

Footprints of adaptive evolution revealed by whole Z chromosomes haplotypes in flycatchers

Krystyna Nadachowska-Brzyska¹  | Reto Burri^{1,2}  | Hans Ellegren¹ 

¹Department of Evolutionary Biology, University of Uppsala, Uppsala, Sweden

²Department of Population Ecology, Friedrich Schiller University Jena, Jena, Germany

Correspondence

Krystyna Nadachowska-Brzyska, Department of Evolutionary Biology, University of Uppsala, Norbyvägen, Uppsala, Sweden.
Email: krystyna.nadachowska-brzyska@ebc.uu.se

Funding information

Swedish Research Council, Grant/Award Number: 2013-8271; Knut och Alice Wallenbergs Stiftelse

Abstract

Detecting positive selection using genomic data is critical to understanding the role of adaptive evolution. Of particular interest in this context is sex chromosomes since they are thought to play a special role in local adaptation and speciation. We sought to circumvent the challenges associated with statistical phasing when using haplotype-based statistics in sweep scans by benefitting from that whole chromosome haplotypes of the sex chromosomes can be obtained by resequencing of individuals of the hemizygous sex. We analyzed whole Z chromosome haplotypes from 100 females from several populations of four black and white flycatcher species (in birds, females are ZW and males ZZ). Based on integrated haplotype score (iHS) and number of segregating sites by length (nSL) statistics, we found strong and frequent haplotype structure in several regions of the Z chromosome in each species. Most of these sweep signals were population-specific, with essentially no evidence for regions under selection shared among species. Some completed sweeps were revealed by the cross-population extended haplotype homozygosity (XP-EHH) statistic. Importantly, by using statistically phased Z chromosome data from resequencing of males, we failed to recover the signals of selection detected in analyses based on whole chromosome haplotypes from females; instead, what likely represent false signals of selection were frequently seen. This highlights the power issues in statistical phasing and cautions against conclusions from selection scans using such data. The detection of frequent selective sweeps on the avian Z chromosome supports a large role of sex chromosomes in adaptive evolution.

KEYWORDS

Ficedula flycatchers, haplotype-based statistics, ongoing selection, sex chromosomes

1 | INTRODUCTION

One of the major goals in evolutionary biology is to understand the processes underlying sequence diversity. In particular, it is commonly asked whether the genetic variation observed within species can be explained by recent or ongoing selection, or if it is mainly

governed by neutral processes (Cutter & Payseur, 2013; Nielsen, 2005; Nielsen, Hellmann, Hubisz, Bustamante, & Clark, 2007; Vitti, Grossman, & Sabeti, 2013). Different types of selection are expected to leave different signatures on the pattern of sequence variation. For selective sweeps, these include low levels of genetic variation around the selected sites, elevated linkage disequilibrium, skewed

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2019 The Authors. *Molecular Ecology* Published by John Wiley & Sons Ltd

allele frequency spectrum, and elevated rates of between-population divergence (Fay & Wu, 2000; Nielsen, 2005; Vitti et al., 2013). The magnitude of the observed signals is depended on the initial frequency of a beneficial allele and the time that has passed since selection started to act. The most pronounced signatures are predicted when selection acts on de novo mutations (hard sweep model, Kim & Stephan, 2002; Przeworski, 2002; Smith & Haigh, 1974).

New mutations are associated with a given set of linked SNPs (forming a haplotype) that can become fixed by hitchhiking with a focal variant under selection. This will generate strong linkage disequilibrium and deplete nucleotide diversity in the neighbouring genomic region. As time passes, we expect to see an excess of high-frequency derived alleles, and, after fixation of the selected allele, an excess of rare alleles that start to accumulate on a homogenized genetic background. Less pronounced signals are predicted when selection starts acting on standing genetic variation, where a focal variant can be associated with several haplotypes present in the population (soft sweep models, Hermisson & Pennings, 2005; Pennings & Hermisson, 2006; Pritchard & Di Rienzo, 2010; Teshima, Coop, & Przeworski, 2006), or in the case of still incomplete sweeps (regardless if they are soft or hard sweeps).

Detecting very recent events of selection, where the beneficial allele(s) is still in low or intermediate frequency may be difficult by methods based on the site frequency spectrum (SFS) or F_{ST} -based scans (Fariello, Boitard, Naya, Sancristobal, & Servin, 2013; Vitti et al., 2013; Voight, Kudravalli, Wen, & Pritchard, 2006). However, haplotype-based statistics that rely on the decay of haplotype homozygosity away from a focal site may pick up signals of ongoing selection much earlier (Vatsiou, Bazin, & Gaggiotti, 2016; Voight et al., 2006). Although haplotype statistics have been used in population genetics for a long time, their application to genome-wide data is still relatively limited and, when it has been done, skewed towards model organisms, in particular humans (Amorim et al., 2017; Didion et al., 2016; Sabeti et al., 2002; Schlamp et al., 2016; Triska et al., 2015; Voight et al., 2006). This is due to the fact that genome-wide haplotype-based statistics can only be estimated on chromosomes of known phase. Typically, phase has to be statistically inferred when analysing diploid organisms and, to be reliable, this requires very large resequencing data sets (Akagi, Hanada, Yaegaki, Gradziel, & Tao, 2016; Gagnaire, Pavey, Normandeu, & Bernatchez, 2013; Nelson, Wallberg, Simões, Lawson, & Webster, 2017; Schlamp et al., 2016; Vijay et al., 2016).

An interesting possibility to circumvent the need for statistical phasing in selection scans is to focus on sex chromosomes and use sequence data from the hemizygote sex; that is males in organisms with male heterogamety (X and Y sex chromosomes) and females in organisms with female heterogamety (Z and W sex chromosomes). This will provide the full X or Z chromosome haplotype for each hemizygous individual and should dramatically facilitate tests for selection both in terms of length and accuracy of the underlying haplotype data. Although limited to a single chromosome, such tests can be of particular relevance given the disproportionate role of sex

chromosomes in adaptation and speciation suggested for many organisms (the fast-X and large-X effects; Meisel & Connallon, 2013; Presgraves, 2008; Qvarnström & Bailey, 2009).

Black-and-white flycatchers of the genus *Ficedula* are small passerine birds that recently (<1–2 million years; Nater, Burri, Kawakami, Smeds, & Ellegren, 2015) radiated into four species: Atlas flycatcher (*F. speculigera*), collared flycatcher (*F. albicollis*), pied flycatcher (*F. hypoleuca*) and semicollared flycatcher (*F. semitorquata*). Following the generation of a chromosome-level assembly of the collared flycatcher genome (Ellegren et al., 2012; Kawakami et al., 2014), attempts have been made to infer the processes behind the heterogeneous genomic landscapes of diversity and differentiation seen in *Ficedula* species (Burri et al., 2015; Dutoit, Burri, Nater, Mugal, & Ellegren, 2017; Dutoit, Vijay, et al., 2017; Ellegren et al., 2012; Kawakami et al., 2017), as well as how demography has shaped these landscapes (Kardos, Qvarnström, & Ellegren, 2017; Nadachowska-Brzyska et al., 2013; Nadachowska-Brzyska, Burri, Smeds, & Ellegren, 2016). Here, we seek to infer the role of adaptive evolution on flycatcher sex chromosomes by benefitting from full Z chromosome haplotypes in genomic resequencing of female birds. We use these data in scans for ongoing and recently completed sweeps in the four mentioned species. Having identified a number of regions on the Z chromosomes that show strong signatures of adaptive evolution based on female haplotypes, we demonstrate that these signals would not have been captured if statistically inferred male haplotypes had been used.

2 | MATERIALS AND METHODS

2.1 | Data and data analysis

Analyses were performed on SNP data obtained from whole genome resequencing of 200 flycatchers (Burri et al., 2015), including collared flycatchers (80 individuals from four populations, 20 individuals per population), pied flycatchers (80 individuals from four populations, 20 individuals per population), Atlas flycatcher (20 individuals from one population) and semicollared flycatcher (20 individuals from one population). Collared flycatchers were sampled in Italy, Hungary, the Czech Republic, and on the Baltic Sea island Öland (Sweden). Pied flycatchers were sampled in Spain, the Czech Republic, Sweden (mainland), and on Öland. Atlas flycatchers were sampled in the Moroccan Atlas Mountains and semicollared flycatchers in Bulgaria.

In all populations 10 males and 10 females were sampled, with the exception of Atlas flycatchers (six females, 14 males), semicollared flycatcher (10 females and nine males) and Italian collared flycatcher (11 females, nine males). Details about sequencing and variant calling are provided in Burri et al. (2015). Briefly, DNA was prepared using Qiagen's Blood and Tissue Kit and sequencing was performed with Illumina paired-end sequencing technology on a HiSeq 2000 instrument at the SNP&SEQ Technology Platform of Uppsala University. An insert size of ~450 bp was used in individually tagged libraries and fragments were sequenced from both ends using 100 cycles. SNP genotyping followed the recommendation of GATK best practices (McKenna et al., 2010). Base quality score

recalibrations (BQSR) were performed for each population separately. The mean genome-wide coverage varied from 12x to 21x per population (Burri et al., 2015 in Table S2). One collared flycatcher and one pied flycatcher from the Öland populations turned out to be hybrids (identified based on samples' heterozygosity pattern, many SNPs in hybrid individuals were heterozygous at places known to be fixed for alternative alleles in the two species) and were excluded from further analysis. We divided the data into one data set consisting only of females and one only of males.

For downstream analysis we only used SNP data from reads mapping to Z-linked scaffolds in the collared flycatcher genome assembly version fAlb15, excluding the 630 kb pseudoautosomal region (PAR; Smeds et al., 2014). In addition to the quality filtering described by Burri et al. (2015), we further filtered the data for high quality SNPs by considering only sites with at least five reads in at least 50% of the individuals in a given population, and at least two reads in other individuals of that population. This additional filter was motivated by the fact that the Z chromosome has lower coverage in females (approximately two times lower since females carry only one Z chromosome). SNPs were polarized using resequencing data from red-breasted flycatcher (*F. parva*) and snowy-browed flycatcher (*F. hyperythra*; Burri et al., 2015). We retained 80,000–16,000 Z-linked SNPs per population.

Sequence data from males were phased using SHAPEIT2 (Delaneau, Zagury, & Marchini, 2013) with window size of 0.5 Mb and 500 iterations, including 200 iterations of burn-in and pruning. Statistical phasing was performed with several data sets. First, we phased 10 males within each population (which had at least 10 males sequenced). Second, we combined all males from all populations and species, and phased them all together. Additionally, we performed phasing within each population including both males and females. Furthermore, we phased a total of 80 males from an additional data set of Öland collared flycatchers (Kardos, Husby, McFarlane, Qvarnström, & Ellegren, 2016).

In all phasing procedures we used recombination rates from a collared flycatcher linkage map with rates estimated in 1 Mb windows (Kawakami et al., 2014). To be able to associate individual SNPs with recombination rate we linearly interpolated recombination rates estimated in neighbouring windows.

2.2 | Haplotype-based statistics

Integrated haplotype score (iHS) (Voight et al., 2006) and cross population extended haplotype homozygosity (XP-EHH) (Sabeti et al., 2007) scans were performed using the selscan software (Szpiech & Hernandez, 2014). We used default parameters with a maximum gap size of 200 kb and an EHH decay cut-off of 0.05. iHS is calculated as the decay of haplotype homozygosity as a function of recombination distance, for which we used the same recombination rate data as above (Kawakami et al., 2014). Additionally, we calculated iHS statistics based on population recombination rate estimates (ρ) available for both collared and pied flycatcher population ($\rho = 3N_e r$ for Z chromosome; Kawakami et al., 2017).

To translate ρ into recombination rate (cM/Mb), we used coalescent-based N_e estimates from Nater et al. (2015). The results for each SNP were normalized in 0.05 frequency bins and summarized in 200 kb windows (by calculating percentage of SNPs with |iHS| of different values in 200 kb windows). We considered a window to be an outlier if at least 10% of the SNPs had |iHS| values (i.e., the Z-score) above 2, but also tested several other criteria (Supporting Information Table S1).

The XP-EHH statistic was calculated between all population pairs of pied flycatcher and collared flycatcher, respectively. For between-species comparisons we selected one population from pied flycatcher (Spain) and one population from collared flycatcher (Italy), and calculated XP-EHH between all possible species pairs including Atlas flycatcher and semicollared flycatcher. We expected to detect two categories of outlier regions. The first category would consist of regions where a signal of selection for a particular population is visible in all between-population comparisons that include the population in question, and is absent in all comparisons that do not include it. We expected to see such a pattern when selection acts only in one population or when the analysed populations are at different stages of a selective sweep (i.e., ranging from the selected allele(s) beginning to rise in frequency to being close to fixation or already fixed). The second category would consist of regions where the signal of selection for a particular population is visible in some between-population comparisons that include that population and in some but not all comparisons that do not include it. We expect to observe such a pattern when selection acts on at least two populations but is absent in other populations or, as in the first category, the studied populations are at different stages of a selective sweep. For example, if the selected site has been fixed in two out of four populations we expect to detect a selection signal in four population pairs; that is in comparisons between a population with the fixed variant and a population with the advantageous allele still segregating.

The results of XP-EHH selection scans were normalized chromosome-wide and summarized in 200 kb windows (by calculating percentage of SNPs with XP-EHH of different values in 200 kb windows). Similar to the iHS selection scans we considered a window to be an outlier if >10% of SNPs had a value of XP-EHH >2 (for one population in pairwise comparison) or XP-EHH <-2 (for the other population). Additionally, we considered a window to be an outlier if selection was detected in at least three between-species comparisons. Additionally, since XP-EHH was calculated for many pairs of populations and at least some signals may be false positives due to multiple testing, we also checked more strict threshold (>15% of SNPs with XP-EHH >2 or XP-EHH <-2).

The number of segregating sites by length (nSL) statistic was calculated using the original implementation from Ferrer-Admetlla, Liang, Korneliusen, and Nielsen (2014). The nSL statistic measures haplotype length as a function of number of segregating sites. It does not need recombination rate information and should be more robust to recombination rate variation. We used a default value for nSL window size of 200 (defined as number of segregating size) but also considered window sizes of 500 and 1,000. The results of nSL

scans were normalized and summarized following the same procedures as in the case of the iHS statistic.

Nucleotide diversity (π), Tajima's D (Tajima, 1989) and Fay and Wu's H (Fay & Wu, 2000) for all populations (including males and females) were estimated in ANGSD (Korneliussen, Albrechtsen, & Nielsen, 2014) and summarized in 200 kb windows.

3 | RESULTS

We used whole genome sequence data from 97 female flycatchers of four closely related species and extracted full Z chromosome haplotypes. These haplotypes were then used as input files in scans for selection based on the iHS, nSL, and XP-EHH statistics. iHS and nSL are primarily expected to capture signatures of ongoing sweeps. XP-EHH can detect recently completed sweeps.

3.1 | Identifying signatures of haplotype structure

iHS selection scans revealed several regions, in all species, with strong haplotype structure in the form of clustering of outlier windows (Figure 1, Table 1). Three genomic regions in particular showed pronounced signals of extended haplotype homozygosity. In Atlas flycatcher, a 4.4 Mb segment of 22 consecutive outlier windows at position 40.2–44.6 Mb was by far the most distinct cluster in this species and also represented the strongest signal seen in any species (21 windows with >30% SNPs with iHS >2 per window). In collared flycatcher, a 5 Mb segment of 25 outlier windows (position 49–54 Mb; 21 windows with >30% SNPs with iHS >2 per window) was found in the Czech Republic population and a 6.2 Mb segment of 31 windows (position 2.6–8.8 Mb; 8 windows with >30% SNPs with iHS >2 per window) was detected in the Hungarian population (Figures 1 and 2). These regions are represented by orange or red colour in Figure 2.

In the analyses presented above we used recombination rate data from a collared flycatcher linkage map. Very similar selection landscapes were observed in iHS selection scans using instead species-specific LD-based recombination map data (Kawakami et al., 2017; Supporting Information Figure S1). In the majority of populations recombination rate did not correlate with iHS statistics. In three populations the correlation was significant but very weak; $0.02 < r < 0.05$ (Öland collared flycatcher populations, Spanish pied flycatcher population and Atlas flycatcher).

The patterns of haplotype structure were predominantly population-specific (Figure 2), as might be expected for ongoing sweeps, with only some clusters of outlier windows partially shared among populations within species. As an example, a significant part of the large outlier cluster at position 3–9 Mb was shared between collared flycatchers from Czech Republic and Hungary. On average, collared flycatcher had twice as many outlier windows (24–67 per population) as pied flycatcher (13–58). However, since collared flycatchers showed much wider regions of extended haplotype homozygosity, both species displayed similar numbers of clusters per population

(means of 12.2 and 13.5, respectively). Atlas flycatcher had six and semicollared flycatcher 25 clusters of outlier windows.

nSL selection scans revealed similar patterns to that seen with iHS (correlation coefficient, r , ranged from 0.3 to 0.8 per population; all $p > 0.001$), meaning that the same genomic regions were typically detected as candidates for sweeps. Comparable numbers of regions having more than 10% of the SNPs with nSL >2 were identified (Table S2). However, the patterns observed with nSL were less pronounced compared to those with iHS, and fewer SNPs were associated with nSL values >3 than with iHS values >3. The number of outlier windows ranged from nine (semicollared flycatcher) to 31 (Czech Republic pied flycatchers). Despite the overall consistency between nSL and iHS, all species showed one or more genomic region/s with a significant signal in one of the statistics but essentially no signal in the other. In particular, in Atlas flycatcher, several narrow regions with a signal in nSL but not iHS could indicate differences in power to detect the locations of sweeps between statistics (Supporting Information Figure S2). The nSL results presented above were calculated based on values obtained with default window size of 200. The nSL results obtained with window size 500 and 1,000 were both highly correlated with the results obtained with default settings (all $r \geq 0.95$, all $p > 0.001$).

Several population genetic statistics that are not based on haplotype data can also be informative about recent selection. However, distributions of nucleotide diversity (π), Tajima's D, and Fay and Wu's H were rather uniform along the Z chromosome (Figure 3). Only some narrow reductions of the statistics were clearly visible (Figure 3). For example, π was reduced in all populations at positions 32 and 62 Mb. Several narrow dips of H were identified at 18–23 Mb in semicollared flycatcher. In general, there was little or no correlation between iHS or nSL and any of the population genetic statistics (Table S3). A very similar pattern was observed for correlations between nSL and the population genetic statistics.

3.2 | XP-EHH

Outlier regions identified by the XP-EHH scans were much smaller than those found by the other haplotype statistics (Figure 4; Table 2). We found 20 outlier windows in collared flycatcher populations, 11 in pied flycatcher populations and eight in the between-species comparisons. Most outlier windows (75%) were population-specific. The majority of signals of selection in collared flycatcher were found in the Öland population (73% of population-specific outlier windows), while we found no outlier windows in the Italian population and only a few in the Czech Republic and Hungarian populations. In pied flycatcher six (out of 11) outlier windows were population-specific, five of which were found in the Spanish population. Most XP-EHH outlier windows did not cluster and we only found two instances of subsequent outlier windows (three-window clusters in the Öland collared flycatcher and the Spanish pied flycatcher populations, respectively). The outlier windows were almost exclusively species-specific, with only one outlier window shared between pied flycatcher and collared flycatcher. With more strict thresholds (>15%

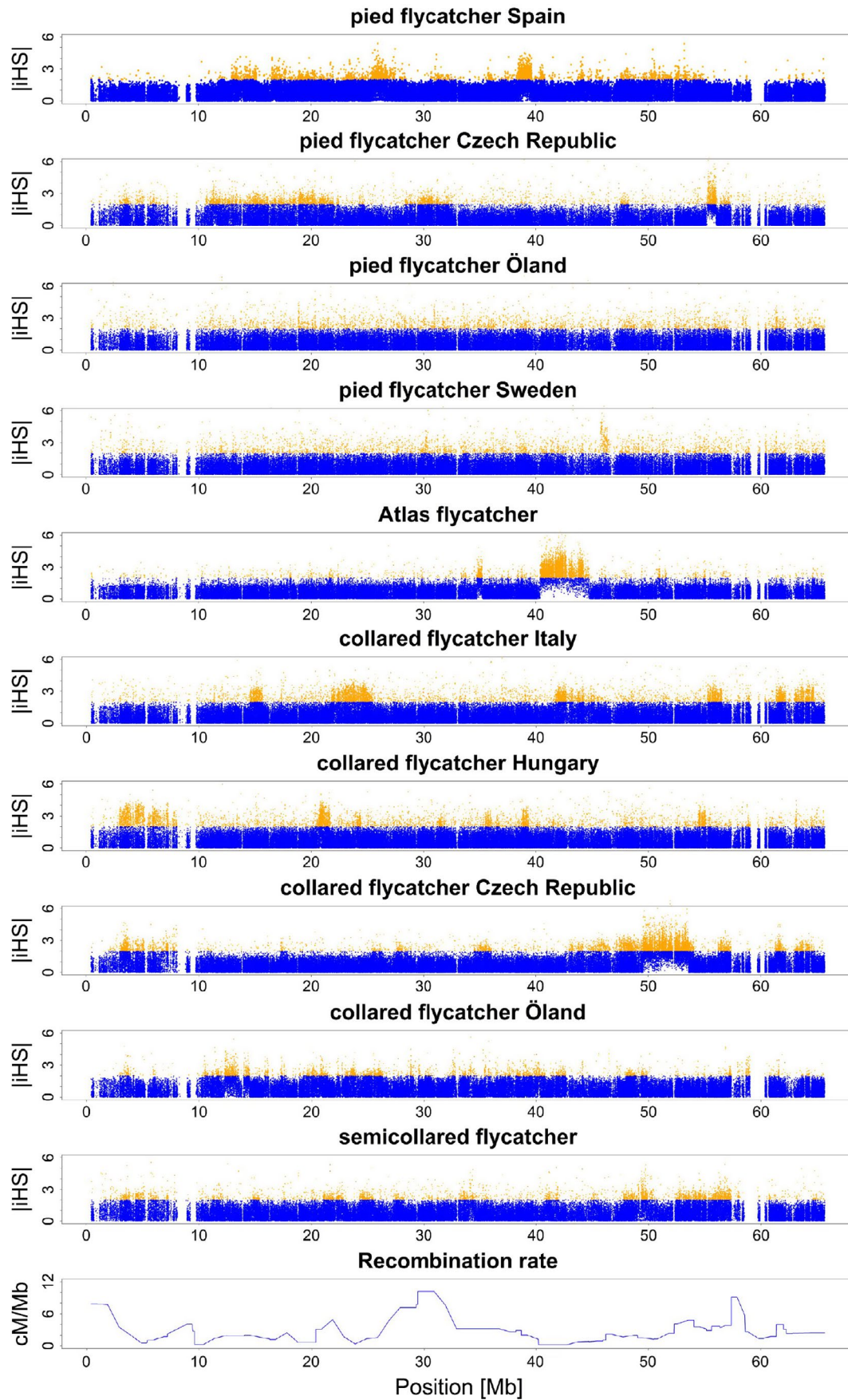
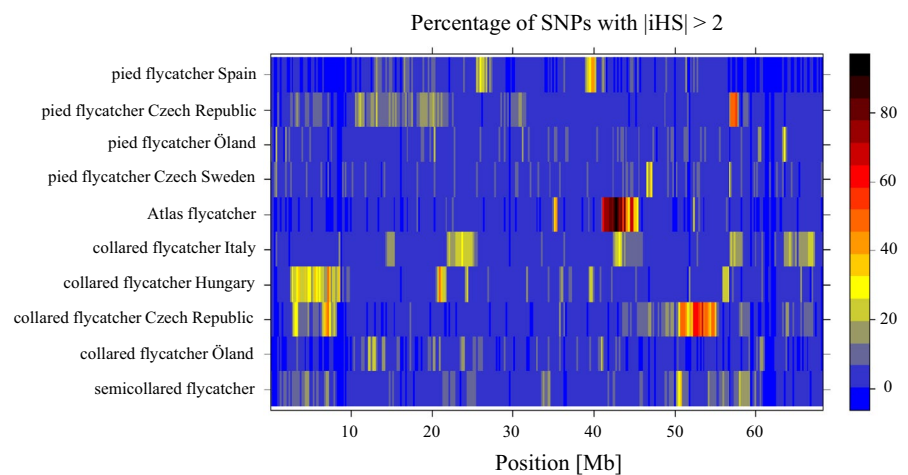


FIGURE 1 Absolute iHS values plotted along the Z chromosome for 10 flycatcher populations. Each dot corresponds to a single SNP. SNPs with $|iHS| > 2$ are shown in yellow [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Statistics for iHS outlier windows. A window was considered to be an outlier if at least 10% of SNPs had $|iHS| > 2$ (equivalent to Z-score equal 2).

Species	Population	Number of windows	Percentage of windows	Number of clusters	Size of the biggest cluster (Mb)
Collared flycatcher	Öland	24	7	14	1.2
	Czech Republic	67	20	17	5
	Hungary	52	15	9	6.2
	Italy	67	20	9	3.6
	Average	52.5	15	12.2	4
Pied flycatcher	Sweden	13	4	9	0.8
	Öland	14	4	11	0.4
	Czech Republic	58	17	23	2.2
	Spain	24	7	11	1.4
	Average	27.25	8	13.5	1.2
Atlas flycatcher		30	9	6	4.4
Semicollared flycatcher		47	14	25	1.6

FIGURE 2 Heat map representing patterns of haplotype structure on Z chromosome among flycatcher populations. Windows are coloured according to the colour bar scale showing the percentage of SNPs with $|iHS| > 2$ [Colour figure can be viewed at wileyonlinelibrary.com]

of SNPs having $XP-EHH > 2$ or < -2), only seven outlier windows were found in both pied and collared flycatcher but the most pronounced population-specific signals were still recovered (Öland collared flycatcher and Spanish pied flycatcher). We found significant correlations between $XP-EHH$ values and Tajima's D as well as Fay and Wu's H (Table S4). There was no correlation between $XP-EHH$ and nucleotide diversity.

3.3 | The influence of statistical phasing on haplotype-based statistics

Selection scans can be facilitated by using full chromosome haplotypes. Moreover, data sets that include full chromosomes obtained without statistical phasing can enable an evaluation of how well statistically inferred haplotype data would capture signatures of selection in the same population. To this end, we performed iHS scans on statistically phased sequence data obtained from the Z chromosome of male flycatchers. We first performed one scan based on 20 haplotypes inferred from 10 males (the same number of individuals as in the female data set) and one scan based on 10 randomly sampled

male haplotypes (the same number of haplotypes as in the female data set). The strongly extended haplotype homozygosities on the Z chromosome observed in females were not recovered in any of the two male data sets (Figure 5 and Supporting Information Figure S3). While the number of outlier windows was in fact larger in the two male data sets than in the female data, outlier windows were much more scattered along the chromosome without distinct clustering. There was almost no overlap with the outlier windows identified in females (Table S5). This indicates that most signals of selection detected with statistically phased data from male samples were likely false positives.

Since phasing 10 individuals constitutes a lower limit for phasing algorithms (set by the software used), we phased all males from all collared flycatcher populations together (92 individuals). Again, the patterns observed in female data were not recovered (Figure 6). Then, we used a separate data set of 80 collared flycatcher males from Öland (Kardos et al., 2016), phased them and compared the iHS patterns with those observed using female collared flycatchers from Öland (Figure 7). Once again haplotype clusters characterised in female data were not detected in this male data set. In addition, in both of these larger male

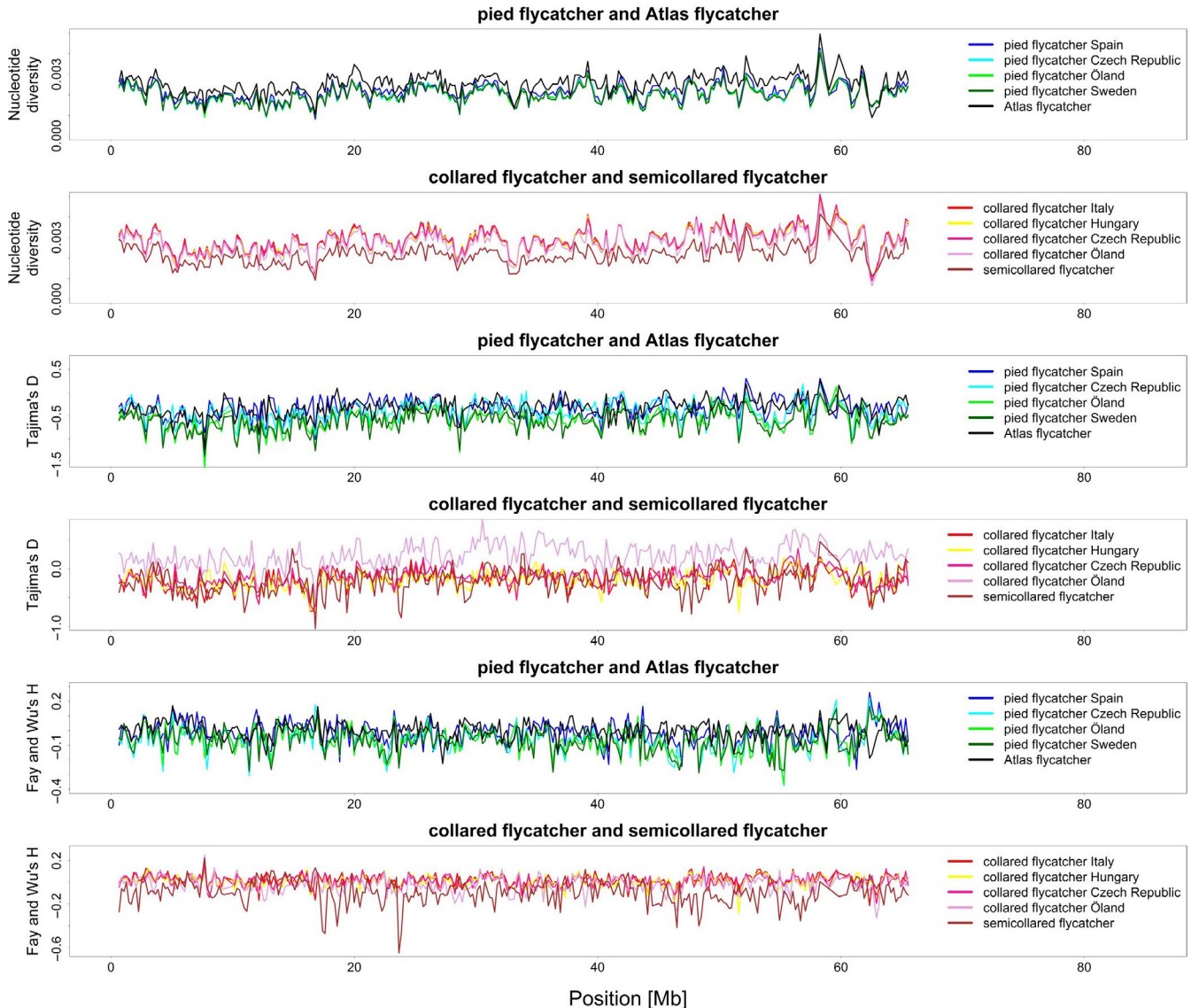


FIGURE 3 Nucleotide diversity, Tajima's D and Fay and Wu's H statistics plotted along Z chromosome for 10 flycatcher populations [Colour figure can be viewed at wileyonlinelibrary.com]

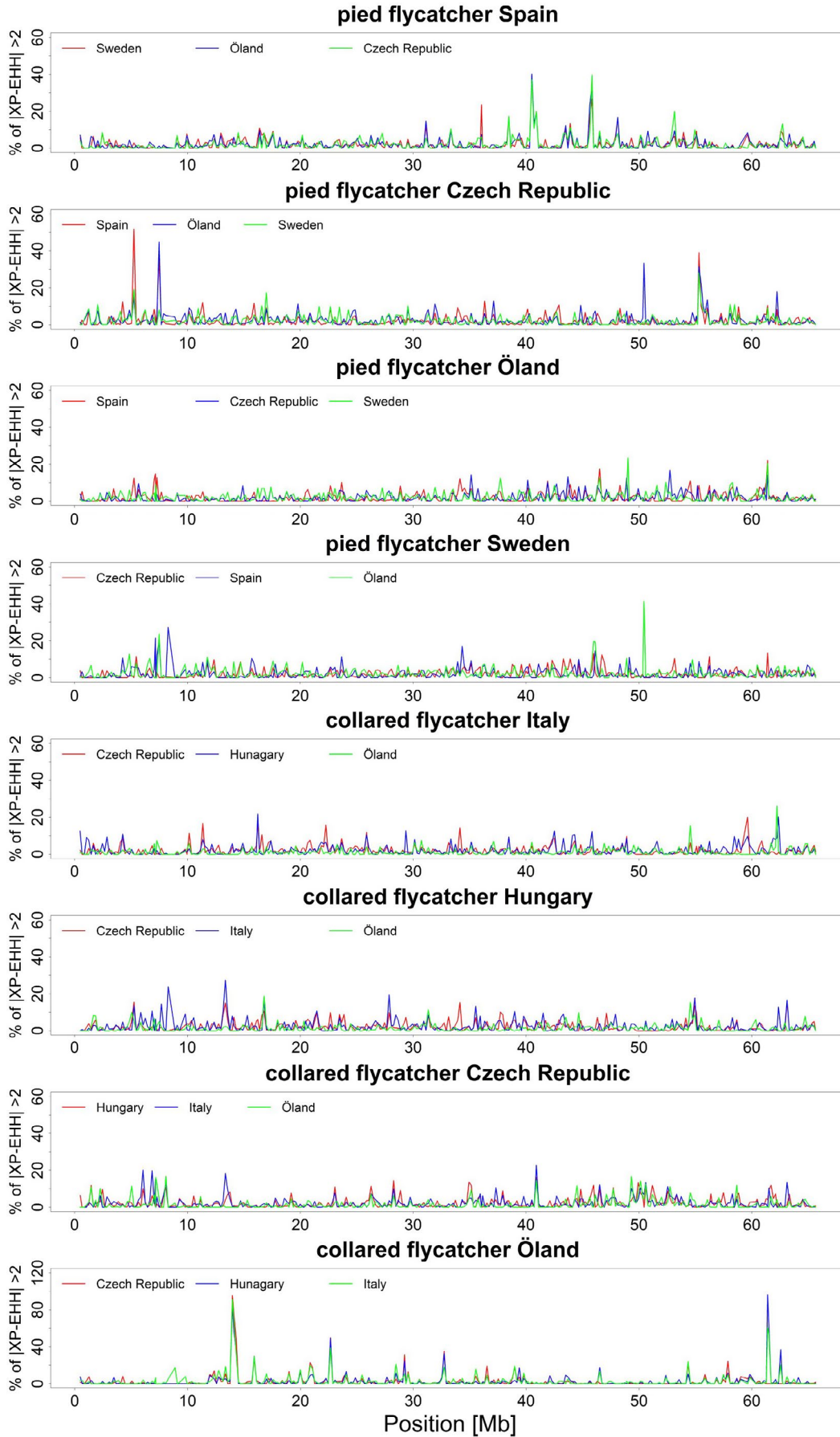
data sets we identified several regions of haplotype structure that were not observed with female data, indicating false positives (Figure 7). Interestingly, when performing statistical phasing of males and females together from each population, the iHS patterns observed in the female data sets were mostly recovered (Supporting Information Figure S4).

4 | DISCUSSION

Haplotype data and statistics include important information about allelic associations that make population genetic analyses more powerful (or at all possible) than when based on genotypes and/

or allele frequencies. This holds for several applications (Browning & Browning, 2011) including genotype imputation (Li, Willer, Ding, Scheet, & Abecasis, 2010; Marchini & Howie, 2010; McCarthy et al., 2016), demographic inference (Hellenthal et al., 2014; Schiffels & Durbin, 2014), and selection scans (Amorim et al., 2017; Sabeti et al., 2002). Statistical phasing constitutes the most common way of obtaining haplotype data. At the same time, statistical phasing is in itself one of the major sources of error and uncertainty in such data (Browning & Browning, 2011). For example, switching errors break up true haplotype blocks and result in underestimation of the number of coalescent events in the recent past. Such errors also lower linkage disequilibrium signals and give rise to false associations of

FIGURE 4 XP-EHH scan results summarized as percentage of SNPs with $|XP-EHH| > 2$ in 200 kb windows. XP-EHH statistic was calculated for each population in pairwise comparison to other conspecific populations. Different comparisons are presented with different colours. Note that the Y-axis for the Öland population has a different scale [Colour figure can be viewed at wileyonlinelibrary.com]



Species	Population	Total number of outlier windows	Number of population-specific outlier windows
Collared flycatcher	Öland	15	11
	Czech Republic	7	3
	Hungary	4	1
	Italy	0	0
	All	20	15
Pied flycatcher	Sweden	4	0
	Öland	3	0
	Czech Republic	5	1
	Spain	6	5
	All	11	6
Atlas flycatcher		4	0
Semicollared flycatcher		5	2

TABLE 2 Statistics for XP-EHH outlier windows. A window was considered to be an outlier if at least 10% of SNPs had a XP-EHH > 2 or XP-EHH < -2 (equivalent to Z-score equal 2)

alleles. This will affect downstream analyses, like, for example, biasing inference of recent N_e in methods that are based on phased genomes (Schiffels & Durbin, 2014). Here we circumvented the challenges associated with statistical phasing in a selective sweep scan by taking advantage of the fact that genomic resequencing will provide the full haplotype of the X/Z sex chromosome when individuals of the heterogametic sex are analysed.

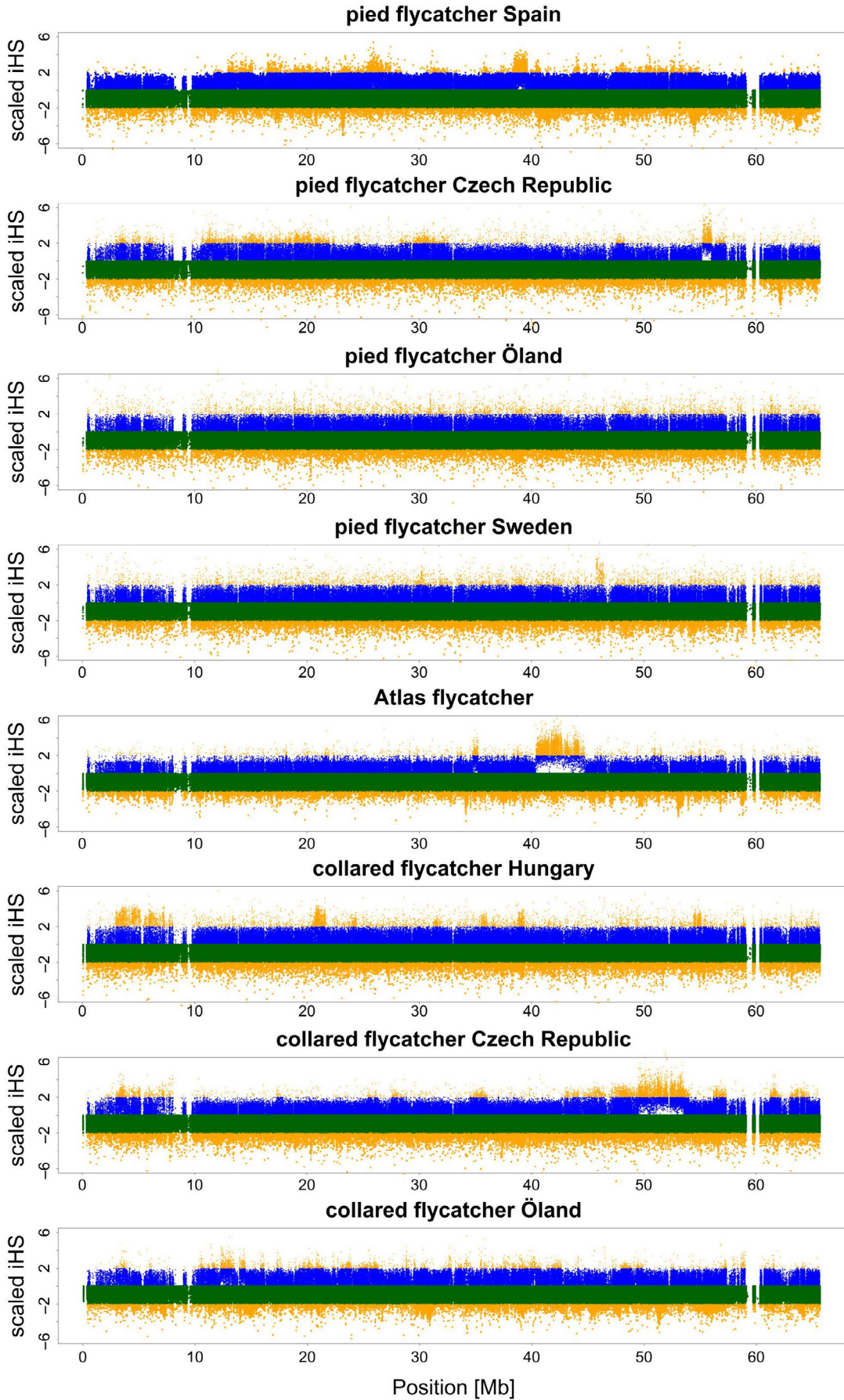
4.1 | Signals of positive selection on the flycatcher Z chromosome

Haplotype-based statistics suggested that the flycatcher Z chromosome is a frequent target for positive selection. The aim of this study was not to identify particular genes subject to selection but rather to provide an overall picture of the landscape of selective sweeps. We found several signals of recently completed sweeps (XP-EHH), both in comparisons between species and between different populations of collared flycatchers and pied flycatchers. The signals were narrow, essentially only single 200 kb windows, and commonly population-/species-specific. The fact that we did not observe shared signals between species is not surprising given that the species are well differentiated, with limited gene flow (Burri et al., 2015; Nater et al., 2015). On the other hand, given low within-species differentiation of flycatcher populations, we may not necessarily have expected population-specific signals. However, such pattern can likely be explained by the demographic history of different populations. Most population-specific selections signals in collared flycatcher were found in the recently colonised and likely bottlenecked Öland island population. Similarly, most signals in pied flycatcher were observed in the most divergent and isolated Spanish population.

We also found many signals of ongoing sweeps (iHS and nSL). Like for completed sweeps they were often population-specific, but, in contrast to the former category, outlier windows clustered and formed extensive haplotype blocks, some of which reached several Mb. This difference can be explained by that after a complete sweep, or at the end of an ongoing sweep, recombination will start to break up the association between selected and linked alleles, and thereby reduce LD and narrow the signal of selection. Additionally, some populations showed more outlier clusters than others, which can be a result of local differences in selection pressure.

We observed a large variance in the size of selection signals indicated by iHS and nSL scans, ranging from 200 kb to several Mb. Such large variance can of course be explained by differences in selection pressure but the difference in strength of selection among selected regions would then have to be very large (Kaplan, Hudson, & Langly, 1989). Thus, the observed pattern is most likely a result of a combination of several factors. For example, chromosomal inversions suppress recombination in heterokaryotypic individuals and can produce strong signals of haplotype structure at the time the inversion segregates in the population. Although the overall karyotype is remarkably similar among avian lineages (Ellegren, 2010), chromosomal inversions are far more common than interchromosomal rearrangements and have been shown to produce distinct phenotypes or to be associated with variation in fitness in several bird species (Huynh, Maney, & Thomas, 2011; Knief et al., 2016; Küpper et al., 2015; Lamichhaney et al., 2015). Importantly, inversions seem to be more common on the avian Z chromosome than on autosomes (Hooper & Price, 2015). This was clearly illustrated in a comparison of chromosomal organisation in collared flycatcher and zebra finch (Kawakami et al., 2014). It may be that at least some of the haplotype blocks identified in

FIGURE 5 Scaled iHS values plotted along the Z chromosome for females (upper part of each panel) and males (phased using 10 males; lower part of each panel) in 10 flycatcher populations. Each dot corresponds to a single SNP. SNPs with iHS > 0 representing female-based |iHS| and iHS < 0 representing male-based |iHS| × (-1). SNPs with |iHS| > 2 are shown in yellow [Colour figure can be viewed at wileyonlinelibrary.com]



our Z chromosome scans are associated with inverted regions. For instance, the distinct signal on positions 40.2–44.6 in Atlas flycatcher, with clear start and end points, could be a candidate. However, with the existing data at hand (short-read sequencing from small insert libraries) it is difficult, if not impossible, to detect inversion breakpoints.

Another possibility is that selection on multiple neighbouring targets may produce a signal of longer haplotypes in comparison to selection acting on a single locus (Chevin, Billiard, & Hospital, 2008; Hermisson & Pennings, 2017). Moreover, everything else being equal, the length of selected haplotypes depends on the recombination rate; selection acting in regions of low recombination will produce longer haplotypes.

Demographic events can influence the power to detect selection. Despite the fact that haplotype based statistics were shown to be relatively robust to complex demographic histories (Ferrer-Admetlla et al., 2014), we cannot exclude that population-/species-specific events influenced the overall pattern observed in the data. For example, the low number of haplotype clusters seen in the Öland population of collared flycatcher may be related to a bottleneck in connection with colonization of the island population. Overall, a difference in the effective population size between pied (lower N_e) and collared flycatcher (higher N_e) may have contributed to the differences in number of haplotype clusters observed in the two species. Additionally, the difference in short-term N_e among populations or species may have influenced the prevalence of hard and soft sweeps among populations (Hermisson & Pennings, 2017).

It has repeatedly been suggested that the X chromosome harbours a disproportionately large number of loci related to natural and sexual selection (Charlesworth, Coyne, & Barton, 1987; Coyne & Orr, 1989). The evidence for such large-X (or large-Z) effects comes from direct studies of hybrid incompatibilities in crosses (Jiggins et al., 2001; Presgraves, 2008), and indirectly from comparisons of relative rates of introgression on sex chromosomes and autosomes (Borge, Lindroos, Nádvořník, Syvänen, & Saetre, 2005; Carling, Lovette, & Brumfield, 2010; Elgvin et al., 2011; Payseur, Krenz, & Nachman, 2004; Starchova, Reif, & Nachman, 2009). More recently, genome scans for selection have indicated that the X chromosome of humans and other great apes may be a hot-spot for selective sweeps (Nam et al., 2015; Sankararaman et al., 2014). Our results from an avian system with female heterogamety provide evidence that the Z chromosome is subject to frequent sweeps. However, since we cannot do the same type of analysis (based on whole chromosome haplotypes) for autosomes, we cannot formally test whether sweeps occur at a higher rate on the Z chromosome than on autosomes (cf. Ellegren, 2009).

In light of the evidence for frequent sweeps documented in this study it may seem paradoxical that recent whole genome sequencing studies did not find clear signatures of recent selection on the

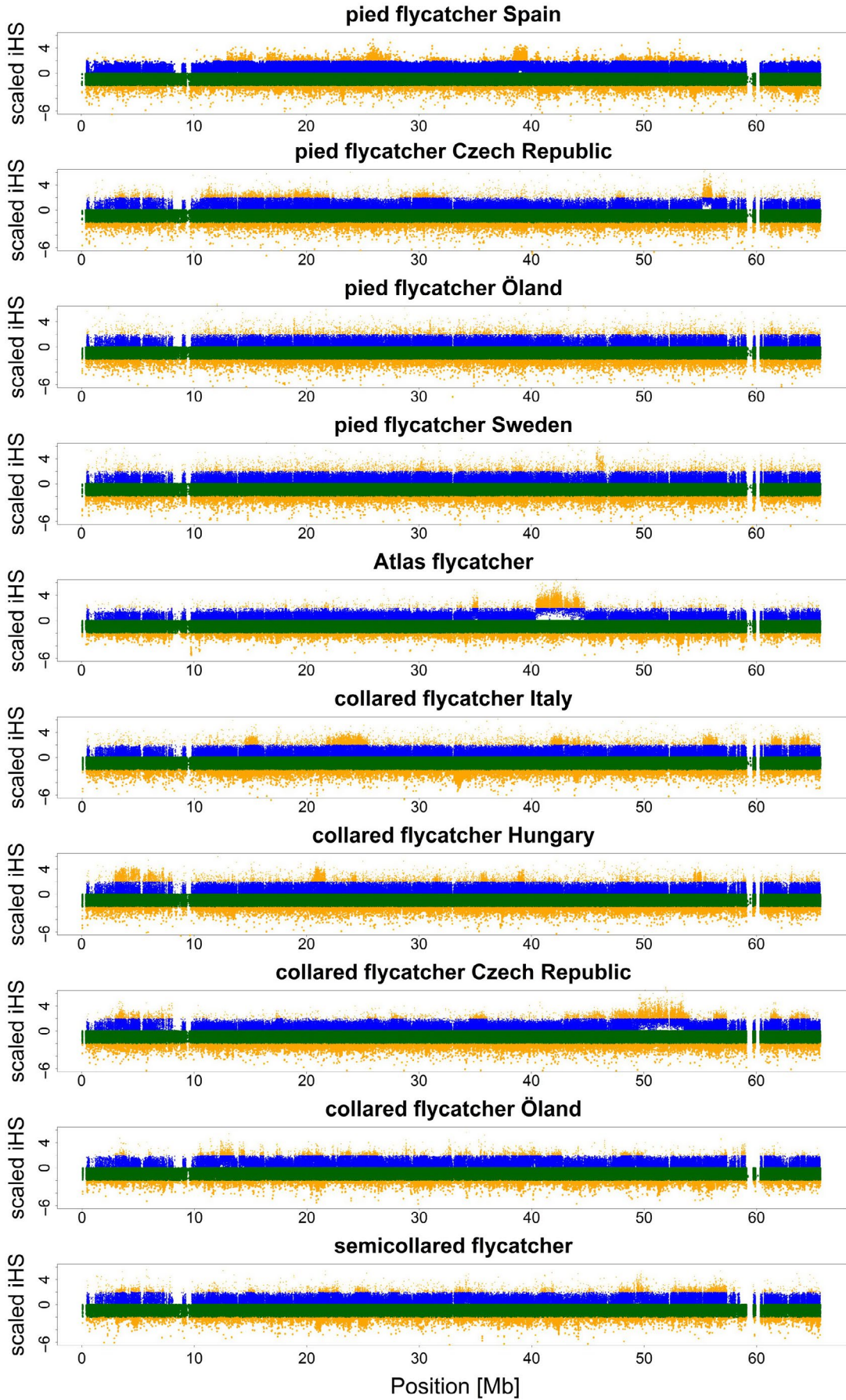
flycatcher Z chromosome, neither in the form of locally elevated F_{ST} in between-species comparisons or locally reduced nucleotide diversity within species (Burri et al., 2015; Ellegren et al., 2012). In contrast, most autosomes showed distinct F_{ST} peaks/nucleotide diversity valleys. There are several possible explanations to these seemingly contradictory observations. First, since N_e of the Z chromosome is smaller than that of autosomes, lineage sorting of the Z chromosome is faster. This manifests in much higher baseline levels of genetic differentiation on the Z chromosome (mean F_{ST} between collared flycatcher and pied flycatcher = 0.62) than on autosomes (0.35; Ellegren et al., 2012), which can potentially mask signals of elevated differentiation driven by selection. Second, it has been suggested that distinct F_{ST} peaks/nucleotide diversity valleys on autosomes are mainly driven by linked selection in regions of low recombination, with a strong role of background selection (e.g., Burri et al., 2015). Recombination rates across the flycatcher Z chromosome are more uniform than on autosomes (Burri et al., 2015; Ellegren et al., 2012). More generally, the much stronger signatures of selection on the Z chromosome found in this study may simply reflect the increased power provided by haplotype-based statistics in detecting selection.

It was somewhat surprising that there was almost no correlation between the Tajima's D and Fay and Wu H statistics, and iHS. However, different statistics vary in power to detect sweeps at different stages. Haplotype-based statistic have much higher power at low/intermediate frequency of the selected allele, while the strongest signals of Tajima's D and Fay and Wu's H are expected when a selected sweep is almost completed (Voight et al., 2006). Additionally, while signals from the site frequency spectrum are expected to last for approximately 0.1 N_e generations, signals based on haplotypes fade away quicker (approximately after 0.01 N_e generations; Przeworski, 2002; Pennings & Hermisson, 2006). It is also possible that Tajima's D and Fay and Wu's H are more affected by demography than haplotype-based statistics.

4.2 | Whole chromosome haplotypes vs. statistical phasing

An important part of this study was the comparison of haplotype-based selection scans using either statistically phased data or whole Z chromosome haplotypes. Sex chromosomes offer an ideal opportunity for such a test, since scans of the same chromosome, sampled from the same population, can be made both using statistically phased data (male samples) and whole chromosome haplotypes (female samples). The results of the comparison were very clear: iHS scans of males based on either the same number of males as females (10), or the same number of phased chromosomes of males and females (10), essentially failed to detect the distinct signals of selection seen in female scans. This is not surprising since

FIGURE 6 Scaled iHS values plotted along the Z chromosome for females and males (phased using 92 males). Each dot corresponds to a single SNP. SNPs with $iHS > 0$ representing female-based $|iHS|$ and $iHS < 0$ representing male-based $|iHS| \times (-1)$. SNPs with $|iHS| > 2$ are shown in yellow [Colour figure can be viewed at wileyonlinelibrary.com]



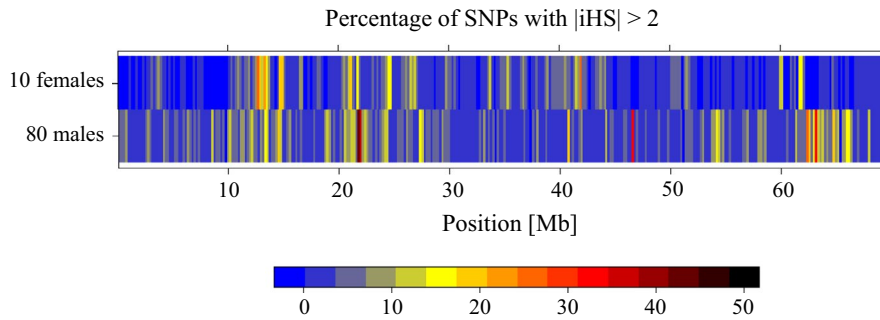


FIGURE 7 Heat map representing patterns of haplotype structure on Z chromosome in the Öland population of collared flycatchers using female and male data sets. Windows are coloured according to a colour bar scale showing the percentage of SNPs with $|iHS| > 2$ [Colour figure can be viewed at wileyonlinelibrary.com]

a sample size of 10 is far below recommended values for statistical phasing (Choi, Chan, Krikness, Telenti & Schork, 2018; Browning & Browning, 2011). However, the same was true when increasing the number of males that were statistically phased to 80 from an independent sample of collared flycatchers, or to 92 when pooling collared flycatchers from the different populations. These sample sizes are on the lower bound of recommendation for statistical phasing (Browning & Browning, 2011). Moreover, the scattered appearance of signatures of selection using male data not replicated in the analyses of whole Z chromosome haplotypes is indicative of a very high rate of false positives. The conclusion here was thus that inference of selection based on results from statistically phased haplotypes may both be misleading in terms providing spurious signals and by missing what are likely to be true signals.

An important implication of our observations is that they raise some concern about conclusions made in several recent studies of demography or in selection scans of natural populations where statistical phasing has been made using small sample size (Akagi et al., 2016; Gagnaire et al., 2013; McGirr, Martin, & Agashe, 2017; Velasco, Hough, Aradhya, & Ross-Ibarra, 2016; Wang, Street, Scofield, & Ingvarsson, 2016). Specifically, low power to detect signals of selective sweeps will underestimate the importance of adaptive evolution. In addition, signals that are detected may be false.

ACKNOWLEDGEMENTS

This work was funded by the Swedish Research Council and the Knut and Alice Wallenberg foundation. We thank members of the Ellegren laboratory for helpful discussions and two anonymous reviewers for valuable suggestions.


AUTHOR CONTRIBUTION

K.N.B. and H.E. conceived of the study and wrote the paper. K.N.B. performed all analyses. R.B. provided scripts.

DATA ACCESSIBILITY

The genome resequencing data are freely available in EMBL-EBI European Nucleotide Archive (<http://www.ebi.ac.uk/ena>) under accession number PRJEB7359. Phased male data sets are available on Dryad repository.

ORCID

Krystyna Nadachowska-Brzyska  <https://orcid.org/0000-0002-8457-310X>

Reto Burri  <https://orcid.org/0000-0002-1813-0079>

Hans Ellegren  <https://orcid.org/0000-0002-5035-1736>

REFERENCES

- Akagi, T., Hanada, T., Yaegaki, H., Gradziel, T. M., & Tao, R. (2016). Genome-wide view of genetic diversity reveals paths of selection and cultivar differentiation in peach domestication. *DNA Research*, 23, 271–282. <https://doi.org/10.1093/dnares/dsw014>
- Amorim, C., Nunes, K., Meyer, D., Comas, D., Cátira, M., Mauro, F., & Hünemeier, T. (2017). Genetic signature of natural selection in first Americans. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 2195–2199. <https://doi.org/10.1073/pnas.1620541114>
- Borge, T., Lindroos, K., Nádvořník, P., Syvänen, A., & Saetre, G. (2005). Amount of introgression in flycatcher hybrid zones reflects regional differences in pre and post-zygotic barriers to gene exchange. *Journal of Evolutionary Biology*, 18, 1416–1424. <https://doi.org/10.1111/j.1420-9101.2005.00964.x>
- Browning, S. R., & Browning, B. L. (2011). Haplotype phasing: Existing methods and new developments. *Nature Reviews Genetics*, 12, 703–714. <https://doi.org/10.1038/nrg3054>
- Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., ... Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. *Genome Research*, 25, 1656–1665. <https://doi.org/10.1101/gr.196485.115>
- Carling, M. D., Lovette, I. J., & Brumfield, R. T. (2010). Historical divergence and gene flow: Coalescent analyses of mitochondrial, autosomal and sex-linked loci in Passerina buntings. *Evolution; International Journal of Organic Evolution*, 64, 1762–1772. <https://doi.org/10.1111/j.1558-5646.2010.00954.x>
- Charlesworth, B., Coyne, J. A., & Barton, N. H. (1987). The relative rates of evolution of sex chromosomes and autosomes. *The American Naturalist*, 130, 113–146. <https://doi.org/10.1086/284701>
- Chevin, L. M., Billiard, S., & Hospital, F. (2008). Hitchhiking both ways: Effect of two interfering selective sweeps on linked neutral variation. *Genetics*, 180, 301–316. <https://doi.org/10.1534/genetics.108.089706>
- Choi, Y., Chan, A. P., Krikness, E., Telenti, A., & Schork, N. J. (2018). Comparison of phasing strategies for whole human genomes. *Plos Genetics*, 14, e1007308. <https://doi.org/10.1371/journal.pgen.1007308>
- Coyne, J. A., & Orr, H. A. (1989). Patterns of speciation in Drosophila. *Evolution*, 43, 362–381. <https://doi.org/10.2307/2410984>
- Cutter, A., & Payseur, B. A. (2013). Genomic signatures of selection at linked sites: Unifying the disparity among species. *Nature Review Genetics*, 14, 262–274. <https://doi.org/10.1038/nrg3425>

- Delaneau, O., Zagury, J., & Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. *Nature Methods*, *10*, 9–10. <https://doi.org/10.1038/nmeth.2307>
- Didion, J. P., Morgan, A. P., Yadgary, L., Bell, T. A., McMullan, R. C., de Solorzano, L. O., ... Weinstock, G. M. (2016). R2d2 drives selfish sweeps in the house mouse. *Molecular Biology and Evolution*, *33*, 1381–1395. <https://doi.org/10.1093/molbev/msw036>
- Dutoit, L., Burri, R., Nater, A., Mugal, C. F., & Ellegren, H. (2017). Genomic distribution and estimation of nucleotide diversity in natural populations: Perspectives from the collared flycatcher (*Ficedula albicollis*) genome. *Molecular Ecology Resources*, *17*, 586–597. <https://doi.org/10.1111/1755-0998.12602>
- 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 B: Biological Sciences*, *284*, 20162756. <https://doi.org/10.1098/rspb.2016.2756>
- Elgvin, T. O., Hermansen, J. O. S., Fijarczyk, A., Bonnet, T., Borge, T., Saether, S. A., ... Sætre, G.-P. (2011). Hybrid speciation in sparrows II: A role for sex chromosomes? *Molecular Ecology*, *20*, 3823–3837. <https://doi.org/10.1111/j.1365-294X.2011.05182.x>
- Ellegren, H. (2009). Genomic evidence for a large-Z effect. *Proceedings of the Royal Society B: Biological Sciences*, *276*, 361–366. <https://doi.org/10.1098/rspb.2008.1135>
- Ellegren, H. (2010). Evolutionary stasis: The stable chromosomes of birds. *Trends in Ecology and Evolution*, *25*(5), 283–291. <https://doi.org/10.1016/j.tree.2009.12.004>
- Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., ... Wolf, J. B. W. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, *491*, 756–760. <https://doi.org/10.1038/nature11584>
- Fariello, M., Boitard, S., Naya, H., Sancristobal, M., & Servin, B. (2013). Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*, *193*, 929–941. <https://doi.org/10.1534/genetics.112.147231>
- Fay, J. C., & Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics*, *155*, 1405–1413.
- Ferrer-Admetlla, A., Liang, M., Korneliussen, T., & Nielsen, R. (2014). On detecting incomplete soft or hard selective sweeps using haplotype structure. *Molecular Biology and Evolution*, *31*, 1275–1291. <https://doi.org/10.1093/molbev/msu077>
- Gagnaire, P. A., Pavé, S. A., Normandeau, E., & Bernatchez, L. (2013). The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by rad sequencing. *Evolution*, *67*, 2483–2497. <https://doi.org/10.1111/evo.12075>
- Hellenthal, G., Busby, G. B. J., Band, G., Wilson, J. F., Capelli, C., Falush, D., & Myers, S. (2014). A genetic Atlas of human. *Science*, *343*, 747–751. <https://doi.org/10.1126/science.1243518>
- Hermisson, J., & Pennings, P. S. (2005). Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. *Genetics*, *169*, 2335–2352. <https://doi.org/10.1534/genetics.104.036947>
- Hermisson, J., & Pennings, P. S. (2017). Soft sweeps and beyond: Understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods in Ecology and Evolution*, *8*, 700–716. <https://doi.org/10.1111/2041-210X.12808>
- Hooper, D. M., & Price, T. D. (2015). Rates of karyotypic evolution in Estrildid finches differ between island and continental clades. *Evolution*, *69*, 890–903. <https://doi.org/10.1111/evo.12633>
- Huynh, L. Y., Maney, D. L., & Thomas, J. W. (2011). Chromosome-wide linkage disequilibrium caused by an inversion polymorphism in the white-throated sparrow (*Zonotrichia albicollis*). *Heredity*, *106*, 537–546. <https://doi.org/10.1038/hdy.2010.85>
- Jiggins, C. D., Linares, M., Naisbit, R. E., Salazar, C., Yang, Z. H., & Mallet, J. (2001). Sex-linked hybrid sterility in a butterfly. *Evolution*, *55*, 1631–1638. [https://doi.org/10.1554/0014-3820\(2001\)055\[1631:slhsia\]2.0.co;2](https://doi.org/10.1554/0014-3820(2001)055[1631:slhsia]2.0.co;2)
- Kaplan, N. L., Hudson, R. R., & Langley, C. H. (1989). The “hitchhiking effect” revisited. *Genetics*, *123*, 887–899.
- Kardos, M., Husby, A., McFarlane, S. E., Qvarnström, 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*, 727–741. <https://doi.org/10.1111/1755-0998.12498>
- Kardos, M., Qvarnström, A., & Ellegren, H. (2017). Inferring individual inbreeding and demographic history from segments of identity by descent in *Ficedula* flycatcher genome sequences. *Genetics*, *205*, 1319–1334. <https://doi.org/10.1534/genetics.116.198861>
- Kawakami, T., Mugal, C. F., Suh, A., Nater, A., Burri, R., Smeds, L., & Ellegren, H. (2017). Whole-genome patterns of linkage disequilibrium across flycatcher populations clarify the causes and consequences of fine-scale recombination rate variation in birds. *Molecular Ecology*, *26*, 4158–4172. <https://doi.org/10.1111/mec.14197>
- 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*, 4035–4058. <https://doi.org/10.1111/mec.12810>
- Kim, Y., & Stephan, W. (2002). Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, *160*, 765–777.
- Knief, U., Hemmrich-Stanisak, G., Wittig, M., Franke, A., Griffith, S. C., Kempnaers, B., & Forstmeier, W. (2016). Fitness consequences of polymorphic inversions in the zebra finch genome. *Genome Biology*, *17*, 1–22. <https://doi.org/10.1186/s13059-016-1056-3>
- Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, *15*, 1–13. <https://doi.org/10.1186/s12859-014-0356-4>
- Küpper, C., Stocks, M., Risse, J. E., dos Remedios, N., Farrell, L. L., McRae, S. B., ... Burke, T. (2015). A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics*, *48*, 79–83. <https://doi.org/10.1038/ng.3443>
- Lamichane, S., Fan, G., Widemo, F., Gunnarsson, U., Thalmann, D. S., Hoepfner, M. P., ... Andersson, L. (2015). Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nature Genetics*, *48*, 84–88. <https://doi.org/10.1038/ng.3430>
- Li, Y., Willer, C. J., Ding, J., Scheet, P., & Abecasis, G. R. (2010). MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic Epidemiology*, *34*, 816–834. <https://doi.org/10.1002/gepi.20533>
- Marchini, J., & Howie, B. (2010). Genotype imputation for genome-wide association studies. *Nature Reviews Genetics*, *11*(7), 499–511. <https://doi.org/10.1038/nrg2796>
- McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., ... Marchini, J. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*, *48*, 1279–1283. <https://doi.org/10.1038/ng.3643>
- McGirr, J. A., Martin, C. H., & Agashe, D. (2017). Novel candidate genes underlying extreme trophic specialization in caribbean pupfishes. *Molecular Biology and Evolution*, *34*, 873–888. <https://doi.org/10.1093/molbev/msw286>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–1303. <https://doi.org/10.1101/gr.107524.110.20>
- Meisel, R. P., & Connallon, T. (2013). The faster-X effect: Integrating theory and data. *Trends in Genetics*, *29*, 537–544. <https://doi.org/10.1016/j.tig.2013.05.009>

- 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, 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, 1058–1072. <https://doi.org/10.1111/mec.13540>
- Nam, K., Munch, K., Hobolth, A., Dutheil, J. Y., Veeramah, K. R., Woerner, A. E., ... Schierup, M. H. (2015). Extreme selective sweeps independently targeted the X chromosomes of the great apes. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 201419306. <https://doi.org/10.1073/pnas.1419306112>
- 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, 1000–1017. <https://doi.org/10.1093/sysbio/syv045>
- Nelson, R. M., Wallberg, A., Simões, Z. L. P., Lawson, D. J., & Webster, M. T. (2017). Genomewide analysis of admixture and adaptation in the Africanized honeybee. *Molecular Ecology*, 26, 3603–3617. <https://doi.org/10.1111/mec.14122>
- Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, 39, 197–218. <https://doi.org/10.1146/annurev.genet.39.073003.112420>
- Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., & Clark, A. G. (2007). Recent and ongoing selection in the human genome. *Nature Reviews Genetics*, 8, 857–868. <https://doi.org/10.1038/nrg2187>
- Payseur, B. A., Krenz, J. G., & Nachman, M. W. (2004). Differential patterns of introgression across the X chromosome in hybrid zone between two species of house mice. *Evolution*, 58, 2064–2078. <https://doi.org/10.1554/03-738>
- Pennings, P. S., & Hermisson, J. (2006). Soft sweeps III: The signature of positive selection from recurrent mutation. *PLoS Genetics*, 2, 1998–2012. <https://doi.org/10.1371/journal.pgen.0020186>
- Presgraves, D. C. (2008). Sex chromosomes and speciation in *Drosophila*. *Trends in Genetics*, 24, 336–343. <https://doi.org/10.1016/j.tig.2008.04.007>
- Pritchard, J. K., & Di Rienzo, A. (2010). Adoption – Not by sooty alone. *Nature Reviews Genetics*, 11, 665–667. <https://doi.org/10.1038/nrg2880>
- Przeworski, M. (2002). The signature of positive selection at randomly chosen loci. *Genetics*, 160(3), 1179–1189. <https://doi.org/10.1554/05-273.1>
- Qvarnström, A., & Bailey, R. I. (2009). Speciation through evolution of sex-linked genes. *Heredity*, 102, 4–15. <https://doi.org/10.1038/hdy.2008.93>
- Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., ... Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419, 832–837. <https://doi.org/10.1038/nature01027.1>
- Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., ... Stewart, J. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449, 913–918. <https://doi.org/10.1038/nature06250>
- Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., ... Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, 507, 354–359. <https://doi.org/10.1038/nature12961>
- Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, 46(8), 919–925. <https://doi.org/10.1038/ng.3015>
- Schlamp, F., Van Der Made, J., Stambler, R., Chesebrough, L., Boyko, A. R., & Messer, P. W. (2016). Evaluating the performance of selection scans to detect selective sweeps in domestic dogs. *Molecular Ecology*, 25, 342–356. <https://doi.org/10.1111/mec.13485>
- Smeds, L., Kawakami, T., Burri, R., Bolivar, P., Husby, A., Qvarnström, A., ... Ellegren, H. (2014). Genomic identification and characterization of the pseudoautosomal region in highly differentiated avian sex chromosomes. *Nature Communications*, 5, 1–7. <https://doi.org/10.1038/ncomms6448>
- Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, 23, 23–35. <https://doi.org/10.1017/S0016672308009579>
- Starchova, R., Reif, J., & Nachman, M. W. (2009). Female heterogamety and speciation: Reduced introgression of the z chromosome between two species of nightingales. *Evolution*, 64, 456–471. <https://doi.org/10.1111/j.1558-5646.2009.00841.x>
- Szpiech, Z. A., & Hernandez, R. D. (2014). selscan: An efficient multi-threaded program to perform EHH-based scans for positive selection. *Molecular Biology and Evolution*, 31, 2824–2827. <https://doi.org/10.1093/molbev/msu211>
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123, 585–595.
- Teshima, K. M., Coop, G., & Przeworski, M. (2006). How reliable are empirical genomic scans for selective sweeps? *Genome Research*, 16(6), 702–712. <https://doi.org/10.1101/gr.5105206>
- Triska, P., Soares, P., Patin, E., Fernandes, V., Cerny, V., & Pereira, L. (2015). Extensive admixture and selective pressure across the Sahel Belt. *Genome Biology and Evolution*, 7, 3484–3495. <https://doi.org/10.1093/gbe/evv236>
- Vatsiou, A. I., Bazin, E., & Gaggiotti, O. E. (2016). Detection of selective sweeps in structured populations: A comparison of recent methods. *Molecular Ecology*, 25, 89–103. <https://doi.org/10.1111/mec.13360>
- Velasco, D., Hough, J., Aradhya, M., & Ross-Ibarra, J. (2016). Evolutionary genomics of peach and almond domestication. *G3 (Bethesda, Md.)*, 6(12), 3985–3993. <https://doi.org/10.1534/g3.116.032672>
- Vijay, N., Bossu, C. M., Poelstra, J. W., Weissensteiner, M. H., Suh, A., Kryukov, A. P., & Wolf, J. B. W. (2016). Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature Communications*, 7, 1–10. <https://doi.org/10.1038/ncomms13195>
- Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting natural selection in genomic data. *Annual Review of Genetics*, 47, 97–120. <https://doi.org/10.1146/annurev-genet-111212-133526>
- Voight, B. F., Kudravalli, S., Wen, X., & Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biology*, 4, 446–458. <https://doi.org/10.1371/journal.pbio.0040072>
- Wang, J., Street, N. R., Scofield, D. G., & Ingvarsson, P. K. (2016). Variation in linked selection and recombination drive genomic divergence during allopatric speciation of European and American aspens. *Molecular Biology and Evolution*, 33, 1754–1767. <https://doi.org/10.1093/molbev/msw051>

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: Nadachowska-Brzyska K, Burri R, Ellegren H. Footprints of adaptive evolution revealed by whole Z chromosomes haplotypes in flycatchers. *Mol Ecol*. 2019;28:2290–2304. <https://doi.org/10.1111/mec.15021>