Received: 28 September 2017 | Revised: 28 March 2018 | Accepted: 2 April 2018

DOI: 10.1111/mec.14789

FROM THE COVER

WILEY MOLECULAR ECOLOGY

Sex-biased gene expression, sexual antagonism and levels of genetic diversity in the collared flycatcher (*Ficedula albicollis*) genome

Ludovic Dutoit[1] | Carina F. Mugal[1] | Paulina Bolívar[1] | Mi Wang[1] | Krystyna Nadachowska-Brzyska[1] | Linnéa Smeds[1] | Homa P. Yazdi[1] | Lars Gustafsson[2] | Hans Ellegren[1]

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

[2]Department of Animal Ecology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

Correspondence
Hans Ellegren, Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden.
Email: Hans.Ellegren@ebc.uu.se

Abstract

Theoretical work suggests that sexual conflict should promote the maintenance of genetic diversity by the opposing directions of selection on males and females. If such conflict is pervasive, it could potentially lead to genomic heterogeneity in levels of genetic diversity an idea that so far has not been empirically tested on a genomewide scale. We used large-scale population genomic and transcriptomic data from the collared flycatcher (*Ficedula albicollis*) to analyse how sexual conflict, for which we use sex-biased gene expression as a proxy, relates to genetic variability. Here, we demonstrate that the extent of sex-biased gene expression of both male-biased and female-biased genes is significantly correlated with levels of nucleotide diversity in gene sequences and that this correlation extends to diversity levels also in intergenic DNA and introns. We find signatures of balancing selection in sex-biased genes but also note that relaxed purifying selection could potentially explain part of the observed patterns. The finding of significant genetic differentiation between males and females for male-biased (and gonad-specific) genes indicates ongoing sexual conflict and sex-specific viability selection, potentially driven by sexual selection. Our results thus indicate that sexual antagonism could potentially be considered as one viable explanation to the long-standing question in evolutionary biology of how genomes can remain so genetically variable in face of strong natural and sexual selection.

KEYWORDS
balancing selection, birds, sex-biased gene expression, sexual antagonism, transcriptomics

1 | INTRODUCTION

A classical problem in evolutionary biology is to explain the wealth of genetic variability, especially multivariate variation for quantitative traits, displayed by most species in the face of selection (Barton & Turelli, 1989; Kruuk & Hill, 2008; Mitchell-Olds, Willis, & Goldstein, 2007). In the general sense, directional selection is expected to reduce genetic diversity at trait loci and in genomic regions linked to such loci. In addition, sexual selection might be a particularly potent force in depleting genetic variance for traits related to male–male competition (Kokko, Brooks, Jennions, & Morley, 2003), although there are several explanations as to how the "lek paradox" might be resolved (Radwan, 2008).

One factor that could potentially contribute to maintenance of genetic diversity is sexual conflict. Here, opposing directions of

selection in males and females for sexually antagonistic alleles (intralocus sexual conflict) should increase the fixation time for positively selected mutation and act towards stable maintenance of segregating variation (Connallon & Clark, 2012, 2014; Jordan & Connallon, 2014), that is, one form of balancing selection. Loci experiencing balancing selection are expected to show an excess of intermediate frequency variants, elevated diversity and longer coalescence times than under neutrality (Charlesworth, 2006; Hudson & Kaplan, 1988; Turelli & Barton, 2004). An excess of trans-species polymorphisms might be observed if coalescent times are deeper than speciation times (Muirhead, Glass, & Slatkin, 2002). However, there is limited empirical support to the idea that balancing selection driven by sexual conflict would have some overall influence on genomewide levels of genetic diversity. This is thus a question that awaits formal testing in different systems using large-scale genomic data sets.

Males and females pursue different strategies for reproduction ultimately resulting in large phenotypic differences between sexes (Arnqvist & Rowe, 2005). Yet, males and females are almost identical at the genetic level, with the exception of sex chromosomes when they are present. It is the high degree of genetic similarity between sexes that introduces sexual conflict. It has become increasingly understood that this conflict might result in, and might ultimately be solved by, sex-biased gene expression (Connallon & Knowles, 2005; Ellegren & Parsch, 2007; Mank, 2017). This could be in many forms, including different expression levels between sexes in one or more tissues, sex-limited expression or temporal variation in gene expression between sexes.

Sex-biased expression can thus be used as a signal of ongoing antagonistic selection or a signature of a recent resolution of sexual conflict. In a seminal paper, Innocenti and Morrow (2010) designed fitness assays in hemiclonal lines of *Drosophila melanogaster* to demonstrate that 8% of the genes with sex-biased expression showed sexually antagonistic fitness effects. This is likely to be an underestimate for several reasons, not least because only a narrow window of time in the lifespan of flies was investigated. It was suggested that genetic variation for fitness might actually be dominated and maintained by sexual antagonism. Such experimental approaches are out of scope in studies of natural populations, but it seems reasonable to assume that a significant proportion of loci with sex-biased gene expression evolve or have evolved under the influence of sexual antagonism in many organisms.

Here, we test whether levels of nucleotide diversity are associated with the extent of sex-biased gene expression in the genome of the collared flycatcher (*Ficedula albicollis*). We use large-scale population genomic data to assess genetic diversity and transcriptomic data from multiple organs to quantify gene expression in males and females. We augment this approach with an analysis of the potential genomic consequences of a specific aspect of sexual conflict: sexually antagonistic viability selection. Our main observation is a positive correlation between levels of nucleotide diversity in gene sequences and sex-biased gene expression. Relaxed purifying selection could potentially play a role but the finding of significant genetic differentiation between males and females for sex-biased genes is consistent with ongoing sexual conflict.

2 | MATERIALS AND METHODS

2.1 | Sampling

We used whole-genome resequencing data from 172 collared flycatchers sampled as breeding pairs at the Baltic Sea islands Gotland (n = 43 males and 43 females) and Öland (n = 86 males). Description of sequencing methods and other methodological details for the Öland population are given in Kardos, Husby, McFarlane, Qvarnstrom, and Ellegren (2016). For the Gotland population, each individual was sequenced with a paired-end approach to a mean coverage of 38.6× (range: 29.5–50.2×) on an Illumina HiSeqX instrument. Read length was 150 bp and insert size ~350 bp. Reads from all individuals were mapped to the collared flycatcher reference genome assembly version FICALB1.5 (GenBank Accession GCA_000247815.2) with BWA mem version 0.7.13 (Li & Durbin, 2009) and processed with SAMTOOLS version 1.3 (Li et al., 2009). The reads were deduplicated with PICARD version 2.0.1 (http://broadinstitute.github.io/picard/), and realigned and recalibrated with GATK version 3.6 (DePristo et al., 2011). Variants were then called with GATK's HaplotypeCaller and GenotypeGVCFs version 3.

In order to reduce the number of false positives and to improve concordance between variants of the two samples, we validated the variants in each of the two populations against a high-quality variant set for flycatchers obtained from sequencing of 200 individuals from 10 populations where different methods for variant calling had been intersected (Burri et al., 2015). Variants that had not been found in these 200 individuals were conservatively excluded.

2.2 | Annotation

We extracted coding sequences for all genes from the ENSEMBL annotation of the flycatcher genome assembly version FICALB1.4 and translated them to the FICALB1.5 version. Genes that span two or more scaffolds were removed, as were overlapping genes on complementary strands. We only used autosomal genes and hence excluded sex-linked genes and genes that had not been assigned to a specific chromosome. In total, 14,300 annotated genes were available and analysed.

2.3 | Transcriptomics

We used RNA-seq data from seven different organs (gonads, brain, kidney, liver, lung, muscle and skin) of four male (NCBI/SRA study ERP001377; coll_01, coll_03, coll_04, coll_05) and four female (coll_06, coll_07, coll_08, coll_10) collared flycatchers described in Ref. (Uebbing, Kunstner, Makinen, & Ellegren, 2013). Reads were quality-checked using FASTQC and duplicates were removed using PICARD v 2.0.1 (http://broadinstitute.github.io/picard/). Reads were then mapped to the FICALB1.5 assembly version using STAR v.2.5.1b (Dobin et al., 2013) with default parameters. Uniquely mapped reads with a mapping quality above 30 were used to assess transcript abundance based on HTSEQ v0.6.1 (Anders, Pyl, & Huber, 2015). The

DESEQ2 package (Love, Huber, & Anders, 2014) in R (R core Team, 2017) was then used to estimate the $\log_2$-fold change in expression between sexes. This $\log_2$-fold change is referred throughout the text as the level of sex-bias for males and females respectively. To increase the signal-to-noise ratio in the statistical analyses and to be able to investigate male-biased genes and female-biased genes independently, we grouped genes into 10 bins of male-biased genes and 10 bins of female-biased genes according to the level of sex-bias for each organ. Note that this grouping was based on all genes such that the bin(s) with the lowest sex-bias essentially comprised unbiased genes. Each bin contained the same number of genes and 10 bins proved empirically to be the number of bins retaining most of the signal while limiting the noise. We chose to not use a discrete classification of sex-bias because filtering criteria and thresholds have major effects on the number of genes defined as sex-biased (Ingleby, Flis, & Morrow, 2014).

2.4 | Population genomics

A site was considered for population genetics analyses if covered by at least one read for all individuals within the population. To calculate population genetic statistics, we computed the allelic frequency spectrum using in-house scripts and the PYTHON package PYVCF 0.4.0 (https://pyvcf.readthedocs.org). Nucleotide diversity, the average number of pairwise differences per site within the sample (π; Nei & Li, 1979), was used as a measure of diversity. It was calculated independently for each location and then taken as the mean between the two samples.

We assessed whether the extent of sex-biased gene expression in gonads had the potential to explain genomewide patterns of diversity across 200-kb windows by regression analysis when accounting for recombination, gene density, intergenic GC content and mutation rate (approximated by the lineage-specific synonymous substitution rate, d_S, for flycatcher estimated from three-species alignments of collared flycatcher, chicken *Gallus gallus* and zebra finch *Taenopygia guttata*). Data for the explanatory variables were taken from Bolívar, Mugal, Nater, and Ellegren (2016) and Dutoit et al. (2017). We calculated the level of sex-bias per window as the average over all genes in that window and were able to retrieve data for 2,190 windows. Explanatory variables with particularly skewed distributions were transformed prior to regression analysis in order to reduce skewness in their distribution. Recombination rate was log-transformed to base 10 after adding a constant 1, and gene density and d_S were square-root-transformed. Associations between variables were assessed by correlation analysis. We then performed a principal component regression (PCR), a method able to handle collinearity between explanatory variables, using axes obtained from a principal component analysis as predictors of the response variable (Mugal, Nabholz, & Ellegren, 2013).

Two population genetics parameters that may signal balancing selection were investigated. First, Tajima's D (Tajima, 1989) was estimated using the allelic frequency spectrum (see above). Second, we quantified the amount of shared polymorphism between collared flycatcher and semicollared flycatcher (*Ficedula semitorquata*) as the number of variant sites in collared flycatchers that were also variable in semicollared flycatchers, divided by the total number of variants across both species in 200-kb windows. Polymorphism data were obtained from Burri et al. (2015).

To further examine the potential explanatory power of relaxed purifying selection, we extracted the ratio of the nonsynonymous to the synonymous substitution rate (d_N/d_S) from Bolívar et al. (2016). Nucleotide diversity was also extracted for 0-fold and 4-fold degenerated sites to obtain the ratio of nonsynonymous to synonymous polymorphisms (p_N/p_S).

We estimated weighted F_{ST} (Weir & Cockerham, 1984; Eq. 6) between males and females within the Gotland sample using VCFtools (Danecek et al., 2011). As we were unable to match the assumptions of normal distribution of the residuals and homoscedasticity (and faced a low signal-to-noise ratio in the per-gene analysis), genes were concatenated into three bins with equal numbers of genes classified as female-biased genes, unbiased, and male-biased based on data from gonads. The average sex-bias for the category of female-biased genes was a $\log_2$-fold change of 2.40 and for male-biased genes a $\log_2$-fold change of 2.23; unbiased genes had on average a small female bias of 0.13. Genes were resampled with replacement 500 times within each bin to estimate the 95% confidence interval of a bootstrap sample.

F_{ST} is zero when the variance within groups is equal to the variance between the groups; this should be the null expectation for two random samples drawn from the same population. To test this assumption and to obtain an empirical background level of F_{ST} in the population, we randomly assigned individuals to two groups 1,000 times and calculated F_{ST} over all genes. We then extracted a 95% confidence interval of the distribution of the bootstrap. We estimated F_{ST} between males and females for all genes, and for genes in the three bins defined above, and resampled genes with replacement 500 times to be able to test whether genomewide F_{ST} between sexes was higher than the null expectation.

3 | RESULTS

We first quantified the extent of sex-biased gene expression in seven different organs of adult collared flycatchers (brain, gonads, kidney, liver, lung, muscle and skin). As expected, gonads showed the highest degree of sex-bias with 49% of genes being differentially expressed between sexes at a significance threshold of $p < 0.05$ (after false discovery rate correction, Benjamini & Hochberg, 1995; Table 1, Figure 1a). More genes were female-biased (n = 3,914, in ovary) than male-biased (n = 3,085, in testis). For all nonreproductive organs, less than 2% of genes showed a significant sex-bias (Table 1, Figure 1c). Similar proportions of sex-biased genes were seen when considering the mean fold change in expression between the sexes and using a fold change as cut-off. Sixty-two per cent of gonadal genes were expressed at least twice as much in one sex than in the other, while this proportion was below 2% in all other organs. Levels

TABLE 1 Description of sex-biased gene expression per organ

Organ	No. of genes expressed	No of genes with significant sex-bias[a]		No of genes with fold change >2	
		Total, females, males	Proportion (%)	Total, females, males	Proportion (%)
Gonads	14,067	6,999, 3,914, 3,085	49.8	8,824, 5,379, 3,445	62.7
Kidney	9,583	139, 20, 119	1.5	46, 6, 40	0.5
Lung	9,719	96, 27, 69	1.0	154, 63, 91	1.6
Liver	9,063	38, 19, 19	0.4	140, 69, 71	1.5
Skin	10,836	38, 4, 34	0.4	49, 9, 40	0.5
Brain	8,862	21, 0, 21	0.2	8, 0, 8	0.1
Muscle	8,010	17, 3, 14	0.1	92, 29, 63	1.1

[a]Significance was based on a model using the negative binomial distribution and corrected estimates of gene counts as implemented in DESEQ2 by Love et al. (2014).

of sex-bias were weakly correlated across organs, with Pearson's correlation coefficients ranging from 0.04 for the gonads–brain comparison to 0.27 for the muscle-kidney comparison.

We used whole-genome resequencing data from 172 collared flycatchers to estimate coding sequence diversity. When binning genes into 10 equally sized categories according to the level of sex-biased expression, we observed strong positive relationships between genetic diversity and level of sex-bias for most organs. In gonads, the relationship was statistically significant both for male-biased genes ($\rho = 0.83$, $p = 0.0056$, Spearman rank correlation) and female-biased genes ($\rho = 0.87$, $p = 0.0027$; Figure 1b). Despite much lower levels of sex-bias (magnitude of sex-bias as well as proportion of sex-biased genes) in nonreproductive organs, the same pattern for male-biased genes were observed in brain (Figure 1d; male-biased: ($\rho = 0.85$, $p = 0.0035$; female-biased: ($\rho = 0.77$, $p = 0.01$) and, to a lower extent, also in kidney, liver, skin, muscle for one or both of the sexes (Supporting information Table S1). Moreover, similar patterns were seen in separate analyses of diversity at nonsynonymous (p_N) and synonymous sites (p_S), respectively (Supporting information Figure S1), again with the strongest effects detected in gonads for both male-biased (nonsynonymous sites) and female-biased (synonymous sites) genes, but with statistically significant effects also in several somatic organs (Supporting information Table S2). These data thus indicate that genes, which show sex-biased expression, are typically associated with higher levels of genetic diversity than unbiased genes. This would be compatible with, but does not prove, the

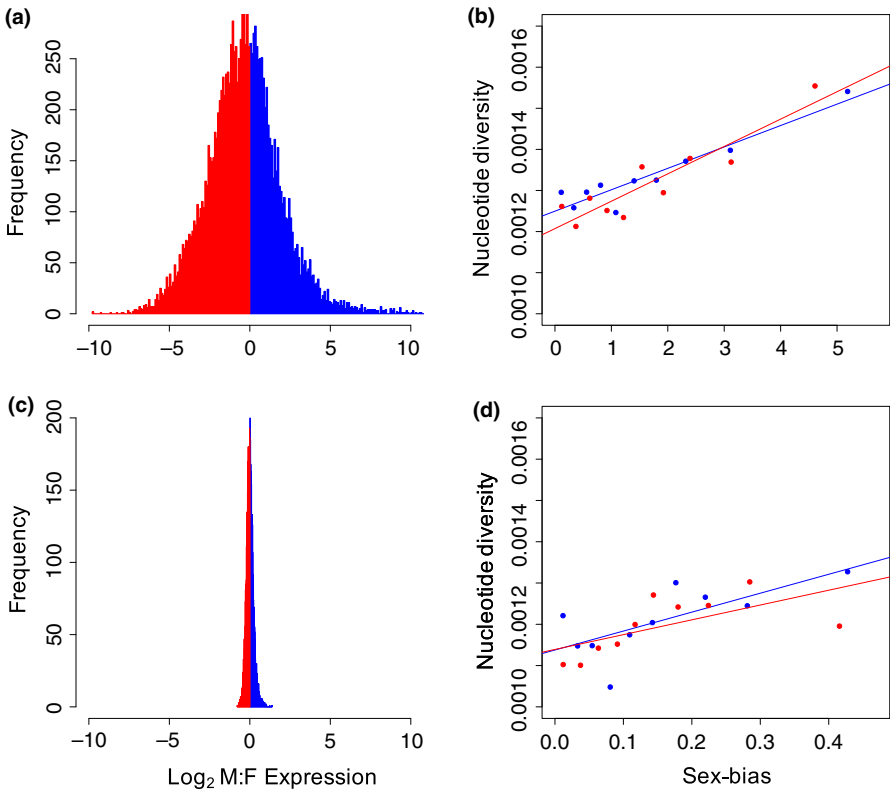

FIGURE 1 Patterns of sex-biased gene expression in gonads (a, b) and brain (c, d). (a) and (c) show the distribution of sex-biased expression among genes with male-biased genes in blue and female-biased in red. (b) and (d) are the relationships between sex-bias ($\log_2$) and nucleotide diversity for male-biased genes (blue) and female-biased genes (red)

action of balancing selection where different alleles are favoured in males and females, respectively, due to sexual antagonism.

Having demonstrated a link between coding sequence diversity and sex-biased gene expression, we asked whether the sex-bias also affects genomewide levels of genetic variation, that is, diversity in genomic regions outside genes including intergenic regions and introns. Several factors can potentially affect nucleotide diversity on a genomic scale and thus need to be taken into account. We performed multiple regression analysis incorporating coding sequence density, GC content, recombination rate, repeat density, the synonymous substitution rate, d_S (meant to approximate the mutation rate) and levels of sex-biased gene expression in gonads as candidate explanatory variables for variation in nucleotide diversity in 200-kb windows across the genome. Sex-biased expression correlated weakly with d_S ($R = 0.10$; $p = 3.24 \times 10^{-6}$) but not with any of the other genomic variables. However, several of these other explanatory variables showed correlations with each other (Supporting information Table S3). To be able to handle the problem of collinearity between explanatory variables, we used a principal component regression (PCR, Figure 2; Supporting information Table S4). PC6 explained most of the variance (8.9%) and was mainly driven by GC content, coding sequence density and recombination rate. This is consistent with previous evidence indicating a strong role of linked selection in governing genomic levels of genetic diversity in flycatchers (Burri et al., 2015; Dutoit et al., 2017); note that GC content is strongly correlated with recombination rate in avian genomes (Backström et al., 2010; Kawakami et al., 2014; Mugal et al., 2013). PC5 explained 6.1% of the variance and was driven by d_S. PC1 explained 4.4% of the variance but without a clearly dominating explanatory variable. PC2 explained 4.2% of the variance and was mainly driven by sex-biased gene expression. PC3 and PC4 explained less variance and were difficult to interpret. A conclusion from the PCR analysis was thus that the higher genetic diversity of genes with sex-biased gene expression extends to be associated with heterogeneity in the overall levels of nucleotide diversity in the collared flycatcher genome.

While the above observations are consistent with balancing selection in sex-biased genes as a cause to elevated genetic diversity, relaxed purifying selection in such genes could also give rise to this pattern. To test this possibility, we performed additional analyses. First, we estimated the ratio of the nonsynonymous to synonymous substitution rate (d_N/d_S). For genes expressed in gonads, d_N/d_S was significantly higher for male-biased but not female-biased, genes compared to unbiased genes as well as to all genes irrespective of where they were expressed (Figure 3). Moreover, d_N/d_S was positively correlated with the degree of sex-bias for male-biased but not female-biased genes (Supporting information Figure S2). Elevated d_N/d_S is often interpreted as a signature of either relaxed purifying selection or adaptive evolution. A way to distinguish between these scenarios is to examine the ratio of the nonsynonymous to synonymous polymorphism (p_N/p_S); relaxed constraint should increase the frequency of segregating nonsynonymous polymorphisms and hence lead to increased p_N/p_S. For male-biased genes, p_N/p_S was positively correlated with the degree of sex-bias ($\rho = 0.71$, $p = 0.028$) while this was not the case for female-biased genes ($\rho = 0.16$, $p = 0.66$; Figure 4). We note that increased p_N/p_S could also be predicted from balancing selection.

Another way to explore the potential role of balancing selection, associated with sexual antagonism, is to investigate the persistence of alleles. Balancing selection should led to deeper coalescence and thereby set the stage for trans-species polymorphisms (Charlesworth, 2006). We assessed the proportion of polymorphic sites within collared flycatchers that also segregate in the semicollared flycatcher;

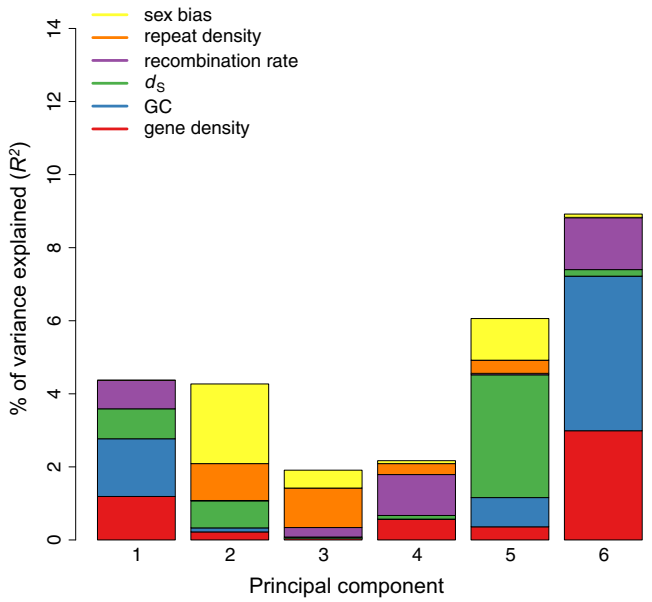

FIGURE 2 Amount of nucleotide diversity in 200-kb windows explained by the different principal components in a principal component regression

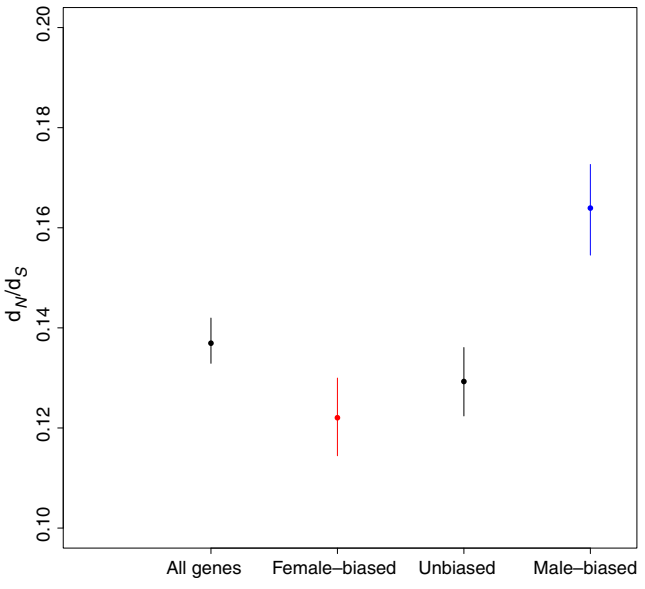

FIGURE 3 Mean d_N/d_S with 95% confidence interval for female-biased, unbiased, and male-biased genes expressed in gonads, as well as for all genes irrespective of where expressed [Colour figure can be viewed at wileyonlinelibrary.com]

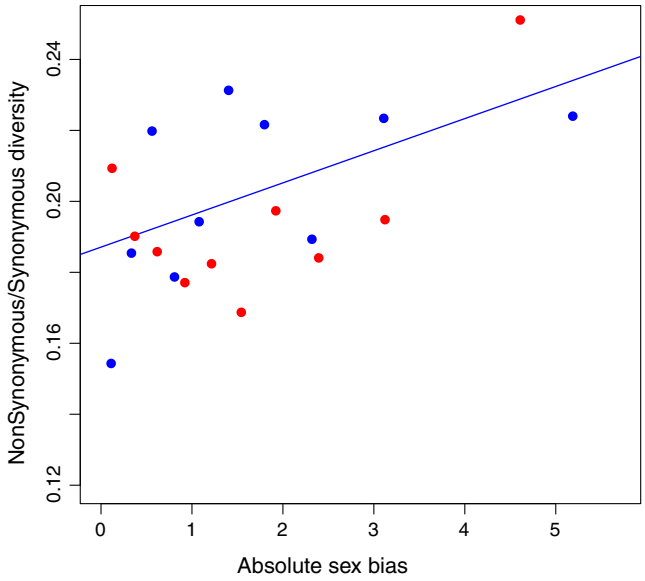

FIGURE 4 Relationship between the ratio of nonsynonymous to synonymous polymorphism (p_N/p_S) and the extent of sex-biased expression in gonads for male-biased (blue; p = 0.028) and female-biased (red; p = 0.66) genes [Colour figure can be viewed at wileyonlinelibrary.com]

the two species diverged 1–1.5 million years ago and have been shown to share many variable sites (Nater, Burri, Kawakami, Smeds, & Ellegren, 2015). Using the same binning approach as above, the proportion of shared variants between the two species was positively correlated with the level of sex-biased expression for male-biased genes (Figure 5a, p = 0.020), but not for female-biased genes (ρ = –0.33, p = 0.35). Consistent with this, Tajima's D was positively correlated with the level of sex-bias for male-biased genes (Figure 5b; (ρ = 0.83, p = 0.0056), but not for female-biased genes (ρ = –0.60, p = 0.07). High estimates of Tajima's D indicate an excess of intermediate frequency alleles.

Finally, if sexual conflict is strong such that contemporary sexual antagonism at loci showing sex-biased expression is associated with viability selection, it may potentially be visible as allele frequency differences between sexes. To test this possibility, we estimated genetic differentiation (F_{ST}) between males and females from the same sample (n = 43 individuals of each sex). To maximize power, we grouped genes into three different categories based on expression patterns in gonads (unbiased, female-biased, and male-biased genes, Figure 6). We also estimated "background" levels of F_{ST} within the population by randomly assigning individuals to two groups and calculating F_{ST} between these groups. The latter confirmed an expectation of a null F_{ST} (mean = -3.55×10^{-6}, 95% CI: -3.60×10^{-4} to 3.87×10^{-4}). However, F_{ST} between males and females across all genes proved to be significantly different from zero (mean = 8.30×10^{-4}, 95% CI 2.95×10^{-4} to 1.45×10^{-3}. When genes were divided into three categories according to sex-biased expression, only F_{ST} for male-biased genes remained significantly different from zero (mean = 1.59×10^{-3}; 95% CI 3.66×10^{-4} to 3.03×10^{-3}), suggesting that these genes might be driving the signal seen for all genes. For female-biased genes (mean $F_{ST} = 3.41 \times 10^{-4}$, 95% CI -2.05×10^{-4} to 9.18×10^{-4}) and unbiased genes (mean $F_{ST} = 5.39 \times 10^{-4}$, 95% CI -3.14×10^{-4} to 1.69×10^{-3}), F_{ST} was not significantly different from zero. When analysing genes specifically expressed in gonads, both testis-specific and ovary-specific genes had significant and non-zero mean F_{ST} estimates (Supporting information Figure S3).

4 | DISCUSSION

This study demonstrates a genomewide association between the extent of sex-biased gene expression and levels of genetic diversity that, to our knowledge, has not previously been documented. The association was seen for diversity both at the level of individual genes and at the level of 200-kb windows across the genome. These observations are consistent with predictions from theory of sexual conflict, namely that sexual antagonism promotes the maintenance of segregating variation with opposing effects on male and female fitness via balancing selection (Balaresque, Toupance, Quintana-

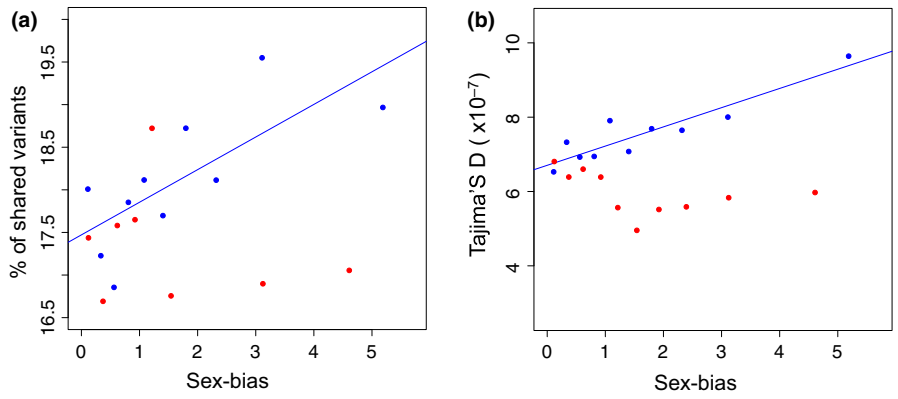

FIGURE 5 (a) Relationship between the proportion of shared polymorphisms between collared flycatcher and semicollared flycatcher and extent of sex-biased gene expression in gonads (blue is male-biased genes with p = 0.02; red is female-biased genes with p = 0.35). (b) Relationship between Tajima's D and extent of sex-biased expression in gonads for male-biased genes (blue, p = 0.0056) and female-biased genes (red, p = 0.07) [Colour figure can be viewed at wileyonlinelibrary.com]

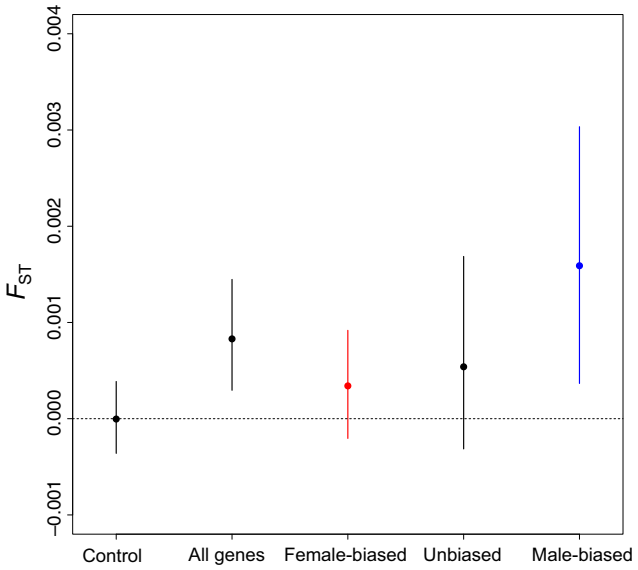

FIGURE 6 F_{ST} between male and female collared flycatchers from the same population, with 95% confidence interval. The control category is F_{ST} calculated after random assignment of individuals in two groups. F_{ST} between males and females for all genes (mean = 8.30×10^{-4}, 95% CI = 2.95×10^{-4} to 1.45×10^{-3}) and for male-biased genes (mean = 1.59×10^{-3}; 95% CI = 3.66×10^{-4} to 3.03×10^{-3}) were significantly different from zero. Mean level of sex-bias was 2.4 for female-biased genes and 2.23 for male-biased genes. Unbiased genes were weakly overexpressed in females (sex-bias = 0.13). The dashed line indicates $F_{ST} = 0$ [Colour figure can be viewed at wileyonlinelibrary.com]

Murcilluis, Crouau-Roy, & Heyer, 2004; Connallon & Clark, 2014; Crow & Kimura, 1970; Kidwell, Clegg, Stewart, & Prout, 1977; Rice, 1984). It relies on the assumption that sex-biased gene expression is an indicator of sexual antagonism. While it is unlikely that all genes showing sex-biased expression are currently exposed to sexually antagonistic selection, we consider sex-biased genes to represent a set of loci that are highly enriched for genes subject to ongoing sexual conflict. But it also means that this may not hold true for any particular gene.

The positive correlation between sex-biased expression and nucleotide diversity seen in collared flycatcher is no evidence for a causal relationship between sexual conflict and genetic diversity. In theory, it could be that some other factor that covaries with sex-biased expression affects diversity levels. Within-genome variation in diversity levels is governed by a complex interplay of several factors (Ellegren & Galtier, 2016). Mutation rate variation and introgression, for example, are two such factors. In flycatchers (Burri et al., 2015; Dutoit et al., 2017), as in many other species (Begun & Aquadro, 1992; Elyashiv et al., 2016; Slotte, 2014), linked selection is a major determinant of local variation in diversity levels. This was confirmed in this study. The extent of linked selection is in turn affected by recombination rate variation and the density of targets of selection (Cutter & Payseur, 2013). The principal component regression analysis that we used was intended to disentangle the effects of different explanatory variables. One principal component was found to be driven by sex-biased expression, and to some extent, it thus strengthens the probability of a causal link between sexual conflict and diversity levels. While sex-biased gene expression is not a major determinant of diversity, the effect appears to be detectable.

Balancing selection, opposite to background selection and selective sweeps (reviewed in Cutter & Payseur, 2013 and Ellegren & Galtier, 2016), can act to maintain and promote genetic diversity at genetically linked sites (Charlesworth, Nordborg, & Charlesworth, 1997; Hudson & Kaplan, 1988; Schierup, Charlesworth, & Vekemans, 2000). Our analyses of the relationship between sex-biased gene expression and genetic diversity in nearby genomic regions were indeed made with this premise. However, relaxed purifying selection could also promote genetic diversity, both directly at functional sites but also indirectly by reducing the effect of background selection. This is an important caveat in our study that should be kept in mind. In our data male-biased genes showed elevated d_N/d_S and elevated p_N/p_S, for which relaxed purifying selection could be an as viable explanation as balancing selection. For female-biased genes, this was not the case. Similarly, the findings that both Tajima's D and the proportion of shared variants between two flycatcher species were positively correlated with the level male-biased expression would be compatible with balancing selection as well as relaxed purifying selection (the latter would locally increase the effective population size (N_e) and thereby coalescence times). Again, this was not observed for female-biased genes.

An intriguing observation was that of nonzero mean F_{ST} estimates between the sexes for genes showing male-biased expression. Intuitively, F_{ST} should be expected to be zero between any (sufficiently large) samples taken from a randomly mating population— that is, where there is no population stratification—and in the absence of frequent sex-specific admixture. Two recent studies have argued that sex-specific viability selection could give rise to small but detectable allele frequency differences between males and females within a generation and have provided genomewide empirical support for such differences in humans and *Drosophila melanogaster* (Cheng & Kirkpatrick, 2016; Lucotte, Laurent, Heyer, Ségurel, & Toupance, 2016; see also Balaresque et al., 2004). The idea is that if some alleles confer a very strong disadvantage to one sex, fewer individuals of that sex carrying those alleles will survive until the time of sampling, introducing allele frequency differences between sexes in the sample. Random assortment of chromosomes at meiosis will then reset these differences such that F_{ST} is again zero at conception.

The pattern of mean F_{ST} between males and females for male-biased genes being significantly different from zero is consistent with sex-specific viability selection and ongoing sexual conflict. The power in our data is far too low to allow analyses at the level of individual genes, and it is difficult to judge what proportion of genes are affected by sex-specific viability selection. Therefore, we cannot test whether the correlation between sex-biased expression and F_{ST} in flycatchers mirrors the quadratic relationship that Cheng and Kirkpatrick (2016) observed in humans and *D. melanogaster*. They found that F_{ST} peaks at intermediate levels of sex-bias (both for male-biased and for female-biased genes) gave rise to a "twin tower" pattern. It was suggested that this can be explained by that sexual

conflict tends to be solved in genes with pronounced sex-biased expression, meaning that viability selection should no longer give rise to allele frequency differences between sexes. While we do not know if this is the case in flycatchers, we note that in our binned data the relationship between nucleotide diversity and extent of sex-biased expression was linear both for male-biased and female-biased genes. Also, we found that, for male-biased genes, sex-biased expression increased linearly with Tajima's D over the whole range of binned sex-biases. This is somewhat surprising because the probability for any form of intralocus conflict over gene expression (i.e., not only viability selection) to be resolved could be expected to be highest for genes with the highest degree of sex-bias. Further work will be needed to better characterize the relationship between nucleotide diversity and sex-bias. For example, it would be valuable to have RNA-seq data from much larger population samples to be able to make analyses at the level of individual genes rather than bins.

An obvious question arising from our results is why the signatures of balancing selection (trans-species polymorphisms, Tajima's D) and nonzero F_{ST} between males and females were seen for genes with male-biased expression but not for genes with female-biased expression. Of course, we cannot exclude that low power or lower effect sizes rendered relationships nonsignificant for female-biased genes. Moreover, the correlation between the extent of sex-biased expression and nucleotide diversity was observed both for male-biased and female-biased genes. Yet, it may very well be that sexual antagonism is more prevalent in male-biased genes than in female-biased genes due to sexual selection (Harrison et al., 2015; Pointer, Harrison, Wright, & Mank, 2013), which in collared flycatchers is visible as distinct sexual dimorphism and a high rate of extrapair paternity (Sheldon & Ellegren, 1999). Harrison et al. (2015) showed that sexual selection leads to rapid turnover of male-biased expression within a clade of birds. Such rapid turnover would set the stage for transient phases of unresolved antagonism that promotes genetic diversity via balancing selection. If so, and somewhat paradoxically given that sexual selection is usually thought of as depleting genetic variability, sexual selection might be particularly important to the maintenance of genetic diversity by ongoing sexual conflict.

The role of balancing selection in maintaining genetic diversity has been a long-standing issue in evolutionary biology (Charlesworth, 2006; Gillespie, 1991; Lewontin, 1974). It is well documented in many organisms that genetic diversity in a limited number of genes involved in coevolutionary arms races is driven by balancing selection (Fijarczyk & Babik, 2015). There has recently also been genomic data in support of balancing selection affecting diversity levels at many loci across the human genome (Andrés et al., 2009; DeGiorgio, Lohmueller, & Nielsen, 2014), but a link to sexual conflict has not been obvious. Our results from flycatchers provide a step in this direction and offer such a link. We note that relaxed purifying selection could potentially explain some, but not all, of the observations related to the observed correlations between sex-biased expression and different measures of diversity. This will need further investigation and our conclusions should therefore be seen as suggestive rather than conclusive. However, it cannot easily be imagined how relaxed purifying selection of male-biased genes would lead to the finding of nonzero F_{ST} between males and females for such genes. In summary, it is therefore the combined picture provided by several observations that leads us to favour the interpretation of the correlation between sex-biased gene expression and genetic diversity to at least in part be driven by sexual antagonism.

ACKNOWLEDGEMENTS

This study was funded by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. Sequencing was performed by the SNP&SEQ Technology Platform at Uppsala University, as part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. Computations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX). We thank three anonymous reviewers for helpful comments, and Göran Arnqvist and members of the Ellegren laboratory for helpful discussions.

AUTHOR CONTRIBUTIONS

H.E. conceived of and supervised the study, and designed research; L.D. performed research; L.D. and H.E. wrote the manuscript; C.F.M., P.B., M.W., K.N.B., L.S. and H.P. analysed data; and L.G. provided samples.

DATA ACCESSIBILITY

Transcriptomics data are available at SRA ERP001377. Genomic resequencing data are available at ENA PRJEB11502. Scripts and other data are available at Dryad https://doi.org/10.5061/dryad.qc5ft8n

REFERENCES

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSEQ—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Andrés, A. M., Hubisz, M. J., Indap, A., Torgerson, D. G., Degenhardt, J. D., Boyko, A. R., … Nielsen, R. (2009). Targets of balancing selection in the human genome. *Molecular Biology and Evolution*, *26*(12), 2755–2764. https://doi.org/10.1093/molbev/msp190

Arnqvist, G., & Rowe, L. (2005). *Sexual conflict*. Princeton, NJ: Princeton University Press. https://doi.org/10.1515/9781400850600

Backström, N., Forstmeier, W., Schielzeth, H., Mellenius, H., Nam, K., Bolund, E., … Ellegren, H. (2010). The recombination landscape of the zebra finch *Taeniopygia guttata* genome. *Genome Research*, *20*(4), 485–495.https://doi.org/10.1101/gr.101410.109

Balaresque, P., Toupance, B., Quintana-Murcilluis , Crouau-Roy, B., & Heyer, E. (2004). Sex-specific selection on the human X chromosome? *Genetical Research*, *83*(3), 169–176. https://doi.org/10.1017/S0016672304006858

Barton, N. H., & Turelli, M. (1989). Evolutionary quantitative genetics: How little do we know? *Annual Review of Genetics*, *23*(1), 337–370. https://doi.org/10.1146/annurev.ge.23.120189.002005

Begun, D. J., & Aquadro, C. F. (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature*, *356*(6369), 519–520. https://doi.org/10.1038/356519a0

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, *57*(1), 289–300.

Bolívar, P., Mugal, C. F., Nater, A., & Ellegren, H. (2016). Recombination rate variation modulates gene sequence evolution mainly via GC-biased gene conversion, not Hill-Robertson interference, in an avian system. *Molecular Biology and Evolution*, *33*(1), 216–227.

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

Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, *2*(4), e64. https://doi.org/10.1371/journal.pgen.0020064

Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, *70*(2), 155–174. https://doi.org/10.1017/S0016672397002954

Cheng, C., & Kirkpatrick, M. (2016). Sex-specific selection and sex-biased gene expression in humans and flies. *PLoS Genetics*, *12*(9), e1006170. https://doi.org/10.1371/journal.pgen.1006170

Connallon, T., & Clark, A. G. (2012). A general population genetic framework for antagonistic selection that accounts for demography and recurrent mutation. *Genetics*, *190*(4), 1477–1489. https://doi.org/10.1534/genetics.111.137117

Connallon, T., & Clark, A. G. (2014). Balancing selection in species with separate sexes: Insights from Fisher's geometric model. *Genetics*, *197*(3), 991–1006. https://doi.org/10.1534/genetics.114.165605

Connallon, T., & Knowles, L. L. (2005). Intergenomic conflict revealed by patterns of sex-biased gene expression. *Trends in Genetics*, *21*(9), 495–499. https://doi.org/10.1016/j.tig.2005.07.006

Crow, J. F., & Kimura, M. (1970). *An introduction to population genetics theory*. New York, NY: Harper and Row.

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

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

DeGiorgio, M., Lohmueller, K. E., & Nielsen, R. (2014). A model-based approach for identifying signatures of ancient balancing selection in genetic data. *PLoS Genetics*, *10*(8), e1004561. https://doi.org/10.1371/journal.pgen.1004561

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

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 B: Biological Sciences*, *284*(1849), 20162756. https://doi.org/10.1098/rspb.2016.2756

Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, *17*(7), 422–433. https://doi.org/10.1038/nrg.2016.58

Ellegren, H., & Parsch, J. (2007). The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics*, *8*(9), 689–698. https://doi.org/10.1038/nrg2167

Elyashiv, E., Sattath, S., Hu, T. T., Strutsovsky, A., McVicker, G., Andolfatto, P., … Sella, G. (2016). A genomic map of the effects of linked selection in *Drosophila*. *PLoS Genetics*, *12*(8), e1006130. https://doi.org/10.1371/journal.pgen.1006130

Fijarczyk, A., & Babik, W. (2015). Detecting balancing selection in genomes: Limits and prospects. *Molecular Ecology*, *24*(14), 3529–3545. https://doi.org/10.1111/mec.13226

Gillespie, J. H. (1991). *The causes of molecular evolution*. Oxford, UK: Oxford University Press.

Harrison, P. W., Wright, A. E., Zimmer, F., Dean, R., Montgomery, S. H., Pointer, M. A., & Mank, J. E. (2015). Sexual selection drives evolution and rapid turnover of male gene expression. *Proceedings of the National Academy of Sciences USA*, *112*(14), 4393–4398. https://doi.org/10.1073/pnas.1501339112

Hudson, R. R., & Kaplan, N. L. (1988). The coalescent process in models with selection and recombination. *Genetics*, *120*(3), 831–840.

Ingleby, F. C., Flis, I., & Morrow, E. H. (2014). Sex-biased gene expression and sexual conflict throughout development. *Cold Spring Harbor Perspectives in Biology*, *7*(1), a017632.

Innocenti, P., & Morrow, E. H. (2010). The sexually antagonistic genes of *Drosophila melanogaster*. *PLoS Biology*, *8*(3), e1000335.

Jordan, C. Y., & Connallon, T. (2014). Sexually antagonistic polymorphism in simultaneous hermaphrodites. *Evolution*, *68*(12), 3555–3569. https://doi.org/10.1111/evo.12536

Kardos, M., Husby, A., McFarlane, S. E., Qvarnstrom, A., & Ellegren, H. (2016). Whole-genome resequencing of extreme phenotypes in collared flycatchers highlights the difficulty of detecting quantitative trait loci in natural populations. *Molecular Ecology Resources*, *16*(3), 727–741. https://doi.org/10.1111/1755-0998.12498

Kawakami, T., Smeds, L., Backström, N., Husby, A., Qvarnström, 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

Kidwell, J. F., Clegg, M. T., Stewart, F. M., & Prout, T. (1977). Regions of stable equilibria for models of differential selection in the two sexes under random mating. *Genetics*, *85*(1), 171–183.

Kokko, H., Brooks, R., Jennions, M. D., & Morley, J. (2003). The evolution of mate choice and mating biases. *Proceedings of the Royal Society of London B: Biological Sciences*, *270*(1515), 653–664. https://doi.org/10.1098/rspb.2002.2235

Kruuk, L. E. B., & Hill, W. G. (2008). Evolutionary dynamics of wild populations: The use of long-term pedigree data. *Proceedings of the Royal Society of London B: Biological Sciences*, *275*(1635), 593–596. https://doi.org/10.1098/rspb.2007.1689

Lewontin, R. C. (1974). *The genetic basis of evolutionary change*. New York: Columbia University Press.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lucotte, E. A., Laurent, R., Heyer, E., Ségurel, L., & Toupance, B. (2016). Detection of allelic frequency differences between the sexes in humans: A signature of sexually antagonistic selection. *Genome Biology and Evolution*, *8*(5), 1489–1500. https://doi.org/10.1093/gbe/evw090

Mank, J. E. (2017). The transcriptional architecture of phenotypic dimorphism. *Nature Ecology & Evolution*, *1*(1), 6. https://doi.org/10.1038/s41559-016-0006

Mitchell-Olds, T., Willis, J. H., & Goldstein, D. B. (2007). Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews Genetics*, *8*(11), 845–856. https://doi.org/10.1038/nrg2207

Mugal, C. F., Nabholz, B., & Ellegren, H. (2013). Genome-wide analysis in chicken reveals that local levels of genetic diversity are mainly governed by the rate of recombination. *BMC Genomics*, *14*(1), 86. https://doi.org/10.1186/1471-2164-14-86

Muirhead, C. A., Glass, N. L., & Slatkin, M. (2002). Multilocus self-recognition systems in fungi as a cause of trans-species polymorphism. *Genetics*, *161*(2), 633–641.

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.

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences USA*, *76*(10), 5269–5273. https://doi.org/10.1073/pnas.76.10.5269

Pointer, M. A., Harrison, P. W., Wright, A. E., & Mank, J. E. (2013). Masculinization of gene expression is associated with exaggeration of male sexual dimorphism. *PLoS Genetics*, *9*(8), e1003697. https://doi.org/10.1371/journal.pgen.1003697

Radwan, J. (2008). Maintenance of genetic variation in sexual ornaments: a review of the mechanisms. *Genetica*, *134*(1), 113–127.

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from: https://www.R-project.org/

Rice, W. R. (1984). Sex chromosomes and the evolution of sexual dimorphism. *Evolution*, *38*(4), 735–742. https://doi.org/10.1111/j.1558-5646.1984.tb00346.x

Schierup, M. H., Charlesworth, D., & Vekemans, X. (2000). The effect of hitch-hiking on genes linked to a balanced polymorphism in a subdivided population. *Genetical Research*, *76*(1), 63–73. https://doi.org/10.1017/S0016672300004547

Sheldon, B., & Ellegren, H. (1999). Sexual selection resulting from extra-pair paternity in collared flycatchers. *Animal Behaviour*, *57*(2), 285–298. https://doi.org/10.1006/anbe.1998.0968

Slotte, T. (2014). The impact of linked selection on plant genomic variation. *Briefings in Functional Genomics*, *13*(4), 268–275. https://doi.org/10.1093/bfgp/elu009

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595.

Turelli, M., & Barton, N. H. (2004). Polygenic variation maintained by balancing selection: Pleiotropy, sex-dependent allelic effects and G × E interactions. *Genetics*, *166*(2), 1053–1079. https://doi.org/10.1534/genetics.166.2.1053

Uebbing, S., Kunstner, A., Makinen, H., & Ellegren, H. (2013). Transcriptome sequencing reveals the character of incomplete dosage compensation across multiple tissues in flycatchers. *Genome Biology and Evolution*, *5*(8), 1555–1566. https://doi.org/10.1093/gbe/evt114

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Dutoit L, Mugal CF, Bolívar P, et al. Sex-biased gene expression, sexual antagonism and levels of genetic diversity in the collared flycatcher (*Ficedula albicollis*) genome. *Mol Ecol*. 2018;27:3572–3581. https://doi.org/10.1111/mec.14789