Received: 8 April 2017 | Revised: 2 May 2017 | Accepted: 15 May 2017

DOI: 10.1111/mec.14197

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Whole-genome patterns of linkage disequilibrium across flycatcher populations clarify the causes and consequences of fine-scale recombination rate variation in birds

Takeshi Kawakami[1,2*] | Carina F. Mugal[1*] | Alexander Suh[1] | Alexander Nater[1,3] | Reto Burri[1,4] | Linnéa Smeds[1] | Hans Ellegren[1]

[1]Department of Evolutionary Biology, Evolutionary Biology Centre (EBC), Uppsala University, Uppsala, Sweden

[2]Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK

[3]Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zürich, Switzerland

[4]Department of Population Ecology, Friedrich Schiller University Jena, Jena, Germany

Correspondence
Hans Ellegren, Department of Evolutionary Biology, Evolutionary Biology Centre (EBC), Uppsala University, Uppsala, Sweden.
Email: hans.ellegren@ebc.uu.se

Funding information
Swedish Research Council, Grant/Award Number: 2007-8731, 2010-5650, 2013-8271, 2014-6325; Marie Sklodowska Curie Actions, Co-fund Project INCA, Grant/Award Number: 600398; Knut and Alice Wallenberg Foundation

Abstract

Recombination rate is heterogeneous across the genome of various species and so are genetic diversity and differentiation as a consequence of linked selection. However, we still lack a clear picture of the underlying mechanisms for regulating recombination. Here we estimated fine-scale population recombination rate based on the patterns of linkage disequilibrium across the genomes of multiple populations of two closely related flycatcher species (*Ficedula albicollis* and *F. hypoleuca*). This revealed an overall conservation of the recombination landscape between these species at the scale of 200 kb, but we also identified differences in the local rate of recombination despite their recent divergence (<1 million years). Genetic diversity and differentiation were associated with recombination rate in a lineage-specific manner, indicating differences in the extent of linked selection between species. We detected 400–3,085 recombination hotspots per population. Location of hotspots was conserved between species, but the intensity of hotspot activity varied between species. Recombination hotspots were primarily associated with CpG islands (CGIs), regardless of whether CGIs were at promoter regions or away from genes. Recombination hotspots were also associated with specific transposable elements (TEs), but this association appears indirect due to shared preferences of the transposition machinery and the recombination machinery for accessible open chromatin regions. Our results suggest that CGIs are a major determinant of the localization of recombination hotspots, and we propose that both the distribution of TEs and fine-scale variation in recombination rate may be associated with the evolution of the epigenetic landscape.

KEYWORDS
CpG island, GC-biased gene conversion, linked selection, population genomics, recombination, transposon

1 | INTRODUCTION

Meiotic recombination plays multiple important roles in DNA sequence evolution, genome integrity and the process of speciation.

*These authors contributed equally to this study.

In the context of population genetics, meiotic recombination contributes to patterns of genomic diversity by shuffling alleles on different haplotypes. This can be a double-edged sword because recombination can not only create novel and potentially advantageous combinations of alleles but can also break up existing advantageous ones. In addition, recombination influences the efficacy of selection at neighbouring sites along a chromosome (i.e., Hill-Robertson interference) (Hill & Robertson, 1966). Moreover, recombination together with the density of targets for selection influences genome-wide variation in genetic diversity via the process of linked selection (Cutter & Payseur, 2013). Both selective sweeps and background selection reduce genetic variation at neutral sites linked to those under natural selection, which manifests as a locally reduced effective population size (N_e). This association between the rate of recombination and local N_e in turn leads to locally accelerated lineage sorting during the process of speciation in regions of low recombination (Charlesworth, 2009). Therefore, meiotic recombination has been suggested to play an important role in shaping the differentiation landscape and linked selection can be seen as a null model in explaining the existence of islands of differentiation (Cruickshank & Hahn, 2014; Wolf & Ellegren, 2017).

Recombination can also affect local nucleotide composition by the process called GC-biased gene conversion (gBGC) (Duret & Galtier, 2009). gBGC is associated with meiotic recombination and leads to the preferential fixation of "strong" bases (G and C) over "weak" bases (A and T). Although this bias is not very strong, gBGC gradually establishes a positive correlation between GC content and recombination rate over time (Backström et al., 2010; Birdsell, 2002; Duret & Arndt, 2008). Recombination may further impact genomic stability by introducing point mutations and structural rearrangements via nonallelic homologous recombination (Sasaki, Lange, & Keeney, 2010; Strathern, Shafer, & McGill, 1995). Moreover, distribution of transposable elements (TEs) is associated with variation in recombination rate due to effective elimination of deleterious TE insertions at high recombination regions and/or suppression of recombination via epigenetic regulation of TE activities (Dolgin & Charlesworth, 2008; Rizzon, Marais, Gouy, & Biemont, 2002). However, several studies suggested a positive association between recombination and insertions of several TE families because of their integration site preferences (Baller, Gao, & Voytas, 2011; Liu et al., 2009; Yoshida et al., 2017), indicating complex interaction between recombination and TE distribution.

Studies in rodents and primates have shown that recombination is genetically regulated by localizing recombination-initiating DNA double-strand breaks (DSBs) to small regions of the genome, known as "recombination hotspots," where recombination rate is 100- to 1,000-fold higher than the genomic average (Baudat, Imai, & de Massy, 2013; Paigen & Petkov, 2010). These recombination hotspots are usually less than a few kilobases (kb) long and contain degenerate DNA sequence motifs of 7–13 nucleotides, which are recognized by the zinc finger protein PR domain-containing 9 (PRDM9) (Myers et al., 2010; Parvanov, Petkov, & Paigen, 2010). However, many nonmammalian species, such as birds, plants and yeasts, lack PRDM9 but still show highly variable rates of recombination across the genome with distinct recombination hotspots (Choi et al., 2013; Pan et al., 2011; Singhal et al., 2015; Smeds, Mugal, Qvarnström, & Ellegren, 2016). In addition, dogs, where PRDM9 was pseudogenized, still show hotspots (Auton et al., 2013; Axelsson et al., 2012; Berglund, Quilez, Arndt, & Webster, 2015). In these species, recombination hotspots tend to coincide with genomic features, such as regions in the proximity of transcription start sites (TSSs) and transcription termination sites (TTSs). It has been suggested that colocalization of recombination hotspots with transcription initiation contributes to an evolutionary stable recombination landscape (Axelsson et al., 2012; Lam & Keeney, 2015; Singhal et al., 2015). However, it is not clear whether transcriptional activity per se or underlying genomic features that are associated with promoter regions, such as CpG islands (CGIs), are responsible for this colocalization.

Studies of ecological adaptation often use multiple pairs of populations in ecologically distinct habitats (e.g., Fraser, Kunstner, Reznick, Dreyer, & Weigel, 2015; Hohenlohe et al., 2010; Westram et al., 2014), and shared patterns of elevated genetic differentiation along a chromosome are often considered as a signature of parallel evolution. However, because of the potential influence of recombination rate variation on several population genetics statistics (e.g., nucleotide diversity [π] and genetic differentiation [F_{ST}]) via the effect of linked selection, conserved patterns of recombination rate variation can lead to concerted evolution of diversity and differentiation landscapes independently of ecological selection (Burri et al., 2015; Roesti, Hendry, Salzburger, & Berner, 2012; Van Doren et al., 2017; Vijay et al., 2017). On the other hand, changes of the recombination landscape can affect the pattern of genetic diversity in a given lineage, mimicking evidence of lineage-specific adaptation. For proper inferences of patterns of genetic diversity, it is therefore important to characterize variation in recombination rate for each population or species under study.

Using linkage map data, we have previously documented heterogeneity in the rate of recombination across the genome of collared flycatcher (*Ficedula albicollis*) at the resolution of 200-kb windows (Kawakami et al., 2014). The precision of pedigree-based recombination rate estimates (hereafter referred to as pedigree recombination rate) is dependent on the number of meiotic crossovers in a given pedigree, and, as a result, there were regions with insufficient resolution in our previous study, particularly at the centre of chromosomes where crossovers rarely occurred. To gain a better insight into variation in recombination rate in avian genomes, we here estimate the population-scaled recombination rate (ρ) based on whole-genome patterns of linkage disequilibrium (hereafter referred to as LD recombination rate) in multiple populations of two closely related species of Old World flycatchers, collared flycatcher and pied flycatcher (*F. hypoleuca*). These two species are estimated to have diverged <1 million years ago with speciation essentially completed, although occasional hybridization still occur (Nadachowska-Brzyska, Burri, Smeds, & Ellegren, 2016; Nadachowska-Brzyska et al., 2013; Nater, Burri, Kawakami, Smeds, & Ellegren, 2015; Wiley, Qvarnstrom, Andersson, Borge, & Saetre, 2009). LD and pedigree recombination rate estimates describe two different timescales of recombination activity. The LD recombination map reflects cumulative events of recombination for hundreds of thousands of generations (Auton &

McVean, 2012; Chan, Jenkins, & Song, 2012). In contrast, the pedigree recombination map reflects the per-generation recombination rate observed in a given pedigree. Therefore, LD recombination rate estimates of the two flycatchers do not only allow us to characterize variation in recombination rate and its regulation at much higher resolution than before but also to analyse the conservation of the recombination landscape and the location of hotspots among populations and species at multiple evolutionary timescales. As the heterogeneous landscape of differentiation in *Ficedula* flycatchers seems mainly to be the result of the extent of linked selection (Burri et al., 2015), species-specific estimates of recombination rate further make it possible to study the impact of changes in the recombination landscape on differentiation.

2 | MATERIALS AND METHODS

2.1 | Samples and SNP selection

We used whole-genome sequences of 77 collared flycatchers from Italy (n = 19), Hungary (n = 20), Czech Republic (n = 19) and on the Baltic Sea island Öland in Sweden (n = 19), and 72 pied flycatchers from Spain (n = 15), Czech Republic (n = 20), mainland Sweden (n = 20) and on Öland (n = 17) (Burri et al., 2015). We also used six additional samples of collared flycatcher collected from the Baltic population, leading to a total of 83 samples of collared flycatcher (sample Accession nos.: u006, u012, u016, u017 and u117). Detailed information about whole-genome sequencing and sequence analysis is described in Burri et al. (2015). Briefly, sequence reads obtained by Illumina HiSeq2000 were mapped to a repeat-masked version of the FicAlb1.5 reference genome of collared flycatcher (Ellegren et al., 2012), at a mean coverage of 14.7 ± 4.5×, to identify SNPs for each population. Of 50,005,568 SNPs initially identified over all the samples, we selected 10,286,186 and 7,111,155 high-quality SNPs in collared flycatcher and pied flycatcher, respectively, by applying the following filters: (i) SNPs with more than two alleles were removed; (ii) individual genotypes with read coverage <5 or >50 were masked and treated as missing genotypes; (iii) SNPs that were genotyped for <90% of individuals per species after applying the second filter were removed; (iv) SNPs with a minor allele frequency (MAF) per population < 2% were removed to exclude singletons; and (v) SNPs on the Z chromosome that were scored as heterozygotes for at least one female were removed. Haplotypes of these SNPs were inferred for each species separately by fastPHASE (Scheet & Stephens, 2006) using default parameters except for the maximum number of EM iteration that was set to 50. The subpopulation option was used to specify four populations within species. Five, 10 and 15 clusters of haplotypes were considered for determining the number of haplotype clusters.

2.2 | Estimation of population recombination rate

LD recombination rate ρ (ρ = $4N_e r$, where N_e denotes the effective population size, and r denotes the recombination rate) was estimated using LDhelmet (Chan et al., 2012). This program uses a coalescent-based, reversible-jump Markov chain Monte Carlo (rjMCMC) simulation to infer historical, sex-averaged recombination rate based on the patterns of LD between pairs of SNPs. To improve the accuracy of coalescent simulations, the program takes into account ancestral allele frequencies and a nucleotide substitution model for inferring ancestral recombination events while allowing for mutations along the branches of coalescent trees. To compute frequencies of ancestral alleles, we polarized the SNP data of the collared flycatcher and pied flycatcher populations based on an ancestral flycatcher reference genome. The ancestral genome was reconstructed using in-group samples from four populations of collared flycatcher, four populations of pied flycatcher and one individual from each of the two out-group species, red-breasted flycatcher (*F. parva*) and snowy-browed flycatcher (*F. hyperythra*) (Burri et al., 2015). We only considered sites that had a minimum variant quality of 15 and minimum mean mapping quality of 20. In addition, we requested a minimum coverage of 3× per site in at least five individuals per population of the in-group species and one individual of each out-group species. The ancestral state was called when at least two of the three groups (in-group species, red-breasted flycatcher and snowy-browed flycatcher) were monomorphic for a given allele. For the polarization of SNPs, we requested a minimum read coverage of 5× and a maximum of 50× in at least 18 individuals per population in at least one population. The prior probability for the inferred ancestral allele was set as 0.97 while prior probabilities of the other three alleles were set as 0.01. For collared flycatcher and pied flycatcher, 25% and 21% of variable sites, respectively, were polarized in this way, whereas prior probabilities at the remaining variable sites were set as 0.25 for each nucleotide. We used the following nucleotide substitution matrix with columns and rows ordered as A, C, G and T, previously established from molecular evolutionary analysis of chicken intergenic regions (Mugal, Arndt, & Ellegren, 2013).

$$\begin{bmatrix} 0.1125 & 0.1478 & 0.5654 & 0.1743 \\ 0.2061 & 0.0000 & 0.1939 & 0.5999 \\ 0.5999 & 0.1939 & 0.0000 & 0.2061 \\ 0.1743 & 0.5654 & 0.1478 & 0.1125 \end{bmatrix}$$

We ran five independent rjMCMC simulations for each population with each run for 2,000,000 iterations after 200,000 iterations of burn-in. Estimated LD recombination rate was averaged over the five runs. Based on the results of simulation analysis (Text S1, Figure S1 Supporting Information), we used a block penalty of 10 to minimize overfit while maintaining resolution of variation in recombination rate.

To characterize variation in recombination rate across genomes, we calculated weighted average of the LD recombination rate in 200-kb windows by taking the number of SNPs into account. Following a previous approach (Auton et al., 2012), the estimated LD recombination rate (ρ = $4N_e r$) in a given window was converted into the unit of per-generation recombination rate (cM/Mb) by dividing it by genomewide N_e. The genomewide N_e was estimated for each population separately by regressing ρ to the pedigree recombination rate of collared flycatcher in 200 kb (Kawakami et al., 2014). Following Auton et al. (2012), robust linear regression was used (without

intercept) to obtain the gradient of the fitted line between these two recombination rates. Using this approach, we estimated genome-wide N_e for four populations of collared flycatcher (Baltic: 52,500, Czech: 142,040, Hungary: 154,109 and Italy: 156,129) and four populations of pied flycatcher (Baltic: 159,103, Czech: 119,910, Sweden: 160,521 and Spain: 125,909). Correlation analyses between population-specific and species-specific estimates of LD recombination rate were performed after applying log-transformation to normalize the data. Nucleotide diversity (π), genetic differentiation (F_{ST}) and population branch statistics (PBS), reported in Burri et al. (2015), were also log-transformed for correlation analysis with LD recombination rate. Because of the short branch lengths within species relative to the branch lengths between species, PBS for Italian population of collared flycatcher and Spanish population of pied flycatcher were used as representatives of respective species.

To evaluate sensitivity and specificity of detecting recombination hotspots, we used the program MaCS (Chen, Marjoram, & Wall, 2009) to simulate sequences of 200 kb with various combinations of effective population size, background recombination rate, the number of haplotypes and hotspot intensity and width (Text S1, Supporting Information). Based on the results of these simulations (Table S1, Supporting Information), recombination hotspots were called if given pairs of SNPs had at least 10 times higher recombination rate than the mean recombination rate of each surrounding 200-kb window, and the width was at least 750 bp from the beginning to the end of hotspots. Recombination hotspots were not called in windows with low SNP density (<1 SNP/kb). LDhelmet occasionally infers narrow spikes with unusually high ρ, which is likely to be a spurious artefact of the rjMCMC procedure (Chan et al., 2012). To distinguish between spurious peaks and true recombination hotspots, we only called recombination hotspots that were detected in all five independent rjMCMC runs.

2.3 | Strength of gBGC

Preferential fixation of "strong" bases (S) over "weak" bases (W) by gBGC leaves footprints mimicking directional selection by skewing the site frequency spectrum at sites with S-W polymorphisms (Duret & Galtier, 2009; Glemin et al., 2015; Mugal, Weber, & Ellegren, 2015). Based on the expected allele frequency shifts at recombination hotspots, we quantified the intensity of recombination at recombination hotspots by using a variant of Fay & Wu's H (Fay & Wu, 2000), which we refer to as H_{gBGC}. Fay & Wu's H is a measure of deviations from the expected site frequency spectrum under a model of neutral molecular evolution. We defined H_{gBGC} as $H_{W\text{-}to\text{-}S} - H_{S\text{-}to\text{-}W}$, where $H_{W\text{-}to\text{-}S}$ was calculated for W-to-S polymorphisms and $H_{S\text{-}to\text{-}W}$ for S-to-W polymorphism. This should give a quantitative measure of the strength of gBGC, with negative values of H_{gBGC} indicating regions under strong influence of gBGC (i.e., high recombination activity).

2.4 | Motif identification

We used HOMER 4.2 (Heinz et al., 2010) and MEME suite 4.11.1 (Bailey, Johnson, Grant, & Noble, 2015) to discover sequence motifs associated with recombination hotspots. In the HOMER de novo motif enrichment analysis, we searched for 6- to 16-bp sequence motifs enriched at 1-kb regions containing recombination hotspots at the centre relative to 1-kb "cold" regions that were at least 10 kb away from neighbouring hotspots and had a GC content similar to the one in hotspots (<2% difference). In the MEME motif discovery analysis, 6- to 20-bp sequence motifs were searched at 1-kb hotspot regions relative to "cold" regions using discriminative motif discovery mode. Only hotspots shared among populations of collared flycatcher were used.

2.5 | Annotation of the recombination landscape

To investigate the distribution of hotspots with respect to annotations of the collared flycatcher genome (Ensembl release 73, assembly version FICALB_1.4), we classified the genome into nine nonoverlapping sequence categories: (i) intergenic regions not belonging to other categories, (ii) TEs, (iii) CpG islands (CGIs) in intergenic regions, (iv) promoter regions with CGIs (2 kb upstream of TSSs), (v) promoter regions without CGIs, (vi) first exons, (vii) other exons, (viii) first introns and (ix) other introns. We defined promoter regions as regions 2 kb upstream of TSSs as promoters are not annotated for the collared flycatcher genome. Collared flycatcher-specific repeat annotations were used for the definition of transposable elements (Smeds et al., 2015). CGIs were identified in the unmasked genome using CPGCLUSTER version 1.0 (Hackenberg et al., 2006) with default parameter settings and a minimum length of at least 50 bp. The density of hotspots was measured as the proportion of base pairs derived from hotspot sequences in a given annotation category. A recombination hotspot that overlapped several categories was assigned values to each of these categories proportional to the length of the overlap. We computed mean LD recombination rate with respect to the distance to the closest gene based on nonoverlapping 5-kb windows. We distinguished between regions upstream of the TSS or downstream of the TTS by considering the orientation of the closest gene. In addition, an average for the transcribed region itself was calculated.

To estimate LD recombination rate in TEs and compare it with the approximate time since transposition of TE copies, we re-estimated LD recombination rate using repeat unmasked data. LD recombination rate was calculated for each copy of TEs using all pairs of adjacent SNPs within the TE copies. We chose 72 LINE subfamilies and 127 LTR retrotransposon subfamilies, which had more than 50 copies within each subfamily in the collared flycatcher genome assembly. Consensus sequences of subfamilies of these TEs were constructed following Smeds et al. (2015) and merged with a library from Vijay et al. (2017) (Data S1, Supporting Information). CpG-corrected Kimura 2-parameter (K2P) genetic distances were calculated between TE copies and the consensus sequences of each subfamily, and these distances were used as a proxy for the time since transposition or "age of TEs" (Kapusta & Suh, 2017). Correlation between mean K2P genetic distance and LD recombination rate was evaluated separately for LINEs and LTR retrotransposons using all copies within LINEs and LTR retrotransposons.

To test whether LD recombination rate was higher in TEs with CGIs than those without CGIs, LINE and LTR retrotransposons were subdivided into four classes: (i) TEs whose ancestral consensus sequences contained CGIs and were identified within 400 bp from the nearest CGIs in the present-day reference genome, (ii) TEs whose ancestral consensus sequences did not contain CGIs and were identified within 400 bp from the nearest CGIs in the present-day reference genome, (iii) TEs whose ancestral consensus sequences contained CGIs and were inserted >400 bp away from CGIs in the present-day reference genome and (iv) TEs whose ancestral consensus sequences did not contain CGIs and were inserted >400 bp away from CGIs in the present-day reference genome. To test the effect of the presence of CGIs in the ancestral consensus sequences on LD recombination rate, we performed a one-tailed Wilcoxon rank-sum test by comparing classes (i) and (ii) with (iii) and (iv). Similarly, to test the effect of CGIs in the current reference genome, we compared classes (i) and (ii) with (iii) and (iv). Accurate genotype calling at repetitive regions, such as TEs, can be difficult and may result in errors in subsequent haplotype phasing, which can consequently lead to overestimation of LD recombination rate. However, the phasing error rate was not different between regions with and without TEs (error rate = 1.48% and 1.54%, respectively; Wilcoxon test, $p = .178$). Density of TEs was measured as the proportion of base pairs derived from TEs in 200-kb windows.

3 | RESULTS

3.1 | LD recombination map of the flycatcher genomes

Whole-genome LD recombination maps were constructed for each population of collared flycatcher separately after applying stringent filters (see Materials and Methods). The mean genomewide LD recombination rate (estimated in 200-kb windows) across all four populations was 3.69 cM/Mb on autosomes and 2.63 cM/Mb on the Z chromosome (Table S2, Supporting Information). The LD recombination maps showed that the autosomal rate was highly variable both within and between chromosomes with a negative nonlinear relationship with chromosome size (Figure S2, Supporting Information). The rate generally increased towards the ends of chromosomes (Figure S3, Supporting Information) and was 1.5–2.1 times higher in distal 5.5-Mb regions than internal regions. Notably, however, the rate sharply declined at the very distal ends of chromosomes (<0.5 Mb) with 29–58% reduction relative to the adjacent 5-Mb region (Figure S4, Supporting Information). Confirming earlier observations based on linkage data (Smeds et al., 2014), the pseudoautosomal region (PAR), which spans 630 kb at the terminal part of the Z chromosome, had about five times higher LD recombination rate (18.98 cM/Mb) than the autosomal average. Recombination was particularly concentrated at the 100-kb region distal to the boundary between PAR and the rest of the Z chromosome (31.46 cM/Mb) (Figure S5, Supporting Information).

The above observations were largely in agreement with those from the existing pedigree recombination map of collared flycatcher (Kawakami et al., 2014) with significant correlation between two estimates of recombination rate (Pearson's correlation coefficient $r = 0.62$, $p < 10^{-16}$) (Figure 1a, S6, Supporting Information). However, estimates of the LD recombination map revealed variation in the recombination rate in low recombination regions, where rare recombination events were not captured in the pedigree-based linkage analysis. For example, the pedigree recombination map has a total of 45.6 Mb of 200-kb windows with an estimated rate of 0 cM/Mb. Our estimates in these windows ranged between 0.01 cM/Mb and 6.98 cM/Mb, with a mean of 2.27 cM/Mb. Given the association between low recombination regions and islands of differentiation (Burri et al., 2015), the improvement of the resolution of recombination rate estimates in such regions represents an important advancement. It should also be noted that the LD

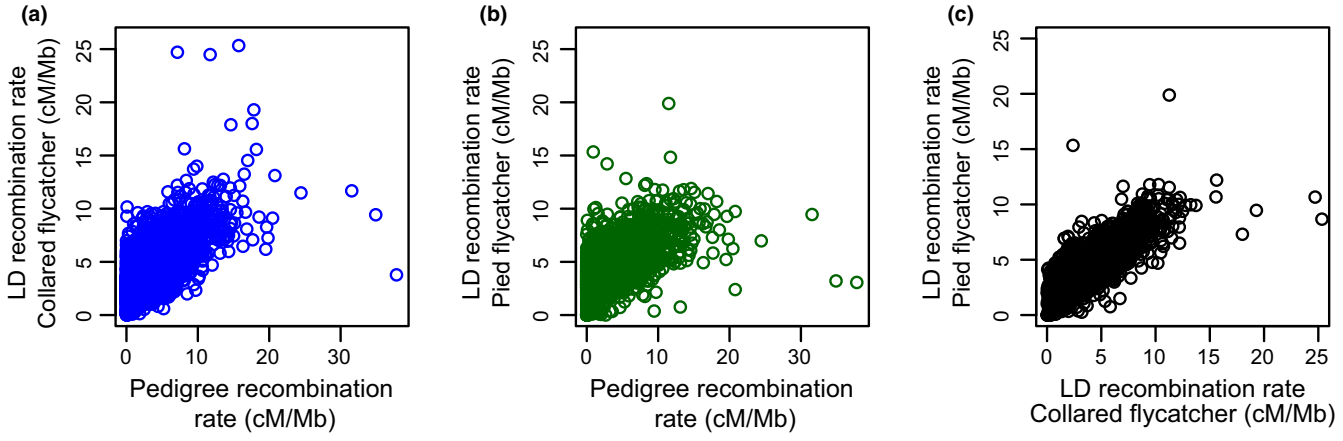

FIGURE 1 The relationship between LD recombination rate and pedigree recombination rate in 200-kb autosomal windows. (a) Correlation between LD recombination rate of collared flycatcher and pedigree recombination rate of collared flycatcher. (b) Correlation between LD recombination rate of pied flycatcher and pedigree recombination rate of collared flycatcher. (c) Correlation between LD recombination rate of collared flycatcher and LD recombination rate of pied flycatcher. [Colour figure can be viewed at wileyonlinelibrary.com]

recombination map provided recombination rate estimate at chromosome ends where the rate was not available in the pedigree map (a total of 89.69 cM in 23.2 Mb).

Next, we estimated LD recombination rate for four populations of pied flycatcher. Mean genomewide LD recombination rate based on 200-kb windows averaged across all four populations was slightly (~1.1 times) higher in pied flycatcher than in collared flycatcher in both autosomes and the Z chromosome (Wilcoxon rank-sum test, $p < 10^{-16}$ for autosomes and $p = .0006$ for the Z chromosome) (Table S1, Supporting Information). Nevertheless, the patterns of interchromosomal and intrachromosomal variation were similar in the two species (Figures S2–S6, Supporting Information). This was reflected by a significant correlation between the per-window LD recombination rates of collared flycatcher and pied flycatcher (Pearson's correlation $r = 0.79$, $p < 10^{-16}$) (Figure 1c), demonstrating conservation of the recombination landscape between species at this scale.

Closer examination of LD recombination rates estimated separately for each of the four collared flycatcher populations revealed small but statistically significant overall differences among populations in autosomes (3.63–3.75 cM/Mb, Kruskal–Wallis test, $p < 10^{-7}$), while the Z chromosome showed no significant differences (2.51–2.85 cM/Mb; $p = .09$) (Table S2, Supporting Information). In pied flycatcher, the LD recombination rate was significantly different among populations both in autosomes (4.02–4.20 cM/Mb) and on the Z chromosome (3.01–3.25 cM/Mb; $p < 10^{-16}$). The per-window LD recombination rate was significantly correlated among the four collared flycatcher populations ($r = 0.88$–0.90, $p < 10^{-16}$ for all pairwise comparisons) as well as among the four populations of pied flycatcher ($r = 0.74$–0.78, $p < 10^{-16}$ for all pairwise comparisons) (Figure 2a). Importantly, interspecific comparisons of populations showed significantly weaker correlation ($r = 0.63$–0.67) than population pairs within species (bootstrap with 1,000 replicates, $p < 10^{-4}$ for all comparisons).

3.2 | The relationship between recombination rate and the diversity landscape

It has previously been shown that variation in genetic diversity within both collared flycatcher and pied flycatcher genomes is correlated with the pedigree-based recombination rate in collared flycatcher (Burri et al., 2015). Here we extend these results by analysing the relationship between species- and population-specific LD recombination rates, and species- and population-specific data on genetic diversity. LD recombination rate and nucleotide diversity (π) showed a significant and positive correlation in all eight populations of collared flycatcher and pied flycatcher in 200-kb windows (Figure 2b). Nucleotide diversity in collared flycatcher populations were more strongly correlated with the LD recombination rate of collared flycatcher populations (Pearson's correlation coefficient $r = 0.40$–0.46) than with the LD recombination rate of pied flycatcher populations ($r = 0.27$–0.32). Somewhat surprisingly, the correlation between nucleotide diversity in pied flycatcher populations and the LD recombination rate of pied flycatcher populations was equally strong with the LD recombination rate of collared flycatcher populations (Figure 2b). The correlation between the pedigree recombination rate in collared flycatcher and nucleotide diversity in any of the eight populations was significantly lower than the correlation between population-specific estimates of the LD recombination rate and genetic diversity (Figure 2b, bootstrap with 1,000 replicates, $p < 10^{-4}$ for all comparisons), suggesting that variation in recombination rate detected in the LD based map was more tightly associated with genetic diversity than data from the pedigree-based map.

Mean genetic differentiation (F_{ST}) between populations of collared flycatcher and pied flycatcher (mean of four populations within species) was negatively correlated with mean LD recombination rate of collared flycatcher and pied flycatcher (mean of four populations within species) (Pearson's correlation coefficient $r = -0.73$ [95% confidence interval: -0.71 to -0.75], $p < 10^{-16}$ in collared flycatcher; $r = -0.63$ [95% confidence interval: -0.61 to -0.65], $p < 10^{-16}$ in pied flycatcher). The correlation using LD recombination rate was stronger than the correlation between pedigree recombination rate of collared flycatcher and F_{ST} (Pearson's correlation coefficient $r = -0.45$ [95% confidence interval = -0.42 to -0.47], $p < 10^{-16}$). Because of the phylogenetic nonindependence in estimates of F_{ST} between several closely related pairs of species and populations, we also used population branch statistics (PBS), which reflects lineage-specific patterns of genetic differentiation (Shriver et al., 2004; Yi et al., 2010). LD recombination rates of four populations of collared flycatcher showed stronger correlations with PBS in the collared flycatcher lineage than with PBS in the pied flycatcher lineage (Figure 2c) (bootstrap with 1,000 replicates, $p < 10^{-4}$ for all four comparisons). Similarly, LD recombination rates of four populations of pied flycatcher showed stronger correlation with PBS in the pied flycatcher lineage than with PBS in the collared flycatcher linage (Figure 2c) (bootstrap with 1,000 replicates, $p < 10^{-4}$ for all four comparisons). This demonstrates that species-specific estimates of recombination rate better describe species-specific patterns of genetic differentiation.

3.3 | Recombination hotspots

To investigate variation in recombination rate at a fine scale, we identified recombination hotspots in each population of collared flycatcher and pied flycatcher using pairwise SNP data. There were 400–1,482 recombination hotspots in the four populations of collared flycatchers and 1,485–3,085 recombination hotspots in the four populations of pied flycatchers (Table 1). Similar to the hotspots identified in other species (e.g., 1–2 kb in humans and chimps, 4.3 kb in dogs; Auton et al., 2012, 2013; Jeffreys, Kauppi, & Neumann, 2001), hotspots were generally narrow (mean width 1.64 kb), with a right-skewed distribution (Figure S7, Supporting Information). Mean LD recombination rate at hotspots was 99.6 cM/Mb in collared flycatchers (27 times the autosomal average) and 152.4 cM/Mb in pied flycatchers (37 times the average). The most extreme

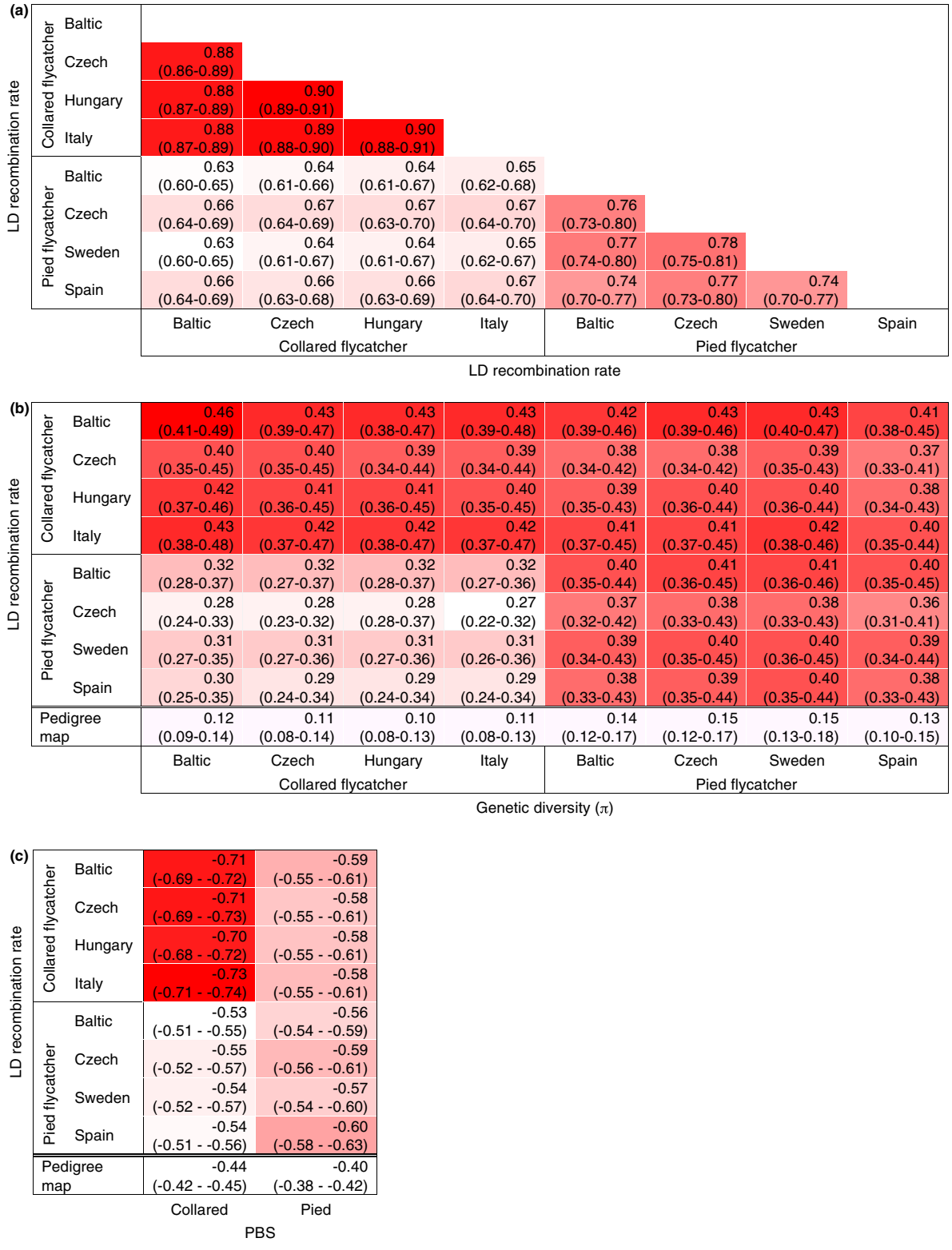

FIGURE 2 Inter- and intraspecific correlations in LD recombination rate and genetic diversity/differentiation. (a) Pearson's correlation coefficient *r* (95% confidence intervals in parentheses) between LD recombination rates of four populations of collared and pied flycatcher, respectively. (b) Pearson's correlation coefficient *r* between LD recombination rate and pairwise genetic diversity (π) among four populations of collared and pied flycatcher, respectively. (c) Pearson's correlation coefficient *r* between LD recombination rate and population branch statistics (PBS). The red shading indicates the strength of *r*. For comparison, correlation coefficients between the pedigree recombination rate and genetic diversity are shown at the bottom of Panel b [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 The number of recombination hotspots detected in four populations of collared flycatcher and four populations of pied flycatcher

	Population	Total	Within	Between	Both
Collared flycatcher	Baltic	400	188	271	83
	Czech	1,274	619	884	265
	Hungary	1,482	678	1,000	322
	Italy	1,160	513	752	239
Pied flycatcher	Baltic	3,047	1,565	2,216	651
	Czech	1,955	1,028	1,469	441
	Sweden	3,085	1,441	2,027	586
	Spain	1,485	777	1,089	312

The number of hotspots shared within species, between species and both within and between species is also shown.

rates observed in single hotspots were 795.8 cM/Mb and 488.2 cM/Mb in collared flycatcher and pied flycatcher, respectively.

Following a previous approach (Singhal et al., 2015), we defined shared hotspots between populations or between species as those whose mid-points were located <3 kb from each other in the genomes of the respective population or species. In collared flycatcher, 51% of hotspots were shared between at least one pair of populations within species, while 39% were shared between at least one pair of populations between species (Figure 3). The corresponding proportions in pied flycatcher were 69% and 45%, respectively. For hotspots not defined as shared between species (i.e., being species-specific), recombination rate in the other species was still elevated, indicating that the extent of hotspot sharing was somewhat dependent on the threshold used for calling hotspots. For hotspots unique to collared flycatcher, the LD recombination rate in the corresponding regions in the pied flycatcher genome was about four times higher than the local background (Figure 3b). Similarly, for hotspots unique to pied flycatcher, the LD recombination rate in collared flycatcher was seven times higher than the local background (Figure 3c). We observed similar patterns for population-specific hotspots within species (Figure S8, Supporting Information).

We next sought to validate the patterns of interspecific variation in the intensity of recombination hotspots by taking advantage of the expected effects of recombination on the site frequency spectrum via gBGC. There were 2,669 and 5,020 hotspots that were called as unique to collared flycatcher or pied flycatcher (hereafter referred to as "collared flycatcher-specific hotspots" and "pied flycatcher-specific hotspots," respectively; Figure 3b and c). Provided that the location of elevated recombination is conserved between species (even though called as hotspot only in one species), we would expect that stronger signatures of gBGC should be observed in both species at hotspots than coldspots. In agreement with this prediction, values of H_{gBGC} were more negative in both species at recombination hotspots than at coldspots regardless of which species were used for identifying hotspots (bootstrap with 1,000 replicates, $p = .001$ and $p = .008$ for collared flycatcher and pied flycatcher, respectively) (Figure 4). In addition, for both species, H_{gBGC} tended to show more negative values when hotspots were identified in the focal species than in the other species (Figure 4), although the differences were not statistically significant (bootstrap with 1,000 replicates, $p = .26$ at collared flycatcher-specific hotspots and $p = .16$ at pied flycatcher-specific hotspots).

3.4 | Association between recombination hotspots and genomic features

We searched the collared flycatcher genome for sequence motifs associated with recombination hotspots using HOMER, which uses ZOOPS scoring (zero or one occurrence per sequence) coupled with hypergeometric enrichment calculations to determine motif enrichment at hotspots (Heinz et al., 2010). Additionally, we used MEME, which discovers longer motifs by calculating position-specific priors to search motifs enriched in hotspot regions based on the expectation-maximization algorithm (Bailey et al., 2015). HOMER discovered 230 motifs that were significantly enriched at recombination hotspots, many of which were mostly composed of G and C nucleotides (mean GC content = 83%, Table S3, Supporting Information). In

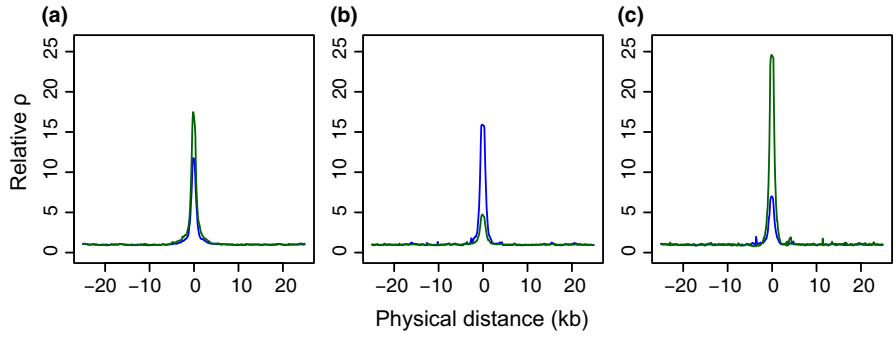

FIGURE 3 LD recombination rate at recombination hotspots in collared flycatcher and pied flycatcher. The recombination rate was standardized by the background recombination rate in the surrounding 20 kb on each side of the hotspot, and then averaged across hotspots within species. (a) Shared hotspots between at least one population of collared flycatcher and one population of pied flycatcher. (b) Hotspots that are unique to collared flycatcher and (c) pied flycatcher. In all three panels, blue lines represent estimates for collared flycatcher and green lines estimates for pied flycatcher [Colour figure can be viewed at wileyonlinelibrary.com]

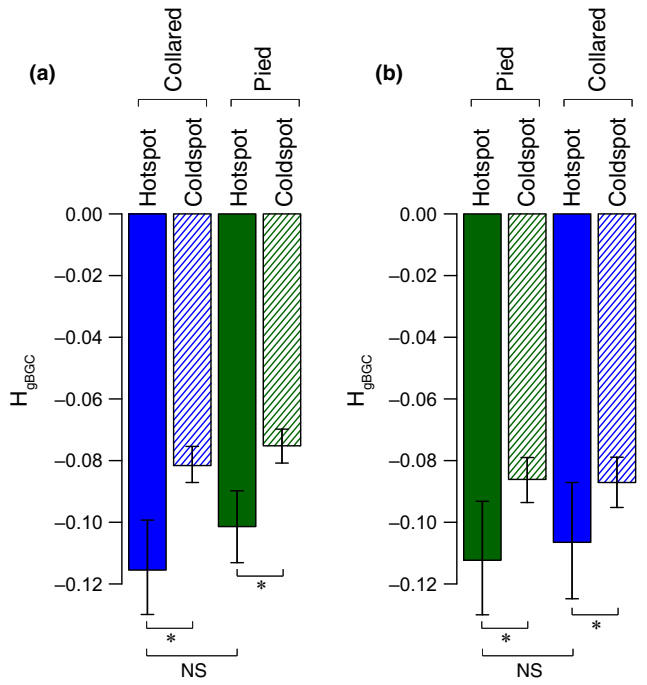

FIGURE 4 Summary of H_{gBGC} at "species-specific" hotspots and coldspots. (a) Collared flycatcher-specific hotspots (corresponding to Figure 5b). (b) Pied flycatcher-specific hotspots (corresponding to Figure 5c). H_{gBGC} was calculated at these species-specific hotspots and coldspots for both collared flycatcher and pied flycatcher. Error bars represent 95% confidence intervals [Colour figure can be viewed at wileyonlinelibrary.com]

addition, the MEME algorithm also identified a GC-rich degenerate 20-mer motif (Motif-3) present in 71% of all hotspots (Figure S9, Supporting Information). Mean GC content at recombination hotspots was 46.3%, higher than the genomewide GC content of 40.7%, in agreement with localized signatures of gBGC. It should be noted that the high proportion of G and C nucleotides at recombination hotspots was not due to a regionally elevated GC content; mean GC content in 5-kb flanking regions of hotspots (39.6%) was lower than the genomic background (Mann–Whitney U test, $p = .0003$). Other highly enriched motifs identified by the MEME algorithm included two CT-rich motifs (Motif-1 and Motif-6), one A-rich motif (Motif-2) and two CAG/CTG repeats (Motif-4 and Motif-5) (Figure S9, Supporting Information).

Consistent with the enrichment of GC-rich motifs in recombination hotspots, the density of hotspots was elevated at CGIs (Figure 5a) both at intergenic CGIs and in promoter regions overlapping with CGIs. In agreement with these observations, the mean LD recombination rate was higher in CGIs than in any other category (Figure 5b). Importantly, promoter regions with CGIs had a higher mean LD recombination rate than promoters without CGIs (7.64 cM/Mb vs. 5.29 cM/Mb). Within genic regions, first exons showed a higher density of recombination hotspots and also a higher LD recombination rate than other exons as well as than introns. Overall, LD recombination rate was high in genic regions and their surrounding upstream and downstream regions, but decreased away from genes (Figure 5c). Similar patterns in the distribution of recombination hotspots and recombination rate were observed in pied flycatcher (Figure S10, Supporting Information).

Regions of the genome containing TEs had high density of recombination hotspot, and intergenic TEs showed higher LD recombination rate than intergenic regions without TEs (Figure 5, S10, Supporting Information). In addition, the density of TEs in 200-kb windows was positively correlated with LD recombination rate (Pearson's correlation coefficient $r = 0.12$, $p < 10^{-16}$) (Figure S11, Supporting Information). This could be either due to colocalization of TEs and recombination hotspots or direct involvement of TEs in recombination events. As CGIs were significantly associated with recombination hotspots (Figure 5), we sought to disentangle these

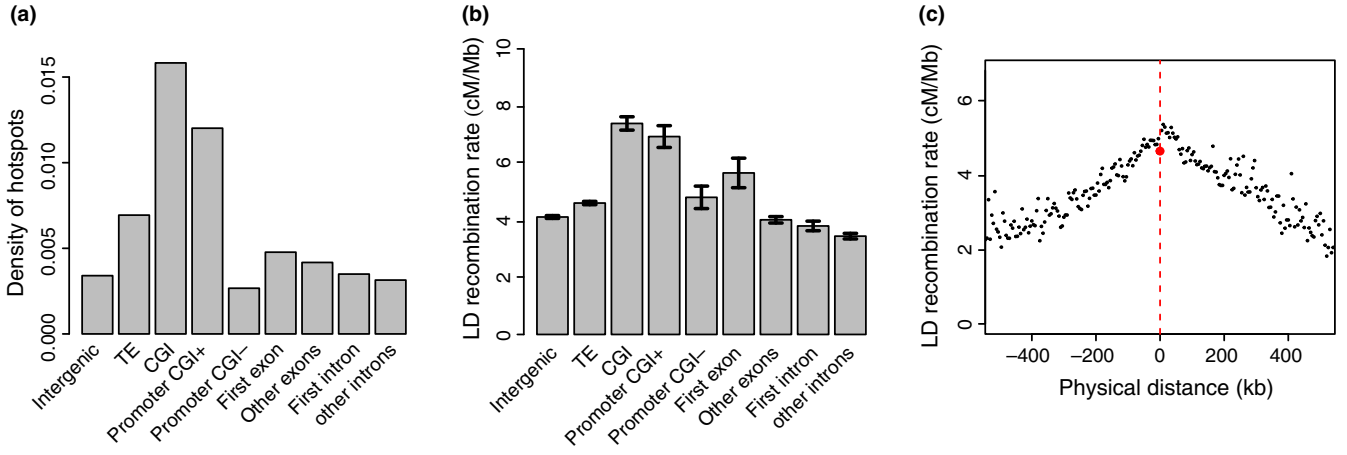

FIGURE 5 Distribution of recombination along the collared flycatcher genome. (a) The density of recombination hotspots and (b) mean LD recombination rate in different annotation categories of the collared flycatcher genome. Error bars represent 95% confidence intervals. Transposable element (TE) and CpG island (CGI). (c) LD recombination rate in regions 500 kb upstream and downstream of genes. Black dots represents mean LD recombination rate based on 5-kb nonoverlapping sliding windows, and the red dot represents the average recombination rate over all genes, including all exons and introns. Corresponding results for pied flycatcher are presented in Figure S10, Supporting Information [Colour figure can be viewed at wileyonlinelibrary.com]

two possibilities by subdividing subfamilies of LINEs and LTRs into four categories each: LINEs and LTRs with or without CGIs in their ancestral consensus sequences, and LINEs and LTRs overlapping with CGIs in the present-day genome sequence. If TEs were directly associated with elevated recombination rate, transposition of TEs carrying CGIs would be expected to locally increase the recombination rate. However, subfamilies of LINEs and LTRs with CGIs in their ancestral sequences did not show elevated local recombination rate (Wilcoxon rank-sum test, $p = 1.00$ for LINEs and $p = .90$ for LTR) (Figure S12, Supporting Information). Regardless of the presence or absence of CGIs in the ancestral sequences, the LD recombination rate was elevated only in copies of LINEs and LTRs overlapping with CGIs in the genome (Wilcoxon rank-sum test, $p < 10^{-16}$ for both LINEs and LTR), arguing for an indirect association between TEs and recombination hotspots.

An indirect association between the distribution of TEs and recombination hotspots could suggest that more recently inserted TEs has had a strong impact on the present recombination landscape than TE insertion events in the far past. To test this prediction, we selected 72 LINE and 127 LTR retrotransposon subfamilies, which had been active in avian genomes at different evolutionary timescales. Consistent with the prediction, the age of LINE and LTR subfamilies, measured as the mean genetic distance between the ancestral sequences and present-day copies within subfamilies, was negatively correlated with recombination rate (Pearson's correlation coefficient $r = -0.556$ for LINE and -0.241 for LTR, $p < 10^{-16}$) (Figure S13, Supporting Information). This provides evidence that young TE subfamilies are more strongly associated with regions of high LD recombination rate than older subfamilies.

4 | DISCUSSION

Genomewide polymorphism data obtained by resequencing of individual genomes have opened up a new avenue for estimating recombination rate at a very fine scale in various organisms (Auton et al., 2012, 2013; Singhal et al., 2015). Using whole-genome polymorphism data of four populations of collared flycatchers and four populations of pied flycatchers, we found highly heterogeneous recombination landscapes in both species. Variation in LD recombination rate at 200-kb resolution was largely concordant with recombination rate data from a collared flycatcher linkage map (Kawakami et al., 2014). However, the LD recombination map could improve the existing pedigree recombination map in at least three aspects.

First, the LD recombination map covered genomic regions that were not represented in the pedigree recombination map due to lack of genetic markers, particularly regions towards the ends of chromosomes. This allowed detection of steep decline in recombination rate at the very ends of chromosomes (Figures S3 and S4, Supporting Information), similar to what has been reported in yeasts, plants and *Drosophila* (Comeron, Ratnappan, & Bailin, 2012; Lam & Keeney, 2015; Yelina et al., 2015). This is potentially because of densely packed chromatin in these highly heterochromatic regions and/or a deleterious effect of recombination in repeat-rich regions (e.g., nonallelic homologous recombination and segregation errors) (Barton, Pekosz, Kurvathi, & Kaback, 2008). Second, the LD recombination map offered much higher resolution of recombination rate variation than the pedigree-based map (Figure S6 Supporting Information). In particular, the LD recombination map allowed for identification of recombination hotspots and thus direct analyses of the association between recombination events and genomic features. Finally, our approach provided a recombination map for the pied flycatcher, which we previously had to assume to be identical to the collared flycatcher map as well as population-specific maps of both species. Using these population- and species-specific maps, we could study the degree of conservation of rates and patterns of recombination within and between species. Below, we discuss the evolution of the recombination landscape in the two closely related *Ficedula* flycatchers from broad chromosome-scale resolution to fine-scale local patterns.

4.1 | Interspecific variation in recombination landscape

The recombination landscape was generally conserved between collared flycatcher and pied flycatcher at the scale of 200-kb windows but less so relative to the conservation between populations within species (Figure 2a). Because variation in recombination rate is an important parameter in shaping genetic diversity and genome differentiation via linked selection (Cruickshank & Hahn, 2014; Wolf & Ellegren, 2017), interspecific variation in recombination rate could impact on genetic differentiation across the genome. The trend of a stronger correlation between LD recombination rate and genetic diversity within each population than across different populations (Figure 2b) is in line with this prediction. Moreover, that the LD recombination rate of collared flycatcher was more strongly correlated with collared flycatcher PBS than pied flycatcher PBS (and *vice versa*) (Figure 2c) supports that there are differences in recombination rate between these closely related species and that they impact on lineage-specific genomic differentiation.

Several recent studies have suggested that a conserved recombination landscape between distantly related birds is the cause of concordant patterns of genetic diversity and differentiation along genomes of different avian species (Dutoit et al., 2017; Van Doren et al., 2017; Vijay et al., 2017). While global conservation of recombination landscape seems indeed likely within flycatcher lineages, our observations demonstrate that small differences in recombination rate between even closely related taxa can be translated into species-specific patterns of genetic diversity and differentiation. This would be particularly so in low recombination regions because only a few recombination events are sufficient to counteract the effects of linked selection (Charlesworth & Campos, 2014).

In addition to the conservation of recombination rate at the scale of 200-kb windows, we also found conservation of the fine-scale localization of recombination events between species, that is, conservation of hotspot location (Figure 3). This was supported by the

strong signatures of gBGC seen in both species at hotspots (Figure 4). Recombination intensity appeared to differ between species at species-specific hotspots, but the differences were not statistically supported in our H_{gBGC} measures, possibly reflecting the recent divergence of the two species. Different recombination intensities at hotspots have been reported in three commercial chicken breeds, where similar associations between recombination and genomic features were observed (Pengelly et al., 2016). One possible explanation for the variation in recombination intensity at hotspots in species without PRDM9 is that differences in gene expression during meiosis results in differences in the likelihood of DSB and thereby recombination frequency between individuals and between species (Adrian & Comeron, 2013; Petes, 2001). In the next section, we further discuss a possible mechanistic link between location of recombination hotspots and regions of transcription regulation, the epigenetic modifications of which are inherently related to gene expression.

4.2 | CGI as a molecular hallmark of recombination hotspots

Hotspot distribution and elevated recombination rate in CGIs (Figure 5, S10, Supporting Information) are consistent with the fact that CGIs are often associated with gene promoter regions in birds (Abe & Gemmell, 2014; Akan & Deloukas, 2008; Singhal et al., 2015; Smeds et al., 2016). These findings are also in good agreement with observations in several nonavian species without functional PRDM9 (Auton et al., 2013; Axelsson et al., 2012; Berglund et al., 2015; Choi et al., 2013). CGIs at actively transcribed gene promoter regions often lack cytosine methylation, and these hypomethylated CGIs are associated with trimethylation of histone subunit H3 at lysine 4 (H3K4me3), which is a hallmark of open chromatin by recruiting various chromatin remodelling factors (Flanagan et al., 2005; Li et al., 2006; Ramirez-Carrozzi et al., 2009; Thomson et al., 2010). It has been suggested that genomic regions with less condensed chromatin have higher recombination rates by allowing the recombination machinery to bind to DNA and initiate DSBs (Berglund et al., 2015; Choi et al., 2013; Comeron et al., 2012; Shilo, Melamed-Bessudo, Dorone, Barkai, & Levy, 2015). It is notable that humans and chimpanzees, which have a functional PRDM9, also showed elevated recombination rates at regions nearby TSSs and CGIs, suggesting that the ancestor of these species and other mammalian species has acquired the PRDM9-dependent regulation of recombination while maintaining the ancestral mechanism of localizing recombination nearby genes and CGIs (Auton et al., 2012). However, a recent study implicated a possible involvement of PRDM9 in recombination in more basal vertebrate lineages (Baker et al., 2017), which warrants further in-depth analysis of the evolutionary origin of the PRDM9-dependent recombination.

We also found that hotspot localization was not limited to promoter-associated CGIs, but intergenic CGIs were also associated with recombination hotspots. It has been shown that artificial insertion of CGIs into a mouse genome induced H3K4ME3 and modified local chromatin state even at sites away from existing promoters, suggesting that structural chromatin modification can be induced by unmethylated CGIs without promoter function (Thomson et al., 2010). Therefore, regardless of their functional significance and relative location to promoters, unmethylated CGIs may be generally targeted by the recombination machinery. Moreover, our results prompt the idea that it is not promoters per se that are associated with recombination, but that only CGI-containing promoters show such an association.

In addition to CGIs, our enrichment analysis identified several other sequence motifs associated with recombination hotspots (Figure S9, Supporting Information). Among these, the CT-rich motifs and the A-rich motif were remarkably similar to the ones identified as associated with recombination in humans and chimpanzees (Auton et al., 2012), finches (Singhal et al., 2015), *Drosophila melanogaster* (Comeron et al., 2012), *Arabidopsis thaliana* (Choi et al., 2013; Shilo et al., 2015) and yeast (Mancera, Bourgon, Brozzi, Huber, & Steinmetz, 2008). In plants, for example, high recombination regions with CT-rich motifs are accompanied by trimethylation of H3K4 and histone variant H2A.Z, which lead to low nucleosome density and increased chromatin accessibility (Choi et al., 2013; Shilo et al., 2015). In plants and yeasts, the A-rich sequence is known to be associated with nucleosome-depleted regions due to their unusual mechanical properties, which disfavour nucleosome formation (Choi et al., 2013; Jansen & Verstrepen, 2011; Segal & Widom, 2009). Characterization of epigenetic patterns at these motifs would represent an important topic for future research in order to investigate whether these motifs can modify patterns of recombination landscape independently or in concert with CGIs in avian genomes.

4.3 | Transposable elements at recombination hotspots

TEs were preferentially located in high recombination regions, which is in contrast to findings in insects and plants (Rizzon et al., 2002; Xu & Du, 2014). The association between TEs and recombination seen in collared flycatcher appears counter-intuitive in the light of the increased efficiency of selection against deleterious insertions of TEs in high recombination regions (Dolgin & Charlesworth, 2008; Petrov, Fiston-Lavier, Lipatov, Lenkov, & González, 2011). To test whether the positive association between TEs and recombination rate is specific to flycatchers or common to other birds, we used LD recombination rate data from zebra finch (Singhal et al., 2015) and similarly found a significantly higher recombination rate at TEs (327.98 ρ/kb [LINE] and 358.21 ρ/kb [LTR]) compared to the genomic average (221.89 ρ/kb). One possible explanation for this association is that both recombination and retrotransposition machineries target accessible genomic regions. Replication of retrotransposons involves RNA intermediates, which are reverse-transcribed and then inserted into new genomic locations via self-encoded proteins (Elbarbary, Lucas, & Maquat, 2016; Kazazian, 2004; Levin & Moran, 2011). For example, LTR retrotransposons in yeast preferentially insert upstream of genes transcribed by RNA polymerase and nucleosome-free regions flanking genes (Baller et al., 2011; Guo & Levin, 2010). Moreover, DNase I hypersensitivity assays of various human cell

lines revealed that up to 80% of LTRs are located in open chromatin regions (Jacques, Jeyakani, & Bourque, 2013). Although insertion site preference is less clear for LINEs, their mechanism of target-primed reverse transcription (TPRT) tends to opportunistically target open chromatin for LINE integration (Luan, Korman, Jakubczak, & Eickbush, 1993; Sandmeyer, Hansen, & Chalker, 1990; Ye, Yang, Hayes, & Eickbush, 2002).

Assuming an indirect relationship between retrotransposons and elevated recombination rate by the shared preference for an accessible chromatin state, it is possible that genomic regions accessible to the recombination machinery and the retrotransposition machinery may shift their location in a concerted way. This assumes that patterns of CpG methylation and other epigenetic marks associated with chromatin state change over time. Such changes have been documented in primates where only 22% of orthologous CpG sites were consistently methylated among humans, chimpanzee, gorilla and orangutan (Hernando-Herraez et al., 2013). The negative relationship between the age of retrotransposons and recombination rate in collared flycatcher supports this hypothesis. Mostly young retrotransposon subfamilies were located in high recombination regions (Figure S13, Supporting Information), which supports the idea that young TEs reflect a more recent state of the epigenetic landscape with higher accessibility to the recombination machinery.

Our results highlight the importance of CGIs as a key molecular feature for the localization of meiotic recombination events in flycatchers. Our analysis thus refines the model proposed for finches (Singhal et al., 2015) and suggests that it is not functional elements per se that play an important role for attracting the recombination machinery, but that the number and location of recombination events depend on the epigenetic status of CGIs irrespective of their genomic location. It is tempting to speculate that the conservation of recombination hotspot locations is a consequence of the conserved distribution of CGIs between the closely related flycatchers, but that the activity of recombination at hotspots can vary between them due to the differences in the epigenetic status of CGIs. The indirect association between TEs and recombination may reflect the long-term history of epigenetic modifications of the genome, where regions of open chromatin accessible to the TE and recombination machineries change over time. Given the significance of recombination rate variation in shaping patterns of genetic diversity and differentiation along a genome via linked selection, it is important to consider the underlying (epi-) genetic changes in the evolution of recombination landscape in population genomics studies. Although global patterns of recombination landscape can be conserved for prolonged time even between distantly related avian lineages, such epigenetic changes can create interspecific variation in recombination rate between closely related species, which can affect species-specific patterns of genomic divergence.

ACKNOWLEDGEMENTS

This work was supported by Swedish Research Council (grant numbers 2007-8731, 2010-5650 and 2013-8271 to HE, and 2014-6325 to TK), Marie Sklodowska Curie Actions, Co-fund Project INCA (grant number 600398 to TK) and the Knut and Alice Wallenberg Foundation (grant type Wallenberg Scholar to HE). We thank Sébastien Renaut and three anonymous reviewers for their constructive comments on an earlier version of the manuscript. Computations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX).

DATA ACCESSIBILITY

Whole-genome resequencing data: ENA (http://www.ebi.ac.uk/ena) under Accession no. PRJEB7359. LD recombination maps: Dryad https://doi.org/10.5061/dryad.hp5h2. Custom Bash and R scripts used in this study are available as Data S1.

AUTHOR CONTRIBUTIONS

T.K. constructed the LD recombination map. T.K. and C.F.M. performed analyses on recombination rate regulation with input from A.S. T.K. and C.F.M. performed population genomics analyses with input from R.B. L.S. and A.N. performed bioinformatics analysis. T.K., C.F.M and H.E. wrote the manuscript with input from the other authors. H.E. contributed to data analysis and conceived of the study.

REFERENCES

Abe, H., & Gemmell, N. J. (2014). Abundance, arrangement, and function of sequence motifs in the chicken promoters. *BMC Genomics*, *15*(1), 900. https://doi.org/10.1186/1471-2164-15-900

Adrian, A. B., & Comeron, J. M. (2013). The *Drosophila* early ovarian transcriptome provides insight to the molecular causes of recombination rate variation across genomes. *BMC Genomics*, *14*, 794. https://doi.org/10.1186/1471-2164-14-794

Akan, P., & Deloukas, P. (2008). DNA sequence and structural properties as predictors of human and mouse promoters. *Gene*, *410*(1), 165–176. https://doi.org/10.1016/j.gene.2007.12.011

Auton, A., Fledel-Alon, A., Pfeifer, S., Venn, O., Segurel, L., Street, T., … McVean, G. (2012). A fine-scale chimpanzee genetic map from population sequencing. *Science*, *336*(6078), 193–198. https://doi.org/10.1126/science.1216872

Auton, A., Li, Y. R., Kidd, J., Oliveira, K., Nadel, J., Holloway, J. K., … Boyko, A. R. (2013). Genetic recombination is targeted towards gene promoter regions in dogs. *PLoS Genetics*, *9*(12), e1003984. https://doi.org/10.1371/Journal.Pgen.1003984

Auton, A., & McVean, G. (2012). Estimating recombination rates from genetic variation in humans. *Methods in Molecular Biology*, *856*, 217–237. https://doi.org/10.1007/978-1-61779-585-5_9

Axelsson, E., Webster, M. T., Ratnakumar, A., Consortium, L., Ponting, C. P., & Lindblad-Toh, K. (2012). Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome. *Genome Research*, *22*(1), 51–63.https://doi.org/10.1101/gr.124123.111

Backström, N., Forstmeier, W., Schielzeth, H., Mellenius, H., Nam, K., & Bolund, E. (2010). The recombination landscape of the zebra finch Taeniopygia guttata genome. *Genome Research*, *20*, 485–495. https://doi.org/10.1101/gr.101410.109

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME suite. *Nucleic Acids Research*, *43*(W1), W39–W49. https://doi.org/10.1093/nar/gkv416

Baker, Z., Schumer, M., Haba, Y., Bashkirova, L., Holland, C., Rosenthal, G. G., & Przeworski, M. (2017). Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. *eLife*, *6*, e24133. https://doi.org/10.7554/eLife24133

Baller, J. A., Gao, J., & Voytas, D. F. (2011). Access to DNA establishes a secondary target site bias for the yeast retrotransposon Ty5. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(51), 20351–20356. https://doi.org/10.1073/pnas.1103665108

Barton, A. B., Pekosz, M. R., Kurvathi, R. S., & Kaback, D. B. (2008). Meiotic recombination at the ends of chromosomes in *Saccharomyces cerevisiae*. *Genetics*, *179*(3), 1221–1235. https://doi.org/10.1534/genetics.107.083493

Baudat, F., Imai, Y., & de Massy, B. (2013). Meiotic recombination in mammals: Localization and regulation. *Nature Reviews Genetics*, *14*(11), 794–806. https://doi.org/10.1038/nrg3573

Berglund, J., Quilez, J., Arndt, P. F., & Webster, M. T. (2015). Germ line methylation patterns determine the distribution of recombination events in the dog genome. *Genome Biology and Evolution*, *7*(2), 522–530. https://doi.org/10.1093/gbe/evu282

Birdsell, J. A. (2002). Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. *Molecular Biology and Evolution*, *19*(7), 1181–1197.

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., … Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula flycatchers*. *Genome Research*, *25*(11), 1656–1665. https://doi.org/10.1101/gr.196485.115

Chan, A. H., Jenkins, P. A., & Song, Y. S. (2012). Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. *PLoS Genetics*, *8*(12), e1003090. https://doi.org/10.1371/journal.pgen.1003090

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, *10*(3), 195–205.

Charlesworth, B., & Campos, J. L. (2014). The Relations between recombination rate and patterns of molecular variation and evolution in *Drosophila*. *Annual Review of Genetics*, *48*, 383–403. https://doi.org/10.1146/annurev-genet-120213-092525

Chen, G. K., Marjoram, P., & Wall, J. D. (2009). Fast and flexible simulation of DNA sequence data. *Genome Research*, *19*(1), 136–142. https://doi.org/10.1101/gr.083634.108

Choi, K., Zhao, X., Kelly, K. A., Venn, O., Higgins, J. D., Yelina, N. E., … Henderson, I. R. (2013). *Arabidopsis* meiotic crossover hot spots overlap with H2A. Z nucleosomes at gene promoters. *Nature Genetics*, *45*(11), 1327–1336.

Comeron, J. M., Ratnappan, R., & Bailin, S. (2012). The many landscapes of recombination in *Drosophila melanogaster*. *PLoS Genetics*, *8*(10), e1002905. https://doi.org/10.1371/journal.pgen.1002905

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*(13), 3133–3157. https://doi.org/10.1111/mec.12796

Cutter, A. D., & Payseur, B. A. (2013). Genomic signatures of selection at linked sites: Unifying the disparity among species. *Nature Reviews Genetics*, *14*(4), 262–274. https://doi.org/10.1038/Nrg3425

Dolgin, E. S., & Charlesworth, B. (2008). The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics*, *178*(4), 2169–2177. https://doi.org/10.1534/genetics.107.082743

Duret, L., & Arndt, P. F. (2008). The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genetics*, *4*(5), e1000071. https://doi.org/10.1371/journal.pgen.1000071

Duret, L., & Galtier, N. (2009). Biased gene conversion and the evolution of mammalian genomic landscapes. *Annual Review of Genomics and Human Genetics*, *10*, 285–311. https://doi.org/10.1146/annurev-genom-082908-150001

Dutoit, L., Vijay, N., Mugal, C. F., Bossu, C. M., Burri, R., Wolf, J., & Ellegren, H. (2017). Covariation in levels of nucleotide diversity in homologous regions of the avian genome long after completion of lineage sorting. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *284*(1849), 20162756. https://doi.org/10.1098/rspb.2016.2756

Elbarbary, R. A., Lucas, B. A., & Maquat, L. E. (2016). Retrotransposons as regulators of gene expression. *Science*, *351*(6274), aac7247. https://doi.org/10.1126/science.aac7247

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backstrom, N., Kawakami, T., … Wolf, J. B. W. (2012). The genomic landscape of species divergence in *Ficedula flycatchers*. *Nature*, *491*(7426), 756–760. https://doi.org/10.1038/nature11584

Fay, J. C., & Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics*, *155*(3), 1405–1413.

Flanagan, J. F., Mi, L. Z., Chruszcz, M., Cymborowski, M., Clines, K. L., Kim, Y. C., … Khorasanizadeh, S. (2005). Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature*, *438*(7071), 1181–1185. https://doi.org/10.1038/nature04290

Fraser, B. A., Kunstner, A., Reznick, D. N., Dreyer, C., & Weigel, D. (2015). Population genomics of natural and experimental populations of guppies (*Poecilia reticulata*). *Molecular Ecology*, *24*(2), 389–408. https://doi.org/10.1111/mec.13022

Glemin, S., Arndt, P. F., Messer, P. W., Petrov, D., Galtier, N., & Duret, L. (2015). Quantification of GC-biased gene conversion in the human genome. *Genome Research*, *25*(8), 1215–1228. https://doi.org/10.1101/gr.185488.114

Guo, Y., & Levin, H. L. (2010). High-throughput sequencing of retrotransposon integration provides a saturated profile of target activity in *Schizosaccharomyces pombe*. *Genome Research*, *20*(2), 239–248. https://doi.org/10.1101/gr.099648.109

Hackenberg, M., Previti, C., Luque-Escamilla, P. L., Carpena, P., Martínez-Aroza, J., & Oliver, J. L. (2006). CpGcluster: A distance-based algorithm for CpG-island detection. *BMC Bioinformatics*, *7*(1), 446. https://doi.org/10.1186/1471-2105-7-446

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., … Glass, C. K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell*, *38*(4), 576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Hernando-Herraez, I., Prado-Martinez, J., Garg, P., Fernandez-Callejo, M., Heyn, H., Hvilsom, C., … Marques-Bonet, T. (2013). Dynamics of DNA methylation in recent human and great ape evolution. *PLoS Genetics*, *9*(9), e1003763. https://doi.org/10.1371/journal.pgen.1003763

Hill, W. G., & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetical Research*, *8*(3), 269–294.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, *6*(2), e1000862. https://doi.org/10.1371/journal.pgen.1000862

Jacques, P.-É., Jeyakani, J., & Bourque, G. (2013). The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genetics*, *9*(5), e1003504. https://doi.org/10.1371/journal.pgen.1003504

Jansen, A., & Verstrepen, K. J. (2011). Nucleosome Positioning in *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews*, *75*(2), 301–320. https://doi.org/10.1128/mmbr.00046-10

Jeffreys, A. J., Kauppi, L., & Neumann, R. (2001). Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genetics*, *29*(2), 217–222. https://doi.org/10.1038/ng1001-217

Kapusta, A., & Suh, A. (2017). Evolution of bird genomes—a transposon's-eye view. *Annals of the New York Academy of Sciences*, *1389*(1), 164–185. https://doi.org/10.1111/nyas.13295

Kawakami, T., Smeds, L., Backstrom, N., Husby, A., Qvarnstrom, A., Mugal, C. F., … Ellegren, H. (2014). A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Molecular Ecology*, *23*(16), 4035–4058. https://doi.org/10.1111/mec.12810

Kazazian, H. H. (2004). Mobile elements: Drivers of genome evolution. *Science*, *303*(5664), 1626–1632. https://doi.org/10.1126/Science.1089670

Lam, I., & Keeney, S. (2015). Nonparadoxical evolutionary stability of the recombination initiation landscape in yeast. *Science*, *350*(6263), 932–937. https://doi.org/10.1126/science.aad0814

Levin, H. L., & Moran, J. V. (2011). Dynamic interactions between transposable elements and their hosts. *Nature Reviews Genetics*, *12*(9), 615–627. https://doi.org/10.1038/nrg3030

Li, H. T., Ilin, S., Wang, W. K., Duncan, E. M., Wysocka, J., Allis, C. D., & Patel, D. J. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature*, *442*(7098), 91–95. https://doi.org/10.1038/nature04802

Liu, S., Yeh, C.-T., Ji, T., Ying, K., Wu, H., Tang, H. M., … Schnable, P. S. (2009). Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. *PLoS Genetics*, *5*(11), e1000733. https://doi.org/10.1371/journal.pgen.1000733

Luan, D. D., Korman, M. H., Jakubczak, J. L., & Eickbush, T. H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition. *Cell*, *72*(4), 595–605. https://doi.org/10.1016/0092-8674(93)90078-5

Mancera, E., Bourgon, R., Brozzi, A., Huber, W., & Steinmetz, L. M. (2008). High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature*, *454*(7203), 479–485. https://doi.org/10.1038/nature07135

Mugal, C. F., Arndt, P. F., & Ellegren, H. (2013). Twisted signatures of GC-biased gene conversion embedded in an evolutionary stable karyotype. *Molecular Biology and Evolution*, *30*(7), 1700–1712. https://doi.org/10.1093/molbev/mst067

Mugal, C. F., Weber, C. C., & Ellegren, H. (2015). GC-biased gene conversion links the recombination landscape and demography to genomic base composition: GC-biased gene conversion drives genomic base composition across a wide range of species. *BioEssays*, *37*(12), 1317–1326. https://doi.org/10.1002/bies.201500058

Myers, S., Bowden, R., Tumian, A., Bontrop, R. E., Freeman, C., MacFie, T. S., … Donnelly, P. (2010). Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science*, *327*(5967), 876–879. https://doi.org/10.1126/science.1182363

Nadachowska-Brzyska, K., Burri, R., Olason, P. I., Kawakami, T., Smeds, L., & Ellegren, H. (2013). Demographic divergence history of pied flycatcher and collared flycatcher inferred from whole-genome re-sequencing data. *PLoS Genetics*, *9*(11), e1003942. https://doi.org/10.1371/journal.pgen.1003942

Nadachowska-Brzyska, K., Burri, R., Smeds, L., & Ellegren, H. (2016). PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula flycatchers*. *Molecular Ecology*, *25*(5), 1058–1072. https://doi.org/10.1111/mec.13540

Nater, A., Burri, R., Kawakami, T., Smeds, L., & Ellegren, H. (2015). Resolving evolutionary relationships in closely related species with whole-genome sequencing data. *Systematic Biology*, *64*(6), 1000–1017. https://doi.org/10.1093/sysbio/syv045

Paigen, K., & Petkov, P. (2010). Mammalian recombination hot spots: Properties, control and evolution. *Nature Reviews Genetics*, *11*(3), 221–233. https://doi.org/10.1038/nrg2712

Pan, J., Sasaki, M., Kniewel, R., Murakami, H., Blitzblau, H. G., Tischfield, S. E., … Keeney, S. (2011). A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell*, *144*(5), 719–731. https://doi.org/10.1016/j.cell.2011.02.009

Parvanov, E. D., Petkov, P. M., & Paigen, K. (2010). Prdm9 controls activation of mammalian recombination hotspots. *Science*, *327*(5967), 835. https://doi.org/10.1126/science.1181495

Pengelly, R. J., Gheyas, A. A., Kuo, R., Mossotto, E., Seaby, E. G., Burt, D. W., … Collins, A. (2016). Commercial chicken breeds exhibit highly divergent patterns of linkage disequilibrium. *Heredity*, *117*(5), 375–382. https://doi.org/10.1038/hdy.2016.47

Petes, T. D. (2001). Meiotic recombination hot spots and cold spots. *Nature Reviews Genetics*, *2*(5), 360–369. https://doi.org/10.1038/35072078

Petrov, D. A., Fiston-Lavier, A.-S., Lipatov, M., Lenkov, K., & González, J. (2011). Population genomics of transposable elements in *Drosophila melanogaster*. *Molecular Biology and Evolution*, *28*(5), 1633–1644. https://doi.org/10.1093/molbev/msq337

Ramirez-Carrozzi, V. R., Braas, D., Bhatt, D. M., Cheng, C. S., Hong, C., Doty, K. R., … Smale, S. T. (2009). A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell*, *138*(1), 114–128. https://doi.org/10.1016/j.cell.2009.04.020

Rizzon, C., Marais, G., Gouy, M., & Biemont, C. (2002). Recombination rate and the distribution of transposable elements in the *Drosophila melanogaster* genome. *Genome Research*, *12*(3), 400–407. https://doi.org/10.1101/gr.210802

Roesti, M., Hendry, A. P., Salzburger, W., & Berner, D. (2012). Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology*, *21*(12), 2852–2862. https://doi.org/10.1111/j.1365-294X.2012.05509.x

Sandmeyer, S. B., Hansen, L. J., & Chalker, D. L. (1990). Integration specificity of retrotransposons and retroviruses. *Annual Review of Genetics*, *24*, 491–518. https://doi.org/10.1146/annurev.ge.24.120190.002423

Sasaki, M., Lange, J., & Keeney, S. (2010). Genome destabilization by homologous recombination in the germ line. *Nature Reviews. Molecular Cell Biology*, *11*(3), 182–195. https://doi.org/10.1038/nrm2849

Scheet, P., & Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics*, *78*(4), 629–644. https://doi.org/10.1086/502802

Segal, E., & Widom, J. (2009). Poly(dA:dT) tracts: Major determinants of nucleosome organization. *Current Opinion in Structural Biology*, *19*(1), 65–71. https://doi.org/10.1016/j.sbi.2009.01.004

Shilo, S., Melamed-Bessudo, C., Dorone, Y., Barkai, N., & Levy, A. A. (2015). DNA crossover motifs associated with epigenetic modifications delineate open chromatin regions in *Arabidopsis*. *The Plant Cell*, *27*(9), 2427–2436. https://doi.org/10.1105/tpc.15.00391

Shriver, M. D., Kennedy, G. C., Parra, E. J., Lawson, H. A., Sonpar, V., Huang, J., … Jones, K. W. (2004). The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Human Genomics*, *1*(4), 274–286.

Singhal, S., Leffler, E. M., Sannareddy, K., Turner, I., Venn, O., Hooper, D. M., … Przeworski, M. (2015). Stable recombination hotspots in birds. *Science*, *350*(6263), 928–932. https://doi.org/10.1126/science.aad0843

Smeds, L., Kawakami, T., Burri, R., Bolivar, P., Husby, A., Qvarnstrom, A., … Ellegren, H. (2014). Genomic identification and characterization of the pseudoautosomal region in highly differentiated avian sex chromosomes. *Nature Communications*, *5*, 5448. https://doi.org/10.1038/ncomms6448

Smeds, L., Mugal, C. F., Qvarnström, A., & Ellegren, H. (2016). High-resolution mapping of crossover and non-crossover recombination events by whole-genome re-sequencing of an avian pedigree. *PLoS Genetics*, *12*(5), e1006044. https://doi.org/10.1371/journal.pgen.1006044

Smeds, L., Warmuth, V., Bolivar, P., Uebbing, S., Burri, R., Suh, A., … Ellegren, H. (2015). Evolutionary analysis of the female-specific avian W

chromosome. *Nature Communications*, *6*, 7330. https://doi.org/10.1038/ncomms8330

Strathern, J. N., Shafer, B. K., & McGill, C. B. (1995). DNA synthesis errors associated with double-strand-break repair. *Genetics*, *140*(3), 965–972.

Thomson, J. P., Skene, P. J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., … Bird, A. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature*, *464*(7291), 1082–1086. https://doi.org/10.1038/nature08924

Van Doren, B. M., Campagna, L., Helm, B., Illera, J. C., Lovette, I. J., & Liedvogel, M. (2017). Correlated patterns of genetic diversity and differentiation across an avian family. *Molecular Ecology*, https://doi.org/10.1111/mec.14083

Vijay, N., Weissensteiner, M. H., Burri, R., Kawakami, T., Ellegren, H., & Wolf, J. B. (2017). Genome-wide patterns of variation in effective population size are shared among populations, species and higher order taxa. *Molecular Ecology*, https://doi.org/10.1111/mec.14195

Westram, A. M., Galindo, J., Alm Rosenblad, M., Grahame, J. W., Panova, M., & Butlin, R. K. (2014). Do the same genes underlie parallel phenotypic divergence in different *Littorina saxatilis* populations? *Molecular Ecology*, *23*(18), 4603–4616. https://doi.org/10.1111/mec.12883

Wiley, C., Qvarnstrom, A., Andersson, G., Borge, T., & Saetre, G. P. (2009). Postzygotic isolation over multiple generations of hybrid descendents in a natural hybrid zone: How well do single-generation estimates reflect reproductive isolation? *Evolution*, *63*(7), 1731–1739. https://doi.org/10.1111/j.1558-5646.2009.00674.x

Wolf, J. B., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, *18*, 87–100. https://doi.org/10.1038/nrg.2016.133

Xu, Y. X., & Du, J. C. (2014). Young but not relatively old retrotransposons are preferentially located in gene-rich euchromatic regions in tomato (*Solanum lycopersicum*) plants. *The Plant Journal*, *80*(4), 582–591. https://doi.org/10.1111/tpj.12656

Ye, J. Q., Yang, Z. Y., Hayes, J. J., & Eickbush, T. H. (2002). R2 retrotransposition on assembled nucleosomes depends on the translational position of the target site. *EMBO Journal*, *21*(24), 6853–6864. https://doi.org/10.1093/emboj/cdf665

Yelina, N. E., Lambing, C., Hardcastle, T. J., Zhao, X., Santos, B., & Henderson, I. R. (2015). DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in *Arabidopsis*. *Genes & Development*, *29*(20), 2183–2202. https://doi.org/10.1101/gad.270876.115

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X. P., Pool, J. E., … Wang, J. (2010). Sequencing of fifty human exomes reveals adaptation to high altitude. *Science*, *329*(5987), 75–78. https://doi.org/10.1126/science.1190371

Yoshida, J., Akagi, K., Misawa, R., Kokubu, C., Takeda, J., & Horie, K. (2017). Chromatin states shape insertion profiles of the piggyBac, Tol2 and Sleeping Beauty transposons and murine leukemia virus. *Scientific Reports*, *7*, 43613. https://doi.org/10.1038/srep43613

SUPPORTING INFORMATION

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

How to cite this article: Kawakami T, Mugal CF, Suh A, et al. Whole-genome patterns of linkage disequilibrium across flycatcher populations clarify the causes and consequences of fine-scale recombination rate variation in birds. *Mol Ecol*. 2017;26:4158–4172. https://doi.org/10.1111/mec.14197