Evidence for ancient selective sweeps followed by differentiation among three species of *Sphyrapicus* sapsuckers

Running title: Genomic divergence in sapsuckers

Libby Natola[1,2] and Darren Irwin[1,2]

[1]Department of Zoology, University of British Columbia, Vancouver, BC, Canada;

[2]Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada;

Corresponding Author: Libby Natola, libby.natola@gmail.com

Abstract

Speciation occurs when gene pools differentiate between populations, but that differentiation is heterogeneous across the genome. Understanding what parts of the genome are more prone to differentiation is a key mission in evolutionary biology because it can identify genomic regions and evolutionary processes central to the speciation process. Here, we study genomic variation among three closely related species of North American woodpecker: red-breasted, red-naped, and yellow-bellied sapsuckers. We construct a new sapsucker reference genome and use whole genome resequencing to measure genetic variation among these species and to quantify how the level of differentiation between these forms varies across the genome. We find that regions of high relative differentiation between species (F_{ST}) tend to have low absolute differentiation between species (π_B), indicating that regions of high relative differentiation often have more recent between-population coalescence times than regions of low relative differentiation do. Most of the high-F_{ST} genomic windows are found on the Z chromosome, indicating this sex chromosome is particularly important in sapsucker speciation. We show that LD is high across the Z chromosome and that each species tends to have a unique Z haploblock, a pattern suggestive of chromosomal inversions. These results are consistent with a model of speciation in which selective sweeps of globally advantageous variants spread among partly differentiated populations, followed by differential local adaptation of those same genomic regions. We propose that sapsucker speciation may occur primarily via this process occurring on the Z chromosomes, followed by epistatic interactions between Z haploblocks.

Introduction

The speciation process is responsible for the generation of all species diversity and is a major topic of study in evolutionary biology. Speciation occurs after some barrier reduces or precludes gene flow between populations, fracturing individuals into multiple isolated groups which then accumulate differences over time (Coyne & Orr 2004, Price 2008). Barriers may be caused by genetic incompatibilities (Huang et al. 2020), behavioral differences (Uy et al. 2018), or geographic barriers to dispersal (Nelson 2006); and they could have existed in the past (Weir & Schluter 2004) or be contemporary (Clark et al. 2010). Often, geographic separation acts as an initial cause of differentiation (Price 2008), allowing genetic and phenotypic differences to evolve that later cause reduced hybrid fitness if populations come into contact.

Many closely related groups in the intermediate stages of speciation today arose during the Pleistocene era, during which fluctuating periods of glaciation geographically isolated conspecific populations, initiating differentiation in what we describe today as species (Howell 1952, Schluter & McPhail 1992, Weir & Schluter 2004). In North American bird populations, for example, a commonly occurring biogeographic pattern is the following: Related forms are often geographically structured into a west coast form, an interior west form, and an eastern form, and these three types are usually now in contact in western Canada (Weir and Schluter 2004). This pattern is a result of the Pleistocene glaciations separating forests into three major southern refugia, which have now expanded into western Canada.

An example of this biogeographic pattern is provided by three species of *Sphyrapicus* woodpeckers: red-breasted, red-naped, and yellow-bellied sapsuckers (*Sphyrapicus ruber, S. nuchalis,* and *S. varius*). Studies conducted using allozymes (Johnson & Zink 1983); mitochondrial DNA loci including Cytochrome-b (Cicero & Johnson 1995), COI (Natola & Burg

2018, Seneviratne et al. 2012), CR (Natola & Burg 2018); a Z-linked locus CHD1Z (Natola & Burg 2018 Seneviratne et al. 2012); and Genotyping-by-Sequencing (GBS) of the nuclear genome (Curtis 2017, Grossen et al. 2016, Natola et al. 2021, Seneviratne et al. 2016) all concur that the three forms are distinct based on genetic differentiation. Molecular dating based on mitochondrial DNA estimates that red-breasted and red-naped sapsuckers diverged 0.32 million years ago (MYA), and their common ancestor diverged from yellow-bellied sapsuckers approximately 1.09 MYA (Weir & Schluter 2004), time frames over which there have been many cycles of glacial and interglacial periods (Menounos et al. 2009). These three species hybridize where they come into contact, even forming a three-species hybrid zone (Natola et al. 2022), but the limited number of hybrids in that zone show that the three forms are stable species. Given the evidence of some current gene flow (Natola et al. 2022) as well as the deep genomic differentiation, it is a safe assumption that these three groups have experienced multiple cycles of geographic separation and hybridization during their path towards full species. Hence these sapsuckers provide an opportunity to examine the differentiation between three forms with a history of cycles of separation and contact.

Genomic data are useful in speciation research because they reflect the evolutionary relationships between species, particularly historical gene flow and population connectivity. The "genomic landscape of differentiation," which describes the peaks and valleys of differentiation between species or populations along the genome, can help us identify regions that are differentiating the most (Burri et al. 2015, Ravinet et al. 2017). There are different ways to measure differentiation, and comparing their genomic landscapes of variation can point to different evolutionary processes. By comparing relative allele frequency differentiation between populations (F_{ST}), nucleotide distance between populations (π_B; often referred to as "D_{xy}"), and

nucleotide diversity within populations (π_W), we can test various differentiation models involving geographic separation, gene flow, and selection (Charlesworth 1998, Cruickshank & Hahn 2014, Irwin et al. 2016, 2018, Turner et al. 2005). Divergent selection acting upon a trait in different populations will cause greater allele frequency differentiation (i.e., F_{ST}) in the loci which encode that trait compared to the rest of the genome. These divergent loci might also cause low fitness of hybrids, creating a barrier between the species. When populations are in contact, such divergent loci will tend to have less gene flow between populations than other parts of the genome do, resulting in comparatively high coalescence time—which is expected to be proportional to nucleotide distance (π_B) when mutation rate is constant. This scenario leads to the expectation that regions of high F_{ST} will tend to also have high π_B (Cruickshank & Hahn 2014, Han et al. 2017).

In contrast to this expectation, some pairs of differentiated populations do not show high π_B in the high F_{ST} regions of the genome (Cruickshank & Hahn 2014). In these cases, high F_{ST} is a result of low within-population nucleotide diversity (π_W) rather than high between-population nucleotide distance. Proposed explanations for such a pattern center on selection (either positive or negative) in low-recombination regions of the genome resulting in especially low genetic diversity. In some cases, there is even a negative relationship between F_{ST} and π_B, a pattern for which two explanations have been proposed. One is that some regions of the genome have experienced selection-driven reductions in diversity in both the common ancestor and the two current populations (Cruickshank & Hahn 2014). Another is that adaptively advantageous mutations can sweep from one population to another across a hybrid zone, thereby lowering π_B, and then subsequent local adaptation can increase F_{ST} (Delmore et al. 2015, Irwin et al. 2016). This latter model is called “sweep-before-differentiation” (Irwin et al. 2016, 2018).

Differentiated yet hybridizing taxa such as the three species of *Sphyrapicus* sapsuckers provide excellent opportunities to examine the potential role of past gene flow in shaping the currently observed genomic landscapes of differentiation.

In addition to the influence of past gene flow and selection, genomic landscapes of differentiation are influenced by structural aspects of the genome (Noor et al. 2001, Payseur & Rieseberg 2016, Rieseberg 2001, Sætre et al. 1997), including sex chromosomes and inversions. In ZW sex chromosome systems such as birds (where females are the heterogametic sex), the Z chromosome often shows higher F_{ST} than the autosomes (Irwin 2018, Qvarnström & Bailey 2009). This is likely due to the smaller effective population size and reduced recombination, supplemented by greater potential for selective differentiation. Chromosomal inversions can also accumulate differentiation more rapidly than the rest of the genome, and they frequently act as isolating barriers (Hooper et al. 2018, Kirkpatrick & Barrett 2015): crossing over in the inverted region of individuals with hybrid karyotypes may cause aneuploid gametes, reducing hybrid fitness (Hooper & Price 2017, Kirkpatrick & Barton 2006). Lack of recombination makes inversions a likely refuge for alleles that contribute to local adaptation (Coyne and Orr 2004), as in dune and non-dune prairie sunflowers (*Helianthus petiolaris*) (Huang et al. 2020, Todesco et al. 2020).

Here, we use long-read sequencing to build a novel *Sphyrapicus* reference genome and use whole genome resequencing (WGS) to measure and examine the genomic landscapes of differentiation between the three hybridizing *Sphyrapicus* sapsucker species. We use these results to determine which parts of the genome have high differentiation and whether these regions tend to have higher, similar, or lower of nucleotide distances compared to the rest of the genome. The results from this analysis will provide insight into the role of gene flow and

selection during the speciation process in this three-species group that is biogeographically representative of many other taxa in North America.

Materials and Methods

Reference genome assembly

To create a *de novo* reference genome we submitted tissues of one female voucher red-breasted sapsucker specimen to the University of Delaware Sequencing & Genotyping Center for Pacific Biosciences Single Molecule Real Time (SMRT) Consensus Long Read (CLR) sequencing, which yielded 172 X coverage. We assembled the SMRT sequences in canu v. 1.9 (Koren et al. 2017), specifying a genome size of 1.3 Gb. The resulting assembly was 1.6 Gb, had an NG50 of 4.7 Mb, and an LG50 of 75. We purged haplotigs in the PurgeHaplotigs pipeline (Roach et al. 2018) and masked repeats in RepeatMasker v. 4.1.1 (Smit et al. 2013) for a resulting 1.28 Gb genome that was 95.6% complete (94.5% single-copy, 1.1% duplicated, 0.9% fragmented, and 3.5% missing) for conserved genes according to BUSCO v 5.1.2 (Manni et al. 2021). We used Satsuma (Grabherr et al. 2010) to combine contigs into scaffolds based on a golden-fronted woodpecker genome assembly (*Melanerpes aurifrons*, Wiley & Miller 2020).

Whole Genome Sequencing

Sample selection

Seventy-eight sapsucker samples were included in our WGS analysis: 10 allopatric individuals from each species, 15 hybrid zone individuals from each species cross, one putative 3-species hybrid, and two Williamson's Sapsuckers (*Sphyrapicus thyroideus*) included as outgroups (Figure 1). We selected hybrid zone individuals to represent a continuum of admixture

proportions between each species pair based on ancestry coefficient values from prior GBS research (L. Natola et al. 2021, 2022, Seneviratne et al. 2016). Samples were tissues from vouchered museum specimens and blood from wild-caught birds collected by Sampath Seneviratne (Seneviratne et al. 2012, 2016) and included both male and female birds (Appendix C.1).

DNA extraction, library preparation

We extracted DNA from all samples using a standard phenol-chloroform extraction procedure. We then quantified DNA concentrations using a Qubit fluorometer and submitted 100 ng DNA to Genome Quebec where technicians prepared shotgun DNA libraries using the NEB Ultra II (New England Biolabs) library preparation kit and then sequenced the libraries using four NovaSeq6000 S4 PE150 lanes.

Read processing

Sequencing resulted in 2.73 billion reads, with an average Phred quality score of 35. We first trimmed the reads using Trimmomatic v 0.38 (Bolger et al. 2014) and aligned reads to the Red-breasted Sapsucker reference genome using BWA v 0.7.17 (Li 2013). Next, we added read group headers, converted to bam format, sorted, and indexed our reads using Picard v 1.97 (http://broadinstitute.github.io/picard/). Using samtools v 1.12 (Li et al. 2009), we merged the single-end and paired-end .bam files and indexed. We used the bamqc function from Qualimap v 2.2.2a (Okonechnikov et al. 2016) to assess .bam file quality. We created a corrected coverage measure from the bamqc files by dividing each scaffold's read coverage per individual divided by that individual's overall mean coverage. With this, we were able to identify scaffolds likely to be on the Z chromosome if known males had approximately twice the normalized coverage

(normalized coverage = 1) as females (normalized coverage = 0.5), or on the W chromosome if known males had approximately zero normalized coverage compared to females' 0.5 X coverage (Appendix C.2). We renumbered all Z scaffolds as their original golden-fronted woodpecker mapped scaffold number + 1000 and all W scaffolds as their original scaffold + 2000 (i.e., all scaffolds on the Z are labelled as between 1000 and 1999 and all W scaffolds are between 2000 and 2999). We then marked PCR duplicates in Picard 1.97 before identifying problematic regions, realigning, and fixing mates using GATK v 3.8 (Mckenna et al. 2010) with scripts adapted from Aguillon et al. (2021). Finally, reads were filtered in GATK by removing those with QD < 2.0, FS > 40.0, MQ < 20.0, and haplotype score > 12.0 settings, then in VCFtools v 0.1.16 (Danecek et al. 2011) with the < 20%, minor allele frequency > 5%, 3 < depth ≤ 50, and biallelic filters. We also created a dataset with all variant and invariant sites sequenced in the WGS dataset per scaffold, using the GATK v 4.0 (Mckenna et al. 2010) HaplotypeCaller tool. Then, we split the data into individual scaffolds and filtered with the same GATK 3.8 filtration protocol as the previous dataset, and similar VCFtools filtration with the exceptions of no minor allele frequency or minimum alleles filters to preserve invariant sites. Finally, we used the GATK 3.8 GenotypeGVCFs tool to include all variant and invariant sites across all individuals into one VCF file for each scaffold.

Analyses

Species differentiation

To quantify mean genomic differentiation among the three species we calculated Weir & Cockerham's (Weir & Cockerham 1984) pairwise weighted average F_{ST} in VCFtools using only allopatric samples. We constructed PCAs of variation among individual genotypes using the R

(R Core Team) package SNPRelate (Zheng et al. 2012). We calculated pairwise F_{ST} for all SNPs in 50 kb windows using VCFtools, then plotted the windowed average F_{ST} values using the package qqman (Turner 2018) in R (R Core Team). We plotted F_{ST} (relative differentiation between populations), π_B (absolute differentiation between populations), and π_W (within population nucleotide diversity) across 50 kb sliding windows using allopatric individuals on all autosome and Z scaffolds > 50 kb; these together are referred to hereafter as the "genomic landscape dataset". The total number of 50 kb windows included in these analyses was 18,550. We created these with scripts from Irwin et al. (2018) adapted in the languages R (R Core Team 2021) and Julia (Bezanson et al. 2017). We separated autosomal and Z chromosome scaffolds and calculated F_{ST} using VCFtools. We calculated the differences between F_{ST} values on the autosomes versus Z chromosomes for each species pair using Welch's two-sample t-tests and plotted the 50kb window means for F_{ST} vs. π_B for each species pair, while indicating the differences between differentiation on autosomes vs. Z chromosomes. We calculated the adjusted Z diversity, $\pi_{W,Z}$*, as π_W /1.1 to account for the higher mutation rate on the Z chromosome (Irwin 2018), and then took the ratio of $\pi_{W,Z}$*/$\pi_{W,A}$ (π_W on the autosomes) to determine if diversity is comparable between Z chromosomes and autosomes. Mathematically, $\pi_{W,Z}$*/$\pi_{W,A}$ values < 0.5625 cannot be explained by even the most extreme asymmetries in variance of male and female reproductive success (e.g., harem or lek mating structures) (Charlesworth 2001, Irwin 2018).

Linkage Disequilibrium

We calculated Linkage Disequilibrium (LD) between within and between each species using the variant sites dataset in plink v 1.9 (Purcell et al. 2007). We used the ten allopatric

individuals per species for within species calculations, and all 20 allopatric samples plus 15 sympatric samples per species pair to calculate between species LD. We used plink to randomly thin scaffolds to fewer than 50,000 SNPs and made LD heatmap figures in R.

Genotype × individual plots

To view scaffold-wide haploblocks, we plotted genotypes of each individual using the fixed SNPs from each species pair per scaffold, and plotted these using R scripts adapted from Irwin et al. (2018). To visualize haploblocks, we extracted fixed SNPs between each species pair using VCFtools. We note that with this method SNPs are included in the three-way comparison that are fixed for differences between only two of the three species. Then, we thinned files to remove sites within 75kb of each other to make plots easier to read. Most of the haplotypes we wanted to visualize were on the Z chromosome, so we removed all females to avoid complexity created by hemizygosity. We repeated these steps for every species pair. To plot the haploblocks, we adapted R scripts from Irwin et al. (2018).

Results

Species differentiation

Our data present clear evidence for strong differentiation between the three species (Figure 2). The weighted average pairwise F_{ST} values calculated between each species confirm that red-breasted and red-naped sapsuckers are the most closely related pair, with an F_{ST} of 0.093. These relationships are reflected in the PCA (Figure 2), which distinguishes first the yellow-bellied split, then the red-naped/red-breasted split. PC1 explains considerably more variation (14.77%) than PC2 (3.17%).

The windowed F_{ST} plots show many windows with low F_{ST} values (< 0.1) between red-breasted and red-naped sapsuckers, and overall higher average F_{ST} values in the species pairs involving yellow-bellied sapsuckers (Figure 3). However, the scaffolds with the highest average pairwise F_{ST} values occur in large peaks that are largely concurrent in each species pair (Figure 3). The most obvious pattern in these plots is the extremely high F_{ST} peaks across most of the scaffolds on the Z chromosome. F_{ST} of all SNPs (9,205,747 sites) on autosomes was 0.28 between red-naped and yellow-bellied, 0.22 between red-breasted and yellow-bellied, and 0.08 between red-breasted and red-naped. On the Z scaffolds (655,683 sites), however, those measures were 0.47, 0.42, and 0.26, respectively. Whereas the autosomes contain 9,205,747 SNPs over 1.14 Gbp, only 8 of these are fixed between red-breasted and red-naped sapsuckers (8.7×10^{-5} % of all variant sites), 26,347 (0.29%) are fixed between red-naped and yellow-bellied, and 20,361 (0.22%) are fixed between red-breasted and yellow-bellied. On the Z chromosomes, however, which contain 655,683 polymorphic sites over 127 Mbp, 16,097 (2.5% of all variant sites), 66,906 (10.2%), and 64,873 (9.9%) are fixed between red-breasted/red-naped, red-naped/yellow-bellied, and red-breasted/yellow-bellied, respectively.

Figure 4 confirms the pattern of high F_{ST} on the Z chromosome scaffolds, and in addition shows high F_{ST} usually co-occurs with regions of low π_W (Figure 4). Generally, the mathematical relationship between F_{ST} and π_W leads to the expectation that F_{ST} will tend to increase as π_W decreases, and that with π_W held constant regions of high F_{ST} will have high π_B. In the sapsucker comparisons, however, genomic regions of high relative differentiation tend to have low π_B. Low π_B in high F_{ST} regions can occur if π_W is proportionally lower than π_B. Biologically, this can be explained by selective sweeps in a common ancestor (lowering π_B) and subsequent sweeps in the daughter species (lowering π_W and driving up F_{ST}) (Cruickshank & Hahn 2014), or a selective

sweep across hybridizing populations (lowering π_B) followed by local differentiation at these loci between populations (causing extremely low π_W, therefore increasing F_{ST}) (Irwin et al. 2016, 2018).

Statistical tests indicate that F_{ST} values among species are indeed much higher on the Z chromosome scaffolds than the autosomes, and that within group differentiation is lower on Z chromosomes (Welch's two-sample t-tests of the genomic landscape dataset including autosome and Z chromosome scaffolds > 50 kb: all P-values < 0.0001; Table 1). In all species comparisons the average F_{ST} among Z chromosome windows was significantly higher than in windows on autosomes, and π_B was significantly lower on the Z than the autosomes (Table 1). These data underscore the role of recurrent selection upon the Z chromosome. Indeed, in all three species, the ratio of adjusted $\pi_{W,Z}$ to $\pi_{W,A}$ was below the theoretical minimum of 0.5625 that can be explained without invoking selection. These differences are illustrated in Figure 5, in which there are many fewer colored dots (Z chromosome windows) than gray dots (autosomal windows) that are high on the y axis. Within species, π_W is consistently significantly lower on the Z chromosome than on autosomes (Table 1). This is shown in the neighbour-joining trees of phylogenetic relationships between sapsuckers in Figure 6, which show shallower branching and proportionally shorter grey boxes on the Z than the autosomes. This indicates the high F_{ST} regions have more recent coalescence times than the low- F_{ST} regions.

Linkage disequilibrium

The linkage disequilibrium plots follow expected patterns for most scaffolds, in which LD is highest along the diagonal edge of the plot, indicating high LD only among loci that are nearest to each other (Figure 7). This pattern holds both within and between species in most

scaffolds. Several scaffolds, however, show aberrant patterns, wherein large regions of the scaffold are in high LD, even regions that are distantly located on the scaffold. Many phenomena can cause these patterns, including structural variation, genomic incompatibilities, or mis-assemblies.

Genotype × individual plots

As shown in Figures 3 and 4, there are strong differences in allele frequencies between the species, and the genotype × individual plots show hybrids have runs of heterozygosity extending across large linked regions (Figure 8). The alternate states of these linked regions can be considered "haploblocks". The genotype × individual plots show that on the autosomes, even regions with stretches of high LD tend to have signs of recombination (Figure 8). However, the Z scaffolds show long stretches of haploblocks that recombine only in what appear to be occasional crossing-over events in hybrids.

Discussion

Sapsucker genomic relationships

We found overall patterns of autosomal differentiation among sapsuckers characterized by generally low F_{ST}, punctuated by regions of high F_{ST} with low π_W and π_B. These patterns are similar to several other sister taxa of North American boreal bird species, including mourning and MacGillivray's warblers (*Geothlypis philadelphia*, *G. tolmiei*), myrtle and Audubon's yellow-rumped warblers (*Setophaga coronata coronata*, *S. c. auduboni*), and Townsend's, hooded, and black-throated green warblers (*Setophaga townsendi*, *S. cintra*, *S. virens*), which all showed low π_B in high F_{ST} regions (Irwin et al. 2018, Wang et al. 2021). However, the sapsucker

genome scans show this pattern to an even stronger degree. Our results agree with existing literature in describing red-breasted and red-naped sapsuckers as sister species, with yellow-bellied positioned as an outgroup to this species pair (Cicero & Johnson 1995, Grossen et al. 2020, Johnson & Zink 1983, L. Natola & Burg 2018). While many loci show remarkably low differentiation between red-breasted and red-naped sapsuckers (Cicero & Johnson 1995, L. Natola & Burg 2018), our genome-wide data show that the two exhibit high differentiation across large genomic distances on the Z chromosome (Figures 3, 4, 7). Our genome-wide mean pairwise F_{ST} calculations indicate higher differentiation between red-naped and yellow-bellied sapsuckers than between red-breasted and yellow-bellied (Figure 4). Lower within-population genetic diversity can lead to the higher relative between-population diversity we observe in the F_{ST} values (Figures 3, 4). This might be a result of the closer geographic distances between red-naped sapsuckers used in this study, or of a species-wide pattern of low within-species genetic diversity.

The genomic landscape and Sphyrapicus evolutionary history

One might expect because red-breasted and red-naped sapsuckers are more closely related their genomes would be similar (have low F_{ST} values) in the regions where each and the yellow-bellied lineages diverged, but regions of high differentiation tend to recur among all species comparisons. For example, in the F_{ST} row of Figure 4, though the red-breasted/red-naped purple line is typically lower than the orange or green line, all line colors tend to peak in similar places. This may indicate that similar regions are under divergent selection among all three species pairs. However, these high F_{ST} regions also tend to show valleys in π_B and π_W. This suggests that regions of high F_{ST} in our data do not reflect a simple divergence-with-gene-flow

model, in which case π_B is expected to be higher than in the surrounding homogenized genome (Cruickshank & Hahn 2014). Neither does it seem to result from selection in allopatry, which typically shows a π_B in high F_{ST} regions similar to π_B in low F_{ST} regions (Cruickshank & Hahn 2014, Irwin 2018, Irwin et al. 2016, T. L. Turner et al. 2005). Therefore, it may be indicative that the current genomic landscape, and regions of high F_{ST} in particular, are shaped by recurrent selection (π_B lower than in low F_{ST} regions) or selective sweeps followed by differentiation to different optima (π_B much lower than in low F_{ST} regions) (Irwin et al. 2016, 2018).

Notably, the geographic landscape was also shaped by recurrent forces at the time sapsuckers began diverging. Pleistocene era glacial fluctuations affected the entire planet, but habitat shifts were especially marked in high latitude regions including northern North America, where kilometers-high ice sheets advanced and receded over the millennia. Sapsuckers, along with other species in this region, were repeatedly isolated into separate refugia with distinct selective pressures before coming back into contact in interglacial periods (Natola & Burg 2018, Weir and Schluter 2004). We suspect that the signatures of recurrent selection or selective sweeps we observe in our sapsucker genome scans are related, whether directly or indirectly, to the recurrent geographic and glacial cycles of the Pleistocene era.

Z chromosome differentiation

The most striking pattern in our data is the differentiation along the Z chromosome scaffolds. These patterns fit with well described patterns of accumulation of differentiation on the Z chromosome in birds and other ZW systems, and the X chromosome in XY sex chromosome species (Dean et al. 2015, Irwin 2018, Qvarnström & Bailey 2009). The large Z effect, in which the Z chromosome is disproportionately involved in causing inviability or

sterility in hybrids (Irwin 2018), and the fast Z effect describing higher rates of functional change on the Z (Mank et al. 2010), are particularly well supported. Lower recombination rates and effective population sizes paired with stronger sexual selective pressures, greater drift, and higher mutation rates are known to drive evolution in loci on the sex chromosomes (Axelsson et al. 2004, Charlesworth 2001, Dean et al. 2015, Irwin 2018, Mank et al. 2010, Presgraves 2018). However, out of the 9.98 million variable sites on the 1.29 Gbp genome, 9.2 million occur on the autosomes, meaning 0.8% of all autosomal sites are variable. Of the 127 Mbp on the Z chromosome, only 655 thousand sites are variable (0.5%). This is consistent more recent coalescence times on the Z.

The Z chromosome is far more differentiated than the autosomes in terms of F_{ST}. In other words, the proportion of variant sites themselves are not higher on the Z than on autosomes, but the relative differentiation (F_{ST}) between haplotypes is far greater on the Z. Astoundingly, more than 10% of SNPs on the Z are fixed between red-naped and yellow-bellied sapsuckers, compared to 0.29% among autosomal SNPs. There is precedent for this pattern in the literature. Irwin (2018) compiled a list of 14 bird species comparisons in which the F_{ST} values were substantially higher on the Z chromosome than autosomes. There are several hypothesized explanations for this phenomenon. The aforementioned faster rate of functional evolution facilitates the faster accumulation of epistatic incompatibilities between population isolates, causing lower fitness in hybrids which further facilitates differentiation between species at these genomic regions (Irwin 2018). However, this explanation operates on the premise that SNPs amass on the Z, whereas our data suggest these staggering F_{ST} and fixed SNP values are the result of low absolute diversity between and within populations, and not an accumulation of SNPs. In other words, F_{ST} between sapsuckers is high because sapsuckers have lost similarities,

not because they have accumulated new differences. Another way to consider these data is the evidence that the coalescence times are more recent in the Z chromosomes than autosomes, despite having higher relative differentiation on the Z (Figure 6).

Postzygotic selection against hybrids can also be invoked with Haldane's rule, in which deleterious recessives are exposed more readily to selection on the hemizygous Z in females. This is proposed to reduce hybrid fitness, particularly if there are incompatibilities between the Z haplotype of one population and the W or mitochondrial DNA of another (Haldane 1922, Irwin 2018, Presgraves 2002). This accumulation of both differentiation and incompatibilities on the Z chromosome is proposed to play a disproportionate role in the speciation process (Irwin 2018), and we might expect patterns of Z differentiation to be most stark on Pleistocene era radiations such as sapsuckers and other incipient boreal species. Because mutations accrue more rapidly and they are more resistant to gene flow on the Z, this would be more apparent in recently diverged species which have experienced recurrent bouts of gene flow.

Origin of differentiation haploblocks

Some data suggest the high differentiation on the Z chromosome could be driven in part by structural changes within the Z. The Z scaffolds have high LD and high differentiation, patterns typical of chromosomal inversions (Figures 3, 4). On scaffolds 1030 and 1044, stretches spanning more than 10 Mbp show regions with an LD close to 1, meaning almost no evidence for recombination in these regions. Our genotype by individual plots show long spans (> 16 Mbp) of almost complete heterozygosity of fixed SNPs in some sympatric individuals, which is often an indication of structural variants such as chromosomal inversions resisting recombination along long chromosomal distances (Figure 8) (da Silva et al. 2019, Huang et al. 2020, Lundberg

et al. 2017). However, there are individuals with shorter blocks of hetero- or homozygosity in our genotype by individual data, possibly a result of successful double crossovers.

Another possible mechanism for the large haploblocks we see on the Z chromosome is a large duplication, deletion, or other form of copy number variation (CNV). Duplicated gene copies have a tendency to differentiate, both in expression and DNA sequence, and a large duplicated region of the genome encompassing many genes could explain long stretches of genome differentiation between populations (Oka 1974, Seehausen et al. 2014, Ting et al. 2004). However, with CNV events in one or more species affecting most of the Z, we would expect to see differentiated read coverage on the high F_{ST} Z scaffolds in at least one species as we do in non-duplicated species. Our qualimap outputs do not indicate any difference in normalized read coverage among birds of different species on these scaffolds. Additionally, a duplication would not create the homozygotes vs heterozygote pattern that is so clear in the fixed SNP data (Figure 8).

Despite the large haploblocks of differentiation and sustained high LD among sapsucker species, a structural variant is not the only possible explanation for the pattern. It could be that the patterns we observe arise from the faster evolution rate on the Z chromosome combined with epistatic interactions. Substitutions which are neutral, slightly deleterious, or even beneficial in one genome may be strongly deleterious on the genomic or ecological background of another population (Irwin 2018). These epistatic interactions would be characterized as Bateson-Dobzhansky-Muller Incompatibilities (BDMIs), in which neutral or beneficial evolutionary changes at different loci in separate populations are incompatible upon secondary contact (Bateson 1909, Dobzhansky 1936, Muller 1940). BDMIs, in reducing recombination between haplotypes, also tend to accumulate more epistatic loci, increasing incompatibility in a “snowball

effect" (Orr 1995). Moreover, these may be exacerbated by the disproportionate accumulation of sexually antagonistic sites on the Z chromosome, because selective sweeps in expression modifiers or preference alleles on the Z can accelerate differentiation in sexual traits and preferences between populations, therefore increasing behavioral incompatibilities and further deepening the divide between the species (Irwin 2018, Pryke et al. 2010). Indeed, we see evidence that differentiation in sexual traits is associated with loci on the Z chromosome (E. Natola 2022).

These patterns of differentiation among sapsucker Z chromosomes could be exacerbated if the various Z haplogroups were each under intense selection unique to a species group. We do see evidence for this scenario in our data. Irwin et al. (2018) attributed patterns of high F_{ST} paired with low π_B and exceptionally low π_W to positive selection causing a selective sweep followed by selective differentiation among populations. In our data, we show π_B is significantly lower on Z chromosomes than autosomes, particularly in regions with high F_{ST} (Figures 4-6). Within species, π_W is significantly lower on the Z chromosome, indicating selective sweeps throughout each Z haploblock, and our $\pi_{W,Z}^*/\pi_{W,A}$ values are lower than the theoretical threshold that could be explained by neutrality, signifying selection must reduce $\pi_{W,Z}^*$ in sapsuckers (Table 1) (Charlesworth 2001, Irwin et al. 2016, 2018). Additionally, sustained homozygosity on haploblocks such as we see within each species on the Z (Figure 8) are evidence of strong positive selection on this region (Wooldridge et al. 2022). We propose these large haplotype blocks are driven by strong, divergent selective forces acting upon the different Z haploblocks, with epistasis maintaining the blocks in sympatry via post-zygotic selection against hybrid forms. While we know of few examples that have examined this question directly, we expect that the patterns we observed in sapsuckerrs could result in large part from a combination of well-

characterized, predictable patterns of Z chromosome differentiation and expansive geographical and environmental fluctuations during the Pleistocene era. We expect this mechanism of speciation driven mainly by recurrent divergent selection on Z haplotypes is not unique to sapsuckers, but rather is likely common in incipient high latitude species with chromosomal sex determination.

Conclusions

During sapsucker differentiation, recurrent selection is implicated by an overall pattern in which regions with high relative differentiation have low pairwise sequence divergence. This pattern is particularly strong on the Z chromosome: relative differentiation is far higher on the Z chromosome than the autosomes, but nucleotide distance is lower on the Z than the autosomes. We identify long haploblocks with unexpectedly high LD between species along Z scaffolds. While this could be a large structural variant encompassing most of the Z chromosome, an alternative is that this pattern might simply result from strong selection on the Z to different species optima combined with epistasis in sympatry. Relative differentiation between sapsucker species along much of the rest of the genome is low, particularly between red-breasted and red-naped sapsuckers. We suggest that sapsucker species were formed and are maintained in large part due to this differentiation along the Z, which was likely driven by recurrent divergent selection in Pleistocene era glacial refugia.

Acknowledgements

We thank Sampath Seneviratne and Jocelyn Hudon for sharing blood and tissue samples from hybrid sapsuckers. Thanks to Carla Cicero (Museum of Vertebrate Zoology), Donald

McAlpine (New Brunswick Museum), Ray Poulin (Royal Saskatchewan Museum), Garth Spellman (Denver Museum of Natural Sciences), Paul Sweet (American Museum of Natural History), Ildiko Szabo (Beaty Biodiversity Museum), Kevin Winker (University of Alaska Museum) for museum tissues. Thank you to Julia Kreiner and Tom Booker for bioinformatic assistance, and to Kathy Martin, Loren Rieseberg, and Dolph Schluter for feedback on drafts.

References

Aguillon, S. M., Walsh, J., & Lovette, I. J. (2021). Extensive hybridization reveals multiple coloration genes underlying a complex plumage phenotype. Proceedings of the Royal Society B: Biological Sciences, 288(1943), 1–10.

Axelsson, E., Smith, N. G. C., Sundstro, H., Berlin, S., & Ellegren, H. (2004). Male-biased mutation rate and divergence in autosomal, Z-linked and W-linked introns of chicken and turkey. Molecular Biology and Evolution, 21(8), 1538–1547. https://doi.org/10.1093/molbev/msh157

Bateson, W. (1909). Heredity and variation in modern lights. In A. Seward (Ed.), *Darwin and Modern Science* (pp. 85–101). Cambridge University Press.

Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. 2017. Julia: a fresh approach to numerical computing. SIAM Review 59:65– 98.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., Suh, A., Dutoit, L., Bure, S., Garamszegi, L. Z., Hogner, S., Moreno, J., Qvarnström, A., Sæther, S., Sætre, G., Török, J., & 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.Freely

Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. Molecular Biology and Evolution, 15(5), 538–543.

Charlesworth, B. (2001). The effect of life-history and mode of inheritance on neutral genetic variability. Genetical Research, 77(2), 153–166.

Cicero, C., & Johnson, N. K. (1995). Speciation in sapsuckers (*Sphyrapicus*): III. Mitochondrial-DNA sequence divergence at the cytochrome-b locus. Auk, 112(3), 547–553.

Clark, R. W., Brown, W. S., Stechert, R., & Zamudio, K. R. (2010). Roads, interrupted dispersal, and genetic diversity in timber rattlesnakes. Conservation Biology, 24(4), 1059–1069. https://doi.org/10.1111/j.1523-1739.2009.01439.x

Coyne, J. A., & Orr, H. A. (2004). Speciation. Sinauer Associates, Inc.

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular Ecology, 23(13), 3133–3157.

da Silva, V. H., Laine, V. N., Bosse, M., Spurgin, L. G., Derks, M. F. L., Oers, K. van, Dibbits, B., Slate, J., Crooijmans, R. P. M. A., Visser, M. E., & Groenen, M. A. M. (2019). The genomic complexity of a large inversion in great tits. Genome Biology and Evolution, 11(7), 1870–1881. https://doi.org/10.1093/gbe/evz106

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McBean, G., Durbin, R., & Group, 100 Genomes Project Analysis. (2011). The variant call format and VCFtools. Bioinformatics, 27(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dean, R., Harrison, P. W., Wright, A. E., Zimmer, F., & Mank, J. E. (2015). Positive selection underlies faster-Z evolution of gene expression in birds. Molecular Biology and Evolution, 32(10), 2646–2656. https://doi.org/10.1093/molbev/msv138

Dobzhansky, T. (1936). Studies on hybrid sterility. II. Localization of sterility factors in

Drosophila pseudoobscura hybrids. Genetics, 21(2), 113–135.

Grabherr, M. G., Russell, P., Meyer, M., Mauceli, E., Alföldi, J., Palma, F. Di, & Lindblad-toh, K. (2010). Genome-wide synteny through highly sensitive sequence alignment: Satsuma. Bioinformatics, 26(9), 1145–1151. https://doi.org/10.1093/bioinformatics/btq102

Grossen, C., Guillaume, F., Keller, L. F., & Croll, D. (2020). Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. Nature Communications, 11(1001), 1–12. https://doi.org/10.1038/s41467-020-14803-1

Haldane, J. B. S. (1922). Sex ratio and unisexual sterility in hybrid animals. Journal of Genetics, 12(2), 101–109. https://doi.org/10.1007/BF02983075

Han, F., Lamichhaney, S., Grant, B. R., Grant, P. R., Andersson, L., & Webster, M. T. (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. Genome Research, 27, 1004–1015. https://doi.org/10.1101/gr.212522.116.1004

Hooper, D. M., Griffith, S. C., & Price, T. D. (2018). Sex chromosome inversions enforce reproductive isolation across an avian hybrid zone. Molecular Ecology, 28(6), 1246–1262. https://doi.org/10.1111/mec.14874

Hooper, D. M., & Price, T. D. (2017). Chromosomal inversion differences correlate with range overlap in passerine birds. Nature Ecology and Evolution, 1(10), 1526–1534. http://dx.doi.org/10.1038/s41559-017-0284-6

Howell, T. R. (1952). Natural history and differentiation in the yellow-bellied sapsucker. Condor, 54(5), 237–282.

Huang, K., Andrew, R. L., Owens, G. L., Ostevik, K. L., & Rieseberg, L. H. (2020). Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype.

Molecular Ecology, 29, 2535–2549. https://doi.org/10.1111/mec.15428

Irwin, D. E. (2018). Sex chromosomes and speciation in birds and other ZW systems. Molecular Ecology, 27(19), 3831–3851. https://doi.org/10.1111/mec.14537

Irwin, D. E., Alcaide, M., Delmore, K. E., Irwin, J. H., & Owens, G. L. (2016). Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. Molecular Ecology, 25(18), 4488–4507.

Irwin, D. E., Milá, B., Toews, D. P. L., Brelsford, A., Kenyon, H. L., Porter, A. N., Grossen, C., Delmore, K. E., Alcaide, M., & Irwin, J. H. (2018). A comparison of genomic islands of differentiation across three young avian species pairs. Molecular Ecology, 27(23), 4839–4855. https://doi.org/10.1111/mec.14858

Johnson, N. K., & Zink, R. M. (1983). Speciation in sapsuckers (*Sphyrapicus*): I. Genetic differentiation. Auk, 100(4), 871–884.

Kirkpatrick, M., & Barrett, B. (2015). Chromosome inversions, adaptive cassettes and the evolution of species' ranges. Molecular Ecology, 24(9), 2046–2055. https://doi.org/10.1111/mec.13074

Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. Genetics, 173(1), 419–434. https://doi.org/10.1534/genetics.105.047985

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu : scalable and accurate long-read assembly via adaptive k -mer weighting and repeat separation. Genome Research, 27(5), 722–736. https://doi.org/10.1101/gr.215087.116.Freely

Li, H. (2013). Aligning sequence reads , clone sequences and assembly contigs with BWA-MEM. AXiv, 1303.3997v(00), 1–3.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & Subgroup, T. 1000 G. P. D. P. (2009). The sequence alignment/map format and SAMtools. Bioinformatics, 25(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lundberg, M., Liedvogel, M., Larson, K., Sigeman, H., Grahn, M., Wright, A., Åkesson, S., & Bensch, S. (2017). Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. Evolution Letters, 1(3), 115-180.

Mank, J. E., Nam, K., & Ellegren, H. (2010). Faster-Z evolution is predominantly due to genetic drift. Molecular Biology and Evolution, 27(3), 661–670. https://doi.org/10.1093/molbev/msp282

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO update: Novel and streamlined workflows along with Bboader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Molecular Biology and Evolution, 38(10), 4647–4654. https://doi.org/10.1093/molbev/msab199

Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20(9), 1297–1303. https://doi.org/10.1101/gr.107524.110.20

Menounos, B., Osborn, G., Clague, J. J., & Luckman, B. H. (2009). Latest Pleistocene and Holocene glacier fluctuations in western Canada. Quaternary Science Reviews, 28(21–22), 2049–2074.

Muller, H. (1940). Bearing of the *Drosophila* work on systematics. In J. Huxley (Ed.), *The New Systematics* (pp. 185–268). Clarendon Press.

Natola, E. (2022). Reproductive Isolation among Sphyrapicus Sapsuckers. University of British Columbia.

Natola, L., & Burg, T. M. (2018). Population genetics and speciation of yellow-bellied, red-naped, and red-breasted sapsuckers (*Sphyrapicus varius*, *S. nuchalis* , and <i>S. ruber). Journal of Heredity, 109(6), 663–674. https://doi.org/10.1093/jhered/esy034%0Ahttps://academic.oup.com/jhered/advance-article/doi/10.1093/jhered/esy034/5053756

Natola, L., Curtis, A., Hudon, J., & Burg, T. M. (2021). Introgression between *Sphyrapicus nuchalis* and *S. varius* sapsuckers in a hybrid zone in west-central Alberta. Journal of Avian Biology, 1–12. https://doi.org/10.1111/jav.02717

Natola, L., Seneviratne, S. S., & Irwin, D. (2022). Population genomics of an emergent tri-species hybrid zone. Molecular Ecology, *August*, 1–12. https://doi.org/10.1111/mec.16650

Nelson, G. (2006). Hawaiian vicariance. Journal of Biogeography, 33(12), 2154–2155. https://doi.org/10.1111/j.1365-2699.2006.01629.x

Noor, M. A. F., Grams, K. L., Bertucci, L. A., & Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. Proceedings of the National Academy of Sciences, 98(21), 12084–12088. https://doi.org/10.1073/pnas.221274498

Oka, H. (1974). Analysis of genes controlling F1 sterility in rice by the use of isogenic lines. Genetics, 77, 521–534.

Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics, 32(2), 292–294.

Orr, H. A. (1995). The population genetics of speciation: The evolution of hybrid

incompatibilities. Genetics, 139(4), 1805–1813.

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. Molecular Ecology, 25(11), 2337–2360. https://doi.org/10.1111/mec.13557

Presgraves, D. C. (2002). Patterns of postzygotic isolation in Lepidoptera. Evolution, 56(6), 1168–1183.

Presgraves, D. C. (2018). Evaluating genomic signatures of "the large X-effect" during complex speciation. Molecular Ecology, 27(19), 3822–3830. https://doi.org/10.1111/mec.14777

Price, T. (2008). Speciation in birds. Roberts and Co.

Pryke, S. R., Evolution, S., & May, N. (2010). Sex chromosome linkage of mate preference and color signal maintains assortative mating between interbreeding finch morph. Evolution, 64(5), 1301–1310.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics, 81(3), 559–575.

Qvarnström, A., & Bailey, R. I. (2009). Speciation through evolution of sex-linked genes. Heredity, 102(1), 4–15.

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. A. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. Journal of Evolutionary Biology, 30(8), 1450–1477. https://doi.org/10.1111/jeb.13047

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. Trends in Ecology & Evolution, 16(7), 351–358.

Sætre, G., Moum, T., Bures, S., Kral, M., Adamjan, M., & Moreno, J. (1997). A sexually selected character displacement in flycatchers reinforces premating isolation. Nature, 387, 589–592. file:///D:/Uni Backup/Research papers/saetre et al 1997_ a sexually selected character displacement in flycatchers reinforces premating isolation.pdf%5Cn%3CGo to ISI%3E://A1997XC52200052

Schluter, D., & McPhail, J. D. (1992). Ecological character displacement and speciation in sticklebacks. The American Naturalist, 140(1), 85–108.

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., & Saetre, G. (2014). Genomics and the origin of species. Nature Reviews Genetics, 15(March), 176–192. https://doi.org/10.1038/nrg3644

Seneviratne, S. S., Davidson, P., Martin, K., & Irwin, D. E. (2016). Low levels of hybridization across two contact zones among three species of woodpeckers (*Sphyrapicus* sapsuckers). Journal of Avian Biology, 47(6), 887–898.

Seneviratne, S. S., Toews, D. P. L., Brelsford, A., & Irwin, D. E. (2012). Concordance of genetic and phenotypic characters across a sapsucker hybrid zone. Journal of Avian Biology, 43(2), 119–130.

Smit, A., Hubley, R., & Green, P. (2013). RepeatMasker Open-4.0.

R Core Team. 2021. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org.

Ting, C.-T., Tsaur, S.-C., Sun, S., Browne, W. E., Chen, Y.-C., Patel, N. H., & Wu, C.-I. (2004). Gene duplication and speciation in Drosophila: Evidence from the Odysseus locus. PNAS, 101(33), 12232–12235.

Todesco, M., Owens, G. L., Bercovich, N., Légaré, J.-S., Soudi, S., Burge, D. O., Huang, K.,

Ostevik, K. L., Drummond, E. B. M., Imerovski, I., Lande, K., Pascual-Robles, M. A., Nanavati, M., Jahani, M., Cheung, W., Staton, S. E., Muños, S., Nielsen, R., Donovan, L. A., … Rieseberg, L. H. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. Nature, 584(September 2019), 602–607. https://doi.org/10.1038/s41586-020-2467-6

Turner, S. D. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. Journal of Open Source Software, 3(25), 2–3. https://doi.org/10.21105/joss.00731

Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in *Anopheles gambiae*. PLoS Biology, 3(9), e285. https://doi.org/10.1371/journal.pbio.0030285

Uy, J. A. C., Irwin, D. E., & Webster, M. S. (2018). Behavioral isolation and incipient speciation in birds. Annual Review of Ecology, Evolution, and Systematics, 49, 1–24.

Wang, S., Ore, M. J., Mikkelsen, E. K., Lee-Yaw, J., Toews, D. P. L., Rohwer, S., & Irwin, D. (2021). Signatures of mitonuclear coevolution in a warbler species complex. Nature Communications, 12(4279). https://doi.org/10.1038/s41467-021-24586-8

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

Weir, J. T., & Schluter, D. (2004). Ice sheets promote speciation in boreal birds. Proceedings. Biological Sciences / The Royal Society, 271(1551), 1881–1887.

Wiley, G., & Miller, M. J. (2020). A highly contiguous genome for the Golden-fronted Woodpecker (Melanerpes aurifrons) via hybrid Oxford Nanopore and short read assembly. G3: Genes, Genomes, Genetics, 10(June), 1829–1836.

Wooldridge, T. B., Kautt, A. F., Lassance, J.-M., McFadden, S., Domingues, V. S., Mallarino, R., & Hoekstra, H. E. (2022). An enhancer of *Agouti* contributes to parallel evolution of cryptically colored beach mice. PNAS, 119(27), e2202862119. https://doi.org/10.1073/pnas.2202862119/-/DCSupplemental.Published

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics, 28(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Author Contributions

LN conceived the project with input from DI. LN sourced samples from museums, extracted DNA, and submitted samples for library preparation. LN prepared raw sequences and ran all analyses. DI converted R scripts to Julia. LN wrote initial drafts with oversight and revisions from DI.

Tables

Table 1 Comparisons between autosomes (A) and Z chromosomes (Z) of the measurements F_{ST} and π_B between species (a), and π_W within species (b). We include adjusted Z diversity ($\pi_{W,Z}$*) to correct for higher mutation rate on Z, and compare nucleotide differentiation on the Z vs autosome. Measurements were calculated as mean values in 50kb sliding windows from the genomic landscape dataset.

A.

Species pair	$F_{ST,A}$	$F_{ST,Z}$	*p* value	$\pi_{B,A}$	$\pi_{B,Z}$	*p* value	$\pi_{B,Z*}$
Red-naped/red-breasted	0.098	0.412	$p < 0.0001$	0.00321	0.00213	$p < 0.0001$	0.00194
Yellow-bellied/red-breasted	0.209	0.546	$p < 0.0001$	0.00478	0.00361	$p < 0.0001$	0.00328
Yellow-bellied/red-naped	0.267	0.573	$p < 0.0001$	0.00483	0.00359	$p < 0.0001$	0.00326

b.

Species	$\pi_{W,A}$	$\pi_{W,Z}$	*p* value	$\pi_{W,Z}$*	$\pi_{W,Z}$*/$\pi_{W,A}$
Red-naped	0.00270	0.00136	$p < 0.0001$	0.00124	0.458
Red-breasted	0.00316	0.00172	$p < 0.0001$	0.00156	0.496
Yellow-bellied	0.00458	0.00258	$p < 0.0001$	0.00235	0.512

Figures

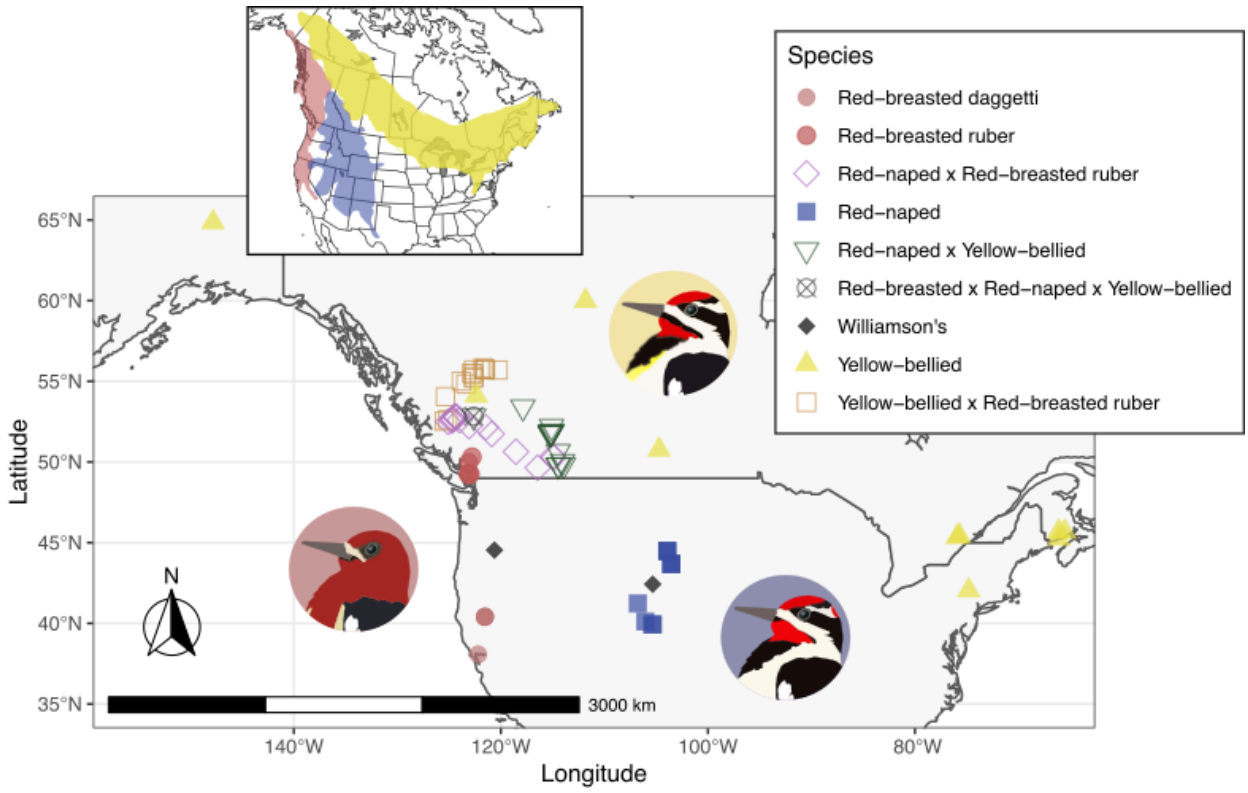


Figure 1 Sampling map of birds included in the WGS sequencing dataset (see Supplemental Table 1 for more details).

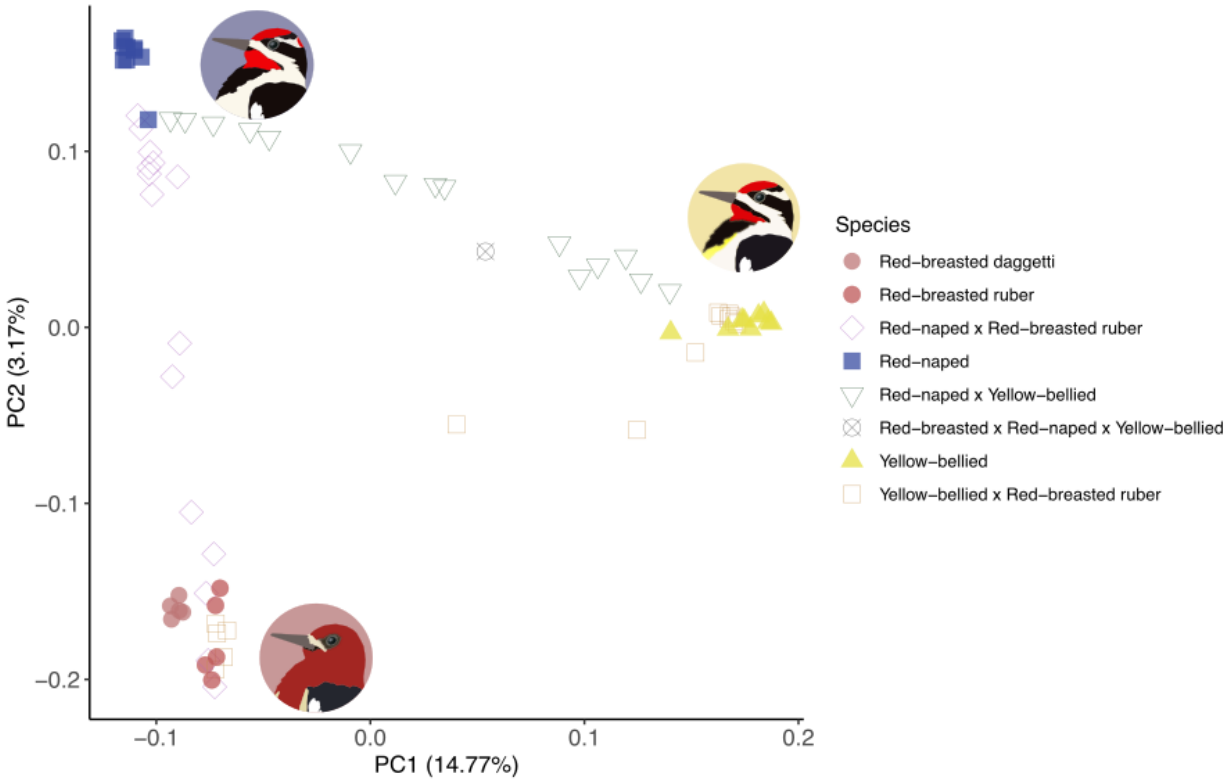


Figure 2 PCA of WGS genotypic data. Yellow dots indicate allopatric yellow-bellied sapsuckers, red dots indicate allopatric red-breasted, blue dots are allopatric red-naped. Orange dots show birds sampled in yellow-bellied × red-breasted sympatry, purple denote red-breasted × red-naped sympatric birds, and green dots show birds sampled in red-naped × yellow-bellied sympatry.

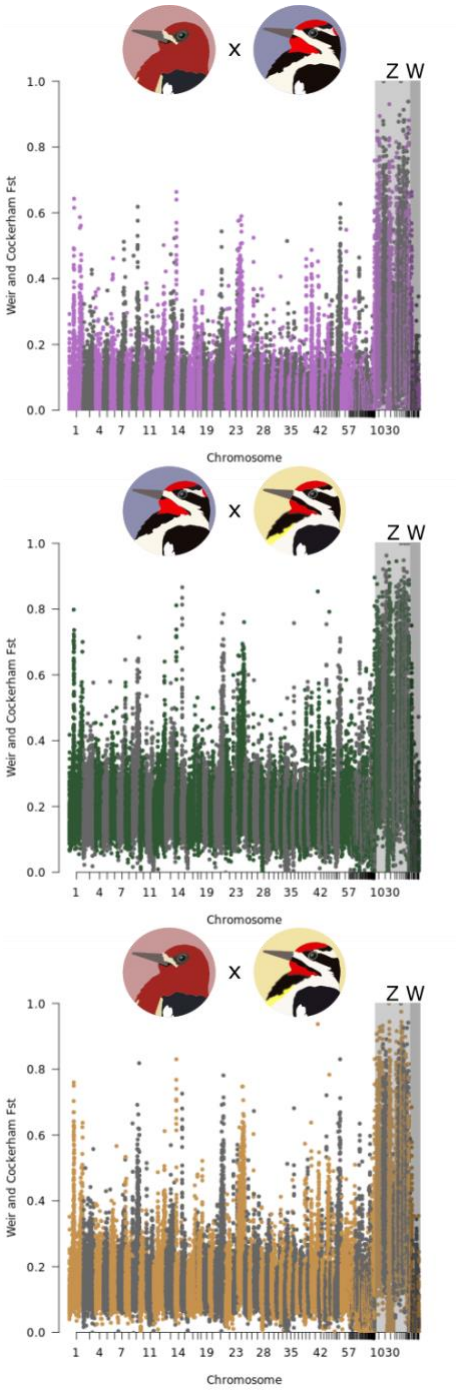


Figure 3 Genomic variation in pairwise F_{ST} between (a) red-breasted and red-naped sapsuckers, (b) red-naped and yellow-bellied sapsuckers, and (c) red-breasted and yellow-bellied sapsuckers. Each dot represents a weighted average over a 25 kb window. Light grey box shading indicates

scaffolds identified as likely to be on the Z chromosome and darker grey box shading designates those on the W chromosome.

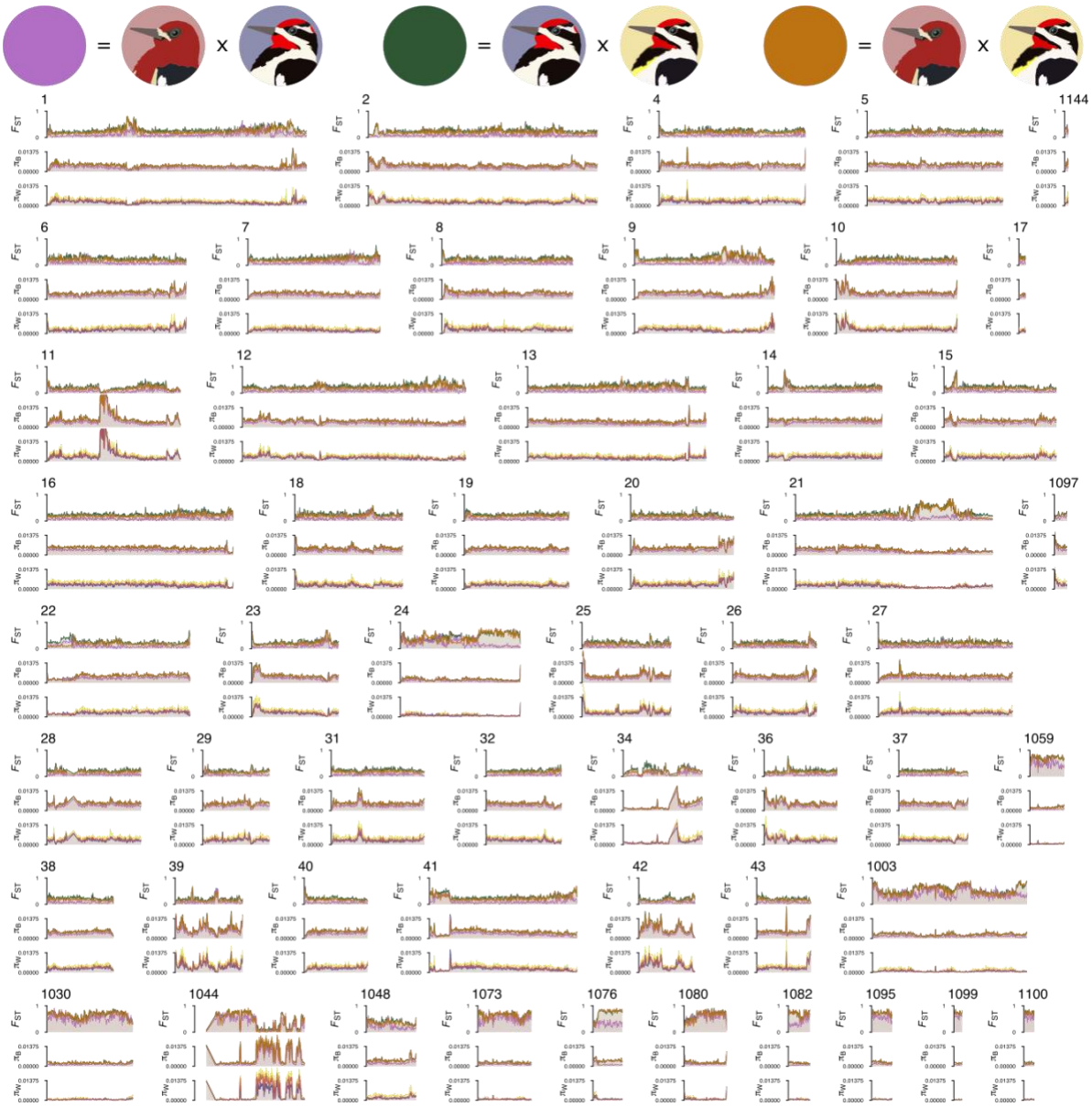


Figure 4 Genomic variation in nucleotide differentiation among sapsuckers for the 45 largest scaffolds and all Z chromosome scaffolds (scaffold IDs > 1000). For each scaffold, graphs show the mean measurement across 50 kb windows for F_{ST} (top), π_B (middle), and π_W (bottom), for all three species/species pairs. The F_{ST} and π_B species pair lines show red-breasted × red-naped in purple, red-naped × yellow-bellied in green, and red-breasted × yellow-bellied in orange. For π_W within species rows indicate red-breasted in red, red-naped in blue, and yellow-bellied in yellow.

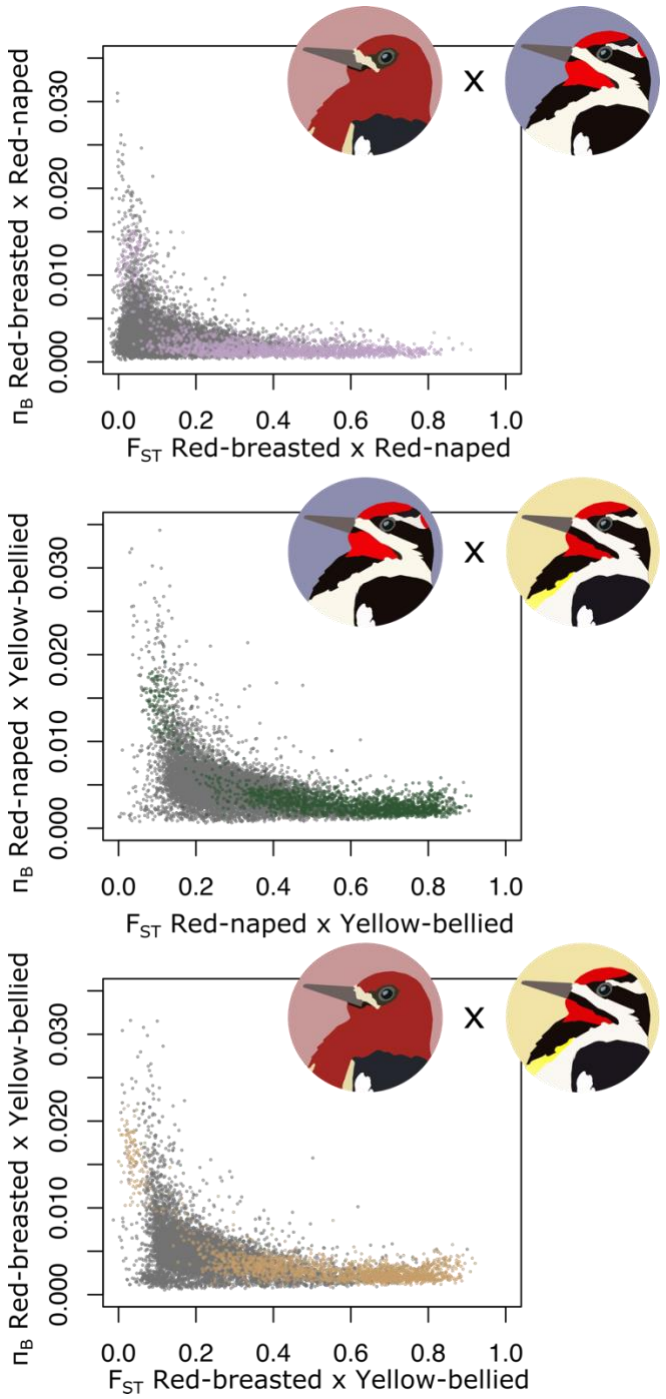


Figure 5 Comparisons of mean F_{ST} and π_B measured in 50 kb sliding windows for red-breasted and red-naped (top), red-naped and yellow-bellied (middle), and red-breasted and yellow-bellied (bottom). Autosomal loci are shown in grey, and scaffolds presumed to be on the Z chromosome are shown in color.

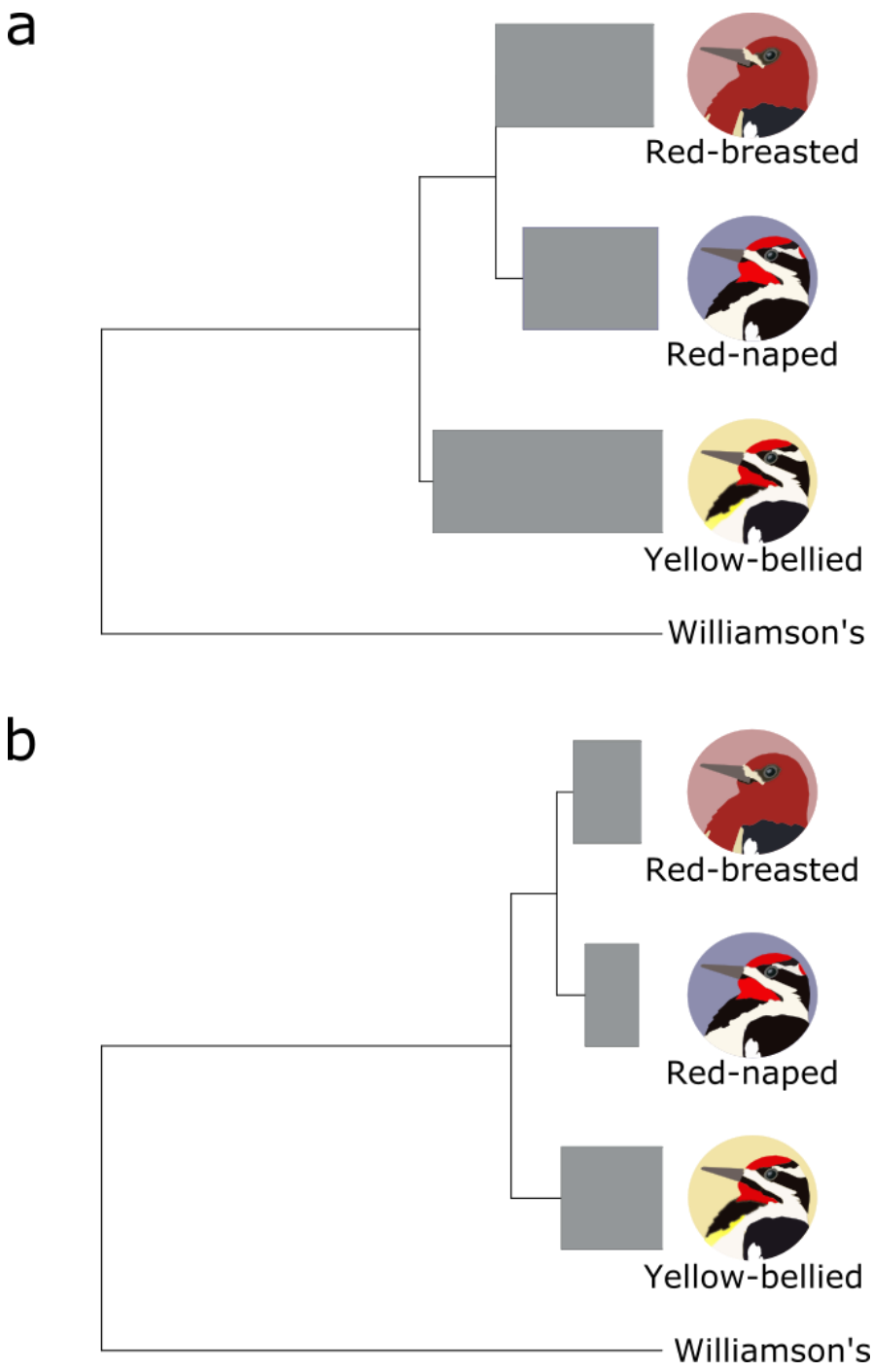
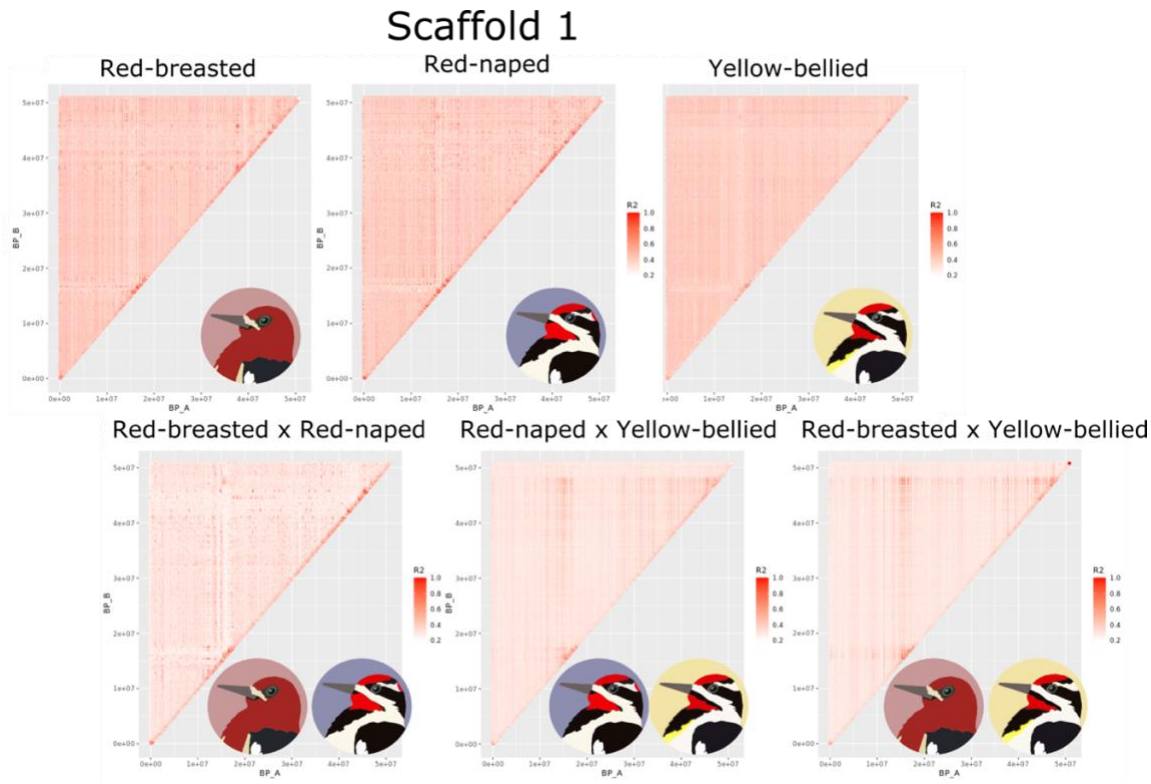
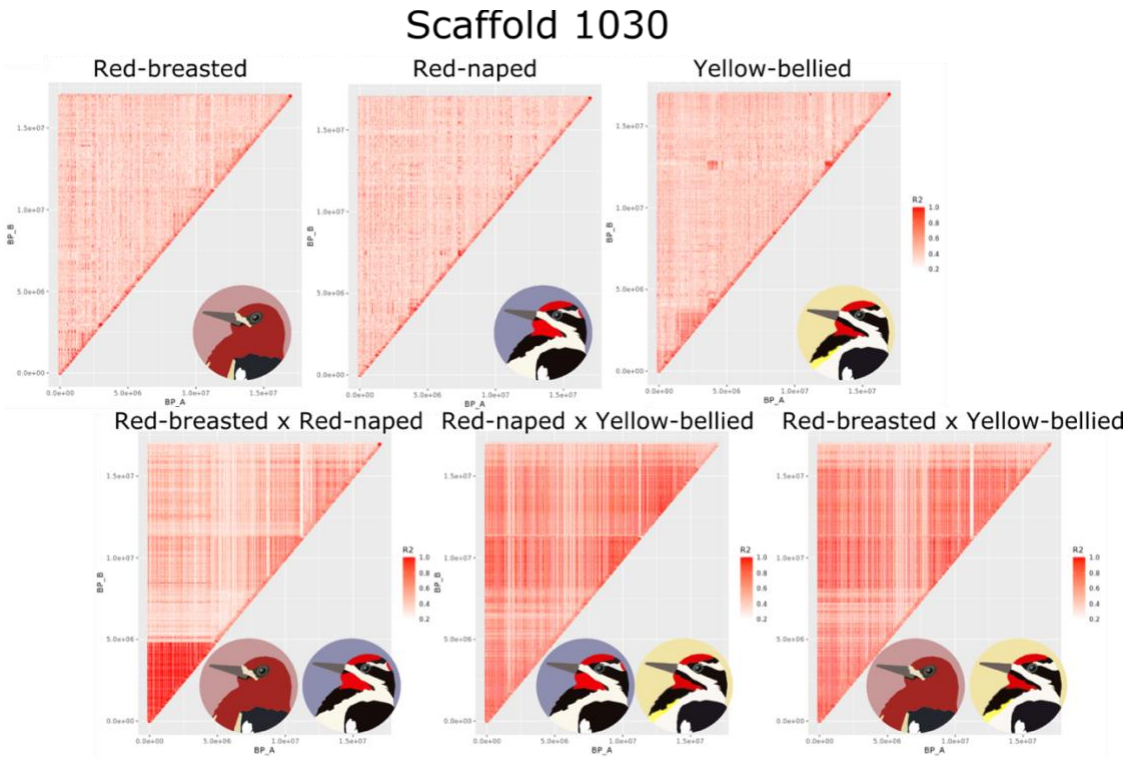


Figure 6 Neighbour-joining trees illustrating phylogenetic relationships among sapsuckers for autosomal (a), and Z-linked regions (b). Tree branch lengths are based on average between-taxon nucleotide distances (π_B) and within-taxon between-individual nucleotide distances (π_W, displayed as length of grey boxes).


Scaffold 1
Red-breasted
Red-naped
Yellow-bellied
Red-breasted x Red-naped
Red-naped x Yellow-bellied
Red-breasted x Yellow-bellied
BP_A
BP_B
R2

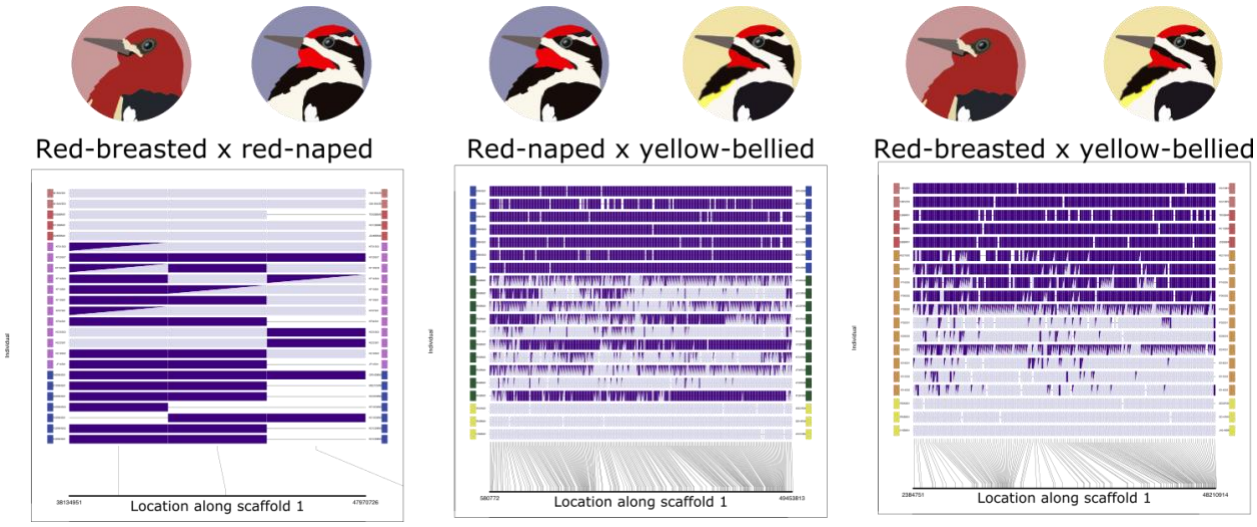
a.


Scaffold 1030
Red-breasted
Red-naped
Yellow-bellied
Red-breasted x Red-naped
Red-naped x Yellow-bellied
Red-breasted x Yellow-bellied
BP_A
BP_B
R2

b.

Figure 7 Heatmap of linkage disequilibrium (LD) across (a) an autosomal and (b) a sex-linked scaffold. High LD is represented as bright red and low LD as white.

a.



b.

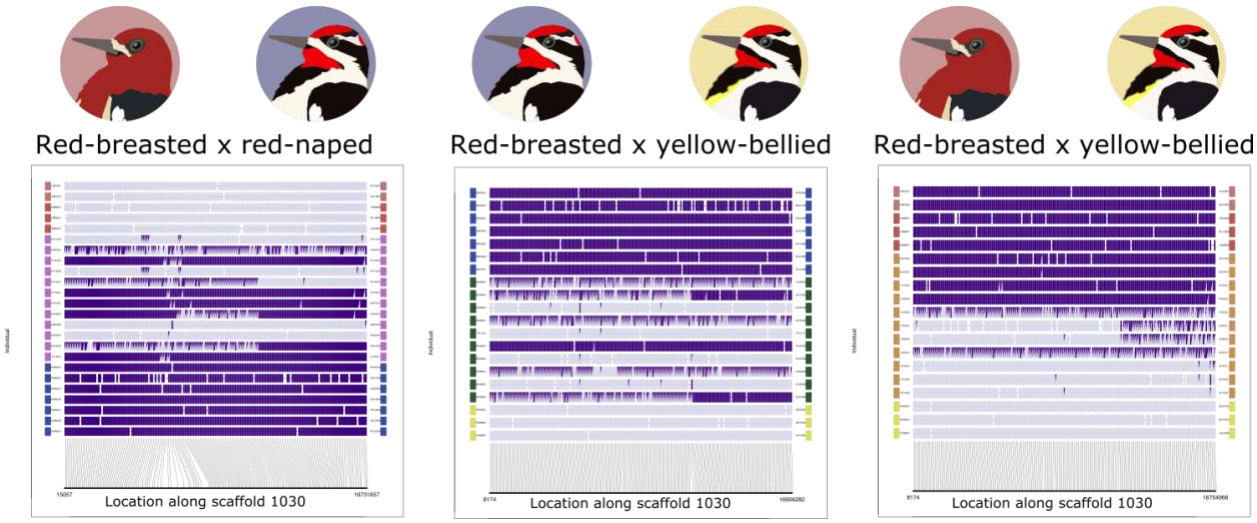


Figure 8 Genotype × individual plots of SNPs (columns) sequenced in each individual (rows) on autosomal scaffold 1 (a) and Z chromosome scaffold 1030 (b). SNPs were included if they are fixed in any pairwise species comparison (they may be fixed in two species but not a third) and thinned to no fewer than 75 Kbp apart for readability. Colored rectangles to the sides of the plot indicate individual was red-breasted *daggetti* (light red), red-breasted *ruber* (dark red), red-breasted × red-naped sympatric (purple), red-naped (blue), red-naped × yellow-bellied sympatric (green), red-breasted × yellow-bellied sympatric (orange), or yellow-bellied (yellow). Each

genotype is depicted as either homozygous (purple rectangle or light purple rectangle) or heterozygous (rectangle split diagonally into a purple and light purple triangle). Lines on the bottom of the plot indicate chromosomal position of each genotype along the scaffold.