Received: 29 January 2018 | Revised: 10 July 2018 | Accepted: 14 July 2018

DOI: 10.1111/mec.14858

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

A comparison of genomic islands of differentiation across three young avian species pairs

Darren E. Irwin[1] | Borja Milá[2] | David P. L. Toews[1,3] | Alan Brelsford[1,4] | Haley L. Kenyon[1,5] | Alison N. Porter[1] | Christine Grossen[1,6] | Kira E. Delmore[1,7] | Miguel Alcaide[1,8] | Jessica H. Irwin[1]

[1]Department of Zoology and Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, Canada

[2]National Museum of Natural Sciences, Spanish National Research Council (CSIC), Madrid, Spain

[3]Cornell Lab of Ornithology & Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York

[4]Department of Evolution, Ecology and Organismal Biology, University of California, Riverside, California

[5]Department of Biology, Queen's University, Biosciences Complex, Kingston, Ontario, Canada

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

[7]Max Planck Institute for Evolutionary Biology, Behavioural Genomics, Plön, Germany

[8]Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, Canada

Correspondence
Darren E. Irwin, Department of Zoology, Vancouver, BC, Canada.
Email: irwin@zoology.ubc.ca

Funding information
Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: 311931

Abstract

Detailed evaluations of genomic variation between sister species often reveal distinct chromosomal regions of high relative differentiation (i.e., "islands of differentiation" in F_{ST}), but there is much debate regarding the causes of this pattern. We briefly review the prominent models of genomic islands of differentiation and compare patterns of genomic differentiation in three closely related pairs of New World warblers with the goal of evaluating support for the four models. Each pair (MacGillivray's/mourning warblers; Townsend's/black-throated green warblers; and Audubon's/myrtle warblers) consists of forms that were likely separated in western and eastern North American refugia during cycles of Pleistocene glaciations and have now come into contact in western Canada, where each forms a narrow hybrid zone. We show strong differences between pairs in their patterns of genomic heterogeneity in F_{ST}, suggesting differing selective forces and/or differing genomic responses to similar selective forces among the three pairs. Across most of the genome, levels of within-group nucleotide diversity (π_{Within}) are almost as large as levels of between-group nucleotide distance ($\pi_{Between}$) within each pair, suggesting recent common ancestry and/or gene flow. In two pairs, a pattern of the F_{ST} peaks having low $\pi_{Between}$ suggests that selective sweeps spread between geographically differentiated groups, followed by local differentiation. This "sweep-before-differentiation" model is consistent with signatures of gene flow within the yellow-rumped warbler species complex. These findings add to our growing understanding of speciation as a complex process that can involve phases of adaptive introgression among partially differentiated populations.

KEYWORDS
genomic differentiation, hybridization, islands of differentiation, Parulidae, speciation, warbler

1 | INTRODUCTION

A central goal of biological science has been to understand how much genetic differentiation occurs between species, how that differentiation is structured across the genome, and what processes generate that structuring. Given the ongoing rapid advances in DNA sequencing technology, such questions are receiving renewed attention, and answers are gradually emerging (Delmore et al., 2018; Ravinet et al., 2017; Samuk et al., 2017; Toews, Campagna, et al., 2016; Wolf & Ellegren, 2016). However, there is currently no consensus regarding the causes of the often-observed pattern of distinct chromosomal regions with strikingly high between-population genetic

differentiation against a background of low differentiation across most of the genome (Cruickshank & Hahn, 2014; Westram & Ravinet, 2017).

Such regions have been referred to as "islands of differentiation," "islands of divergence" or "islands of speciation." Differentiation is typically measured using the statistic F_{ST} (Weir & Cockerham, 1984; Wright, 1949), a measure of relative differentiation in allele frequencies between two populations. In many studies of genomic differentiation between closely related species, F_{ST} is high (near one) in distinct regions of particular chromosomes, but close to zero over much of the rest of the genome. The cause of this pattern is under much debate (Cruickshank & Hahn, 2014; Nachman & Payseur, 2012; Nosil & Feder, 2012; Ravinet et al., 2017). In the absence of selection, genomic variation in F_{ST} due to genetic drift alone is well approximated by a χ^2 distribution under a variety of simulated demographic histories (Lewontin & Krakauer, 1973; Whitlock & Lotterhos, 2015). Observed F_{ST} distributions often differ strongly from that expectation, with much of the genome showing low F_{ST} and a small subset showing highly elevated F_{ST}, leading most researchers to invoke explanations involving natural selection.

Selection-based explanations for islands of differentiation are all founded on the concept that physically linked loci tend to show similar patterns of variation (Wu, 2001). The stronger the physical linkage and the stronger the selection, the more similar the patterns. Hence, neutral loci linked to selected loci are said to be under indirect selection. Positive selection favouring particular alleles causes genetic hitchhiking at nearby neutral sites (Fay & Wu, 2000; Feder & Nosil, 2010; Via, 2009; Via & West, 2008). Negative selection against novel mutations (i.e., background selection) can result in reduced effective population size, lowering diversity at linked neutral sites (Charlesworth, Morgan, & Charlesworth, 1993).

1.1 | Four models of genomic islands of differentiation

Four prominent selection-based models for the formation of genomic islands of differentiation have been proposed. These models differ in (a) the role of gene flow vs. geographic isolation and (b) the type of selection driving the process, and whether that selection occurs repeatedly. Real cases of population differentiation are likely more complex than each scenario described by these simple models, and combinations of these processes could occur. Nonetheless, considering these simple models and the patterns they predict can help us to exclude some scenarios and better infer history from empirical data. Distinguishing these models requires understanding how F_{ST} is related to between- and within-population nucleotide distances (Cruickshank & Hahn, 2014; Han et al., 2017; Irwin, Alcaide, Delmore, Irwin, & Owens, 2016). Each of these is calculated as the average proportion of nucleotides that differ between two homologous sequences. We refer to these metrics as $\pi_{Between}$ (abbreviated as π_B) and π_{Within} (π_W), following similar terminology as Charlesworth (1998); we suggest these symbols better convey the relationship of the two metrics compared to the often-used "d_{XY}" and "π." Assuming a constant mutation rate, both $\pi_{Between}$ and π_{Within} have expected values that are proportional to the coalescence time of two sequences (Charlesworth, 1998; Slatkin, 1991; Figure 1). These have a simple relationship with F_{ST} (see equations in Figure 1; Charlesworth, 1998; Hudson, Slatkin, & Maddison, 1992), such that F_{ST} can be interpreted as the proportion of between-population coalescent time that is greater than within-population coalescent time. Asking about the causes of high-F_{ST} regions is the same as asking about how the ratio of within-population to between-population coalescent times becomes low.

The first model, "divergence-with-gene-flow" (Figure 2a), envisions a locus that contributes to reproductive isolation between differentiating populations that are in physical and genetic contact (Feder & Nosil, 2010; Via, 2009; Wu, 2001). Due to selection at that locus, the chromosomal region closely linked to that locus has reduced gene flow between the two populations. Physically unlinked parts of the genome can flow more freely. Here, $\pi_{Between}$ is predicted to be higher in regions of high F_{ST} than in regions of low F_{ST} (Cruickshank & Hahn, 2014). The reduced gene flow at regions that cause reproductive isolation leads to both higher F_{ST} and higher average coalescent times; other parts of the genome can move between the two populations, keeping both F_{ST} and $\pi_{Between}$ low.

In contrast, the "selection-in-allopatry" model (Figure 2b) relies on within-population selection to explain regions of high relative differentiation between populations (Burri, 2017a; Burri et al., 2015; Cruickshank & Hahn, 2014; Nachman & Payseur, 2012; Noor & Bennett, 2009; Vijay et al., 2017). After a species is divided into two populations, selection on distinct regions of the genome can lead to lower within-population diversity (π_{Within}) in those regions, hence higher F_{ST}. In pure selection-in-allopatry, $\pi_{Between}$ is expected to be similar on average between areas of high F_{ST} and low F_{ST}, because selection in the two isolated populations decreases within-group diversity but has no effect on expected between-group nucleotide distance (Cruickshank & Hahn, 2014; Han et al., 2017). The selection-in-allopatry model is sometimes referred to as "linked selection," but we avoid that term because both linkage and selection are also central to the divergence-with-gene-flow model (Burri et al., 2015; Feder & Nosil, 2010; Nachman & Payseur, 2012) and the other models considered here.

Neither of these above two models explain a commonly observed pattern: that of regions of high F_{ST} having low $\pi_{Between}$ compared to other parts of the genome (Cruickshank & Hahn, 2014; Van Doren et al., 2017; Zhang et al., 2017). This led to the development of the "recurrent selection" model (Figure 2c; Cruickshank & Hahn, 2014; Nachman & Payseur, 2012). Here, certain regions of the genome experienced more intense selection (either positive or negative) in the common ancestor, reducing standing variation in those regions before that common ancestor split into the two current populations. Then, those same regions experienced diversity-reducing selection in two allopatric daughter populations. The effect is that the current regions of higher F_{ST} (due to recent selection) also tend to have lower $\pi_{Between}$ (due to selection in the common ancestor).

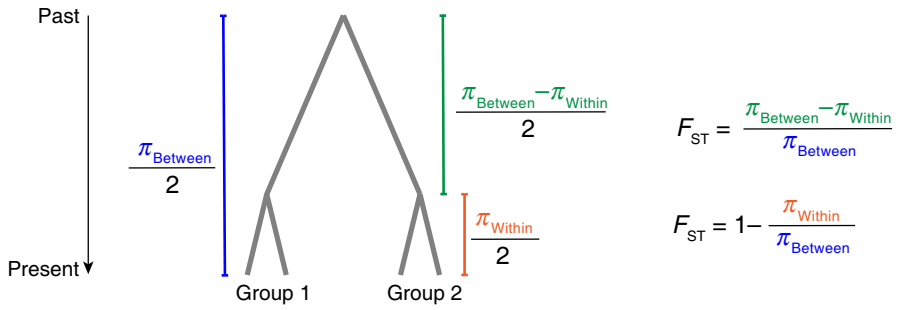

FIGURE 1 Illustration of how relative differentiation (F_{ST}) is related to average pairwise nucleotide distance between sequences within the same population (π_{Within}) and average pairwise nucleotide distance between sequences in different populations ($\pi_{Between}$). If mutation rate is constant, then average pairwise nucleotide distances are proportional to average pairwise coalescent times. The figure shows the relationships of four sequences. Note that the equations shown here are appropriate for large and equal sample sizes, whereas the actual calculation of F_{ST} as used in this study corrects for small and/or unequal sample size, following Weir and Cockerham (1984)

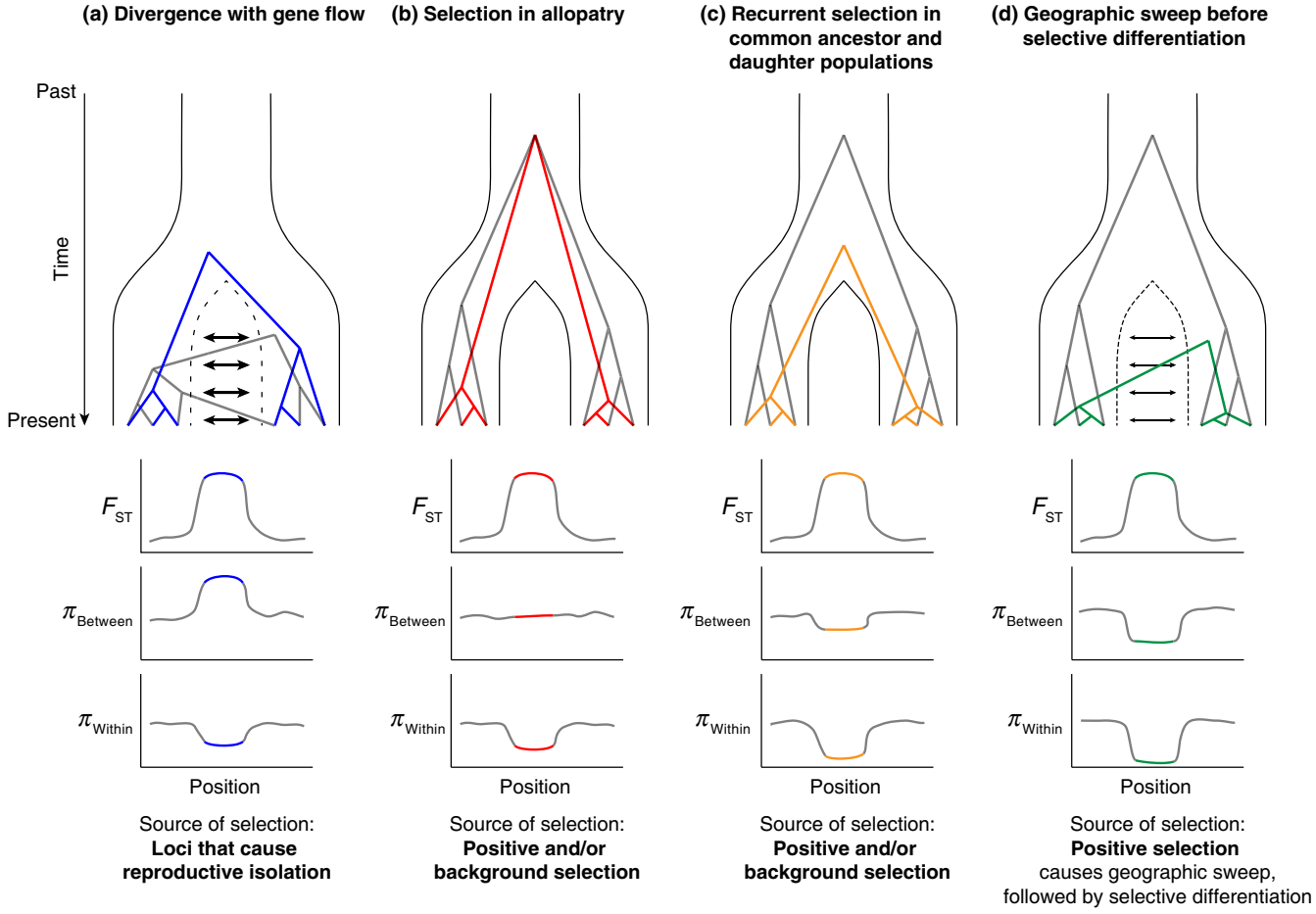

FIGURE 2 Illustrations of four models for the formation of genomic islands of relative differentiation. Each of the top panels illustrates a single population splitting into two over time, along with example gene genealogies (of three individuals in each population) for a neutral part of the genome (in grey) and a part of the genome subject to selection (in colour). Also shown are variation in F_{ST}, $\pi_{Between}$, and π_{Within} across an example chromosome, with the region under selection in the middle and neutral regions on either side. Note the differing relationships between $\pi_{Between}$ and F_{ST}. Finally, sources of selection in each model are listed. These illustrations build on those of Cruickshank and Hahn (2014) and Irwin et al. (2016)

While the "recurrent selection" model explains some of the patterns observed, it was envisioned in the context of no gene flow between daughter populations (Cruickshank & Hahn, 2014). Observed associations between islands of differentiation and low $\pi_{Between}$ in population pairs where current hybridization is known (Delmore et al., 2015; Irwin et al., 2016) led us to develop a fourth

model: "geographic-sweep-before-selective-differentiation," which we refer to in brief as "sweep-before-differentiation" (Figure 2d). In a geographically structured species complex, advantageous alleles could spread throughout the whole species (e.g., across hybrid zones) faster than neutral parts of the genome, reducing $\pi_{Between}$ in those regions. Subsequent selection in local populations could then reduce π_{Within}, driving up F_{ST}.

These four models emphasize different types of selection. In the divergence-with-gene-flow model, loci causing reproductive isolation are the source of selection. The selection-in-allopatry and recurrent selection models envision either positive or background selection as the cause of reductions of within-population diversity (Cruickshank & Hahn, 2014). In contrast, sweep-before-differentiation emphasizes positive selection, as only that can lead to geographic sweeps. This distinction is important because the influence of background selection is often viewed as being more dependent than directional selection on structural features of the genome, such as the recombination landscape, location of centromeres and gene density (e.g., Burri et al., 2015).

1.2 | Comparisons among species pairs

Comparisons of genomic differentiation landscapes across multiple pairs of populations or species can help distinguish among these models for the formation of islands of differentiation (Delmore et al., 2018; Ellegren & Wolf, 2017; Feulner et al., 2015; Irwin et al., 2016; Van Doren et al., 2017). Similarities in the differentiation landscape among different pairs can implicate phylogenetically conserved factors in shaping differentiation. Factors that are thought to be relatively conserved over long spans of evolutionary time include the recombination landscape (at least in birds; Singhal et al., 2015), the locations of centromeres, which may be associated with low recombination (Ellegren et al., 2012; Vijay et al., 2017), and the landscape of background selection (Burri, 2017a,b). In contrast, differences can implicate factors that are idiosyncratic to each pair, such as the form and targets of directional selection and the stochasticity of mutation (Ravinet et al., 2017).

Here, we compare patterns of differentiation across the genome in three pairs of wood warblers within the family Parulidae, using identical sequencing and analytical methodology (Figure 3). Each pair has a hybrid zone in western Canada that formed after the last major glaciation, such that the three pairs have similar geographical and historical contexts. The pair that is most distantly related to the others is the MacGillivray's and mourning warblers (*Geothlypis tolmiei/G. philadelphia*; Irwin, Brelsford, Toews, MacDonald, & Phinney, 2009; Kenyon, Toews, & Irwin, 2011). The more closely related two pairs are the Townsend's and black-throated green warblers (*Setophaga townsendi/S. virens*; Kenyon, Alcaide, Toews, & Irwin, 2017; Toews, Brelsford, & Irwin, 2011) and the Audubon's and myrtle warblers (*Setophaga [coronata] auduboni/S. [coronata] coronata*; Brelsford & Irwin, 2009; Hubbard, 1969). To enable a close look at gene flow and genomic differentiation within a large and geographically variable species complex, we also examine genomic patterns in the other members of the yellow-rumped warbler complex (see Figure 3 caption for citations and details): black-fronted warblers (in Mexico; *S. [coronata] nigrifrons*) and Goldman's warblers (in Guatemala; *S. [coronata] goldmani*).

We ask whether each species pair displays islands of differentiation and whether the genomic landscapes of relative and absolute differentiation are similar among the three species pairs. Close similarities in the landscape of relative differentiation would implicate phylogenetically conserved factors, such as conserved recombination landscapes, in explaining differentiation landscapes. We ask whether peaks of relative differentiation tend to have absolute nucleotide distance that is high (supporting pure divergence-with-gene-flow), average (consistent with pure selection-in-allopatry) or low (supporting recurrent selection and/or sweep-before-differentiation) compared to other parts of the genome. Finally, we analyse the yellow-rumped warbler complex in detail to infer whether there has been recent gene flow between geographically differentiated groups, an element of the sweep-before-differentiation model.

2 | MATERIALS AND METHODS

In previous work, we used a standard genotyping-by-sequencing (GBS) protocol (Alcaide, Scordato, Price, & Irwin, 2014; Elshire et al., 2011) to survey genomic variation within each of the three species complexes studied here (Figure 3): MacGillivray's/mourning warblers (Porter, 2015), Townsend's/black-throated green warblers (Kenyon et al., 2017) and Audubon's/myrtle/black-fronted/Goldman's warblers (Toews, Brelsford, Grossen, Milá, & Irwin, 2016). The analyses of GBS reads differed among those studies. Here, we have re-analysed these GBS reads using a standardized analysis pipeline. See those earlier studies for details of the GBS library preparation process; in brief, the restriction enzyme PstI was used to cut the genomes at specific recognition sites, and from each cut site, roughly 100 bp was sequenced in each direction. Homologous sequences were then compared to identify single nucleotide polymorphisms (SNPs).

To avoid potential biases due to variation in sample size, for the calculations of within- and between-group variation, we used the same number of individual male birds (14), whenever possible, from each major focal population (see Supporting Information Table S1 for details). For Audubon's warblers, we included two samples of 14 individuals each: northern Audubon's warblers and southern Audubon's warblers (which have sometimes been treated as the different subspecies *memorabilis*; Milá, Toews, Smith, & Wayne, 2011; Moore, 1946). Individuals were sampled far from hybrid zones to avoid recent introgression (see locations in Supporting Information Table S1). We mapped sequencing reads to the zebra finch (*Taeniopygia guttata*) genome (version 3.2.4; Warren et al., 2010), which is evolutionarily equidistant from each of the warblers in this study.

Read demultiplexing, trimming, mapping to the reference genome, realigning around indels and genotype calling were all conducted according to the protocol of Irwin et al. (2016). These initial processing steps were conducted on GBS reads from 573 individuals

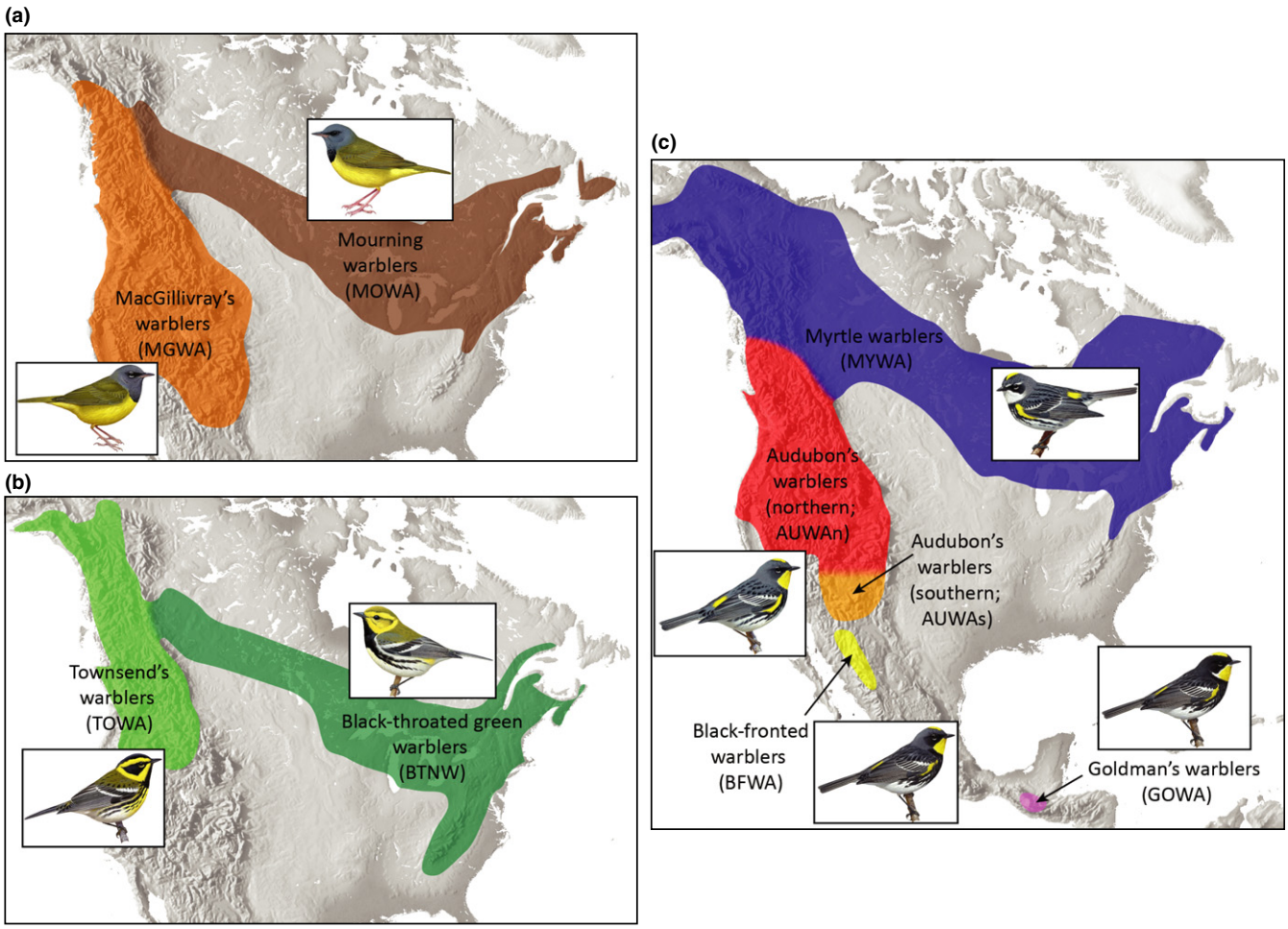

FIGURE 3 Breeding ranges of the three species pairs under study: (a) MacGillivray's and mourning warblers; (b) Townsend's and black-throated green warblers; and (c) Audubon's and myrtle warblers. The latter pair are members of the yellow-rumped warbler species complex (Brelsford et al., 2011; Milá, Smith, & Wayne, 2007; Toews, Brelsford, et al., 2016), which also consists of black-fronted warblers and Goldman's warblers. Audubon's and myrtle warblers are considered different species by the International Ornithologists' Union (Gill & Donsker, 2018). We show Audubon's warblers as consisting of northern populations that have mitochondrial DNA similar to that of myrtle warblers, and southern populations that have mitochondrial DNA similar to those of black-fronted warblers. The colours and taxon abbreviations (e.g., AUWAn for northern Audubon's warblers) used in this figure will be used to refer to the same groups in other figures. Abbreviations are standard codes for these species as used by bird banders. Illustrations of warbler species were reproduced with permission from the Handbook of the Birds of the World Alive (del Hoyo, Elliott, Sargatal, Christie, & de Juana, 2018)

(including many from hybrid zones and hence not included in the present analysis). We used the *SelectVariants* command in GATK to select a focal set of 117 individuals (Supporting Information Table S1) for analysis. This data set included both variant and invariant sites.

Using VCFTOOLS (Danecek et al., 2011) and custom scripts, we then applied a series of filters to determine the set of nucleotide sites used in the analysis and prepare the data for analysis in R (R Core Team, 2014). We removed indels, sites with more than 2 alleles, sites where more than 60% of the 117 individuals had missing genotypes, and sites with "mapping quality" (MQ) <20 or heterozygosity above 60% (to avoid paralogs). We then imported the matrix of genotypes into R (R Core Team, 2014), where subsequent analysis was performed using custom scripts based on those of Irwin et al. (2016).

Between-group relative differentiation F_{ST}, between-group absolute nucleotide differentiation π_B and within-group nucleotide diversity π_w were calculated for nonoverlapping windows of size 10,000 nucleotides for which we had good sequence data (i.e., that passed the filtering described above). This methodology was chosen because it results in the same amount of information per window and the same locations of boundaries between windows in all of the taxa in our analysis, rendering windows directly comparable. Calculations were performed using a custom R script modified slightly from Irwin et al. (2016; note that π_B was referred to as "D_{xy}" in that study, and π_W was referred to as "π"). Windowed (i.e., multilocus) F_{ST} was calculated using only variant sites (i.e., SNPs) following Weir and Cockerham (1984) by summing numerators of their $\hat{\theta}$ equation across sites and dividing by the sum of the denominators. Both π_B and π_W were calculated using

all sequenced sites, both variant and invariant, as invariant sites are essential in calculating true nucleotide distances. We calculated π_B as $p_1(1-p_2) + p_2(1-p_1)$, where p_1 and p_2 are the frequencies of a given allele in the first and second groups. We have now made a slight modification in how we estimate within-group nucleotide diversity π_W, in that we now correct for limited sample size by calculating π_W as $\left(\frac{2n}{2n-1}\right)(2p(1-p))$, where n is the number of individuals with called genotypes and p is the sample frequency of a given allele; the π_W values reported for the analysis of Irwin et al. (2016) are biased slightly downward due to the lack of the $2n/(2n-1)$ correction.

Because sex chromosomes can show different rates and patterns of evolution than autosomes (reviewed by Irwin, 2018; Wright & Mank, 2013), we focused our comparisons of differentiation statistics on autosomes only; however, we provide an analysis of variation in the Z chromosome (male birds have two copies of the Z, and females have one) in the online supplementary material.

To test for patterns of allelic variation that are not well explained by a bifurcating phylogeny of populations, and hence may indicate introgression, we used ABBA-BABA comparisons (Durand, Patterson, Reich, & Slatkin, 2011; Zhang, Dasmahapatra, Mallet, Moreira, & Kronforst, 2016). These were used for examining patterns of past introgression among populations in the yellow-rumped warbler complex. Each comparison involves four groups (designated P_1, P_2, P_3 and P_4), in which P_1 and P_2 are the putative closest relatives (e.g., southern and northern Audubon's warblers), P_3 is a more distantly related group (e.g., myrtle warblers) that may have exchanged genes with P_1 and/or P_2, and P_4 is a distant outgroup (e.g., Townsend's warblers). If two alleles at a locus are designated A and B, the number of loci that follow an ABBA pattern (for P_1, P_2, P_3 and P_4 consecutively) is expected to equal the number of loci that follow a BABA pattern. If the difference in the number of ABBA and BABA patterns (i.e., the D statistic) is significantly different from zero, we can conclude that the bifurcating tree does not completely explain the genetic variation. In that case, introgression between P_3 and P_1 and/or P_2 can account for the difference. We calculated the D statistic using population allele frequencies according to equation 2 of Durand et al. (2011), after first converting allele frequencies to refer to the allele that is rare (frequency <0.5) in the outgroup (P_4) (Zhang et al., 2016). We calculated D for each chromosome and then conducted a sign test (using chromosomes as replicates) to test for significant deviation from zero.

To summarize geographic variation among samples of yellow-rumped warblers, we conducted principal components analysis (PCA) using custom scripts in version 3.1.2 of R (R Core Team 2014), employing the "pca" command from the PCAMETHODS package (Stacklies, Redestig, Scholz, Walther, & Selbig, 2007), with method "svdImpute" to account for missing genotypes. We first filtered the SNPs such that loci with more than 30% of individuals having missing data were excluded. We centred but did not scale genotypic values, such that each nucleotide mismatch had equal weighting.

3 | RESULTS

Following filtering, our genotype-by-sequencing analysis of 117 individuals from the three species complexes produced a data set of 480,714 variable and 10,868,483 invariant nucleotide sites, together accounting for roughly 0.87% of the genome (assuming a genome size of 1.3 billion base pairs; Kapusta, Suh, & Feschotte, 2017). When grouped into windows of 10,000 sequenced base pairs, there were 1118 windows across the whole genome (1071 windows on autosomes; 47 on the Z chromosome), with an average window length of 1.2 million base pairs.

The three species pairs differ strongly in the structuring of relative differentiation (F_{ST}) across the genome (Figure 4). Heterogeneity in F_{ST} is strong in the Audubon's/myrtle warbler comparison, with low F_{ST} across most of the genome being punctuated occasionally by regions of much higher F_{ST}. This heterogeneity is more moderate in the Townsend's/black-throated green warbler comparison, and weakest in the MacGillivray's/mourning warbler comparison. Figure 4b illustrates a curious pattern: Across the autosomal genome, MacGillivray's/mourning warblers have the highest median windowed F_{ST} and Audubon's/myrtle warbler the lowest, but the opposite is true for the upper 1% of windowed F_{ST} values (Audubon's/myrtle warblers have 1% of windows with an F_{ST} above 0.58; this F_{ST} threshold is 0.38 for Townsend's/black-throated green warblers and 0.26 for MacGillivray's/mourning warblers). In MacGillivray's/mourning warblers, F_{ST} peaks are more apparent when window size is reduced to 5,000 (rather than 10,000) sequenced base pairs (see Supporting Information Figure S1).

There is some similarity among different pairs in the location of differentiation peaks, but this is a weak pattern (Figure 4c). The comparison of the two most closely related pairs (Audubon's/myrtle warblers and Townsend's/black-throated green warblers) shows a significant positive relationship between F_{ST} in one pair and F_{ST} in the other pair (Spearman's rank correlation: $r_s = 0.189$, $p = 4.8 \times 10^{-10}$), but this relationship explains only 3.6% (square of the correlation coefficient) of the variation in F_{ST} in one pair by F_{ST} in the other pair. Comparisons with the more distant pair (MacGillivray's/mourning warblers) show no positive relationships (MacGillivray's/mourning warblers and Audubon's/myrtle warblers: $r_s = -0.063$, $p = 0.040$; MacGillivray's/mourning warblers and Townsend's/black-throated green warblers: $r_s = -0.020$, $p = 0.504$; see also Supporting Information Figure S1).

This dissimilarity in genomic variation in F_{ST} among the species pairs contrasts with the pattern of variation in both absolute nucleotide differentiation (π_B) and within-group nucleotide diversity (π_W). Variation in each of these across the autosomal genome shows strong similarity among species pairs, with variation in any pair explaining 26%–50% of variation in π_B in another pair, or 27%–55% of variation in π_W (Supporting Information Figure S2; see caption for statistical tests). Like F_{ST}, the patterns of variation in both π_B and π_W across the genome show the most similarity between the most closely related two species pairs, and much lower similarity with the more distant species pair.

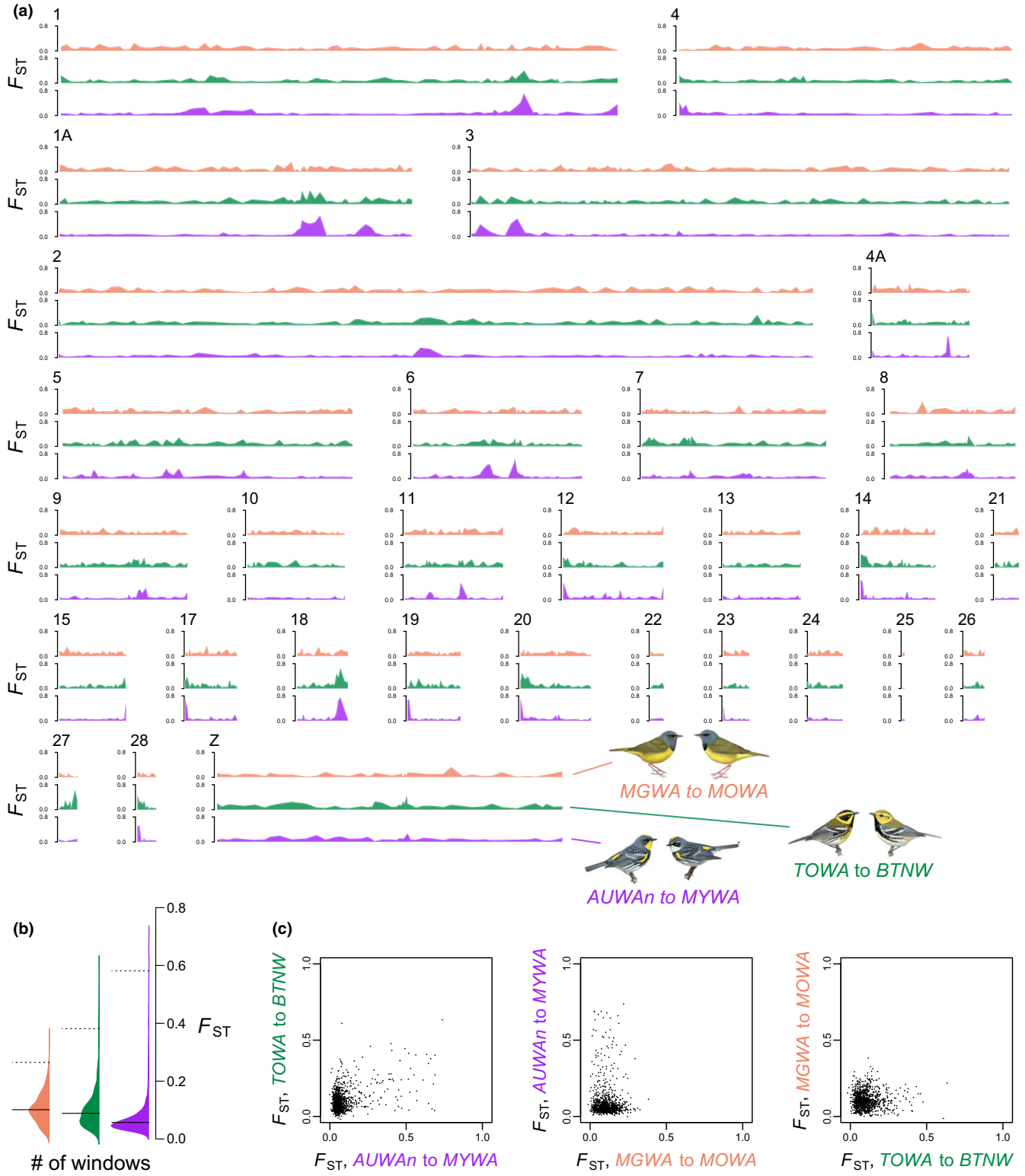

FIGURE 4 Three warbler species pairs differ in their patterns of relative differentiation (F_{ST}) across the genome, based on 11,349,197 nucleotide positions (of which 480,714 are variable). (a) Variation across each chromosome in F_{ST} between MacGillivray's and mourning warblers (top, in orange), Townsend's and black-throated green warblers (middle, green), and northern Audubon's and myrtle warblers (bottom, purple). (b) Distributions of windowed F_{ST} across autosomes in each species pairs, with positions of medians (solid black line) and upper 1 percentile (dotted black line). (c) Per-window F_{ST} shows little or no association between species pairs. Each dot represents an autosomal window of 10,000 nucleotide sites with genotypic information. Bird illustrations were reproduced with permission from the Handbook of the Birds of the World Alive (del Hoyo et al., 2018)

Figure 5 shows a comparison of windowed averages of each of these statistics (F_{ST}, π_B and π_W) across the whole genome in the Audubon's/myrtle warbler comparison (see Supporting Information Figure S3 for a detailed example of a single chromosome). Most regions of high relative differentiation (i.e., F_{ST} peaks) do not tend to have high absolute nucleotide distances (i.e., π_B), but rather tend to have especially low within-group nucleotide diversity (π_W). See Supporting Information Figures S4 and S5 for similar figures for the other two species pairs.

Within each of the three species pairs, most of the genome shows levels of within-taxon diversity (π_W) that are only slightly smaller than between-taxon absolute nucleotide distance (π_B; Figure 6). Although there is much variation in both π_B and π_W across the genome, the great majority of genomic windows cluster near the 1:1 line, the expectation in a single panmictic population. However, a small subset of genomic regions shows much reduced within-group variation compared to between-group variation; these are the regions with high F_{ST} (i.e., the points coloured more blue in Figure 6). This pattern is most apparent in the Audubon's/myrtle warbler comparison, moderate in the Townsend's/black-throated green warbler comparison, and weak in the MacGillivray's/mourning warbler comparison, despite the latter pair having higher average absolute distances (as measured by π_B) across the genome.

Using values of π_B to represent distances between taxa and π_W to represent variation within taxa, we constructed a phylogeny (Figure 7; consistent with that of Lovette et al., 2010). Distances

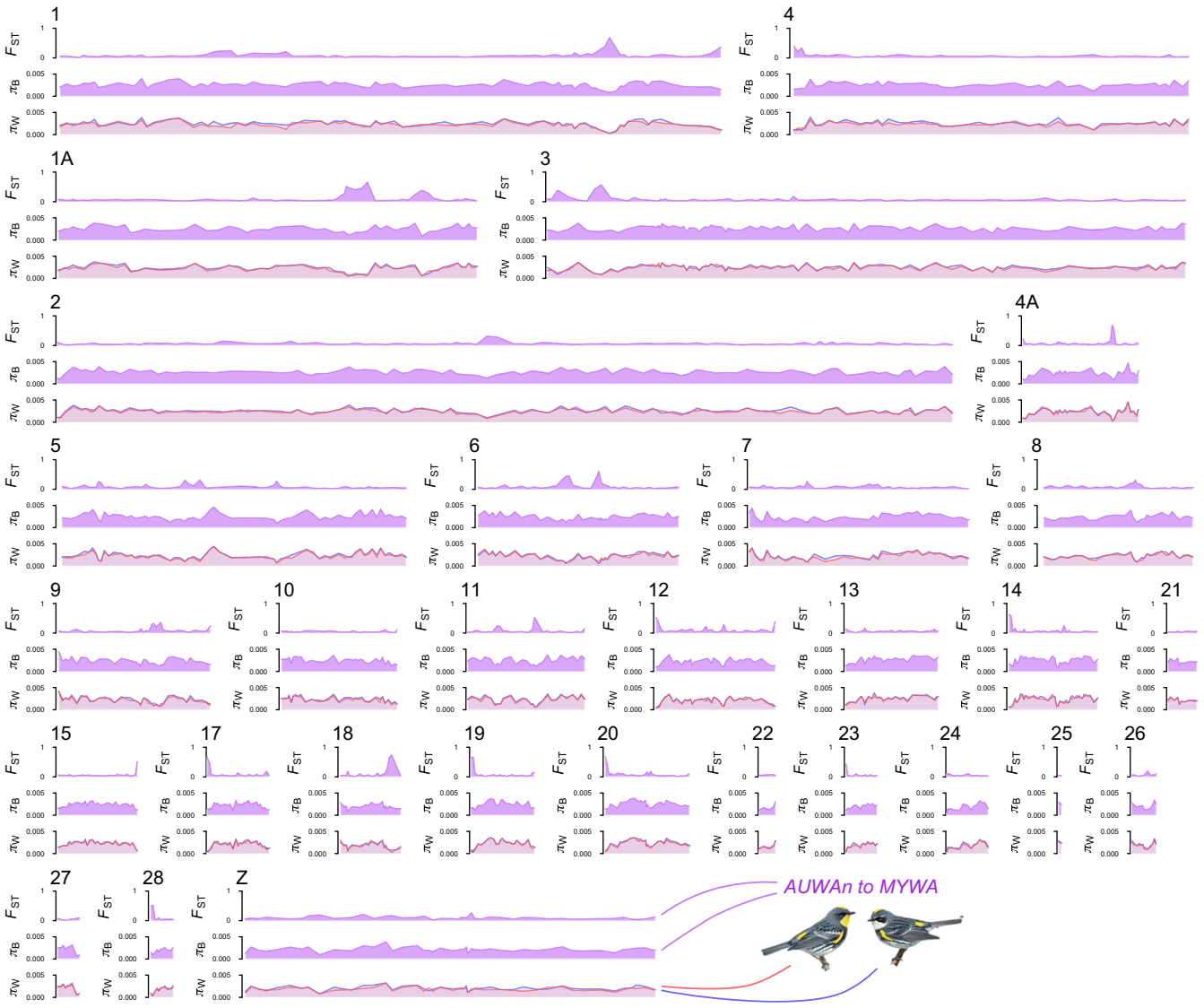

FIGURE 5 Genomic variation in patterns of nucleotide differentiation between northern Audubon's and myrtle warblers. For each chromosome, graphs show per-window relative nucleotide differentiation (F_{ST}, top), absolute nucleotide distance (π_B, middle) and within-group nucleotide diversity (π_W). For F_{ST} and π_B the purple lines represent the comparison of northern Audubon's and myrtle warblers, whereas for π_W red represents northern Audubon's warblers and blue represents myrtle warblers. Bird illustrations were reproduced with permission from the Handbook of the Birds of the World Alive (del Hoyo et al., 2018)

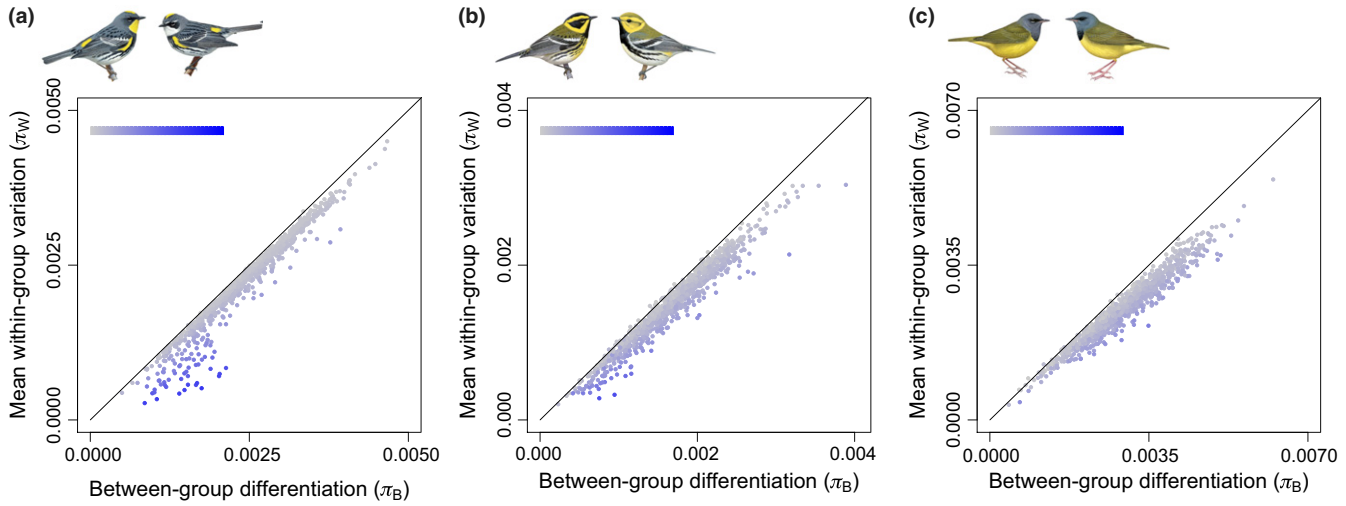

FIGURE 6 For each of the three species pairs, most autosomal regions (each dot represents an autosomal window) show levels of mean within-group nucleotide variation (π_W) that are almost as high as between-group nucleotide distance (π_B), with the exception of the few high-F_{ST} regions (illustrated with increasing blue colour) that have greatly reduced π_W compared to π_B. The black line illustrates equality between π_W and π_B, expected under no differentiation. Regions of high F_{ST} tend to have low π_B in both the myrtle/northern Audubon's warbler pair (a) and the Townsend's/black-throated green warbler pair (b), whereas the MacGillivray's/mourning warbler pair (c) has few if any high-F_{ST} regions at the spatial scale of the genomic windows used in this analysis (10,000 sequenced base pairs, with average window size of 1.2 Mb). Note that the three graphs differ in their axis scales. Bird illustrations were reproduced with permission from the Handbook of the Birds of the World Alive (del Hoyo et al., 2018)

between individuals within each taxon are almost as large as those between taxa within a pair, indicating recent shared ancestry and/or high gene flow (or their combination) at most of the genome. We explore the potential for recent gene flow across hybrid zones by investigating one species complex, the yellow-rumped warblers, in more detail.

3.1 | The yellow-rumped warbler species complex

Among most yellow-rumped warbler populations, the phylogeny in Figure 7 shows similar levels of within-population (π_W) and between-population (π_B) nucleotide distances, suggesting recent common ancestry and/or high gene flow. The one exception is Goldman's

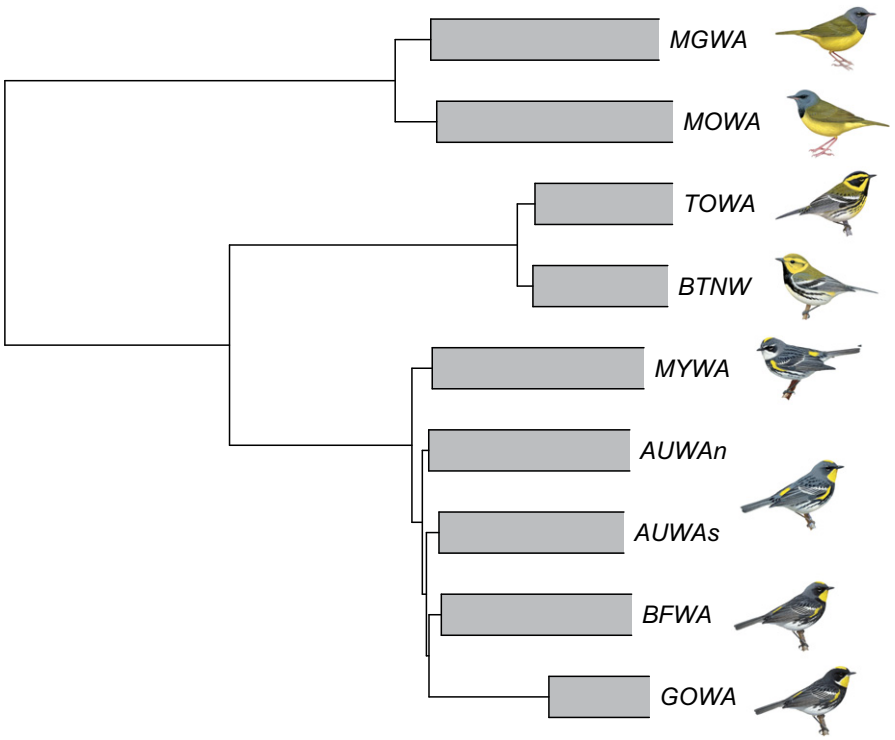

FIGURE 7 Relationships as illustrated using a neighbour-joining tree based on autosomal average between-taxon nucleotide distances (π_B) and within-taxon between-individual nucleotide distances (π_W, indicated by horizontal length of grey boxes). Bird illustrations were reproduced with permission from the Handbook of the Birds of the World Alive (del Hoyo et al., 2018)

warbler (GOWA), for which within-group pairwise distance is only 47% of its pairwise distance to Black-fronted warbler. The small population size and geographic isolation has likely led to loss of genetic diversity and short coalescence times. The most divergent group of yellow-rumped warblers with respect to π_B are the myrtle warblers (MYWA), which forms a basal split with a clade containing all other taxa in the yellow-rumped warbler complex (Figure 7).

Principal components analysis (PCA) of the whole yellow-rumped warbler data set (Figure 8a) clearly separates Goldman's warbler from the rest, due to Goldman's warblers having high genotypic similarity to one another compared to pairs of individuals from elsewhere in the complex. The other populations are also separated from one another by PCA (Figure 8b), with black-fronted warblers, Audubon's warblers, and myrtle warblers forming discrete clusters in genotype space. Within Audubon's warblers, there is subtle geographic differentiation in genotypes, with a cluster of southern Audubon's warblers (from Arizona and New Mexico; orange symbols in Figure 8b) showing slight differences compared to a cluster of northern Audubon's warblers from Idaho and Oregon (the cluster of red symbols close to the orange symbols), and then a larger jump to a cluster of northern Audubon's warblers from south-central British Columbia (the cluster of seven red symbols to the right of the others), in the direction of myrtle warblers. Patterns of relative differentiation between yellow-rumped warbler populations as quantified by F_{ST} (Figure 8c; Table 1) are similar to those shown in the PCA plots. In comparisons among these taxa, only comparisons involving myrtle warblers have genomic windows with high F_{ST}; comparisons of black-fronted, southern Audubon's and northern Audubon's warblers do not show any high (e.g., >0.5) F_{ST} windows (Supporting Information Figures S6 and S7).

As the arrangement of populations in the PCA suggests historical gene flow between myrtle and northern Audubon's warblers, we used ABBA-BABA tests (Durand et al., 2011; Green et al., 2010; Zhang et al., 2016) to more specifically test for such genetic exchange. Results were strongly indicative of greater historical genetic exchange between myrtle warblers and northern Audubon's warblers compared to between myrtle warblers and southern Audubon's warblers (mean value of the D statistic following Zhang et al. (2016) using either Townsend's warblers or black-throated green warblers as outgroup: averaged across chromosomes: 0.077; 31 out of 31 chromosomes show positive D; sign test: $p = 9.3 \times 10^{-10}$). Results also indicated greater historical genetic exchange between myrtle warblers and northern Audubon's warblers compared to between myrtle warblers and black-fronted warblers ($D = 0.047$; 28 out of 31 chromosomes show positive D; sign test: $p = 4.6 \times 10^{-6}$).

3.2 | Associations between relative and absolute differentiation

Having established in each of the taxon pairs that only a small subset of the genome had strong relative differentiation (F_{ST}), we now address whether there is a relationship between relative differentiation and absolute nucleotide distance (π_B), as different models of the process of differentiation (Figure 1) make alternative predictions. Graphs of F_{ST} vs. π_B (Figure 9; see also Figures 5, 6 and Supporting Information Figure S8) show that almost none of the high-F_{ST} genomic windows have high π_B, the pattern that would be expected under a pure divergence-with-gene-flow model (Cruickshank & Hahn, 2014). Rather, at least two of the taxon pairs (Audubon's/myrtle warbler and Townsend's/black-throated green warblers) tend to have low π_B in windows with high F_{ST}, consistent with models of sweep-before-differentiation (Figure 2d) and/or recurrent selection in the common ancestor as well as more recently in the daughter populations (Figure 2c).

4 | DISCUSSION

4.1 | Patterns of genomic differentiation

The three pairs of western/eastern taxa examined in this study consist of phenotypically well-differentiated taxa that originally were each described as distinct species. Given this, it is remarkable that most of their genomes show little relative differentiation between the taxa within each pair, with only a small subset showing strong differentiation. The hybrid zones found within each pair suggest that gene flow may be preventing differentiation, although recent speciation would also be consistent with low differentiation. Previous analyses have concluded that the narrowness of the hybrid zones and the patterns of variation within them (e.g., high linkage disequilibrium) indicate moderate selection against hybrids (Brelsford & Irwin, 2009; Irwin et al., 2009; Kenyon et al., 2017, 2011; Toews et al., 2011). The peaks of genomic differentiation within each pair further support the conclusion of moderate reproductive isolation between the forms, yet the fact that these peaks of genomic differentiation are so narrow and rare in the genome suggests that the concept of reproductive isolation applies to just a small part of the genome. In a sense, the taxa within each pair are recognizable as clearly distinct species at only a small fraction of their genomes, an observation consistent with Wu's "genic view of speciation" (Wu, 2001).

Given that such a small proportion of the genomes of each pair show clear differentiation, these three pairs provide insight into the relatively early stages of the process of genomic differentiation during speciation (Supporting Information Figure S9 shows that comparisons of more distantly related species show high F_{ST} across the entire genome). Specific regions of high F_{ST} are in most cases different between the three pairs, although there is a small amount of similarity (3.6% of variation explained) observed between the most closely related pairs (Audubon's/myrtle warblers and Townsend's/black-throated green warblers). These results suggest that causal factors common to all three pairs have had little influence in determining which regions have become differentiated. Thus, factors that are hypothesized to be relatively consistent in their operation over long spans of evolutionary time, such as low recombination (Singhal et al., 2015), locations of centromeres (Ellegren et al., 2012; Vijay et al., 2017) and consistent selection pressures (e.g., background selection), do not well explain the observed differences among the species pairs

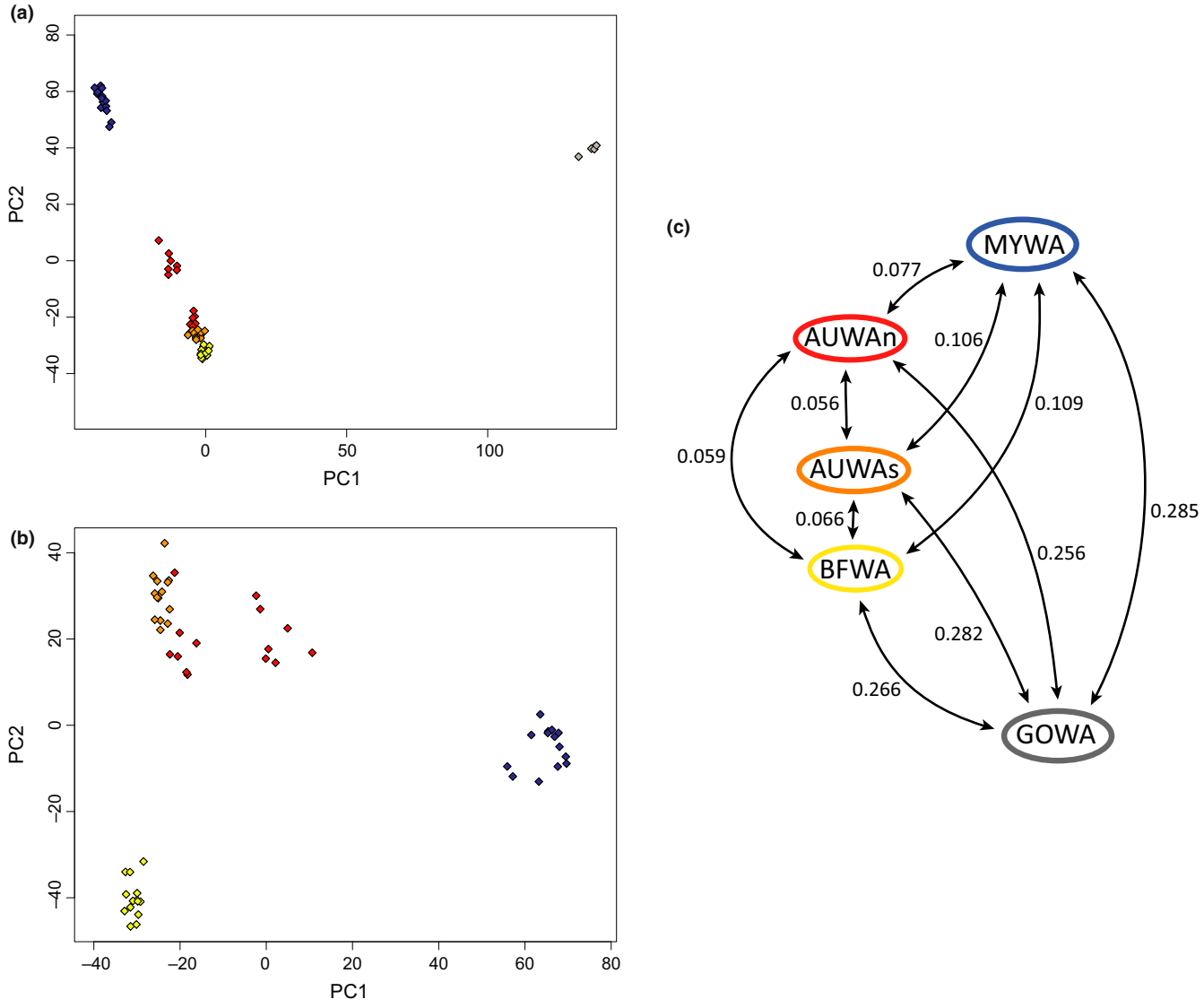

FIGURE 8 Whole-genome principal components analysis of (a) all individuals in the yellow-rumped warbler analysis, and (b) with Goldman's warbler removed; and (c) diagram of whole-genome F_{ST} values between groups of yellow-rumped warblers. Colours indicate sampling groups: myrtle warblers (blue), northern Audubon's warblers (red), southern Audubon's warblers (orange), black-fronted warblers (yellow) and Goldman's warblers (grey). Variation explained by axes is 7.4% by PC1 and 5.3% by PC2 in (a), and 6.1% by PC1 and 3.1% by PC2 in (b)

TABLE 1 Whole-genome F_{ST} (above diagonal) and π_B (below diagonal) between taxa in the study, as well as π_W (on diagonal, in bold) for each taxon

	MGWA	MOWA	TOWA	BTNW	MYWA	AUWAn	AUWAs	BFWA	GOWA
MGWA	**0.002477**	0.109	0.714	0.715	0.641	0.647	0.657	0.657	0.717
MOWA	0.00293	**0.002555**	0.713	0.714	0.640	0.647	0.656	0.657	0.715
TOWA	0.00724	0.00731	**0.001499**	0.101	0.576	0.585	0.603	0.598	0.709
BTNW	0.00721	0.00728	0.00166	**0.001466**	0.571	0.581	0.599	0.594	0.710
MYWA	0.00709	0.00716	0.00466	0.00463	**0.002300**	0.077	0.106	0.109	0.285
AUWAnn	0.00702	0.00710	0.00460	0.00458	0.00243	**0.002188**	0.056	0.059	0.256
AUWAs	0.00693	0.00701	0.00453	0.00451	0.00242	0.00222	**0.002000**	0.066	0.282
BFWA	0.00701	0.00709	0.00460	0.00457	0.00245	0.00226	0.00218	**0.0020666**	0.266
GOWA	0.00712	0.00719	0.00469	0.00466	0.00255	0.00234	0.00227	0.00229	**0.0010808**

Note. For key to taxon abbreviations, see Figure 3. Grey shading indicates comparisons within a species complex.

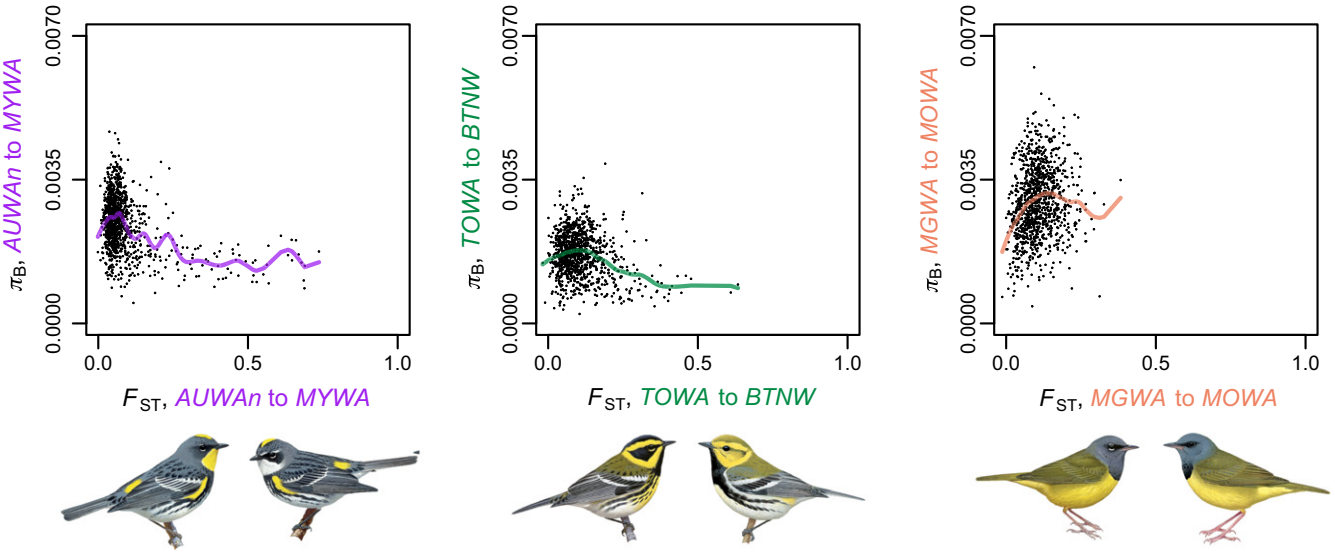

FIGURE 9 Relationships between relative (F_{ST}) and absolute (π_B) nucleotide distance between northern Audubon's and myrtle warblers (left), Townsend's and black-throated green warblers (centre) and MacGillivray's and mourning warblers (right). Each dot represents a single autosomal window of 10,000 sequenced base pairs. Coloured lines show cubic spline fits of π_B to F_{ST} (using the "smooth.spline" function in R, with smoothing parameter equal to one). Regions with high F_{ST} tend to have low π_B in both the myrtle/northern Audubon's warbler comparison (Spearman's rank correlation: $r_s = -0.256$, $p < 2.2 \times 10^{-16}$) and the Townsend's/black-throated green warbler comparison ($r_s = -0.118$, $p = 1.1 \times 10^{-4}$). MacGillivray's/mourning warblers have a slight positive correlation between F_{ST} and π_B ($r_s = 0.208$, $p = 7.5 \times 10^{-12}$), although this largely results from associations in low-F_{ST} windows, as the few windows with higher F_{ST} tend to have moderate or low π_B (see the orange spline). Bird illustrations are from del Hoyo et al. (2018)

in patterns of genomic variation in F_{ST}. Rather, the differences appear best explained by a combination of differing selective forces, differing genomic responses (e.g., due to different mutations) to similar selective forces, and/or different amounts of gene flow in the three cases of speciation.

Contrasting with the pattern in relative differentiation (F_{ST}), variation in absolute nucleotide distance (π_B) across the genome shows much similarity among the species pairs. This is in part because ancient polymorphisms are to some degree shared among the taxon pairs, such that more variable regions in the common ancestor of all three tend to produce higher levels of π_B in comparisons of any two daughter taxa. These variable levels of ancient variation across the genome could have been influenced by variation in mutation rates. However, the high degree of variation in π_B, which ranges over more than an order of magnitude among genomic windows, is unlikely to be fully explained by mutation rate variation and ancestral polymorphism. Rather, selective sweeps in ancestral populations and/or across contact zones within specific pairs of taxa have likely reduced levels of π_B in some windows. According to the modelling that has been done to date (Matthey-Doret & Whitlock, 2018; Zeng & Charlesworth, 2011; Zeng & Corcoran, 2015), the effects of background selection alone (selection against deleterious mutations resulting in loss of variation) are rather modest in strength, and unlikely to be strong enough to explain the exceptionally low π_B in some regions compared to others (see also Irwin et al., 2016).

Across-genome variation in within-group diversity (π_W) also shows strong similarity among species pairs, as expected given that over most of the genome of each pair, levels of π_W are similar to levels of π_B. However, a small subset of windows within each taxon pair shows unexpectedly low levels of π_W given the level of π_B. These correspond to the F_{ST} peaks.

4.2 | Evaluating the models of genomic islands of differentiation

Turning to the major models for the evolution of F_{ST} peaks, a pure divergence-with-gene-flow-model (Figure 2a) is not supported, due to the observation that the F_{ST} peaks do not tend to have elevated π_B (Cruickshank & Hahn, 2014). This is true for all three taxon pairs (Figure 9). Rather, the F_{ST} peaks tend to have low π_B in the Audubon's/myrtle warbler pair and the Townsend's/black-throated green warbler pair, and moderate π_B in the MacGillivray's/mourning warbler pair. A pure selection-in-allopatry model (Figure 2b) is not supported in two of the pairs, due to the low π_B in F_{ST} peaks. Areas of moderately low π_B can be explained by recurrent selection in the common ancestor of a species pair (Figure 2c; Cruickshank & Hahn, 2014), whereas regions with extremely low π_B are better explained by the sweep-before-differentiation model (Figure 2d; Delmore et al., 2015; Irwin et al., 2016), in which globally advantageous alleles move between geographically differentiated incipient species and subsequently become differentiated. A full evaluation of the evidence for the latter two models (Figure 2c,d) will require future development of theory and modelling, but we note that this will likely rely in part on analysis of the distribution of π_B: This

should be narrower, with fewer low π_B values, under the recurrent selection model, and wider with more low π_B values under the sweep-before-differentiation model. Given that there are hybrid zones within each of our species pairs, we presently focus on evaluating whether patterns of gene flow are consistent with the sweep-before-differentiation model. The presence of hybrid zones does not necessarily indicate that there is gene flow between populations; hence, we have used ABBA-BABA tests and principal components analysis of the yellow-rumped warbler to more explicitly test for gene flow. Our results confirm that northern Audubon's warblers are more genomically similar to myrtle warblers than are the more southern groups (southern Audubon's warblers and black-fronted warblers), a pattern consistent with past gene flow between myrtle and Audubon's warblers.

The initial geographic sweep phase of the sweep-before-differentiation model is essentially equivalent to the well-known phenomenon of adaptive introgression, which has been invoked to explain patterns of variation in such organisms as *Helianthus* sunflowers (Whitney, Randell, & Rieseberg, 2010), *Anopheles* mosquitos (Rosenzweig, Pease, Besansky, & Hahn, 2016), *Heliconius* butterflies (Pardo-Diaz et al., 2012; Zhang et al., 2016), *Drosophila* fruit flies (Llopart, Herrig, Brud, & Stecklein, 2014), *Lonchura* finches (Stryjewski & Sorenson, 2017), mice (Song et al., 2011; Staubach et al., 2012), snowshoe hares (Jones et al., 2018), ibex (Grossen et al., 2014) and humans (Huerta-Sánchez et al., 2014; Jeong et al., 2014). Similarly, parts of the genome may have adaptively introgressed between closely related western and eastern forms of birds. Given the many cycles of glaciation and forest re-expansion and contact that have occurred in North America (Lovette, 2005; Weir & Schluter, 2004), many currently hybridizing western and eastern related forms of birds have likely experienced previous periods of population contact and hybridization. During these periods of partial genetic exchange, global selective sweeps would have reduced variation (i.e., π_B and π_W) at some parts of the genome, and subsequent selection (e.g., during a period of subsequent geographic isolation) could have led to increased F_{ST}. A genomic region that is globally advantageous can still be imperfect under local conditions, such that subsequent mutations and/or recombination can result in increased fitness locally.

We emphasize that the four models for the formation of genomic islands of differentiation (Figure 2) are not mutually exclusive, and that actual cases of speciation may contain complex histories with elements of each. Patterns of low π_B in F_{ST} peaks in the context of hybridization and gene flow have focused our attention on the sweep-before-differentiation model, but elements of the other models can still play a role. In particular, a plausible scenario is this: Ancient selective sweeps could have played a role in reducing diversity at some genomic regions in the past, resulting in low π_B, and later those same regions could play a role in causing reproductive isolation, thereby producing and maintaining high F_{ST}. This scenario combines the sweep-before-differentiation and divergence-with-gene-flow models in a way that leads to a prediction of low π_B in F_{ST} peaks, consistent with observation in many systems (Cruickshank & Hahn, 2014; Delmore et al., 2015; Irwin et al., 2016; Van Doren et al., 2017; Zhang et al., 2017). These considerations show that lack of high π_B in F_{ST} peaks is not compelling evidence against gene flow during population differentiation.

4.3 | Insights from the yellow-rumped warbler species complex

The geographically differentiated set of populations in the yellow-rumped warbler complex provides us with a particularly rich opportunity to examine patterns of genomic differentiation and gene flow. In pairwise comparisons of different forms within the yellow-rumped warbler complex, only comparisons involving myrtle warblers tend to show peaks of highly elevated F_{ST} compared to the genomic background (see the particularly high skew in the F_{ST} distribution in Figure 4b). However, average genomewide π_B is only slightly higher in comparisons involving myrtle warblers than comparisons involving any two of the other forms within the complex, and the F_{ST} peaks tend to have quite low π_B and exceptionally low π_W. These patterns suggest that the differentiation of myrtle warblers from the other forms is a result of strong selection in a small subset of the genome over a relatively short span of time, rather than accumulation of genomewide differences during a long period of pure geographic isolation. Admixture mapping in the hybrid zone has indicated that some of these differentiation peaks are strongly associated with colour differences between these two taxa; for example, the differentiation peak on chromosome 20 (Figure 5, on the left side of that chromosome) is associated with the presence of white eye spots and lines in myrtle warblers and their absence in Audubon's warblers (Brelsford, Toews, & Irwin, 2017).

This genomic differentiation pattern in myrtle warblers contrasts strongly with that seen in Goldman's warblers, the form that is presently geographically separated from the rest within a small range in Guatemala. While average F_{ST} is high between Goldman's and either Audubon's or black-fronted warblers, there are few if any F_{ST} peaks (see also Toews, Brelsford, et al., 2016). These patterns are likely to be partly a result of genetic drift in a small isolated population causing allele frequencies to differ more strongly from the average of other populations, which have much larger population sizes. The loss of genetic variation in Goldman's warblers leads to strong covariation in genotypes compared to those of other taxa, causing Goldman's warbler to appear far from the other taxa in a principal components analysis, despite it not having a particularly large π_B compared to the other populations. This observation leads us to advise caution in interpreting principal components analyses of genomic data, as the patterns can be driven as much by shared loss of variation in small populations as by large nucleotide distances between populations. We also note that the high average F_{ST} of Goldman's warbler compared to the other populations could be misinterpreted as indicating an especially large span of evolutionary time separating it from the other taxa. In fact, average pairwise coalescence time (i.e., proportional to π_B) for two individuals from different forms of yellow-rumped warbler is about the same for all pairs of taxa in the

complex; the high F_{ST} is primarily a result of particularly low within-group diversity in Goldman's warbler. We do however note that Goldman's warblers differ from the others in a number of phenotypic traits that appear adaptive (Milá, Wayne, Smith, 2008), suggesting a role for natural selection as well as drift in shaping their genomic patterns.

Turning to the remaining taxa in the complex, we see clear genetic differentiation between black-fronted warblers and Audubon's warblers, as well as differences across the range of Audubon's warblers. The low genomewide average F_{ST} among these populations, the similar values of π_B and π_W between and within them, and lack of evidence for strong peaks of differentiation indicates recent ancestry and/or moderate gene flow; yet the clear genetic clustering indicates that gene flow is limited enough for some differentiation. The northern Audubon's warblers show evidence for greater genetic exchange with myrtle warblers than the more southern populations do.

Altogether, these results suggest that populations of yellow-rumped warblers across North America can be viewed as a set of geographically differentiated forms that have resulted from a complex history of population division, range expansions, adaptation and gene flow. During periods of population contact, some alleles have been able to move between forms. Globally advantageous alleles could spread throughout the species complex, reducing diversity in linked genomic regions. Alleles advantageous over a portion of the overall geographic range could sweep to fixation over that part of the range, reducing variation at that part of the genome, and leaving high relative differentiation between that part of the range and other parts where that allele did not spread.

5 | CONCLUSIONS

The three pairs of hybridizing warblers studied here provide windows into the early stages of genomic differentiation during the speciation process. Most of their genomes show little differentiation, whereas in each case a small fraction of the genome is highly differentiated in terms of allele frequencies at variable sites. The three species pairs differ strongly in the degree to which they display genomic peaks of differentiation and the locations of their islands of differentiation. These results imply the causes of genomic differentiation are mostly specific to each speciation event rather than being highly predictable from evolutionarily conserved factors (e.g., the recombination landscape). In two of the pairs (Audubon's/myrtle warblers; and Townsend's/black-throated green warblers), islands of high relative differentiation (F_{ST}) tend to have low absolute nucleotide distance (π_B) and exceptionally low within-group diversity (π_W), a pattern best explained by some form of recurrent diversity-reducing selection. We propose that some of this pattern is due to selective sweeps of globally advantageous alleles between geographically differentiated groups, explaining low π_B, followed by subsequent adaptation within each group, explaining especially low π_W and high F_{ST}. However, a full quantitative evaluation of support for this sweep-before-differentiation model (Delmore et al., 2015; Irwin et al., 2016) in comparison with a simpler model of recurrent selection (due to either positive or negative selection) in a panmictic ancestral population (Cruickshank & Hahn, 2014) will depend on future advances in theory and modelling.

ACKNOWLEDGEMENTS

For stimulating conversation on the topic of genomic differentiation, speciation and gene flow, we thank Bret Payseur, Loren Rieseberg, Dolph Schluter and Mike Whitlock. Armando Geraldes, Silu Wang and other members of the Irwin laboratory provided helpful comments on the manuscript. We thank the Burke Museum for providing DNA samples of four Audubon's warblers from Oregon. Greg Owens and Greg Baute provided some of the scripts used in processing GBS reads (see Irwin et al. (2016) for details). Major research funding was provided by the Natural Sciences and Engineering Research Council of Canada (Discovery Grant 311931 to DEI). We are grateful for the various governmental agencies and the UBC Animal Care Committee for providing research permits; for detailed acknowledgements regarding field assistance, permitting for field sampling and fellowship support see the Acknowledgement sections of Brelsford and Irwin (2009); Brelsford, Milá, and Irwin (2011); Irwin et al. (2009); Kenyon et al. (2017); Porter (2015); Toews, Brelsford, et al. (2016); Toews et al. (2011). We thank Rosemary Gillespie and seven anonymous reviewers for comments on the manuscript.

AUTHOR CONTRIBUTIONS

All authors contributed to data collection and conceptual development. D.E.I. conducted the analysis and wrote the study with input from all authors.

DATA ACCESSIBILITY

GBS reads have been deposited at NCBI SRA under the Accession no. PRJNA471352 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA471352); within this accession are data for 6 sets (i.e., plates) of samples: SRR7172622 contains reads from *Geothlypis tolmiei* and *G. philadelphia*; SRR7172623 and SRR7172620 contain reads from *Setophaga townsendi* and *S. virens*; and SRR7172621, SRR7172618 and SRR7172619 contain reads from the *S. coronata* species complex. Metadata for individuals used in the analysis, scripts containing code for all steps used in the analysis and production of figures, and files containing processed data at several stages are deposited in the Dryad Digital Repository (https://doi.org/10.5061/dryad.4j2662g). In that repository, we have included a file containing custom-built functions (written in R) that may be useful in other analyses.

ORCID

Darren E. Irwin http://orcid.org/0000-0002-1050-5841
Borja Milá https://orcid.org/0000-0002-6446-0079
David P. L. Toews http://orcid.org/0000-0002-9763-0476

REFERENCES

Alcaide, M., Scordato, E. S. C., Price, T. D., & Irwin, D. E. (2014). Genomic divergence in a ring species complex. *Nature*, *511*(7507), 83–85. https://doi.org/10.1038/nature13285

Brelsford, A., & Irwin, D. E. (2009). Incipient speciation despite little assortative mating: The yellow-rumped warbler hybrid zone. *Evolution*, *63*(12), 3050–3060. https://doi.org/10.1111/j.1558-5646.2009.00777.x

Brelsford, A., Milá, B., & Irwin, D. E. (2011). Hybrid origin of Audubon's warbler. *Molecular Ecology*, *20*(11), 2380–2389. https://doi.org/10.1111/j.1365-294X.2011.05055.x

Brelsford, A., Toews, D. P. L., & Irwin, D. E. (2017). Admixture mapping in a hybrid zone reveals loci associated with avian feather coloration. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1866), 20171106. https://doi.org/10.1098/rspb.2017.1106

Burri, R. (2017a). Interpreting differentiation landscapes in the light of long-term linked selection. *Evolution Letters*, *1*, 118–131. https://doi.org/10.1002/evl3.14

Burri, R. (2017b). Linked selection, demography and the evolution of correlated genomic landscapes in birds and beyond. *Molecular Ecology*, *26*, 3853–3856. https://doi.org/10.1111/mec.14167

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

Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, *15*(5), 538–543. https://doi.org/10.1093/oxfordjournals.molbev.a025953

Charlesworth, B., Morgan, M. T., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*, *134*(4), 1289–1303. https://doi.org/10.1111/j.0014-3820.2002.tb00188.x

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

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

del Hoyo, J., Elliott, A., Sargatal, J., Christie, D. A., & de Juana, E. (Eds) (2018). *Handbook of the birds of the world alive*. Barcelona, Spain: Lynx Edicions.

Delmore, K. E., Hübner, S., Kane, N. C., Schuster, R., Andrew, R. L., Câmara, F., … Irwin, D. E. (2015). Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. *Molecular Ecology*, *24*(8), 1873–1888. https://doi.org/10.1111/mec.13150

Delmore, K. E., Lugo Ramos, J. S., Van Doren, B. M., Lundberg, M., Bensch, S., Irwin, D. E., & Liedvogel, M. (2018). Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds. *Evolution Letters*, *49*(4522), 1–33. https://doi.org/10.1002/evl3.46

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, *28*(8), 2239–2252. https://doi.org/10.1093/molbev/msr048

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

Ellegren, H., & Wolf, J. B. W. (2017). Parallelism in genomic landscapes of differentiation, conserved genomic features and the role of linked selection. *Journal of Evolutionary Biology*, *30*(8), 1516–1518. https://doi.org/10.1111/jeb.13113

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, *6*(5), 1–10. https://doi.org/10.1371/journal.pone.0019379

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

Feder, J. L., & Nosil, P. (2010). The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*, *64*(6), 1729–1747. https://doi.org/10.1111/j.1558-5646.2009.00943.x

Feulner, P. G. D., Chain, F. J. J., Panchal, M., Huang, Y., Eizaguirre, C., Kalbe, M., … Milinski, M. (2015). Genomics of divergence along a continuum of parapatric population differentiation. *PLoS Genetics*, *11*(2), e1004966. https://doi.org/10.1371/journal.pgen.1004966

Gill, F., & Donsker, D. (2018). IOC world bird list (v 8.1). https://doi.org/10.14344/ioc.ml.8.1 May 5, 2018. Retrieved from http://www.worldbirdnames.org/

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., … Paabo, S. (2010). A draft sequence of the Neandertal genome. *Science*, *328*(5979), 710–722. https://doi.org/10.1126/science.1188021

Grossen, C., Keller, L., Biebach, I., Zhang, W., Tosser-Klopp, G., Ajmone, P., … Croll, D. (2014). Introgression from domestic goat generated variation at the major histocompatibility complex of Alpine Ibex. *PLoS Genetics*, *10*(6), e1004438. https://doi.org/10.1371/journal.pgen.1004438

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*, 1–12. https://doi.org/10.1101/gr.212522.116

Hubbard, J. P. (1969). The relationships and evolution of the *Dendroica coronata* complex. *The Auk*, *86*, 393–432. https://doi.org/10.2307/4083405

Hudson, R. R., Slatkin, M., & Maddison, W. P. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics*, *132*(2), 583–589. https://doi.org/PMC1205159

Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B. M., Vinckenbosch, N., … Nielsen, R. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, *512*(7513), 194–197. https://doi.org/10.1038/nature13408

Irwin, D. E. (2018). Sex chromosomes and speciation in birds and other ZW systems. *Molecular Ecology*. 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. https://doi.org/10.1111/mec.13792

Irwin, D. E., Brelsford, A., Toews, D. P. L., MacDonald, C., & Phinney, M. (2009). Extensive hybridization in a contact zone between MacGillivray's warblers *Oporornis tolmiei* and mourning warblers *O. philadelphia* detected using molecular and morphological analyses. *Journal of Avian Biology*, *40*(5), 539–552. https://doi.org/10.1111/j.1600-048X.2009.04687.x

Jeong, C., Alkorta-Aranburu, G., Basnyat, B., Neupane, M., Witonsky, D. B., Pritchard, J. K., … Di Rienzo, A. (2014). Admixture facilitates genetic adaptations to high altitude in Tibet. *Nature Communications*, *5*, 1–7. https://doi.org/10.1038/ncomms4281

Jones, M. R., Mills, L. S., Alves, P. C., Callahan, C. M., Alves, J. M., Lafferty, D. J. R., … Good, J. M. (2018). Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science*, *360*(6395), 1355–1358. https://doi.org/10.1126/science.aar5273

Kapusta, A., Suh, A., & Feschotte, C. (2017). Dynamics of genome size evolution in birds and mammals. *Proceedings of the National Academy of Sciences*, *114*(8), E1460–E1469. https://doi.org/10.1073/pnas.1616702114

Kenyon, H. L., Alcaide, M., Toews, D. P. L., & Irwin, D. E. (2017). Cultural isolation is greater than genetic isolation across an avian hybrid zone. *Journal of Evolutionary Biology*, *30*, 81–95. https://doi.org/10.1111/jeb.12989

Kenyon, H. L., Toews, D. P. L., & Irwin, D. E. (2011). Can song discriminate between Macgillivray's and Mourning Warblers in a narrow hybrid zone? *The Condor*, *113*(3), 655–663. https://doi.org/10.1525/cond.2011.100182

Lewontin, R. C., & Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, *74*(1), 175–195.

Llopart, A., Herrig, D., Brud, E., & Stecklein, Z. (2014). Sequential adaptive introgression of the mitochondrial genome in *Drosophila yakuba* and *Drosophila santomea*. *Molecular Ecology*, *23*(5), 1124–1136. https://doi.org/10.1111/mec.12678

Lovette, I. J. (2005). Glacial cycles and the tempo of avian speciation. *Trends in Ecology and Evolution*, *20*(2), 57–59. https://doi.org/10.1016/j.tree.2004.11.011

Lovette, I. J., Pérez-Emán, J. L., Sullivan, J. P., Banks, R. C., Fiorentino, I., Córdoba-Córdoba, S., … Bermingham, E. (2010). A comprehensive multilocus phylogeny for the wood-warblers and a revised classification of the Parulidae (Aves). *Molecular Phylogenetics and Evolution*, *57*(2), 753–770. https://doi.org/10.1016/j.ympev.2010.07.018

Matthey-Doret, R., & Whitlock, M. C. (2018). Background selection and the statistics of population differentiation: Consequences for detecting local adaptation. *bioRxiv*, 326256. https://doi.org/10.1101/326256

Milá, B., Smith, T. B., & Wayne, R. K. (2007). Speciation and rapid phenotypic differentiation in the yellow-rumped warbler *Dendroica coronata* complex. *Molecular Ecology*, *16*(1), 159–173. https://doi.org/10.1111/j.1365-294X.2006.03119.x

Milá, B., Toews, D. P. L., Smith, T. B., & Wayne, R. K. (2011). A cryptic contact zone between divergent mitochondrial DNA lineages in southwestern North America supports past introgressive hybridization in the yellow-rumped warbler complex (Aves: *Dendroica coronata*). *Biological Journal of the Linnean Society*, *103*(3), 696–706. https://doi.org/10.1111/j.1095-8312.2011.01661.x

Milá, B., Wayne, R. K., & Smith, T. B. (2008). Ecomorphology of migratory and sedentary populations of the Yellow-rumped Warbler (*Dendroica coronata*). *The Condor*, *110*(2), 335–344.

Moore, R. T. (1946). The status of *Dendroica auduboni nigrifrons* in the United States. *The Auk*, *63*, 241–242. https://doi.org/10.2307/4080016

Nachman, M. W., & Payseur, B. A. (2012). Recombination rate variation and speciation: Theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 409–421. https://doi.org/10.1098/rstb.2011.0249

Noor, M. A. F., & Bennett, S. M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, *103*(6), 439–444. https://doi.org/10.1038/hdy.2009.151

Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: Causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1587), 332–342. https://doi.org/10.1098/rstb.2011.0263

Pardo-Diaz, C., Salazar, C., Baxter, S. W., Merot, C., Figueiredo-Ready, W., Joron, M., … Jiggins, C. D. (2012). Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genetics*, *8*(6), e1002752. https://doi.org/10.1371/journal.pgen.1002752

Porter, A. N. (2015). *An analysis of ecological traits as reproductive barriers between the MacGillivray's (Geothlypis tolmiei) and Mourning (G. philadelphia) warblers*. M.Sc. thesis, University of British Columbia. Retrieved from https://open.library.ubc.ca/cIRcle/collections/ubctheses/24/items/1.0167279

R Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., … Westram, A. M. (2017). Interpreting the genomic landscape of speciation: Finding barriers to gene flow. *Journal of Evolutionary Biology*, *30*, 1450–1477. https://doi.org/10.1111/jeb.13047

Rosenzweig, B. K., Pease, J. B., Besansky, N. J., & Hahn, M. W. (2016). Powerful methods for detecting introgressed regions from population genomic data. *Molecular Ecology*, *25*(11), 2387–2397. https://doi.org/10.1111/mec.13610

Samuk, K., Owens, G. L., Delmore, K. E., Miller, S. E., Rennison, D. J., & Schluter, D. (2017). Gene flow and selection interact to promote adaptive divergence in regions of low recombination. *Molecular Ecology*, *26*(17), 4378–4390. https://doi.org/10.1111/mec.14226

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

Slatkin, M. (1991). Inbreeding coefficients and coalescence times. *Genetics Research*, *58*, 167–175. https://doi.org/10.1017/S0016672308009671

Song, Y., Endepols, S., Klemann, N., Richter, D., Matuschka, F. R., Shih, C. H., … Kohn, M. H. (2011). Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Current Biology*, *21*(15), 1296–1301. https://doi.org/10.1016/j.cub.2011.06.043

Stacklies, W., Redestig, H., Scholz, M., Walther, D., & Selbig, J. (2007). pcaMethods—a Bioconductor package providing PCA methods for incomplete data. *Bioinformatics*, *23*, 1164–1167.

Staubach, F., Lorenc, A., Messer, P. W., Tang, K., Petrov, D. A., & Tautz, D. (2012). Genome patterns of selection and introgression of haplotypes in natural populations of the house mouse (*Mus musculus*). *PLoS Genetics*, *8*(8), e1002891. https://doi.org/10.1371/journal.pgen.1002891

Stryjewski, K. F., & Sorenson, M. D. (2017). Mosaic genome evolution in a recent and rapid avian radiation. *Nature Ecology & Evolution*, *1*, 1912–1922. https://doi.org/10.1038/s41559-017-0364-7

Toews, D. P. L., Brelsford, A., Grossen, C., Milá, B., & Irwin, D. E. (2016). Genomic variation across the Yellow-rumped Warbler species complex. *The Auk*, *133*(4), 698–717. https://doi.org/10.1642/AUK-16-61.1

Toews, D. P. L., Brelsford, A., & Irwin, D. E. (2011). Hybridization between Townsend's *Dendroica townsendi* and black-throated green warblers *D. virens* in an avian suture zone. *Journal of Avian Biology*, *42*(5), 434–446. https://doi.org/10.1111/j.1600-048X.2011.05360.x

Toews, D. P. L., Campagna, L., Taylor, S. A., Balakrishnan, C. N., Baldassarre, D. T., Deane-Coe, P. E., … Winger, B. M. (2016). Genomic approaches to understanding population divergence and speciation in birds. *The Auk*, *133*(1), 13–30. https://doi.org/10.1642/AUK-15-51.1

Van Doren, B. M., Campagna, L., Helm, B., Illera, J. C., Lovette, I. J., & Liedvogel, M. (2017). Correlated patterns of genetic diversity and differentiation across an avian family. *Molecular Ecology*, *38*(1), 42–49. https://doi.org/10.1111/mec.14083

Via, S. (2009). Natural selection in action during speciation. *Proceedings of the National Academy of Sciences*, *106*(Supplement_1), 9939–9946. https://doi.org/10.1073/pnas.0901397106

Via, S., & West, J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, *17*(19), 4334–4345. https://doi.org/10.1111/j.1365-294X.2008.03921.x

Vijay, N., Weissensteiner, M., Burri, R., Kawakami, T., Ellegren, H., & Wolf, J. B. W. (2017). Genome-wide patterns of variation in genetic diversity are shared among populations, species and higher order taxa. *Molecular Ecology*, *38*(1), 42–49. https://doi.org/10.1111/mec.14195

Warren, W. C., Clayton, D. F., Ellegren, H., Arnold, A. P., Hillier, L. W., Künstner, A., … Wilson, R. K. (2010). The genome of a songbird. *Nature*, *464*(7289), 757–762. https://doi.org/10.1038/nature08819

Weir, B., & Cockerham, C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.2307/2408641

Weir, J. T., & Schluter, D. (2004). Ice sheets promote speciation in boreal birds. *Proceedings of the Royal Society B: Biological Sciences*, *271*(1551), 1881–1887. https://doi.org/10.1098/rspb.2004.2803

Westram, A. M., & Ravinet, M. (2017). Land ahoy? Navigating the genomic landscape of speciation while avoiding shipwreck. *Journal of Evolutionary Biology*, *30*(8), 1522–1525. https://doi.org/10.1111/jeb.13129

Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of F_{ST}. *The American Naturalist*, *186*(S1), S24–S36. https://doi.org/10.1086/682949

Whitney, K. D., Randell, R. A., & Rieseberg, L. H. (2010). Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. *New Phytologist*, *187*(1), 230–239. https://doi.org/10.1111/j.1469-8137.2010.03234.x

Wolf, J. B. W., & Ellegren, H. (2016). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, *18*(2), 87–100. https://doi.org/10.1038/nrg.2016.133

Wright, S. (1949). The genetical structure of populations. *Annals of Eugenics*, *15*(1), 323–354. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x

Wright, A. E., & Mank, J. E. (2013). The scope and strength of sex-specific selection in genome evolution. *Journal of Evolutionary Biology*, *26*(9), 1841–1853. https://doi.org/10.1111/jeb.12201

Wu, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, *14*(6), 851–865. https://doi.org/10.1046/j.1420-9101.2001.00335.x

Zeng, K., & Charlesworth, B. (2011). The joint effects of background selection and genetic recombination on local gene genealogies. *Genetics*, *189*(1), 251–266. https://doi.org/10.1534/genetics.111.130575

Zeng, K., & Corcoran, P. (2015). The effects of background and interference selection on patterns of genetic variation in subdivided populations. *Genetics*, *201*(4), 1539–1554. https://doi.org/10.1534/genetics.115.178558

Zhang, W., Dasmahapatra, K. K., Mallet, J., Moreira, G. R. P., & Kronforst, M. R. (2016). Genome-wide introgression among distantly related *Heliconius* butterfly species. *Genome Biology*, *17*(1), 25. https://doi.org/10.1186/s13059-016-0889-0

Zhang, D., Song, G., Gao, B., Cheng, Y., Qu, Y., Wu, S., … Lei, F. (2017). Genomic differentiation and patterns of gene flow between two long-tailed tit species (*Aegithalos*). *Molecular Ecology*, *26*(23), 6654–6665. https://doi.org/10.1111/mec.14383

SUPPORTING INFORMATION

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

How to cite this article: Irwin DE, Milá B, Toews DPL, et al. A comparison of genomic islands of differentiation across three young avian species pairs. *Mol Ecol*. 2018;00:1–17. https://doi.org/10.1111/mec.14858