotide diversity and nucleotide divergence due to the same genomic regions experiencing linked selection in ancestor and descendant populations.

islands and the genomic background. It is logical that the divergence with gene flow model could explain the formation of a subset of islands for species in parapatry or partial sympatry if loci within these regions promote reproductive isolation in the presence of gene flow (Figure 1A). In this case, we would expect weaker relationships between  $\pi$  and  $d_{xy}$  in  $F_{ST}$  islands than in the genomic background because some regions have higher  $d_{xy}$  than the genomic background (loci promoting reproductive isolation) while others have lower  $d_{xy}$  than the genomic background (loci under recurrent selection). Instead, we find that the median  $\pi \sim d_{xy}$  correlation in islands is slightly higher than the background (Figure S14A), although these distributions are not significantly different (Mann–Whitney *U*-test;  $p = 0.126$ ). For allopatric species, relationships between  $\pi$  and  $d_{xy}$  could be weakened if some islands have formed through selection in allopatry (no selection in the ancestral population) instead of recurrent selection (selection in the ancestral population). Again,

we find that correlations are instead stronger in differentiation islands (Figure S14B; Mann–Whitney *U*-test;  $p = 0.005$ ). Sliding window tests of introgression between *H. smithii* and *H. aethiopica* and *H. dimidiata* are consistent with low levels of admixture between partially sympatric species across the genome (mean  $\pm$  standard deviation  $f_d = 0.011 \pm 0.024$  between *H. smithii* and *H. aethiopica* and  $f_d = 0.0035 \pm 0.0055$  between *H. smithii* and *H. dimidiata*). We find a significant, albeit weak negative relationship between  $F_{ST}$  and  $f_d$  for *H. smithii* and *H. aethiopica* ( $\rho = -0.38$ ;  $p = 4 \times 10^{-9}$ ) and no relationship between  $F_{ST}$  and  $f_d$  for *H. smithii* and *H. dimidiata* ( $p = 0.24$ ). Evidence for a negative genome-wide relationship between introgression and differentiation between *H. smithii* and *H. aethiopica* supports polygenic barriers to gene flow between species (Martin et al. 2019). While rare introgression between these species could act to reduce or erase genomic differentiation islands, we nonetheless find signal consistent with strong genome-wide effects on reproductive



**FIGURE 7** | Distributions of statistical summaries of the site frequency spectrum (SFS) in  $F_{ST}$  islands (colours) compared to the genomic background (grey) in *H. rustica* (A), *H. aethiopica* (B), *H. neoxena* (C), *H. javanica* (D), *H. smithii* (E) and *H. dimidiata* (F). Left panels in (A)–(F) show distributions of Tajima's  $D$  and centre panels show distributions of Fay & Wu's  $H$ . Dashed vertical lines depict the minimum values of Tajima's  $D$  and Fay & Wu's  $H$  from null distributions generated using permutations in each species. Density distributions and statistical comparisons are based on mean values in non-overlapping 50 kb windows. Right panels (boxplots) in (A)–(F) compare  $F_{ST}$  between the genomic background (grey) and regions identified as outliers of both Tajima's  $D$  and Fay & Wu's  $H$  (colours; 'SFS outliers'). Percentages of  $F_{ST}$  islands that are also SFS outliers are labelled above panels for each species. Summaries of Fisher's exact tests to test whether  $F_{ST}$  islands are enriched for SFS outliers are labelled; \*\*\* $p < 2.2 \times 10^{-16}$ ; \*\* $p < 0.00001$ ; NS, not significant.

isolation, and our tests here broadly support a dominant role of recurrent selection in shaping genomic landscapes, whether or not speciation has occurred in the presence of gene flow.

Signatures of recurrent evolution of differentiation islands could be caused by positive selection (i.e., repeated episodes of genetic hitchhiking in ancestors and descendants) or by repeated background selection against deleterious variation in low recombination regions (Noor and Bennett 2009; Sella et al. 2009; Via 2012; Cruickshank and Hahn 2014). We used measures of the site frequency spectrum in six *Hirundo* species to further test the hypothesis that a subset of differentiation islands in the genomic landscape are the result of divergent positive selection. Distributions of Tajima's  $D$  are lower in islands than the genomic background in all species (Figure 7A–F). Fay & Wu's  $H$  is also skewed towards lower values in  $F_{ST}$  islands (Figure 7). We identified regions that are both Tajima's  $D$  and Fay & Wu's  $H$  outliers ('SFS outliers') using a permutation approach (see Methods) in order to test whether  $F_{ST}$  islands are enriched for signatures of divergent positive selection. Indeed, between 2% and 22.3% of  $F_{ST}$  islands overlap with SFS outliers among species, corresponding to significant enrichment of islands for signatures of selection based on both Tajima's  $D$  and Fay & Wu's  $H$  in each species (Fisher's exact tests;  $p$ -values  $< 1.5 \times 10^{-11}$ ), with the exception of *H. smithii* (Figure 7E;  $p = 0.052$ ). Moreover,  $F_{ST}$  is significantly higher in SFS outliers than non-outlier (i.e., background) regions in all species (Figure 7A–F; Mann–Whitney  $U$ -tests;

$p$ -values  $< 1.2 \times 10^{-12}$ ). We find that  $9.8 \pm 13.3\%$  (range = 0%–53.8%) of  $F_{ST}$  islands with SFS outliers are shared among pairs of species, with low degrees of overlap driven by small numbers of SFS outliers in *H. javanica* and *H. smithii*. Together, these results support that a substantial proportion of genomic differentiation between *Hirundo* species has evolved through recurrent selective sweeps, with divergent selection producing frequencies of derived variants beyond what would be expected under background selection and genetic drift alone.

#### 4 | Discussion

Clarifying how evolutionary processes shape genomic divergence during speciation is essential to our broader understanding of the origins of biological diversity and how it is maintained in nature. In this study, we analysed population genomic data from swallows in the genus *Hirundo* to investigate the drivers of genomic landscapes of divergence between species existing across a continuum of evolutionary divergence and with variation in geographic range overlap, demographic history and gene flow (Figure 1). Species in this radiation share remarkably correlated landscapes of genetic diversity and divergence (Figures 2–4), pointing to the role of shared processes during genomic divergence. We find evidence that highly correlated genomic landscapes have been shaped largely by recurrent selection in ancestral and descendant populations since (and likely

prior to) divergence from a common ancestor ~5 million years ago (Figures 5–7). Indeed, the conservation of genomic landscapes at both shallow and ancient timescales argues for profound impacts of selection in shaping recurrent patterns of genomic divergence, as we would not expect such strong correlation between landscapes under mutation and genetic drift alone (Burri et al. 2015; Jiang et al. 2023; Rodrigues et al. 2024). Moreover, our findings align with predictions from models describing the evolution of genomic differentiation through linked selection, wherein the accumulation and spread of genetic changes across the genome is strongly influenced by interactions between background selection or genetic hitchhiking and recombination rate variation (Begun and Aquadro 1992; Charlesworth 1998; Noor and Bennett 2009; Nachman and Payseur 2012). Collectively, our findings from *Hirundo* add to a growing body of evidence supporting the importance of incorporating non-neutral evolutionary processes in our predictive frameworks for genomic divergence and the evolution of new species.

Relatedly, our findings support the hypothesis that the evolutionary stability of recombination landscapes in songbirds predicts conserved patterns of genome-wide heterogeneity in genetic diversity and divergence among species. Evidence from great tits (Van Oers et al. 2014), finches (Singhal et al. 2015), crows (Vijay et al. 2016) and flycatchers (Kawakami et al. 2017) indicates remarkable degrees of conservation in recombination rate variation among closely related populations and species, even at fine genomic scales. We find strong correlations among *Hirundo* recombination landscapes based on analysis of multiple species (Figure S9; Table S3), similar to observations in these other songbirds. The conservation of recombination rate variation among species likely plays a prominent role in the presence of many shared islands of differentiation, in particular in regions of the genome with low recombination where the effects of linked selection are most pronounced (Cutter and Payseur 2013; Burri et al. 2015). Genome-wide relationships between recombination rate and population genetic summary statistics in *Hirundo* are consistent with this hypothesis, with higher  $F_{ST}$  between populations and lower  $\pi$  within populations on average in genomic regions with low recombination (Figures 5 and S12; Data S5). These patterns underscore the diversity-reducing effects of selection in genomic regions with low recombination, where direct effects of selection on loci are most effectively spread to neutral loci through genetic linkage.

As described above, genome-wide patterns in *Hirundo* broadly fit a model of recurrent linked selection shaped by variation in recombination rate (Figures 5, S12 and S13), as opposed to being strongly predicted by alternative models of divergence with gene flow or selection in allopatry (Figure 1A). This is illustrated further by a consistent signature of lower  $d_{xy}$  in genomic islands of differentiation across species due to reduced ancestral genetic diversity and thus rapid coalescence in these regions compared to elsewhere in the genome (Figure 6; Tables S4 and S6). The opposite pattern (i.e., higher  $d_{xy}$  in differentiation islands than neutral regions) would be expected if differential gene flow largely explained the formation of differentiation islands between lineages. By examining nucleotide diversity and divergence in islands of differentiation between *Hirundo* populations and species at various stages of divergence, our results emphasise the significance of recurrent selection in shaping correlated

landscapes of genetic variation in this system. Indeed,  $d_{xy}$  is significantly reduced in  $F_{ST}$  islands between populations at the earliest stages of speciation (e.g., between barn swallow subspecies; Tables S4 and S6) in a manner reminiscent of species at late or complete stages of the speciation process (e.g., between *H. atrocaerulea* and others). Recurrent selection appears to be a pervasive driver of differentiation islands across systems (e.g., Cruickshank and Hahn 2014; Stankowski et al. 2019; Shang et al. 2023; Glover et al. 2024) and in birds in particular (e.g., Ellegren et al. 2012; Burri et al. 2015; Delmore et al. 2015; Irwin et al. 2016, 2018; Van Doren et al. 2017; Battey 2020; Schield et al. 2021; Jiang et al. 2023). Observations from across a continuum of divergence in *Hirundo* thus add to accumulating evidence for the relative importance of recombination rate (which often spans several orders of magnitude across the genome) to heterogeneous genomic divergence in comparison to gene flow (Nachman and Payseur 2012; Burri et al. 2015; Wolf and Ellegren 2017). Importantly, neither our findings nor those from previous studies argue that differential gene flow is irrelevant in shaping genomic differentiation. Rather, in many cases, we may have little power to detect the effects of gene flow against the dominant effects of recurrent linked selection, particularly at more advanced stages of the speciation process.

What information, then, do our findings provide about the evolution of reproductive isolation? Original interpretations of genomic islands of differentiation were that these form because they contain loci responsible for pre- or postzygotic isolation ('speciation genes'; Wu 2001; Turner et al. 2005; Via and West 2008; Cruickshank and Hahn 2014). In this scenario, described by the divergence with gene flow model, regions of the genome containing speciation genes contribute barriers to gene flow while regions unrelated to reproductive isolation homogenised by gene flow between populations. Empirical examples matching these predictions provide perhaps the clearest evidence for the formation of genomic islands of differentiation due to the direct involvement of loci in reproductive isolation (Nosil et al. 2009; Nadeau et al. 2012; Poelstra et al. 2014), yet systems fitting the assumptions of the divergence with gene flow model may be comparatively rare (especially because virtually no natural system will fit the assumption of selective neutrality outside of barrier loci). By contrast, recurrent linked selection can explain the formation of heterogeneous genomic landscapes of differentiation without the need for differential gene flow across the genome (Noor and Bennett 2009; Cruickshank and Hahn 2014; Wolf and Ellegren 2017). This has led to interpretations that differentiation islands formed through recurrent selection are distinct from those involved in reproductive isolation or that a causal role in reproductive isolation cannot be ascribed to these regions per se. These are non-mutually exclusive mechanisms, however, and it is clear recurrent linked selection can predominately explain heterogeneous genomic divergence even in systems where gene flow is present (e.g., Burri et al. 2015; Irwin et al. 2018; Stankowski et al. 2019; Glover et al. 2024), including *Hirundo* (Scordato et al. 2017, 2020; Schield et al. 2021; Schield, Carter, et al. 2024). Indeed, evidence for highly correlated genomic landscapes despite extensive historical introgression in *Hirundo* (Figure S4) highlights the plausibility that certain loci show consistently elevated differentiation because they have been repeatedly involved in reproductive isolation during diversification. Moreover, our finding of a negative



genome-wide relationship between recent introgression and  $F_{ST}$  between partially sympatric *H. smithii* and *H. aethiopica* supports the presence of genomically widespread barrier loci that have contributed to the formation of differentiation islands. However, parsing differentiation caused by reproductive isolation from other sources of selection remains a major challenge and ultimately requires additional data to link sources of prezygotic and postzygotic isolation to their underlying genetic basis.

Evolutionary patterns in barn swallows may help to better understand connections between the effects of recurrent selection on genetic divergence and sources of reproductive isolation across *Hirundo* more broadly. Barn swallow subspecies diverged from a common ancestor very recently (Zink et al. 2006; Smith et al. 2018) yet exhibit variation in plumage traits used in sexual signalling due to divergent sexual selection (Scordato and Safran 2014; Romano et al. 2017; Lotem et al. 2022). Consistent with recent common ancestry, genome-wide differentiation and divergence are extremely shallow between barn swallow populations (Safran, Scordato, et al. 2016; Schield et al. 2021; Schield, Carter, et al. 2024). Genomic regions underlying sexual plumage traits, however, exhibit concentrated islands of differentiation due to divergent selection and promote reproductive isolation in hybrid zones (Schield, Carter, et al. 2024). Based on these patterns, we would predict that divergence between barn swallow populations fits the divergence with gene flow model (and indeed multiple reproductive isolation loci exhibit both high  $F_{ST}$  and  $d_{xy}$ ), yet genome-wide signatures are nonetheless strongly consistent with the recurrent selection model (Schield et al. 2021; Schield, Carter, et al. 2024). Thus, signatures of recurrent selection are dominant at even the earliest stages of speciation in this system. By examining the patterns of linked selection without additional information about barriers to gene flow, it would be possible to overlook genuine signals that specific highly differentiated loci are directly involved in reproductive isolation. These loci are concentrated on the Z chromosome and previous evidence supports that extreme reductions in Z-linked genetic diversity are a consequence of extra-pair mating and reproductive skew against males due to sexual selection in barn swallows (Schield et al. 2021). As a result, the Z chromosome exhibits much higher differentiation than would be expected in the absence of selection (Charlesworth 2001; Pool and Nielsen 2007), supporting that the Z chromosome exists at a more advanced stage of the speciation process. We find that Z-linked differentiation is consistently higher than autosomal differentiation across *Hirundo* (Figure 2; Data S2), raising the possibility that sex-linked traits have repeatedly played a disproportionate role in reproductive isolation throughout diversification, which has been suggested to explain elevated sex-linked differentiation generally (Irwin 2018).

As a final consideration, we were motivated to disentangle the effects of alternative forms of selection in the recurrently evolving genomic landscape across *Hirundo*. We have shown that the patterns of genetic diversity and divergence are consistent with models of linked selection, with potential contributions from both background selection (Charlesworth et al. 1993) and genetic hitchhiking due to positive selection (Smith and Haigh 1974). Both forms of linked selection leave local reductions in  $N_e$  and associated increases in differentiation between populations, though the magnitude of their effects and relevance in

shaping genomic landscapes are subjects of debate (Begun and Aquadro 1992; Charlesworth et al. 1993; Stephan 2010; Corbett-Detig et al. 2015; Schrider 2020). Evidence consistent with linked selection based on summary statistics (i.e.,  $F_{ST}$ ,  $d_{xy}$  and  $\pi$ ) alone cannot necessarily distinguish between these alternative processes. We therefore incorporated additional tests (Tajima's  $D$  and Fay & Wu's  $H$ ) designed to measure signatures of positive selection, specifically from skews in site frequency spectra. This approach has been used effectively in previous studies (e.g., Burri et al. 2015; Glover et al. 2024) to reveal the relative contribution of genetic hitchhiking to genomic divergence. Our results indicate a fairly substantial influence of divergent positive selection on genomic divergence, with an enrichment of differentiation islands for signatures of recurrent hitchhiking across *Hirundo* species (Figure 7). These results together support the hypothesis that certain traits and their underlying genetic basis have been repeatedly targeted by selection due to their roles in adaptation and speciation. Still, a large proportion of highly differentiated regions between species are consistent with background selection (at least based on our chosen parameter thresholds), highlighting the co-occurrence of some differentiation islands across species through recurrent purifying selection modulated by recombination rate variation. Collectively, our findings underscore that both genetic hitchhiking and background selection are important processes shaping genomic divergence during speciation in swallows.

#### Author Contributions

D.R.S. designed the research, generated data, performed the research, analysed data and wrote the paper. J.K.C. generated data, performed research, analysed data and edited the paper. M.G.A., K.F. and D.K.H. performed research, analysed data and contributed to the writing of the paper. R.J.S. designed the research, generated data and wrote and edited the paper.

#### Acknowledgements

We are grateful for tissue loans from Denver Museum of Nature and Science, Field Museum, Louisiana State University Museum of Natural Science, University of Alaska Museum of the North, University of Washington Burke Museum and Yale Peabody Museum. We thank Elizabeth Scordato, Iris Levin, Kayleigh Keller and members of the Safran Lab for helpful discussion. D.R.S. thanks Blood Incantation for auditory inspiration during the analysis and writing stages of this study. This work was supported by National Science Foundation (NSF) postdoctoral research fellowship grant DBI-1906188 to D.R.S., NSF postdoctoral research fellowship grant DBI-2409958 to K.F., NSF CAREER grant DEB-1149942 to R.J.S. and NSF grant IOS-DEB-1856266 to R.J.S.

#### Conflicts of Interest

The authors declare no conflicts of interest.

#### Data Availability Statement

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/hirundo\\_divergence\\_landscape](https://github.com/drewschield/hirundo_divergence_landscape).

#### References

Bathey, C. J. 2020. "Evidence of Linked Selection on the Z Chromosome of Hybridizing Hummingbirds." *Evolution* 74: 725–739.

- Begun, D. J., and C. F. Aquadro. 1992. "Levels of Naturally Occurring DNA Polymorphism Correlate With Recombination Rates in *D. melanogaster*." *Nature* 356: 519–520.
- Bendall, E. E., R. K. Bagley, V. C. Sousa, and C. R. Linnen. 2022. "Faster-Haplodiploid Evolution Under Divergence-With-Gene-Flow: Simulations and Empirical Data From Pine-Feeding Hymenopterans." *Molecular Ecology* 31: 2348–2366.
- Benjamini, Y., and Y. Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society, Series B* 57: 289–300.
- Bolger, A. M., M. Lohse, and B. Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30: 2114–2120.
- Broyles, G. G., B. M. Myers, N. R. Friedman, et al. 2023. "Evolutionarily Labile Dispersal Behavior and Discontinuous Habitats Enhance Population Differentiation in Island Versus Continentally Distributed Swallows." *Evolution* 77: 2656–2671.
- 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, et al. 2015. "Linked Selection and Recombination Rate Variation Drive the Evolution of the Genomic Landscape of Differentiation Across the Speciation Continuum of *Ficedula* Flycatchers." *Genome Research* 25: 1656–1665.
- Carter, J. K., P. Innes, A. M. Goebel, et al. 2020. "Complete Mitochondrial Genomes Provide Current Refined Phylogenomic Hypotheses for Relationships Among Ten *Hirundo* Species." *Mitochondrial DNA Part B Resources* 5: 2881–2885.
- Charlesworth, B. 1998. "Measures of Divergence Between Populations and the Effect of Forces That Reduce Variability." *Molecular Biology and Evolution* 15: 538–543.
- Charlesworth, B. 2001. "The Effect of Life-History and Mode of Inheritance on Neutral Genetic Variability." *Genetical Research* 77: 153–166.
- Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. "The Effect of Deleterious Mutations on Neutral Molecular Variation." *Genetics* 134: 1289–1303.
- Chifman, J., and L. Kubatko. 2014. "Quartet Inference From SNP Data Under the Coalescent Model." *Bioinformatics* 30: 3317–3324.
- Chifman, J., and L. Kubatko. 2015. "Identifiability of the Unrooted Species Tree Topology Under the Coalescent Model With Time-Reversible Substitution Processes, Site-Specific Rate Variation, and Invariable Sites." *Journal of Theoretical Biology* 374: 35–47.
- Cook, D. E., and E. C. Andersen. 2017. "VCF-Kit: Assorted Utilities for the Variant Call Format." *Bioinformatics* 33: 1581–1582.
- Corbett-Detig, R. B., D. L. Hartl, and T. B. Sackton. 2015. "Natural Selection Constrains Neutral Diversity Across a Wide Range of Species." *PLoS Biology* 13: e1002112.
- Cruickshank, T. E., and M. W. Hahn. 2014. "Reanalysis Suggests That Genomic Islands of Speciation Are due to Reduced Diversity, Not Reduced Gene Flow." *Molecular Ecology* 23: 3133–3157.
- Cutter, A. D., and B. A. Payseur. 2013. "Genomic Signatures of Selection at Linked Sites: Unifying the Disparity Among Species." *Nature Reviews. Genetics* 14: 262–274.
- Danecek, P., A. Auton, G. Abecasis, et al. 2011. "The Variant Call Format and VCFtools." *Bioinformatics* 27: 2156–2158.
- del Hoyo, J., A. Elliott, and D. A. Christie, eds. 2004. *Handbook of the Birds of the World*. Vol. 9. Lynx Edicions.
- Delmore, K. E., S. Hübner, N. C. Kane, et al. 2015. "Genomic Analysis of a Migratory Divide Reveals Candidate Genes for Migration and Implicates Selective Sweeps in Generating Islands of Differentiation." *Molecular Ecology* 24: 1873–1888.
- 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." *Molecular Phylogenetics and Evolution* 56: 409–418.
- Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. "Testing for Ancient Admixture Between Closely Related Populations." *Molecular Biology and Evolution* 28: 2239–2252.
- Ellegren, H., L. Smeds, R. Burri, et al. 2012. "The Genomic Landscape of Species Divergence in *Ficedula* Flycatchers." *Nature* 491: 756–760.
- Enard, D., P. W. Messer, and D. A. Petrov. 2014. "Genome-Wide Signals of Positive Selection in Human Evolution." *Genome Research* 24: 885–895.
- Fay, J. C., and C.-I. Wu. 2000. "Hitchhiking Under Positive Darwinian Selection." *Genetics* 155: 1405–1413.
- Feder, J. L., S. P. Egan, and P. Nosil. 2012. "The Genomics of Speciation-With-Gene-Flow." *Trends in Genetics* 28: 342–350.
- Feder, J. L., and P. Nosil. 2010. "The Efficacy of Divergence Hitchhiking in Generating Genomic Islands During Ecological Speciation." *Evolution* 64: 1729–1747.
- Feder, J. L., P. Nosil, A. C. Wacholder, S. P. Egan, S. H. Berlocher, and S. M. Flaxman. 2014. "Genome-Wide Congealing and Rapid Transitions Across the Speciation Continuum During Speciation With Gene Flow." *Journal of Heredity* 105: 810–820.
- Flaxman, S. M., J. L. Feder, and P. Nosil. 2013. "Genetic Hitchhiking and the Dynamic Buildup of Genomic Divergence During Speciation With Gene Flow." *Evolution* 67: 2577–2591.
- Formenti, G., M. Chiara, L. Poveda, et al. 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.
- Frankham, R. 2012. "How Closely Does Genetic Diversity in Finite Populations Conform to Predictions of Neutral Theory? Large Deficits in Regions of Low Recombination." *Heredity* 108: 167–178.
- Gautier, M., A. Klassmann, and R. Vitalis. 2017. "Rehh 2.0: A Reimplementation of the R Package Rehh to Detect Positive Selection From Haplotype Structure." *Molecular Ecology Resources* 17: 78–90.
- Glover, A. N., V. C. Sousa, R. D. Ridenbaugh, S. B. Sim, S. M. Geib, and C. R. Linnen. 2024. "Recurrent Selection Shapes the Genomic Landscape of Differentiation Between a Pair of Host-Specialized Haplodiploids That Diverged With Gene Flow." *Molecular Ecology* 33: e17509.
- Han, F., S. Lamichhaney, B. R. Grant, P. R. Grant, L. Andersson, and M. T. Webster. 2017. "Gene Flow, Ancient Polymorphism, and Ecological Adaptation Shape the Genomic Landscape of Divergence Among Darwin's Finches." *Genome Research* 27: 1004–1015.
- Harr, B. 2006. "Genomic Islands of Differentiation Between House Mouse Subspecies." *Genome Research* 16: 730–737.
- Hermisson, J., and P. S. Pennings. 2017. "Soft Sweeps and Beyond: Understanding the Patterns and Probabilities of Selection Footprints Under Rapid Adaptation." *Methods in Ecology and Evolution* 8: 700–716.
- 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 Genetics* 6: e1000862.
- Hudson, R. R., M. Slatkin, and W. P. Maddison. 1992. "Estimation of Levels of Gene Flow From DNA Sequence Data." *Genetics* 132: 583–589.
- Hund, A. K., J. K. Hubbard, T. Albrecht, et al. 2020. "Divergent Sexual Signals Reflect Costs of Local Parasites." *Evolution* 74: 2404–2418.
- Irwin, D. E. 2018. "Sex Chromosomes and Speciation in Birds and Other ZW Systems." *Molecular Ecology* 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." *Molecular Ecology* 25: 4488–4507.
- Irwin, D. E., B. Milá, D. P. L. Toews, et al. 2018. "A Comparison of Genomic Islands of Differentiation Across Three Young Avian Species Pairs." *Molecular Ecology* 27: 4839–4855.
- Jiang, Z., G. Song, X. Luo, D. Zhang, F. Lei, and Y. Qu. 2023. "Recurrent Selection and Reduction in Recombination Shape the Genomic Landscape of Divergence Across Multiple Population Pairs of Green-Backed Tit." *Evolution Letters* 7: 99–111.
- Kamm, J. A., J. P. Spence, J. Chan, and Y. S. Song. 2016. "Two-Locus Likelihoods Under Variable Population Size and Fine-Scale Recombination Rate Estimation." *Genetics* 203: 1381–1399.
- Kawakami, T., C. F. Mugal, A. Suh, et al. 2017. "Whole-Genome Patterns of Linkage Disequilibrium Across Flycatcher Populations Clarify the Causes and Consequences of Fine-Scale Recombination Rate Variation in Birds." *Molecular Ecology* 26: 4158–4172. <https://doi.org/10.1111/mec.14197>.
- Korunes, K. L., and K. Samuk. 2021. "Pixy: Unbiased Estimation of Nucleotide Diversity and Divergence in the Presence of Missing Data." *Molecular Ecology Resources* 21: 1359–1368.
- Kozlov, A. M., D. Darriba, T. Flouri, B. Morel, and A. Stamatakis. 2019. "RAxML-NG: A Fast, Scalable and User-Friendly Tool for Maximum Likelihood Phylogenetic Inference." *Bioinformatics* 35: 4453–4455.
- Lewis, P. O. 2001. "A Likelihood Approach to Estimating Phylogeny From Discrete Morphological Character Data." *Systematic Biology* 50: 913–925.
- 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, et al. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25: 2078–2079.
- Lombardo, G., N. Rambaldi Migliore, G. Colombo, et al. 2022. "The Mitogenome Relationships and Phylogeography of Barn Swallows (*Hirundo rustica*)." *Molecular Biology and Evolution* 39: msac113.
- Lotem, A., Y. Vortman, and R. J. Safran. 2022. "The Evidence for Divergent Sexual Selection Among Closely Related Barn Swallow Populations Is Strong." *Evolution* 76: 2204–2211.
- Malinsky, M., M. Matschiner, and H. Svardal. 2021. "Dsuite-Fast D-Statistics and Related Admixture Evidence From VCF Files." *Molecular Ecology Resources* 21: 584–595.
- Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, et al. 2013. "Genome-Wide Evidence for Speciation With Gene Flow in *Heliconius* Butterflies." *Genome Research* 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." *Molecular Biology and Evolution* 32: 244–257.
- Martin, S. H., J. W. Davey, C. Salazar, and C. D. Jiggins. 2019. "Recombination Rate Variation Shapes Barriers to Introgression Across Butterfly Genomes." *PLoS Biology* 17: e2006288.
- McKenna, A., M. Hanna, E. Banks, et al. 2010. "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data." *Genome Research* 20: 1297–1303.
- Møller, A. P. 1988. "Female Choice Selects for Male Sexual Tail Ornaments in the Monogamous Swallow." *Nature* 332: 640–642.
- Nachman, M. W., and B. A. Payseur. 2012. "Recombination Rate Variation and Speciation: Theoretical Predictions and Empirical Results From Rabbits and Mice." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 367: 409–421.
- Nadeau, N. J., A. Whibley, R. T. Jones, et al. 2012. "Genomic Islands of Divergence in Hybridizing *Heliconius* Butterflies Identified by Large-Scale Targeted Sequencing." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 367: 343–353.
- Nei, M., and W.-H. Li. 1979. "Mathematical Model for Studying Genetic Variation in Terms of Restriction Endonucleases." *Proceedings of the National Academy of Sciences* 76: 5269–5273.
- 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.
- Nosil, P., and J. L. Feder. 2012. "Genomic Divergence During Speciation: Causes and Consequences Introduction." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 367: 332–342.
- Nosil, P., D. J. Funk, and D. Ortiz-Barrientos. 2009. "Divergent Selection and Heterogeneous Genomic Divergence." *Molecular Ecology* 18: 375–402.
- Poelstra, J. W., N. Vijay, C. M. Bossu, et al. 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.
- Quinlan, A. R., and I. M. Hall. 2010. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics* 26: 841–842.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
- Renaut, S., C. J. Grassa, S. Yeaman, et al. 2013. "Genomic Islands of Divergence Are Not Affected by Geography of Speciation in Sunflowers." *Nature Communications* 4: 1827.
- Riesch, R., M. Muschick, D. Lindtke, et al. 2017. "Transitions Between Phases of Genomic Differentiation During Stick-Insect Speciation." *Nature Ecology & Evolution* 1: 1–13.
- Rodrigues, M. F., A. D. Kern, and P. L. Ralph. 2024. "Shared Evolutionary Processes Shape Landscapes of Genomic Variation in the Great Apes." *Genetics* 226: iyae006.
- Romano, A., A. Costanzo, D. Rubolini, N. Saino, and A. P. Møller. 2017. "Geographical and Seasonal Variation in the Intensity of Sexual Selection in the Barn Swallow *Hirundo rustica*: A Meta-Analysis." *Biological Reviews* 92: 1582–1600.
- 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, et al. 2016. "Genome-Wide Differentiation in Closely Related Populations: The Roles of Selection and Geographic Isolation." *Molecular Ecology* 25: 3865–3883.
- Safran, R. J., Y. Vortman, B. R. Jenkins, et al. 2016. "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.
- Schild, D. R., C. E. Brown, S. B. Shakya, G. M. Calabrese, R. J. Safran, and F. H. Sheldon. 2024. "Phylogeny and Historical Biogeography of the Swallow Family (Hirundinidae) Inferred From Comparisons of Thousands of UCE Loci." *Molecular Phylogenetics and Evolution* 197: 108111.
- Schild, D. R., J. K. Carter, E. S. C. Scordato, et al. 2024. "Sexual Selection Promotes Reproductive Isolation in Barn Swallows." *Science* 386: ead8766.
- Schild, D. R., E. S. C. Scordato, C. C. R. Smith, et al. 2021. "Sex-Linked Genetic Diversity and Differentiation in a Globally Distributed Avian Species Complex." *Molecular Ecology* 30: 2313–2332.



- Schiffels, S., and R. Durbin. 2014. "Inferring Human Population Size and Separation History From Multiple Genome Sequences." *Nature Genetics* 46: 919–925.
- Schrider, D. R. 2020. "Background Selection Does Not Mimic the Patterns of Genetic Diversity Produced by Selective Sweeps." *Genetics* 216: 499–519.
- Scordato, E. S. C., and R. J. Safran. 2014. "Geographic Variation in Sexual Selection and Implications for Speciation in the Barn Swallow." *Avian Research* 5: 1.
- Scordato, E. S. C., C. C. R. Smith, G. A. Semenov, et al. 2020. "Migratory Divides Coincide With Reproductive Barriers Across Replicated Avian Hybrid Zones Above the Tibetan Plateau." *Ecology Letters* 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." *Molecular Ecology* 26: 5676–5691.
- Secomandi, S., G. R. Gallo, M. Sozzoni, et al. 2023. "A Chromosome-Level Reference Genome and Pangenome for Barn Swallow Population Genomics." *Cell Reports* 42: 111992.
- Seehausen, O., R. K. Butlin, I. Keller, et al. 2014. "Genomics and the Origin of Species." *Nature Reviews Genetics* 15: 176–192.
- Sella, G., D. A. Petrov, M. Przeworski, and P. Andolfatto. 2009. "Pervasive Natural Selection in the Drosophila Genome?" *PLoS Genetics* 5: e1000495.
- Shang, H., D. L. Field, O. Paun, et al. 2023. "Drivers of Genomic Landscapes of Differentiation Across a Populus Divergence Gradient." *Molecular Ecology* 32: 4348–4361.
- Singhal, S., E. M. Leffler, K. Sannareddy, et al. 2015. "Stable Recombination Hotspots in Birds." *Science* 350: 928–932. <https://doi.org/10.1126/science.aad0843>.
- Slatkin, M. 1991. "Inbreeding Coefficients and Coalescence Times." *Genetical Research* 58: 167–175.
- Smeds, L., A. Qvarnström, and H. Ellegren. 2016. "Direct Estimate of the Rate of Germline Mutation in a Bird." *Genome Research* 26: 1211–1218.
- Smith, C. C. R., S. M. Flaxman, E. S. C. Scordato, et al. 2018. "Demographic Inference in Barn Swallows Using Whole-Genome Data Shows Signal for Bottleneck and Subspecies Differentiation During the Holocene." *Molecular Ecology* 27: 4200–4212.
- Smith, J. M., and J. Haigh. 1974. "The Hitch-Hiking Effect of a Favourable Gene." *Genetical Research* 23: 23–35.
- Spence, J. P., and Y. S. Song. 2019. "Inference and Analysis of Population-Specific Fine-Scale Recombination Maps Across 26 Diverse Human Populations." *Science Advances* 5: eaaw9206.
- Stankowski, S., M. A. Chase, A. M. Fuiten, M. F. Rodrigues, P. L. Ralph, and M. A. Streisfeld. 2019. "Widespread Selection and Gene Flow Shape the Genomic Landscape During a Radiation of Monkeyflowers." *PLoS Biology* 17: e3000391.
- Stephan, W. 2010. "Genetic Hitchhiking Versus Background Selection: The Controversy and Its Implications." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 365: 1245–1253.
- Swofford, D. L. 2003. *PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods)*. Version 4. Sinauer Associates.
- Tajima, F. 1989. "Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism." *Genetics* 123: 585–595.
- Terhorst, J., J. A. Kamm, and Y. S. Song. 2017. "Robust and Scalable Inference of Population History From Hundreds of Unphased Whole Genomes." *Nature Genetics* 49: 303–309. <https://doi.org/10.1038/ng.3748>.
- Turbek, S. P., D. R. Schield, E. S. C. Scordato, et al. 2022. "A Migratory Divide Spanning Two Continents Is Associated With Genomic and Ecological Divergence." *Evolution* 76: 722–736.
- Turner, A. 2018. "Swallows and Martins (Hirundinidae)." In *Handbook of the Birds of the World*, edited by J. del Hoyo, A. Elliott, D. A. Christie, and E. de Juana. Lynx Edicions.
- Turner, A. K., and C. Rose. 1989. *Swallows & Martins: An Identification Guide and Handbook*. Houghton Mifflin.
- Turner, T. L., and M. W. Hahn. 2007. "Locus-And Population-Specific Selection and Differentiation Between Incipient Species of *Anopheles gambiae*." *Molecular Biology and Evolution* 24: 2132–2138.
- Turner, T. L., M. W. Hahn, and S. V. Nuzhdin. 2005. "Genomic Islands of Speciation in *Anopheles gambiae*." *PLoS Biology* 3: e285.
- Van der Auwera, G. A., M. O. Carneiro, C. Hartl, et al. 2013. "From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline." *Current Protocols in Bioinformatics* 43: 10–11.
- Van Doren, B. M., L. Campagna, B. Helm, J. C. Illera, I. J. Lovette, and M. Liedvogel. 2017. "Correlated Patterns of Genetic Diversity and Differentiation Across an Avian Family." *Molecular Ecology* 26: 3982–3997.
- Van Oers, K., A. W. Santure, I. De Cauwer, et al. 2014. "Replicated High-Density Genetic Maps of Two Great Tit Populations Reveal Fine-Scale Genomic Departures From Sex-Equal Recombination Rates." *Heredity* 112: 307–316.
- Via, S. 2009. "Natural Selection in Action During Speciation." *Proceedings of the National Academy of Sciences* 106: 9939–9946.
- Via, S. 2012. "Divergence Hitchhiking and the Spread of Genomic Isolation During Ecological Speciation-With-Gene-Flow." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 367: 451–460.
- Via, S., and J. West. 2008. "The Genetic Mosaic Suggests a New Role for Hitchhiking in Ecological Speciation." *Molecular Ecology* 17: 4334–4345.
- Vijay, N., C. M. Bossu, J. W. Poelstra, et al. 2016. "Evolution of Heterogeneous Genome Differentiation Across Multiple Contact Zones in a Crow Species Complex." *Nature Communications* 7: 13195.
- Vijay, N., M. Weissensteiner, R. Burri, T. Kawakami, H. Ellegren, and J. B. W. Wolf. 2017. "Genomewide Patterns of Variation in Genetic Diversity Are Shared Among Populations, Species and Higher-Order Taxa." *Molecular Ecology* 26: 4284–4295.
- Wang, J., N. R. Street, D. G. Scofield, and P. K. Ingvarsson. 2016. "Natural Selection and Recombination Rate Variation Shape Nucleotide Polymorphism Across the Genomes of Three Related *Populus* Species." *Genetics* 202: 1185–1200.
- Weir, B. S., and C. C. Cockerham. 1984. "Estimating F-Statistics for the Analysis of Population Structure." *Evolution* 38: 1358–1370.
- Wilkins, M. R., D. Shizuka, M. B. Joseph, J. K. Hubbard, and R. J. Safran. 2015. "Multimodal Signalling in the North American Barn Swallow: A Phenotype Network Approach." *Proceedings of the Royal Society B: Biological Sciences* 282: 20151574.
- Winkler, D. W., S. M. Billerman, and I. J. Lovette. 2020. "Swallows (Hirundinidae), Version 1.0." In *Birds of the World*, edited by S. M. Billerman, B. K. Keeney, P. G. Rodewald, and T. S. Schulenberg. Cornell Lab of Ornithology.
- Wolf, J. B. W., and H. Ellegren. 2017. "Making Sense of Genomic Islands of Differentiation in Light of Speciation." *Nature Reviews. Genetics* 18: 87–100.
- Wu, C. 2001. "The Genic View of the Process of Speciation." *Journal of Evolutionary Biology* 14: 851–865.



Zink, R. M., A. Pavlova, S. Rohwer, and S. V. Drovetski. 2006. "Barn Swallows Before Barns: Population Histories and Intercontinental Colonization." *Proceedings of the Royal Society B: Biological Sciences* 273: 1245–1251.

## Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** Results of ABBA-BABA tests with different configurations of species assigned to P1, P2 and P3 tips. *p*-values are adjusted based on Benjamini–Hochberg correction for multiple testing. **Data S2:** Mean and standard deviation population differentiation (FST), nucleotide divergence ( $d_{xy}$ ) and nucleotide diversity ( $\pi$ ) among species and subspecies across the whole genome, and on autosomes, the Z chromosome, specifically. All statistics were calculated in non-overlapping 1 Mb windows. **Data S3:** Spearman correlation coefficients between genome-wide landscapes of relative differentiation (FST) between pairs of *Hirundo* species. Correlation coefficients were calculated based on mean values in non-overlapping 1 Mb windows. **Data S4:** Spearman rank correlation coefficients between genomic landscapes of nucleotide diversity and divergence. Branch lengths correspond to the total phylogenetic distance between landscapes in substitutions per site. Correlation and branch length values correspond to Figure 4 in the main text. **Data S5:** Statistical summaries of relationships between population genetic summary statistics and genomic features. **Figure S1:** Breeding and year-round distributions of *Hirundo* species included in this study, based on records in the Handbook of the Birds of the World (del Hoyo et al. 2004; Winkler et al. 2020) and obtained from the IUCN. Following recommendations from Broyles et al. (2023), we exclude the range of the hill swallow (*H. domicola*; previously *H. t. domicola*) from the Pacific swallow (*H. javanica*; previously *H. t. javanica*). Arrows in the map for *H. atrocaerulea* point to its fragmented montane distribution in Africa. **Figure S2:** Phylogenetic relationships among *Hirundo* species and subspecies estimated using maximum likelihood (A) and coalescent-based species tree (B) inference based on a concatenated matrix of genome-wide SNPs with no missing data. Nodal values indicate bootstrap support. The outgroup *Petrochelidon pyrrhonota* is not shown. **Figure S3:** Phylogenetic relationships and branch lengths among *Hirundo* species and subspecies estimated using maximum likelihood based on a concatenated matrix of genome-wide SNPs with no missing data. Nodal values indicate bootstrap support. Branch lengths are represented as the number of substitutions per site. **Figure S4:** Results of ABBA-BABA tests of introgression between *Hirundo* species. The grid summarises the results of various four-taxon topology arrangements to test for introgression between pairs of species (P2 and P3 taxa in respective analyses). Shaded red squares indicate that introgression was detected based on significant D statistics summarising the ratio of ABBA to BABA derived allele patterns. Darker red indicates stronger evidence of introgression (see legend). Tree topologies are included to provide evolutionary context for comparisons and to guide inferences of introgression between ancestral branches. We interpret evidence for introgression between one taxon and all members of a clade sharing a more recent common ancestor as being consistent with introgression between ancestral populations (i.e., internal branches). We collapsed these sets of tests to denote instances of ancestral introgression; these are summarised by bolded borders surrounding a series of squares in the grid (e.g., significant positive D statistics between *H. atrocaerulea* and all members of the barn swallow clade: *H. smithii*, *H. nigrita*, *H. albigularis*, *H. angolensis*, *H. aethiopica* and *H. rustica*). Using this procedure, we infer six introgression events among ancestral branches, denoted as A–F in the grid and with corresponding arrows between branches on the tree to the left. **Figure S5:** Correlation between mean genome-wide nucleotide divergence ( $d_{xy}$ ) and relative differentiation (FST) for pairs of *Hirundo* species (points). Grey points are values for pairs of barn swallow subspecies (i.e., within-species comparisons). Closed blue points are FST values between species within the barn swallow clade (*H. rustica*, *H. aethiopica*, *H. angolensis*, *H. albigularis*, *H. nigrita* and *H. smithii*). Open blue points are values for all other between-species comparisons (e.g., *H. rustica* vs. *H. atrocaerulea*). **Figure S6:** Genome-wide variation in relative

population differentiation (FST) between *Hirundo* subspecies and species, shown as genome scans in 1 Mb sliding windows with a 100 kb step size across chromosomes (alternating white and grey vertical bands). Recombination rate in *H. rustica* is shown at bottom for context. **Figure S7:** Genome-wide variation in absolute nucleotide divergence ( $d_{xy}$ ) between *Hirundo* subspecies and species, shown as genome scans in 1 Mb sliding windows with a 100-kb step size across chromosomes (alternating white and grey vertical bands). Recombination rate in *H. rustica* is shown at bottom for context. **Figure S8:** Genome-wide variation in nucleotide diversity ( $\pi$ ) within *Hirundo* subspecies and species, shown as genome scans in 1 Mb sliding windows with a 100 kb step size across chromosomes (alternating white and grey vertical bands). Recombination rate in *H. rustica* is shown at bottom for context. **Figure S9:** Variation in per-generation recombination rate among *Hirundo* species illustrated by patterns across chromosome 1A, chromosome 4 and the Z chromosome. Recombination rates are shown as genome scans in 1 Mb sliding windows with a 100 kb step size. Lines are coloured according to specific species values with matching colour labels. **Figure S10:** Schematic representation of the phylogenetic distance between genomic landscapes of divergence ( $d_{xy}$ ) and diversity ( $\pi$ ), measured as the total branch length (blue) in substitutions per site to the common ancestor of respective landscapes (blue point). Branch length calculations correspond to interpretations of genomic landscape correlations as a function of phylogenetic distance in the main text (i.e., Figure 4). **Figure S11:** A Genome-wide relationship between exon density (measured as the proportion of sites within annotated exons per window) and recombination rate measured in non-overlapping 1 Mb windows (points). The Spearman correlation coefficient ( $\rho$ ) and *p*-value are labelled above. B Spearman correlation coefficients (points) within chromosomes between exon density, recombination rate and FST as a function of chromosome length (bp). The dashed horizontal line represents no correlation between variables. **Figure S12:** Genome-wide relationships between genetic diversity and divergence, recombination rate and exon density illustrated by *H. rustica* – *H. aethiopica* (A), *H. neoxena* – *H. javanica* (B), *H. aethiopica* – *H. smithii* (C), *H. rustica* – *H. javanica* (D) and *H. smithii* – *H. dimidiata* (E) species pairs. Panels to the left show phylogenetic distance between pairs of species as shaded branches and summarise the present-day geographic arrangement of species pairs (allopatric vs. partial sympatry). The centre panel in A–E shows correlations between mean  $\pi$  and  $d_{xy}$  for each pair of species, with labels for Spearman correlation coefficients ( $\rho$ ). The right two panels show correlations between mean  $\pi$  and recombination rate and exon density, respectively, with summaries of multiple linear regression (MLR) to test the effects ( $r^2$  and  $t$ ) of genomic features on summary statistics. All statistical comparisons are based on mean values in non-overlapping 1 Mb windows. In all panels  $***p < 2.2 \times 10^{-16}$ ;  $*p < 0.05$ ; NS, not significant. **Figure S13:** Genome-wide relationships between genetic divergence and recombination rate, and exon density illustrated by *H. rustica* – *H. aethiopica* (A), *H. neoxena* – *H. javanica* (B), *H. aethiopica* – *H. smithii* (C), *H. rustica* – *H. javanica* (D) and *H. smithii* – *H. dimidiata* (E) species pairs. Panels to the left show phylogenetic distance between pairs of species as shaded branches and summarise the present-day geographic arrangement of species pairs (allopatric vs. partial sympatry). The right two panels in A–E show correlations between  $d_{xy}$  and recombination rate and exon density, respectively, with summaries of multiple linear regression (MLR) to test the effects ( $r^2$  and  $t$ ) of genomic features on summary statistics. All statistical comparisons are based on mean values in non-overlapping 1 Mb windows. In all panels  $***p < 2.2 \times 10^{-16}$ ;  $*p < 0.05$ ; NS, not significant. **Figure S14:** Comparison of Spearman correlation coefficients ( $\rho$ ) between  $\pi$  and  $d_{xy}$  in FST islands versus the genomic background in species pairs in parapatry/partial sympatry (A) versus allopatry (B). Statistical summaries of Mann–Whitney *U*-tests are labelled;  $***p < 0.01$ ; NS, not significant. **Table S1:** Samples used in the study, museum accessions, locality details, sex and whole genome mean read depth, assuming a 1.2 Gbp genome size. **Table S2:** Present-day geographic arrangement of *Hirundo* species pairs based on accounts in the Handbook of the Birds of the World (Winkler et al. 2020) for the breeding and year-round distribution of each species. Species with no geographic overlap are considered to be in strict allopatry. Species with adjacent or partly overlapping distributions are considered

parapatric/partially sympatric. **Table S3:** Spearman rank correlation coefficients ( $\rho$ ) between recombination landscapes of *Hirundo* species based on mean recombination rate in 1 Mb non-overlapping windows. **Table S4:** Mean  $\pm$  standard deviation genetic differentiation ( $F_{ST}$ ), nucleotide divergence ( $d_{xy}$ ) and nucleotide diversity (mean  $\pi$ ) in the genomic background and islands of differentiation between pairs of *Hirundo* subspecies and species. **Table S5:** Statistical comparison of nucleotide diversity ( $\pi$ ) in genomic islands of differentiation ( $F_{ST}$ ) versus the genome background. One-way ANOVA and Tukey post hoc test results comparing distributions from the genome background, all  $F_{ST}$  islands and  $F_{ST}$  islands outside of centromere regions. **Table S6:** Statistical comparison of nucleotide divergence ( $d_{xy}$ ) in genomic islands of differentiation ( $F_{ST}$ ) versus the genome background. One-way ANOVA and Tukey post hoc test results comparing distributions from the genome background, all  $F_{ST}$  islands and  $F_{ST}$  islands outside of centromere regions.