# 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;

**Abstract**

Speciation occurs when gene pools differentiate between populations, but that differentiation is often highly heterogeneous across the genome. Understanding what parts of the genome are more prone to differentiation can inform us about genomic regions and evolutionary processes that may be central to the speciation process. Here, we study genomic variation among three hybridizing species of North American woodpecker: red-breasted, red-naped, and yellow-bellied sapsuckers (*Sphyrapicus ruber, S. nuchalis,* and *S. varius*). We use whole genome resequencing to measure genetic variation among these species and to quantify how the level of differentiation varies across the genome. We find that regions of high relative differentiation between species ($F_{ST}$) tend to have low absolute differentiation between species ($\pi_B$), 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 differentiation and potentially speciation. 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 have occurred primarily via this process occurring on the Z chromosomes, resulting in genetic incompatibilities involving divergent Z chromosomes.

## 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), ecological differences, or geographic barriers to dispersal (Nelson, 2006); and they could have existed in the past (J. T. 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 between groups that we describe today as species (Howell, 1952; Schluter & McPhail, 1992; J. T. Weir & Schluter, 2004). In North American bird populations, for example, a commonly occurring biogeographic pattern is the following: related groups (i.e., different species, subspecies, or clearly differentiated populations) are often geographically structured into a west coast group, an interior west group, and an eastern group, and these three types are often 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*; Figure 1). 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 about 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). Today, the species inhabit distinct habitats and climatic niches. All three species are associated with forested habitats containing quaking aspen groves (*Populus tremuloides*), but the yellow-bellied sapsuckers are found primarily in deciduous early-successional and riparian habitats, red-naped sapsuckers are most closely tied to higher elevation and more arid ponderosa pine (*Pinus ponderosa*) forests, and red-breasted sapsuckers inhabit coastal coniferous forests with high annual precipitation levels (Walters et al. 2020 a, b, c). There is a great deal of overlap in the areas in which these species can persist (Billerman et al. 2016, Natola & Burg 2018, Natola et al. 2023). These species hybridize where they come into contact, even forming a three-species hybrid zone (Natola et al. 2022) in central British Columbia, 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, incorporating 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 most differentiated between species and might thereby have played a large role in speciation. 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 ($\pi_B$; often referred to as "$D_{xy}$"), and nucleotide diversity within populations ($\pi_W$), we can test various differentiation models involving geographic separation, gene flow, and selection (Charlesworth, 1998; Cruickshank & Hahn, 2014; Irwin et al., 2016, 2018; T. L. Turner et al., 2005). There is a straightforward mathematical relationship between these aspects of diversity, with $F_{ST} = (\pi_B - \pi_W) / \pi_B$ when sample sizes are large and equal (Charlesworth 1998, Nei 1973). Holding $\pi_W$ constant, higher $\pi_B$ means higher $F_{ST}$; holding $\pi_B$ constant, higher $\pi_W$ means lower $F_{ST}$. Hence, we naively expect high-$F_{ST}$ regions to be those with high $\pi_B$ and low $\pi_W$. To what degree this is true in real datasets depends on the details of the differentiation process.

Divergent selection acting upon a trait in different populations tends to cause greater allele frequency differentiation (i.e., $F_{ST}$) in the loci which encode that trait compared to the rest of the genome. This process however is affected by the recombination landscape of the genome, such that a large genomic region of low recombination can be influenced by the combined effects of selection on multiple loci with the region. When genomic regions become

differentiated through the combined effects of selection and low recombination, divergent loci might 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 ($\pi_B$) when mutation rate is constant. This scenario leads to the expectation that regions of high $F_{ST}$ will tend to also have high $\pi_B$ (Cruickshank & Hahn, 2014; Han et al., 2017) (Figure 2A).

In contrast to this expectation, some pairs of differentiated populations do not show high $\pi_B$ 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 ($\pi_W$) 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 (Figure 2B). In some cases, there is even a negative relationship between $F_{ST}$ and $\pi_B$, 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) (Figure 2C). Another is that adaptively advantageous mutations can sweep from one population to another across a hybrid zone, thereby lowering $\pi_B$, and then subsequent local adaptation can increase $F_{ST}$ (Delmore et al. 2015, Irwin et al. 2016) (Figure 2D). 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 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 nucleotide distances compared to the rest of the genome. We also examine genotypes of hybrids to infer whether there is evidence of recombination in highly differentiated regions, which would suggest that they are not inversions and that hybrids can reproduce successfully. The results from this analysis 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

*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, Supplemental Table S1). 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 (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 similar proportions of both male and female birds (Supplemental Table S1).

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 Illumina NovaSeq6000 S4 PE150 lanes.

Read processing

Sequencing resulted in 2.73 billion reads, with an average Phred quality score of 35. We used Trimmomatic v 0.38 (Bolger et al., 2014) to trim reads, and we then used BWA v 0.7.17 (Li, 2013) to align reads to our previously published red-breasted sapsucker reference genome (Natola et al. 2022). This reference genome is that of a female red-breasted sapsucker collected in Juneau, Alaska, and was assembled using long-read sequences and high sequencing depth (172x), and contigs were assembled into scaffolds based on a golden-fronted Woodpecker (*Melanerpes aurifrons*) reference genome (Wiley & Miller, 2020). After aligning reads, 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 paired-end and un-paired bam files and indexed. We used the bamqc function from Qualimap v 2.2.2a (Okonechnikov et al., 2016) to assess quality of alignments. We created a corrected coverage measure from the bamqc files by dividing each scaffold's read coverage per individual 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. 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 (Polymerase Chain Reaction) duplicates in Picard 1.97 before identifying problematic regions, realigning, and fixing mate-pair coordinates using GATK v 3.8 (Mckenna et al., 2010) with scripts adapted from Aguillon et al. (2021). Finally, SNPs (Single Nucleotide Polymorphisms) 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 site-wise genomic differentiation among the three species we calculated Weir & Cockerham's (1984) pairwise weighted average $F_{ST}$ in VCFtools using only allopatric samples, as we know differentiation is lower in hybrid zones and not representative of birds across the majority of their distribution. To visualize clusters of genomic similarity among the sapsuckers we constructed a PCA (Principal Component Analyses) of variation among individual genotypes using the SNPRelate package (Zheng et al., 2012) in the R programming language (R Core Team 2021). 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 (S. D. Turner, 2018) in R. We plotted $F_{ST}$ (relative differentiation between populations), $\pi_B$ (absolute differentiation between populations), and $\pi_W$ (within population nucleotide diversity) across 50 kb non-overlapping 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 R and the Julia programming language (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. $\pi_B$ for each species pair, while indicating the differences between differentiation on autosomes vs. Z chromosomes. We calculated Spearman's rank correlation between $F_{ST}$ and $\pi_B$ for autosomes and Z chromosomes separately in R. We calculated the adjusted Z diversity, $\pi_{W,Z}$*, as $\pi_W$ /1.1 to account for the higher mutation rate on the Z chromosome (Irwin 2018), and then took the ratio of $\pi_{W,Z}$*/$\pi_{W,A}$ ($\pi_W$ 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).

Genotype × individual plots

To visualize Z-chromosome genotypes across many variant sites and individuals we created genotype-by-individual plots using the fixed (species-diagnostic, $F_{ST}$ = 1) SNPs from each species pair per scaffold, and plotted these using Julia scripts that call the GenomicDiversity.jl package (Irwin et al. 2025). We removed all females to avoid complexity created by hemizygosity. 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. To make plots easier to read, we thinned SNPs to only show SNPs with no missing genotypes and that are 100kb from other included SNPs.

## Results

Our data present clear evidence for strong differentiation between the three species. In the PCA (Figure 3), the major axis of variation (PC1) accounts for 14.77% of the variation and separates yellow-bellied sapsuckers from the other two species. The second axis of variation (PC2) accounts for 3.17% of the variation and separates red-breasted and red-naped sapsuckers. The weighted average pairwise $F_{ST}$ values calculated for each species pair reflect these relationships: all three species pairs have clear differentiation (Table 1), and $F_{ST}$ is highest in comparisons involving yellow-bellied sapsuckers.

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 4). The scaffolds with the highest average pairwise $F_{ST}$ values occur in large peaks that are largely concurrent in each species pair (Figure 4, Supplemental Figures S1, S2). The most obvious pattern in these plots is the extremely high $F_{ST}$ (many windowed values close to 1) across most of the Z chromosome. The weighted Weir and Cockerham $F_{ST}$ of all SNPs (9,205,747 sites) on autosomes is 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. In contrast, on the Z scaffolds (655,683 sites) those measures are 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 (0.000087 % 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. In contrast, on the Z chromosome scaffolds, 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 our samples of red-breasted/red-naped, red-naped/yellow-bellied, and red-breasted/yellow-bellied sapsuckers, respectively.

The mathematical relationship between $F_{ST}$ and $\pi_W$ (see Introduction) leads to the expectation that $F_{ST}$ will tend to increase as $\pi_W$ decreases, and that with $\pi_W$ held constant regions of high $F_{ST}$ will have high $\pi_B$. The sapsucker comparisons show the opposite relationship, such that genomic regions of high $F_{ST}$ tend to have low $\pi_B$ with Spearman's correlations indicating strong negative relationships between $F_{ST}$ and $\pi_B$ in all comparisons of species, autosomes, and Z chromosomes ($-0.264 \leq \rho \leq -0.544$, all *p*-values < 0.0001; Figure 5). Low $\pi_B$ in high $F_{ST}$ regions can occur if $\pi_W$ is proportionally lower than $\pi_B$. Biologically, this can be explained by selective sweeps in a common ancestor (lowering $\pi_B$) and subsequent sweeps in the daughter species (lowering $\pi_W$ and driving up $F_{ST}$) (Cruickshank & Hahn, 2014) (Figure 2C), or a selective sweep across hybridizing populations (lowering $\pi_B$) followed by local differentiation at these loci between populations (causing extremely low $\pi_W$, therefore increasing $F_{ST}$) (Irwin et al., 2016, 2018) (Figure 2D).

Windowed $F_{ST}$ values among sapsucker species are much higher on the Z chromosome scaffolds than on the autosomes, within-group variation ($\pi_W$) is lower, and between-group nucleotide distance ($\pi_B$) is lower on the Z chromosome (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, Figure 5). 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 $\pi_{W,A}$ is below the theoretical minimum of 0.5625 that can be explained without invoking selection (Charlesworth, 2001; Irwin, 2018). Within species, $\pi_W$ is consistently significantly lower on the Z chromosome than on autosomes (Table 1). This is shown in the neighbor-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.

As we have seen in Figures 3 and 4, there are strong differences in allele frequencies between the species, particular in the Z chromosome where windowed $F_{ST}$ is often close to 1. In Figure 7 we use genotype-by-individual plots of two Z-chromosome scaffolds (together accounting for more than 47.0 Mbp) to illustrate the clear differentiation of these regions and to examine whether there is evidence of recombination in hybrids. The SNPs shown for each species pair have been filtered for those that are fixed between the species, for which there are no missing genotypes, and that are separated by at least 100,000 bp of chromosomal distance. Despite these strong filters, there are abundant SNPs that have fixed differences for each species pair. This allows a close look for evidence of recombination in hybrids, for which there is much evidence—this can be seen in hybrid individuals which have some long segments of heterozygosity and then other long segments of homozygosity. These can be explained as the result of past crossing-over and recombination during meiosis in individuals that have two Z chromosomes (i.e., generally males in birds) of different types.

## Discussion

### *Sapsucker relationships*

We find overall patterns of autosomal differentiation among sapsuckers characterized by generally low $F_{ST}$, punctuated by regions of high $F_{ST}$ with low $\pi_W$ and $\pi_B$. 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 show low $\pi_B$ in high $F_{ST}$ regions (Irwin et al., 2018; Wang et al., 2021). 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 sapsuckers 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. 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. Lower within-population genetic diversity can lead to the higher relative between-population diversity we observe in the $F_{ST}$ values. 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 differentiation 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 (Supplemental Figure S2). 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 $\pi_B$ and $\pi_W$. This suggests that regions of high $F_{ST}$ in our data do not reflect a

simple divergence-with-gene-flow model, in which case $\pi_B$ is expected to be higher than in the surrounding homogenized genome (Cruickshank & Hahn, 2014). Neither is it the expected pattern from a model of selection in allopatry, in which $\pi_B$ in high $F_{ST}$ regions is similar to $\pi_B$ 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 ($\pi_B$ lower than in low $F_{ST}$ regions; Cruickshank & Hahn, 2014), likely via between-populations selective sweeps followed by differentiation between populations ($\pi_B$ much lower than in low $F_{ST}$ regions) (Irwin et al., 2016, 2018).

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 (L. Natola & Burg, 2018). 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 high 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; Henderson & Brelsford, 2020; Irwin, 2018; Qvarnström & Bailey, 2009). The large Z effect, in which the Z chromosome is disproportionately involved in causing inviability or sterility in hybrids (Storchová et al. 2010; reviewed by Irwin 2018), and the fast Z effect describing higher rates of functional change on the Z (Ellegren, 2009; 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 of loci on the sex chromosomes (Axelsson et al., 2004; Charlesworth, 2001; Dean et al., 2015; Irwin, 2018; Mank et al., 2010; Presgraves, 2018). However, considering the three sapsucker species as a whole, the Z chromosome shows less variation than the autosomes do: 0.8% of all autosomal sites are variable, whereas only 0.5% of Z chromosome sites are variable. While the Z chromosome having a smaller effective population size than autosomes can explain some of the reduction in diversity on the Z compared to the autosomes, the ratio of adjusted $\pi_{W,Z}$ to $\pi_{W,A}$ is below the theoretical minimum of 0.5625 that can be explained without invoking selection (Charlesworth, 2001; Irwin, 2018). Add to these considerations the finding that $\pi_B$ is also lower on the Z than on autosomes, and we reach the conclusion that the Z chromosome has more recent coalescent times that the autosomes, both within and between species, a pattern likely driven by diversity-reducing selection at multiple timepoints in the past.

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. 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 high $F_{ST}$ values with many fixed SNPs are the result of extremely low within population diversity ($\pi_W$) compared to moderately low between-population nucleotide distance ($\pi_B$), and not an accumulation of SNPs. These results provide evidence that 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 (reviewed by 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. In an earlier analysis of patterns of genetic and phenotypic variation in a three-way hybrid zone of these sapsucker species, Natola et al. (2022) concluded that there is reduced fitness of hybrids, and the

present results suggest that incompatibilities on the Z chromosome play a role in that lowered hybrid fitness.

The fixed differences for divergent haploblocks between these sapsucker species might suggest that they are due to chromosomal inversions (or other structural variants), which can result in suppressed recombination and hence accumulation of differentiation between inversion types (da Silva et al., 2019; Huang et al., 2020). We investigated whether the highly differentiated Z chromosomes might be due in part to inversions by inspecting genotype-by-individual plots for two scaffolds covering large regions on the Z (Figure 7). Both of these scaffolds show much evidence for recombination in hybrid birds. Considering this finding together with earlier conclusions that the narrowness of hybrid zones and lack of buildup of hybrids is due to low fitness of hybrids (Natola et al. 2022, 2023), we conclude that the high differentiation across these Z chromosome scaffolds is primarily due to selection against recombinants, rather than nonrecombining inversions.

Hence the strong Z chromosome differentiation between the three sapsucker species may have arisen from a faster functional evolution rate on the Z chromosome (Ellegren, 2009; Mank et al., 2010) 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 (Hoekstra et al., 2006; Steiner et al., 2007). 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, 2010). Indeed, we see evidence that differentiation in sexual traits is associated with loci on the Z chromosome (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 $\pi_B$ and exceptionally low $\pi_W$ to positive selection causing a selective sweep followed by selective differentiation among populations. In our data, we show $\pi_B$ is significantly lower on Z chromosomes than autosomes, particularly in regions with high $F_{ST}$. Within species, $\pi_W$ 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 (Charlesworth, 2001; Irwin et al., 2016, 2018). Additionally, sustained homozygosity on haploblocks such as we see within each species on the Z 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 sapsuckers 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. Recombination in hybrids casts doubt on inversions or other structural variants being the cause of the high Z chromosome divergence. Rather, this pattern might simply result from strong selection on the Z to different species optima combined with epistasis in sympatry causing low fitness of mixed genotypes at the Z chromosome. 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. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada (Discovery Grants RGPIN-2017-03919, RGPAS-2017-507830, RGPIN-2023-04300), the American Ornithological Society Hesse Award, and the Linnean Society LinnéSys Grant. All protocols were approved by the University of British Columbia Animal Care Committee.

**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 most analyses. DI converted R scripts to Julia. LN wrote initial drafts with oversight and revisions from DI.

**Data Accessibility**

All raw reads will be accessioned on SRA pending publication. All analysis scripts are available in a GitHub repository and will be accessioned for archiving in dryad upon publication acceptance.

## 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.

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

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.

Ellegren, H. (2009). Genomic evidence for a large-Z effect. *Proceedings of the Royal Society B: Biological Sciences*, 276: 361-366.

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

Henderson, E. C. & Brelsford, A. (2020). Genomic differentiation across the speciation continuum in three hummingbird species pairs. *BMC Evolutionary Biology*, 20:113. https://doi.org/10.1186/s12862-020-01674-9

Hoekstra, H. E., Hirschmann, R. J., Bundey, R. A, & Crossland, J. P. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science*, *313*(*5783*), 101-104. https://doi.org/10.1126/science.1126121

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.

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

Irwin, D., Bensch, S., Charlebois, C., David, G., Geraldes, A., Gupta, S.K., Harr, B., Holt, P., Irwin, J.H., Ivanitskii, V.V., Marova, I.M., Niu, Y., Seneviratne, S., Singh, A., Wu, Y., Zhang, S., & Price, T.D. (2025). The distribution and dispersal of large haploblocks in a superspecies. *Molecular Ecology*, *in press*.

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

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., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence

alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

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

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.

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

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, 70: 3321-3323.

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

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. (2010). Sex chromosome linkage of mate preference and color signal maintains assortative mating between interbreeding finch morph. *Evolution*, *64*(5), 1301–1310.

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

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.

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

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.

Steiner, C. C., Weber, J. N., & Hoekstra, H. E. (2007). Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLOS Biology*. *5*(*9*), e219. https://doi.org/10.1371/journal.pbio.0050219.

Storchová, R., Reif, J., & Nachman, M.W. (2010). 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

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.

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

**Tables**

Table 1 Comparisons between autosomes (A) and Z chromosomes (Z) of the measurements $F_{ST}$ and $\pi_B$ between species (a), and $\pi_W$ 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 means of values from 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**

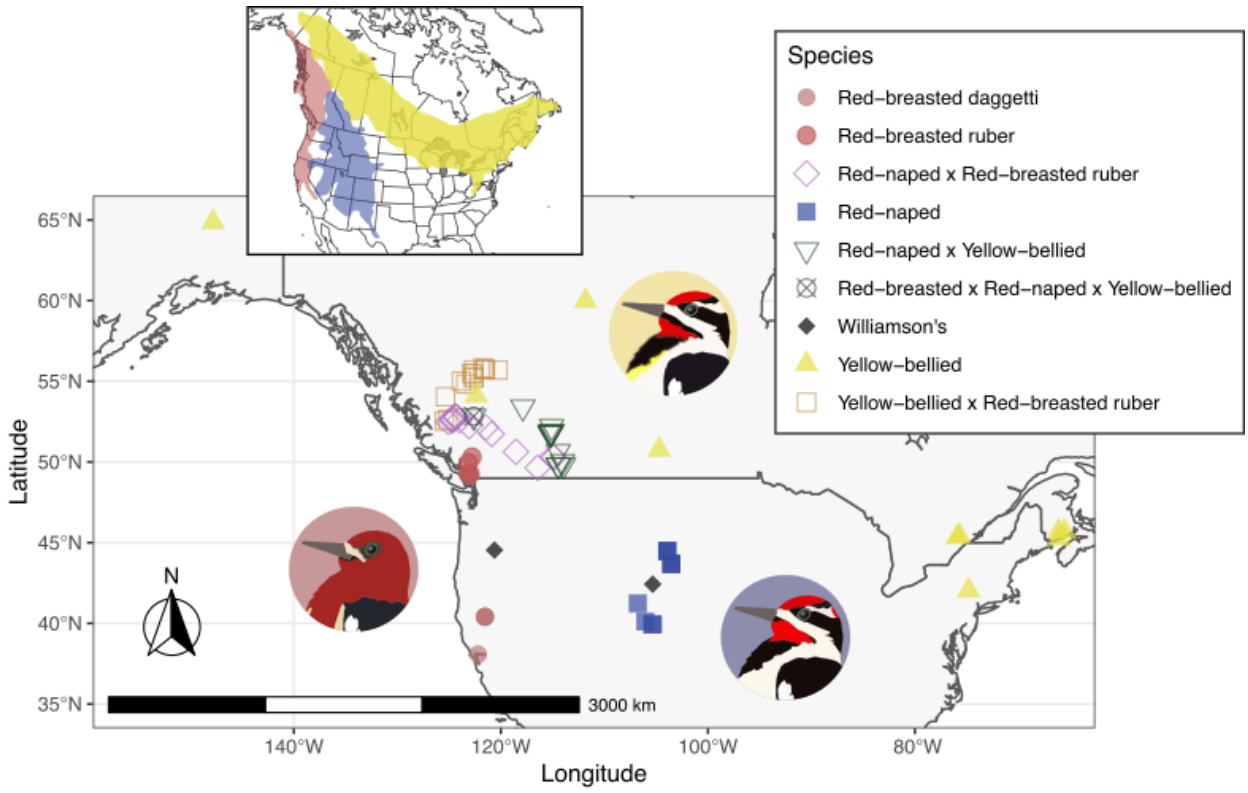



Figure 1 Sampling map of birds included in the WGS sequencing dataset (see Supplemental Table S1 for more details).

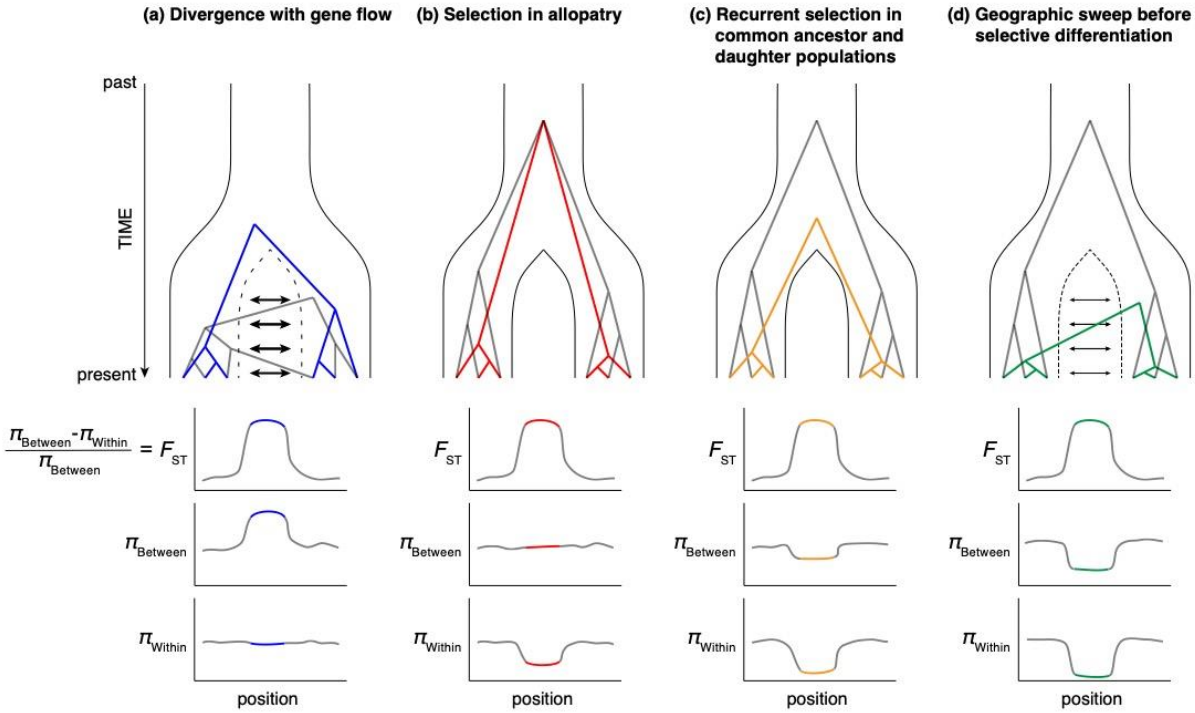



Figure 2 Illustration of four models of how selection can cause genomic regions to have high relative differentiation ($F_{ST}$), along with the expected effects on between-group nucleotide distance ($\pi_{Between}$) and within-group nucleotide diversity ($\pi_{Within}$). Adapted from Irwin et al. (2018).

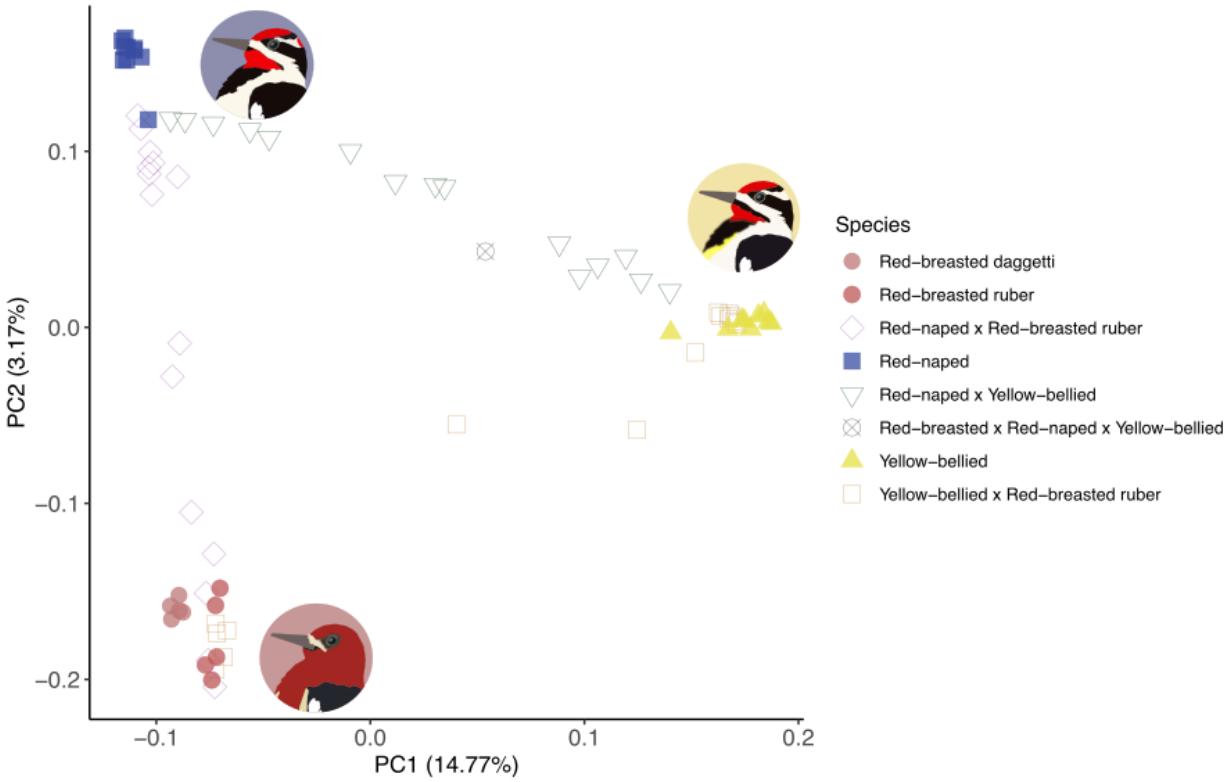



Figure 3 PCA of WGS genotypic data. Yellow dots indicate allopatric yellow-bellied sapsuckers, red dots indicate allopatric red-breasted, blue dots are allopatric red-naped. Orange dots show birds sampled in yellow-bellied × red-breasted sympatry, purple denote red-breasted × red-naped hybrid zone birds, and green dots show birds sampled in red-naped × yellow-bellied sympatry.

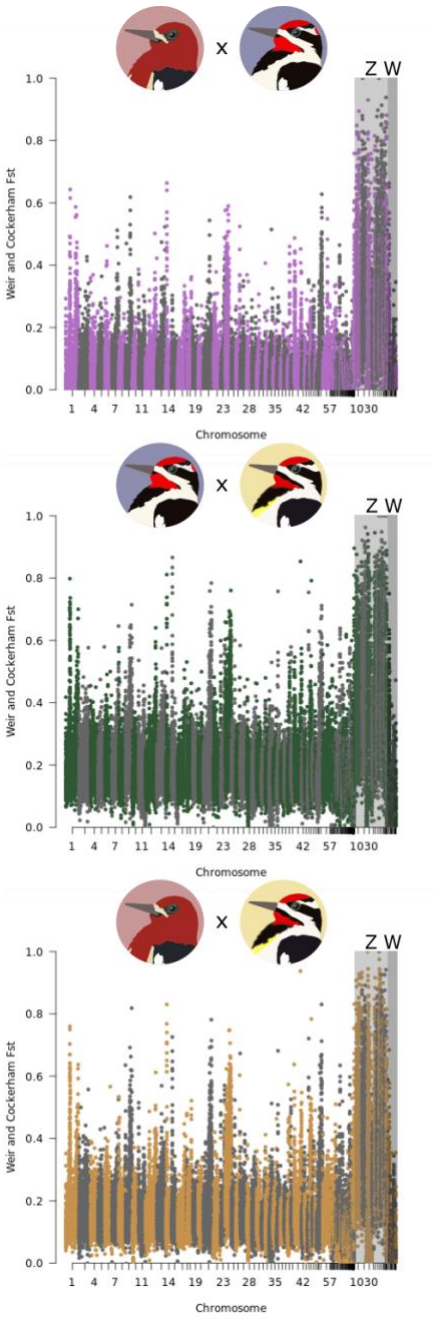



Figure 4 Genomic variation in pairwise $F_{ST}$ between (a) red-breasted and red-naped sapsuckers, (b) red-naped and yellow-bellied sapsuckers, and (c) red-breasted and yellow-bellied sapsuckers. Each dot represents a weighted average over a 25 kb window. Light grey box shading indicates scaffolds identified as likely to be on the Z chromosome and darker grey box shading designates those on the W chromosome.

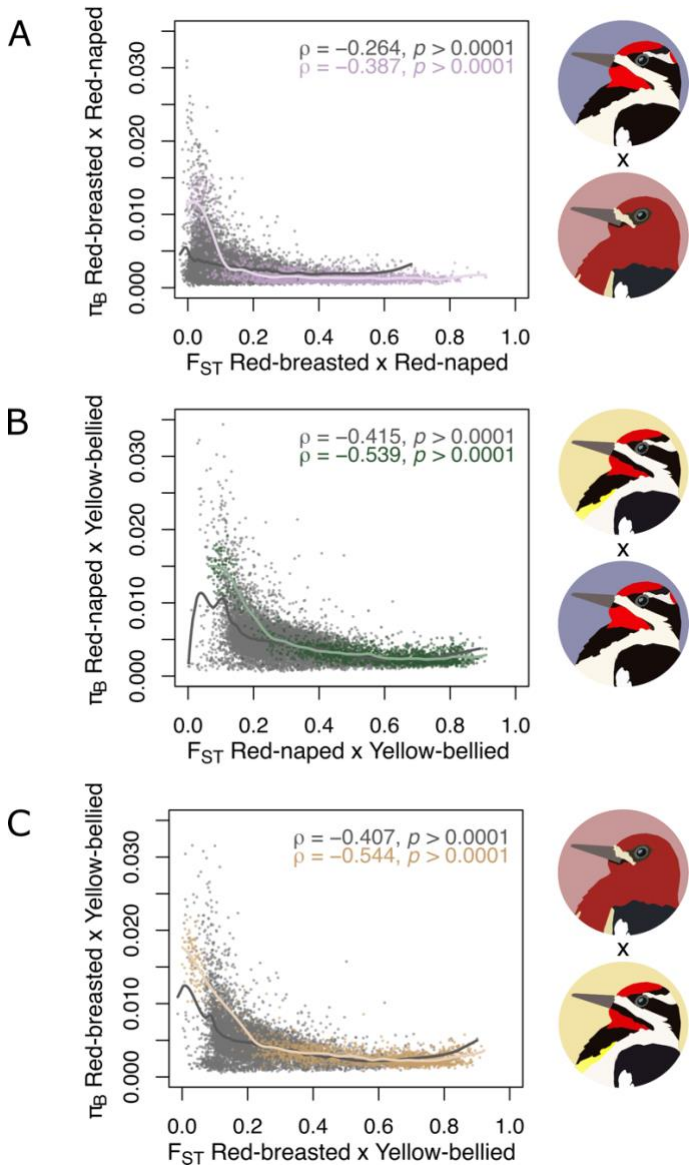



Figure 2 Comparisons of mean $F_{ST}$ and $\pi_B$ measured in 50 kb sliding windows for red-breasted and red-naped (A), red-naped and yellow-bellied (B), and red-breasted and yellow-bellied (C). Autosomal loci are shown in grey, and scaffolds presumed to be on the Z chromosome are shown in color. Spearman's correlation rho and p values are shown for each species pair and chromosome category, with autosomal loci values written in grey text and Z chromosomal loci in color.

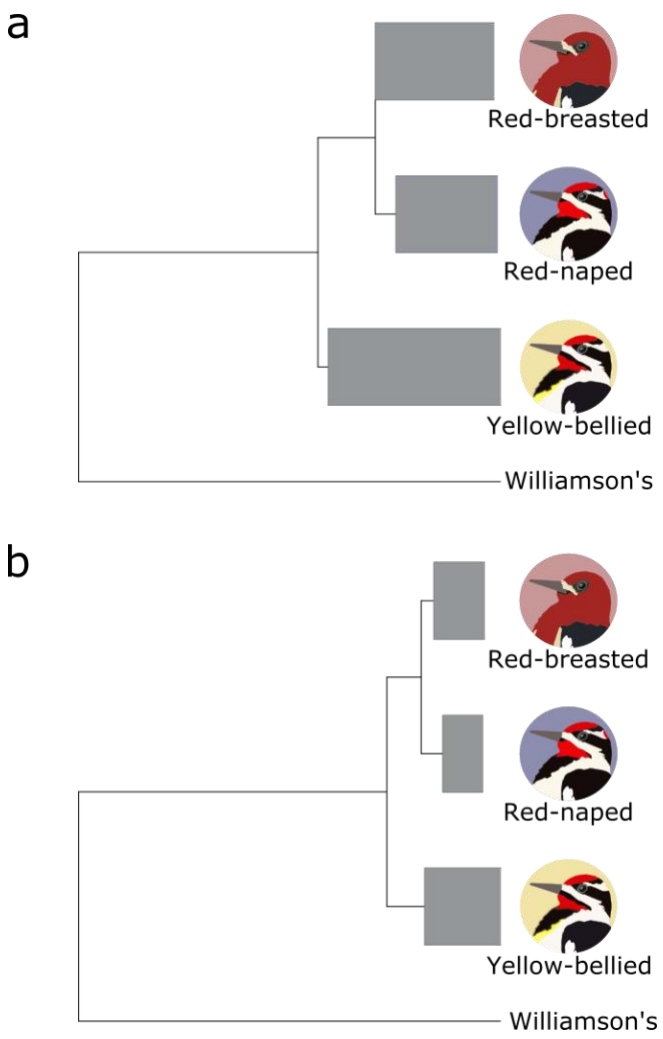



Figure 6 Neighbor-joining trees illustrating phylogenetic relationships among sapsuckers for autosomal (a), and Z-linked regions (b). Tree branch lengths are based on average between-taxon nucleotide distances ($\pi_B$) and within-taxon between-individual nucleotide distances ($\pi_W$, displayed as length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

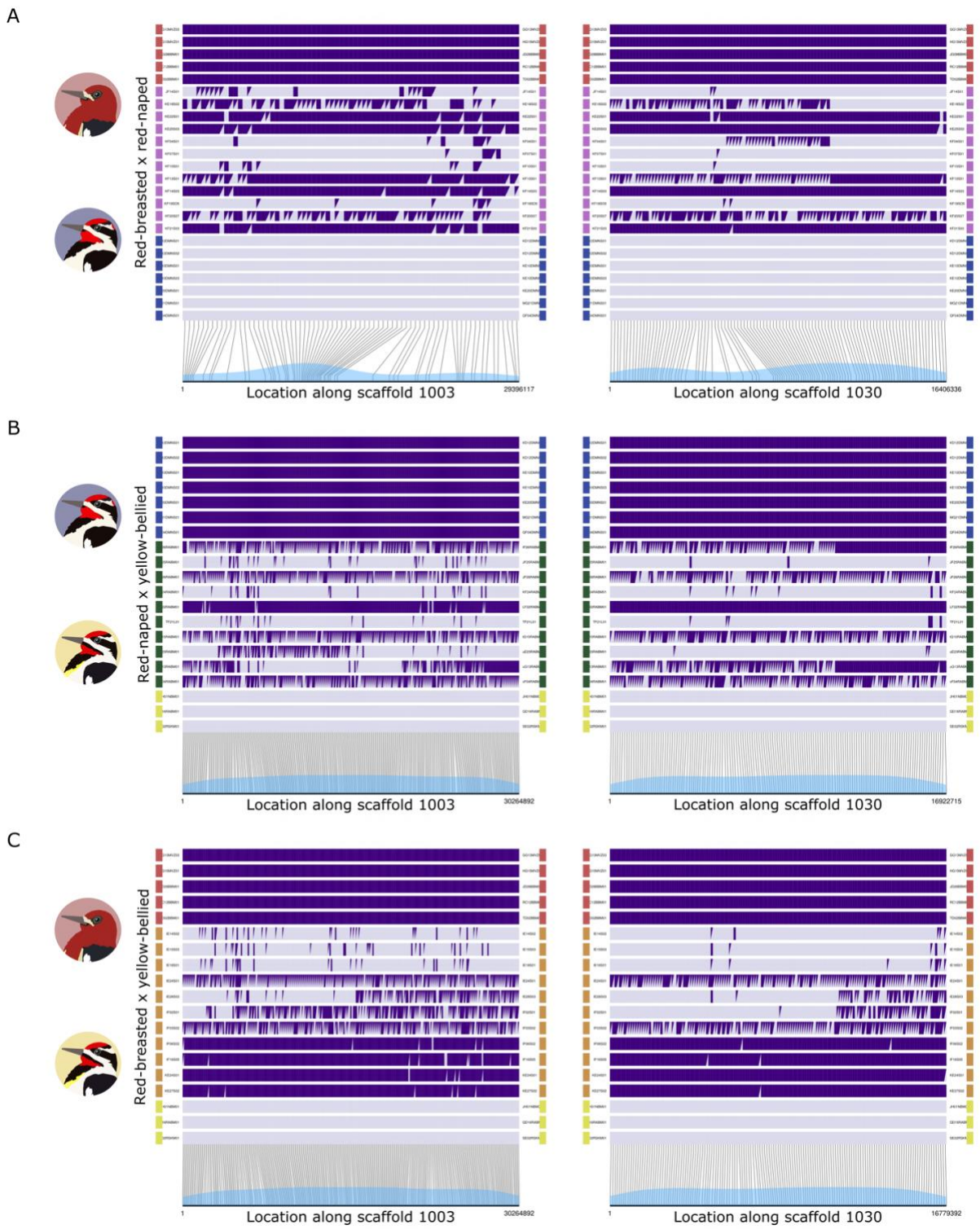



Figure 7 Genotype by individual plots showing strong Z chromosome differentiation between pairs of *Sphyrapicus* sapsucker species and clear evidence of recombination in hybrids. Panel A

shows red-breasted / red-naped sapsuckers, panel B shows red-naped / yellow-bellied sapsuckers, and panel C shows red-breasted / red-naped sapsuckers. Two Z-chromosome scaffolds are shown: scaffold 1003 on left and scaffold 1030 on right. Rows within each panel correspond to individuals, with colors on left and right indicating the species or hybrid group they belong to: red-breasted (red), red-breasted × red-naped hybrid zone (purple), red-naped (blue), red-naped × yellow-bellied hybrid zone (green), red-breasted × yellow-bellied hybrid zone (orange), or yellow-bellied sapsuckers (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). Alleles colored dark or light purple can differ between panels A, B, and C, with dark purple assigned to the first group in each comparison. Lines on the bottom of the plot indicate chromosomal position of each genotype along the scaffold. Only males are included on these plots, since females are hemizygous for the Z chromosome.

Supplement

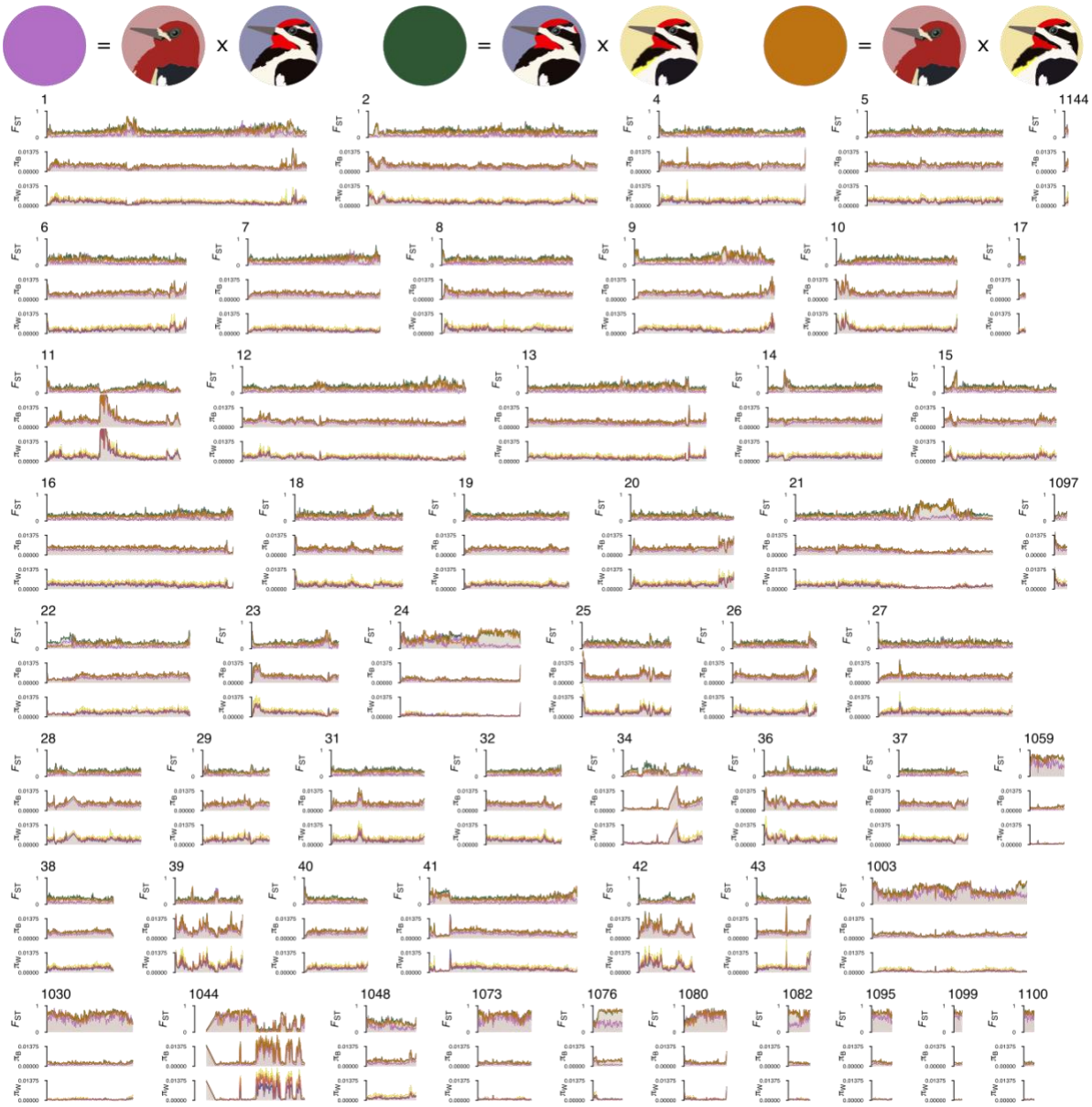
Supplemental Table S1 Sample metadata, ID includes Irwin Lab ID code, Museum/Band ID includes either museum accession code or CWS bird band record. Museum specifies which museum provided the tissue, or if it was trapped as part of Irwin Lab activities. Species use AOU four-letter species codes RBSA (red-breasted sapsucker), with d indicating *daggeti* and r indicating *ruber* subspecies, YBSA (yellow-bellied sapsucker), RNSA (red-naped sapsucker), and WISA (Williamson's sapsucker) in allopatry; and YBxRB (yellow-bellied x red-breasted), RNxYB (red-naped x yellow-bellied), RBxRN (red-breasted x red-naped), and RNxYBxRB (red-naped x yellow-bellied x red-breasted) in sympatry. Loc refers to the state or province code where the bird was trapped. Date, Lat, and Long indicate the date and location of sample capture. Sex indicates sex ID as determined by genotypes.

| ID | Museum/Band ID | Museum | Species Zone | Loc | Date | Lat | Long | Sex |
|---|---|---|---|---|---|---|---|---|
| EA22MVZ01 | MVZ:Bird:181742 | MVZ | RBSAd | CA | 2005-01-22 | 38.0807 | -122.1911 | F |
| GG13MVZ01 | MVZ:Bird:191404 | MVZ | RBSAd | CA | 2007-07-13 | 40.4126 | -121.5308 | F |
| GG13MVZ02 | MVZ:Bird:191399 | MVZ | RBSAd | CA | 2007-07-13 | 40.4126 | -121.5308 | F |
| GG13MVZ03 | MVZ:Bird:191401 | MVZ | RBSAd | CA | 2007-07-13 | 40.4126 | -121.5308 | M |
| HF14AMN01 | PRS 3051 | AMNH | YBSA | NY | 2008-06-14 | 42.0413 | -74.8075 | F |
| HG15MVZ01 | MVZ:Bird:182366 | MVZ | RBSAd | CA | 2006-07-15 | 40.3788 | -121.5660 | M |
| IE10S04 | 2251-49108 | Irwin Lab | YBxRB | BC | 2009-05-10 | 55.6969 | -120.2700 | F |
| IE14S02 | 2251-49115 | Irwin Lab | YBxRB | BC | 2009-05-14 | 55.7217 | -121.5736 | M |
| IE15S03 | 2251-49119 | Irwin Lab | YBxRB | BC | 2009-05-15 | 55.7503 | -121.5214 | M |
| IE18S01 | 2251-49122 | Irwin Lab | YBxRB | BC | 2009-05-18 | 55.8267 | -121.7233 | M |
| IE21S02 | 2251-49132 | Irwin Lab | YBxRB | BC | 2009-05-21 | 55.7117 | -121.4925 | F |
| IE24S01 | 2251-49134 | Irwin Lab | YBxRB | BC | 2009-05-24 | 55.4953 | -122.8206 | M |
| IE28S03 | 2251-49147 | Irwin Lab | YBxRB | BC | 2009-05-28 | 55.1958 | -122.6292 | M |

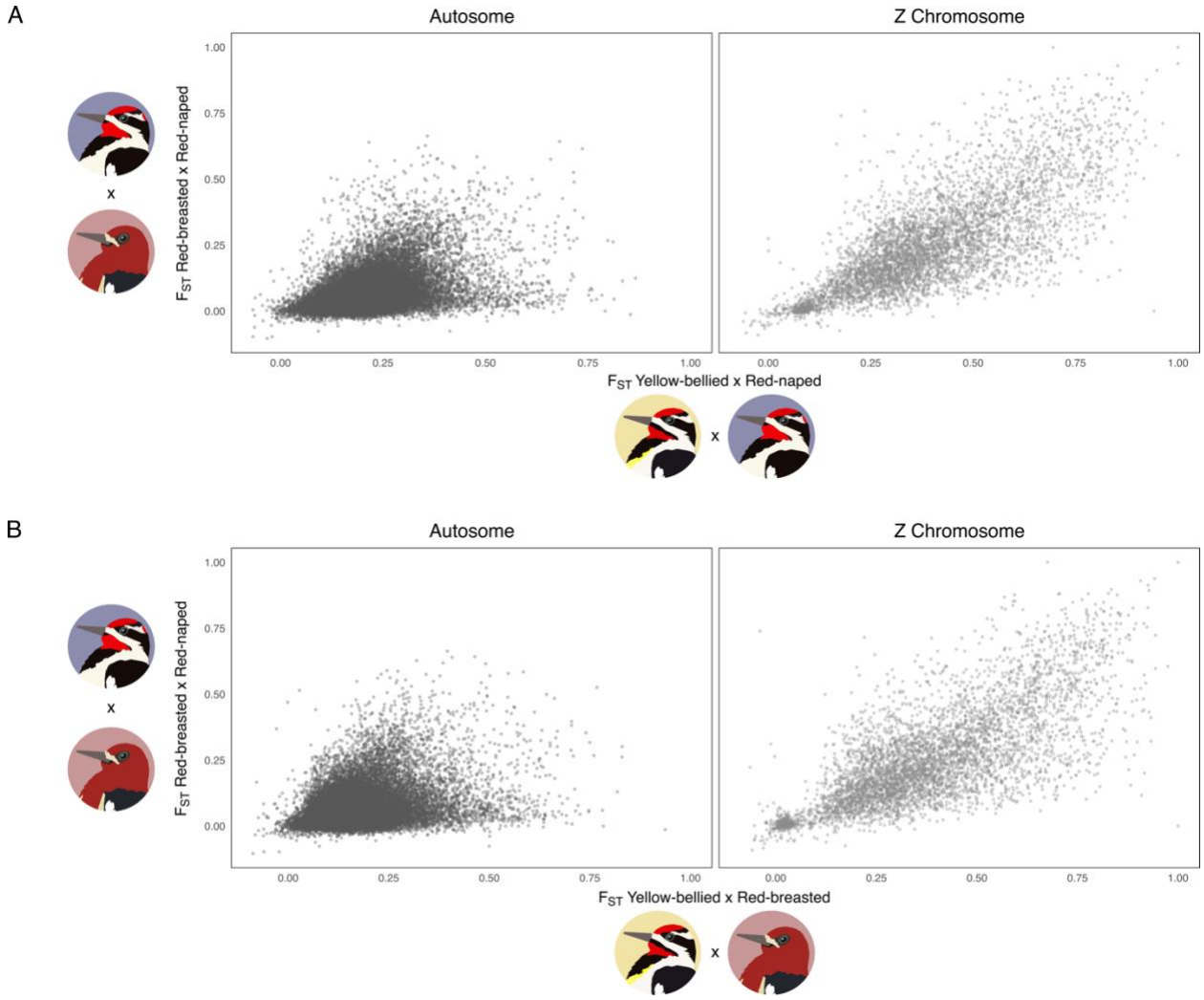
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| IF02S01 | 2251-49149 | Irwin Lab | YBxRB | BC | 2009-06-02 | 55.7008 | -122.6267 | M |
| IF03S02 | 2251-49151 | Irwin Lab | YBxRB | BC | 2009-06-03 | 55.2686 | -122.6028 | M |
| IF03S04 | 2251-49152 | Irwin Lab | YBxRB | BC | 2009-06-03 | 55.4611 | -122.6400 | F |
| IF06S02 | 2251-49155 | Irwin Lab | YBxRB | BC | 2009-06-06 | 54.8333 | -123.3836 | M |
| IF13S02 | 2251-49171 | Irwin Lab | YBxRB | BC | 2013-06-09 | 55.0061 | -123.8108 | F |
| IF16S05 | 1841-90618 | Irwin Lab | YBxRB | BC | 2009-06-16 | 54.0481 | -125.3497 | M |
| IF26RABM01 | Z09.8.5 | RABM | RNxYB | AB | 2009-06-26 | 51.8912 | -115.0333 | M |
| JD28BBM01 | TB001883 | BBM | RBSAr | BC | 2010-04-28 | 50.3068 | -122.7821 | M |
| JF14S01 | 1841-90626 | Irwin Lab | RBxRN | BC | 2010-06-14 | 52.1469 | -121.4581 | M |
| JF25RABM01 | Z10.3.9 | RABM | RNxYB | AB | 2010-06-25 | 51.8722 | -115.1365 | M |
| JF26RABM01 | Z10.3.13 | RABM | RNxYB | AB | 2010-06-26 | 51.8059 | -115.2022 | M |
| JF27NBM01 | NBM- 013432 | NBM | YBSA | NB | 2010-06-27 | 45.6878 | -66.1117 | F |
| JF30RABM01 | Z10.13.14 | RABM | RNxYB | AB | 2010-06-30 | 51.8202 | -115.1865 | F |
| JH01NBM01 | NBM- 012066 | NBM | YBSA | NB | 2010-08-01 | 45.6834 | -65.5333 | M |
| KD12DMNS01 | DMNS:Bird:45554 | DMNS | RNSA | SD | 2011-04-12 | 44.4833 | -103.9333 | M |
| KD12DMNS02 | DMNS:Bird:45556 | DMNS | RNSA | SD | 2011-04-12 | 44.4833 | -103.9333 | M |
| KE10DMNS01 | DMNS:Bird:45539 | DMNS | RNSA | SD | 2011-05-10 | 43.6821 | -103.5363 | M |
| KE10DMNS02 | DMNS:Bird:45540 | DMNS | RNSA | SD | 2011-05-10 | 43.6821 | -103.5363 | F |
| KE10DMNS03 | DMNS:Bird:45541 | DMNS | RNSA | SD | 2011-05-10 | 43.6821 | -103.5363 | M |
| KE18S02 | 2251-49198 | Irwin Lab | RBxRN | BC | 2011-05-18 | 52.2206 | -123.0758 | M |
| KE20DMNS01 | DMNS:Bird:34830 | DMNS | RNSA | CO | 2011-05-20 | 40.1146 | -106.0476 | M |
| KE20S02 | 1841-90646 | Irwin Lab | RBxRN | BC | 2011-05-20 | 52.8581 | -124.4383 | F |
| KE22S01 | 1841-90649 | Irwin Lab | RBxRN | BC | 2011-05-22 | 52.5933 | -123.9911 | M |
| KE24S01 | 1841-90653 | Irwin Lab | YBxRB | BC | 2011-05-24 | 52.5331 | -125.4647 | M |
| KE25S03 | 1841-90659 | Irwin Lab | RBxRN | BC | 2011-05-25 | 52.4903 | -125.0964 | M |
| KE27S02 | 1841-90662 | Irwin Lab | YBxRB | BC | 2011-05-27 | 52.6217 | -125.1247 | M |
| KF04S01 | 2251-49187 | Irwin Lab | RBxRN | BC | 2011-06-04 | 51.7361 | -120.8856 | M |
| KF07S01 | 2251-49310 | Irwin Lab | RBxRN | BC | 2011-06-07 | 50.3497 | -115.0022 | M |
| KF07S03 | 2251-49312 | Irwin Lab | RBxRN | BC | 2011-06-07 | 50.2753 | -114.8914 | F |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| KF10S01 | 2251-49317 | Irwin Lab | RBxRN | BC | 2011-06-10 | 49.6467 | -116.4442 | M |
| KF11S02 | 2251-49320 | Irwin Lab | RBxRN | BC | 2011-06-11 | 50.6364 | -118.5450 | F |
| KF13S01 | 2251-49327 | Irwin Lab | RBxRN | BC | 2011-06-13 | 52.7661 | -124.3747 | M |
| KF14S03 | 1841-90675 | Irwin Lab | RBxRN | BC | 2011-06-14 | 52.7853 | -124.2825 | M |
| KF19SO5 | 2251-49346 | Irwin Lab | RBxRN | BC | 2011-06-19 | 52.6317 | -124.6400 | M |
| KF20S07 | 2251-49353 | Irwin Lab | RBxRN | BC | 2011-06-20 | 52.6594 | -124.6256 | M |
| KF21S03 | 1841-90676 | Irwin Lab | RBxRN | BC | 2011-06-21 | 52.5617 | -124.7444 | M |
| KF24RABM01 | Z11.4.30 | RABM | RNxYB | AB | 2011-06-24 | 53.4005 | -117.8685 | M |
| KG01DMNS01 | DMNS:Bird:45552 | DMNS | RNSA | SD | 2011-07-01 | 44.4833 | -103.9333 | F |
| KG07RABM01 | Z11.5.10 | RABM | RNxYB | AB | 2011-07-07 | 51.7774 | -115.2274 | F |
| LE29RABM01 | Z12.2.20 | RABM | RNxYB | AB | 2012-05-29 | 49.7995 | -114.4678 | F |
| LE29RABM02 | Z12.2.7 | RABM | RNxYB | AB | 2012-05-29 | 49.8215 | -114.4246 | F |
| LE30RABM01 | Z12.2.27 | RABM | RNxYB | AB | 2012-05-30 | 50.0243 | -114.0447 | F |
| LF22RABM01 | Z12.3.9 | RABM | RNxYB | AB | 2012-06-22 | 50.6437 | -114.4661 | M |
| MG13RABM01 | Z13.5.7 | RABM | YBSA | AB | 2013-07-06 | 59.9779 | -111.8291 | F |
| MG21DMNS01 | DMNS:Bird:52692 | DMNS | RNSA | CO | 2011-05-21 | 39.9259 | -105.3978 | M |
| ND10L01 | TB001343 | BBM | RBSAr | BC | 2014-04-10 | 49.2422 | -122.9569 | F |
| ND16DMNS01 | DMNS:Bird:43558 | DMNS | RNSA | CO | 2014-04-16 | 39.9337 | -105.3024 | F |
| QD16RABM01 | Z17.10.38 | RABM | YBSA | ON | 2016-04-16 | 45.3443 | -75.9244 | M |
| QD21RABM01 | Z17.10.37 | RABM | YBSA | ON | 2016-04-21 | 45.4180 | -75.7026 | F |
| QF04DMNS01 | DMNS:Bird:47794 | DMNS | RNSA | WY | 2016-06-04 | 41.2333 | -106.7745 | M |
| QF16DMNS01 | DMNS:Bird:46075 | DMNS | WISA | WY | 2016-06-16 | 42.4241 | -105.3365 | M |
| RC06BBM01 | TB001886 | BBM | RBSAr | BC | 2017-03-10 | 49.2236 | -123.1080 | F |
| RC12BBM01 | TB001887 | BBM | RBSAr | BC | 2017-03-12 | 49.2930 | -123.1403 | M |
| RE03NBM01 | NBM- 013265 | NBM | YBSA | NB | 2017-05-03 | 45.2419 | -66.0980 | F |
| RF19DMNS01 | DMNS:Bird:47932 | DMNS | WISA | OR | 2017-06-19 | 44.5388 | -120.6212 | F |
| SE02RSKM01 | RSKM_BIRD_A-9388 | RSKM | YBSA | SK | | 50.7554 | -104.7088 | M |
| TD02BBM01 | TB001884 | BBM | RBSAr | BC | 2019-04-02 | 49.9572 | -123.1202 | M |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| TF15L03 | 1951-58115 | Irwin Lab | YBSA | BC | 2019-06-15 | 54.1097 | -122.3853 | F |
| TF21L01 | 1951-58120 | Irwin Lab | RNxYB | BC | 2019-06-21 | 52.8331 | -122.6220 | M |
| TF21L02 | 1951-58121 | Irwin Lab | RNxYBxRB | BC | 2019-06-21 | 52.8331 | -122.6220 | F |
| tG15RABM01 | Z94.13.11 | RABM | RNxYB | AB | 1994-07-15 | 51.8300 | -115.2200 | M |
| uE23RABM01 | Z95.9.1 | RABM | RNxYB | AB | 1995-05-23 | 52.2000 | -115.1000 | M |
| uG13RABM01 | Z95.12.4 | RABM | RNxYB | AB | 1995-07-13 | 51.7700 | -115.2200 | M |
| vF04RABM01 | Z96.18.9 | RABM | RNxYB | AB | 1996-06-04 | 49.8200 | -113.9500 | M |
| YE19UAM01 | UAM:Bird:28585 | UAM | YBSA | AK | 2011-07-10 | 64.8570 | -147.8043 | F |



Supplemental Figure S1 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), $\pi_B$ (middle), and $\pi_W$ (bottom), for all three species/species pairs. The $F_{ST}$ and $\pi_B$ species pair lines show red-breasted × red-naped in purple, red-naped × yellow-bellied in green, and red-breasted × yellow-bellied in orange. For $\pi_W$ within species rows indicate red-breasted in red, red-naped in blue, and yellow-bellied in yellow.




Supplemental Figure S2 Scatterplots illustrating the relationship between the $F_{ST}$ of more distantly related sapsuckers (x-axis) and the $F_{ST}$ of the sister species (y-axis) for each 25kb window, paneled separately for autosomal scaffolds (left panels) and Z chromosome scaffolds (right panels). The top panels (A) show how differentiation in Red-breasted and Red-naped sapsuckers compares to differentiation among Yellow-bellied and Red-naped in the same region. The bottom panels (B) show comparisons between Red-breasted and Red-naped $F_{ST}$ versus Yellow-bellied and Red-breasted $F_{ST}$.