Evidence of Linked Selection on the Z Chromosome of Hybridizing Hummingbirds

C. J. Battey
University of Washington Department of Biology
Current address: University of Oregon Institute of Ecology and Evolution
cbattey2@uoregon.edu

::::November 2019 ::::

https://onlinelibrary.wiley.com/doi/abs/10.1111/evo.13888

1 Abstract

Levels of genetic differentiation vary widely along the genomes of recently diverged species. What processes cause this variation? Here I analyze geographic population structure and genome-wide patterns of variation in the Rufous, Allen's, and Calliope Hummingbirds (*Selasphorus rufus/sasin/calliope*) and assess evidence that linked selection on the Z chromosome drives patterns of genetic differentiation in a pair of hybridizing species. Demographic models, introgression tests, and genotype clustering analyses support a reticulate evolutionary history consistent with divergence during the late Pleistocene followed by gene flow across migrant Rufous and Allen's Hummingbirds during the Holocene. Relative genetic differentiation (F_{st}) is elevated and within-population diversity (π) depressed on the Z chromosome in all interspecific comparisons. The ratio of Z to autosomal within-population diversity is much lower than that expected from population size effects alone, and Tajima's D is depressed on the Z chromosome in *S. rufus* and *S. calliope*. These results suggest that conserved structural features of the genome play a prominent role in shaping genetic differentiation through the early stages of speciation in northern *Selasphorus* hummingbirds, and that the Z chromosome is a likely site of genes underlying behavioral and morphological variation in the group.

2 Introduction

Populations differentiate over time through a combination of mutation, drift, and selection, but the relative importance of these factors in shaping modern biodiversity is contentious. Are differences among species and populations shaped primarily by isolation and drift, or by selection? Variations on this debate occur at all levels of biological hierarchy – from population genetic studies asking if genetic variation is adequately explained by neutral processes (Kern and Hahn, 2018; Kimura, 1968), to phylogeographic studies that ask if speciation has proceeded with or without gene flow (Nosil, 2008), to macroevolutionary analyses that seek to estimate the ratio of sympatric to allopatric speciation by studying range overlap across whole taxonomic classes (Phillimore et al., 2008). At the level of the genome, a similar question arises: what process explains variation in differentiation across the genome?

Recent studies analyzing whole genome data from, among others, hominins (Slon et al., 2018), bears (Kumar et al., 2017), fruit flies (Cooper et al., 2018), butterflies (Heliconius Genome Consortium, 2012), fish (Schumer et al., 2018), and songbirds (Toews et al., 2016) indicate that hybridization is relatively common in animals and is not restricted to recently diverged sister lineages. In some cases morphologically and behaviorally differentiated populations hybridize to the extent that the vast majority of the genome appears to be homogenized with only a few small regions differentiated across species (Ellegren et al., 2012; Toews et al., 2016; Fontaine et al., 2015). Two explanations are typically offered for this pattern: selection against hybrid ancestry in certain regions of the genome leading to local reduction in gene flow in "genomic islands of speciation" (Nosil et al., 2009), or reductions in diversity caused by linked selection within isolated populations leading to local elevation of relative differentiation in regions of low recombination (Cruickshank and Hahn, 2014). Though the debate around which of these scenarios is more common has sometimes set them in opposition, both imply that the core processes driving variation in differentiation across the genome are natural selection and recombination.

Studies of genetic variation within species also support a prominent role for selection and recombination. In the Passenger and Band-Tailed Pigeons, regions of elevated nucleotide diversity (π) occur in parts of the genome with high recombination, and the strength of this relationship appears to vary with effective population size (Murray et al., 2017). Because selection should be more effective in larger populations and will remove diversity from larger chunks of the chromosome in regions with low recombination, these results suggest that linked selection could explain much of the variation in diversity across the genome of these species. If this hypothesis is true we would expect to see similar patterns of differentiation and diversity across multiple species pairs sharing a similar recombination landscape, as is thought to occur in birds (Singhal et al., 2015).

One region in which the impacts of selection should be particularly prominent are the sex chromosomes. In birds, which have a ZW sex chromosome system (fe-

males are ZW), some Z-linked alleles have been linked to sexually selected plumage traits involved in mate selection (Toews et al., 2016; Saether et al., 2007; Campagna et al., 2017). The avian Z chromosome is also known to have a faster mutation rate (around 1.1x; Irwin (2018); Carmichael et al. (2000)), and a lower recombination rate than most autosomes (Kawakami et al., 2014, 2017). Both purifying selection and disruptive selection may then be more common and are expected remove diversity from larger genomic regions on sex chromosomes (Betancourt et al., 2004; Charlesworth, 1994). However, sex chromosomes also have lower population sizes than autosomes (3/4 for the Z; 1/4 for the W), and their diversity can be strongly effected by differential variance in reproductive success between males and females (Charlesworth, 2001) or changing population sizes (Pool and Nielsen, 2007).

Here I investigate the impacts of geographic isolation and linked selection in generating patterns of genetic differentiation in *Selasphorus* hummingbirds in western North America and assess evidence for linked selection in driving regions of elevated relative differentiation, with a particularly strong effect on the Z chromosome. I first test for population structure and isolation-by-distance by comparing genetic variation in individuals caught on the breeding range and during migration to determine if populations retain geographic structure through migration. I then estimate the phylogeographic history of the group by fitting demographic models to whole-genome sequencing data, and use a genome scan approach to identify specific regions of the genome differentiated between and within species. Last I test for variation in genetic diversity on the Z chromosome and autosomes, and compare relative divergence and within-population diversity across multiple population pairs representing a spectrum of divergence times.

2.1 Natural History of Northern *Selasphorus* Hummingbirds

The Rufous (*S. rufus*) and Allen's (*S. sasin*) hummingbirds are the most recently diverged members of the "bee hummingbird" clade that colonized temperate North America roughly 5-10 million years ago (McGuire et al., 2014; Licona-Vera and Ornelas, 2017). Both species are small-bodied with rufous sides, pale bellies, and (for males) an iridescent orange-red gorget. In the field most individuals are indistinguishable, with reliable morphological traits visible without capture limited to a completely rufous back vs a mixed green and rufous back in adult males and a slightly narrower outer tail feather in *S. sasin* (Pyle et al., 1997). The species also have distinctive flight-display routines used in courting females (Myers et al., 2019; Clark, 2014). Females and juveniles are generally indistinguishable unless they can be caught and measured.

S. rufus are obligate migrants that breed in riparian and wet conifer forests across the Pacific Northwest from southern Oregon to central Alaska, extending east to western Montana and the Canadian Rockies. *S. sasin* includes two subspecies, with the migratory *S. s. sasin* breeding in a narrow strip along the California coast from Los Angeles to the Oregon border and the sedentary *S. s. sedentarius* occupying

the Channel Islands and parts of urban southern California year-round (Healy and Calder, 2006). Migratory populations of both species winter primarily in central Mexico, but the specific regions and timing of seasonal movements on the wintering grounds are poorly described. In recent years Rufous Hummingbirds have also been increasingly observed wintering along the Gulf Coast of the southern US (Hill et al., 1998).

Phylogenetic studies of the Trochilidae (McGuire et al. 2014) and Bee Hummingbird clades (Licona-Vera and Ornelas, 2017) identify the rufous and Allen's hummingbirds as sister to the *calliope* hummingbird (*S. calliope*) in the "Northern *Selasphorus*" clade, which diverged from a primarily Central American group including the broad-tailed hummingbird (*S. platycercus*) 4.5 – 2.5 million years ago. A recent study by Myers et al. (2019) documented morphological and behavioral evidence of a hybrid zone between *S. sasin* and *S. rufus* stretching from northern California through southern Oregon. One previous study examined microsatellite variability in rufous hummingbirds across British Columbia and found evidence of a modestly structured population (Bailey et al. 2013), but no study has examined structure across the full breeding range.

3 Methods

3.1 Sampling

I gathered tissue samples from natural history specimens of 85 *S. rufus*, 6 *S. sasin sasin* from the San Francisco Bay Area, 5 *S. sasin sedentarius* from the Channel Islands, and 8 *S. calliope* (Supplementary Table S2-S5). All *S. sasin* and *S. calliope* were collected during the breeding season, primarily by researchers at the Museum of Vertebrate Zoology for studies including Clark (2014, 2015). *S. rufus* samples included 50 individuals collected between March and June along a north-south axis from Oregon to Alaska, and 35 individuals collected during fall migration along an east-west axis from southern California through Texas. DNA was extracted with a Qiagen DNEasy extraction kit and quantified with a qubit BR kit. All samples were prepared for reduced-representation library sequencing via the double-digestion restriction-associated digest protocol (ddRADseq; (Peterson et al., 2012)); using the digestion enzymes sbf1 and msp1 and a size-selection window of 350-900bp. Pooled libraries were then sequenced for 150bp single-end reads on two lanes of a Hiseq 2500 at the UC Berkeley Vincent J Coates sequencing lab. *S. rufus* and *S. sasin* samples were split across the two lanes, while all *S. calliope* samples were on a single lane with other *S. rufus* and *Selasphorus* samples not included in this study.

40 individuals were selected for low-coverage whole-genome sequencing based on fragment size and preliminary population structure analyses of ddRAD data. These samples included 4 *S. sasin sasin*, 4 *S. sasin sedentarius*, 7 *S. calliope*, and 25 *S. rufus* (8 breeding and 17 migrants).

WGS library prep followed the standard illumina protocol for 450-550bp fragments using Truseq nano prep kits, but employed a Bioruptor rather than a Covaris for the initial sonication step. Fragment lengths and concentrations were assayed throughout the library preparation using a bioanalyzer. Twenty uniquely barcoded individuals were pooled in equimolar amounts for each sequencing lane, and samples were sequenced for paired-end 150bp reads on a Hiseq 3000. *S. rufus* and *S. calliope* samples were split across lanes, while all *S. sasin* were sequenced on a single lane with breeding *S. rufus* samples. We found no evidence of sequencing lane effects in downstream analyses (that is, *S. rufus* samples sequenced on the same lane do not cluster in tree-based analyses and retain high F_{st} across distance population pairs).

3.2 Sequence Assembly

ddRAD libraries were demultiplexed using the "-s 1" function of pyRAD (v 1.8; (Eaton, 2014)). I first checked for contamination by aligning all reads to a concatenated fasta file including the human genome and twenty common bacterial and fungal genomes and removing any reads that aligned to this contaminant set at mapping quality > 10 (no samples returned over 0.5% contaminant alignment). Remaining reads were then aligned to a high-quality reference genome of the Anna's Hummingbird (*Calypte anna*) based on a combination of illumina and Pac-Bio long-read sequencing (*Calypte anna*, (Korlach et al., 2017)) using the program Bowtie2 (Langmead and Salzberg, 2012). Reads with a mapping quality below 30 were dropped from the analysis. I used STACKS v2.0 (Catchen et al., 2013) to assemble aligned reads into orthologous loci and call SNPs across individuals.

SNP calling for whole-genome sequence data followed the best practices workflow described in the GATK documentation (https://software.broadinstitute.org/gatk/best-practices/). AdapterRemoval v2 (Lindgreen, 2012) was used to trim adapters and merge overlapping paired reads, and all reads were aligned to the contaminant reference described above. Reads aligning to the contaminant reference with a MQ > 10 were dropped. I then aligned all reads to the Korlach et al. 2017 *C. anna* reference genome using Bowtie2. I sorted and indexed bam files with samtools and marked optical and PCR duplicates using the "MarkDuplicates" tool in Picard (http://broadinstitute.github.io/picard/). Two approaches were used to prepare variant alignments for for downstream inference - static SNP calling with the "UnifiedGenotyper" tool in GATK (McKenna et al., 2010), and genotype likelihood estimation with ANGSD (Korneliussen et al., 2014). Called SNPs were used to confirm sexes of individuals (by comparing read depth on contigs mapping to the Z chromosome), estimate population trees, and infer single-species demographic histories; while genotype likelihoods were used to calculate windowed summary statistics across the genome.

The use of two approaches for SNP calling in whole-genome data was necessary because quality filtering genotypes with low coverage sequencing data likely eliminates some true SNP's such that the effective sequence length used for SNP calling is

lower than the length of the reference genome. Consequently analyses relying on full sequence information (e.g. π) are biased downward in static calls if the entire genome is assumed to be accessible. In both cases sites in the top 2.5% of average read depth were dropped to avoid including paralogs in the alignment. After manually inspecting diversity estimates and read depths across the genome, two contigs on chromosome 4 were also dropped as they appear to include paralogs or copy number variants that were not removed by previous filters. Static calls were filtered in vcftools (Danecek et al., 2011) to retain only biallelic sites with no missing data and a site quality of at least 30. Genotype likelihoods and site frequency estimates were calculated from bam files in ANGSD using the options "-remove_bads -unique_only -minMapQ 20 -minQ 20 -only_proper_pairs 1", following those implemented in (Delmore et al., 2018).

3.3 Phylogenetics and Population clustering

I first estimated a population tree for the concatenated sequence matrix of ddRAD samples in RAxML v8.0 (Stamatakis, 2014), using a GTRGAMMA model of sequence evolution and estimating uncertainty with 100 bootstrap replicates. I then used the static called SNP alignment from whole-genome resequencing to estimate neighbor-joining trees in the R package 'APE' (Paradis et al., 2004), using a k80 model when calculating genetic distances. A single neighbor-joining tree was first estimated for the entire concatenated genotype matrix to describe average genome-wide patterns of relationship. To estimate how well different topologies were supported across the genome I then constructed local neighbor-joining trees from SNPs extracted in 50kb windows across the genome, and used TWISST (Martin and Van Belleghem, 2017) to quantify the proportion of subtrees supporting each species tree topology. Last, a population tree was estimated from whole-genome SNP's in Treemix (Pickrell and Pritchard, 2012) with *S. sasin sasin* and *S. sasin sedentarius* modeled as separate populations and allowing one migration edge. *S. calliope* was set as the outgroup for treemix analyses and SNP's were grouped in bins of 1000 to account for linkage disequilibrium.

I then focused on describing population structure in *S. rufus* and *S. sasin* with principal components analysis (PCA) and model-based clustering in the program *structure* (Pritchard et al., 2000). PCA's were conducted using the R package 'Adegenet' (Jombart, 2008) on the covariance matrix of SNP's across *rufus* and *sasin* ddRAD samples, with missing data replaced by the mean allele frequency across all samples. After finding that the dominant feature of *S. rufus* spatial structure is a hybrid cline with *S. sasin* I created separate maps showing mean PC1 coordinate of breeding *rufus* + *sasin* samples and breeding *rufus* + migrant *rufus* samples; rescaling point colors according to each subset of individuals to visualize both the strong structure between *rufus* and *sasin* and the weak east-west cline during migration.

Structure analyses are biased by inclusion of singletons (Linck and Battey, 2017) and by unequal sample sizes across populations (Lawson et al., 2018), so I first filtered

out all sites with a minor allele count lower than 3 using custom R scripts (see <link removed for anonymity>) and then ran *structure* on *sasin* + breeding *rufus* samples. *Structure* analyses were run using the admixture model with correlated allele frequencies for 100,000 MCMC steps with 10,000 steps of burn-in for five replicates under different starting seeds at K values of 1-4. I used structure harvester (Earl and vonHoldt, 2012); (`http://taylor0.biology.ucla.edu/structureHarvester/`) to rank K values by second-order change in marginal likelihood (Evanno et al., 2005), and show the top-ranked model in figure 3.

To test for isolation by distance within *S. rufus*, I assessed correlations between genetic and geographic distance matrices using a Mantel test. Genetic distances were estimated assuming a k80 model of sequence evolution in the R package 'ape' (Paradis et al., 2004). As a secondary test of IBD, I also used linear regression to test for correlations between the first principal component of the allele frequency matrix and latitude (for breeding samples) or longitude (for migrants). Results from all analyses were plotted using the R packages 'ggplot2' (Wickham, 2016), 'maps' (Becker et al., 2013), and 'cowplot' (Wilke, 2016) in R v3.4.1 (Team and Others, 2013).

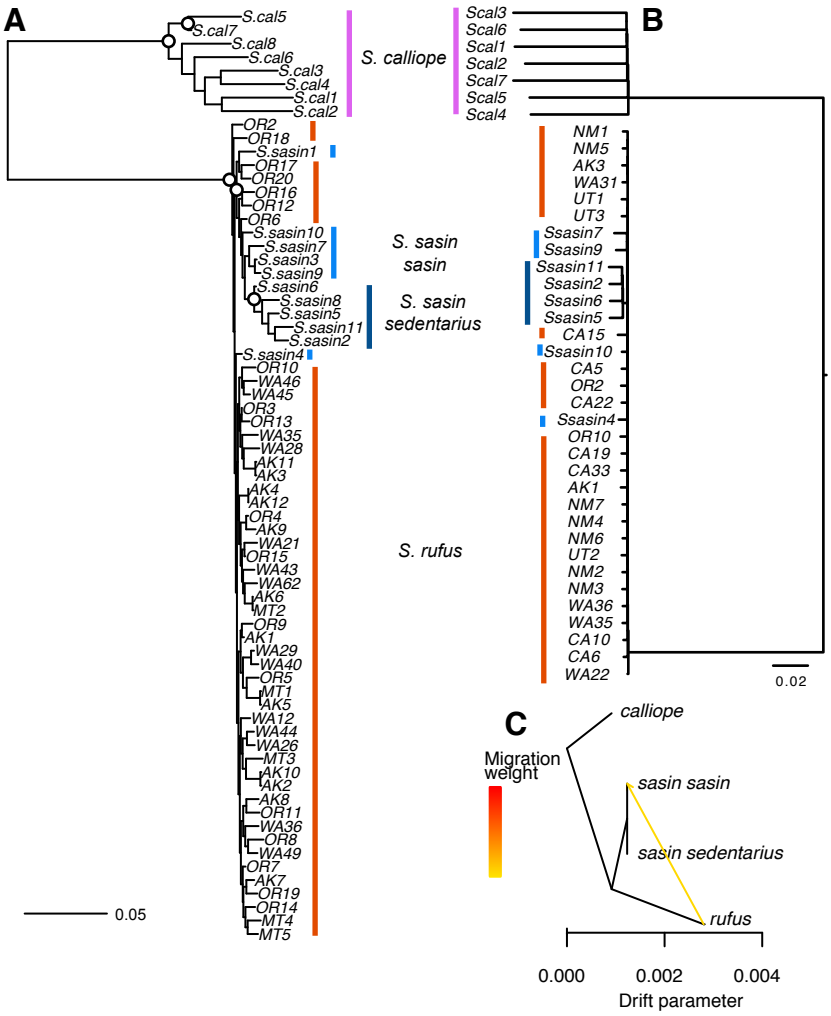

Figure 1: A: RAxML tree of concatenated ddRAD sequences from breeding individuals. Circles indicate nodes with over 90% bootstrap support. B: neighbor-joining tree of whole-genome SNP's. C: treemix population tree with one admixture edge.

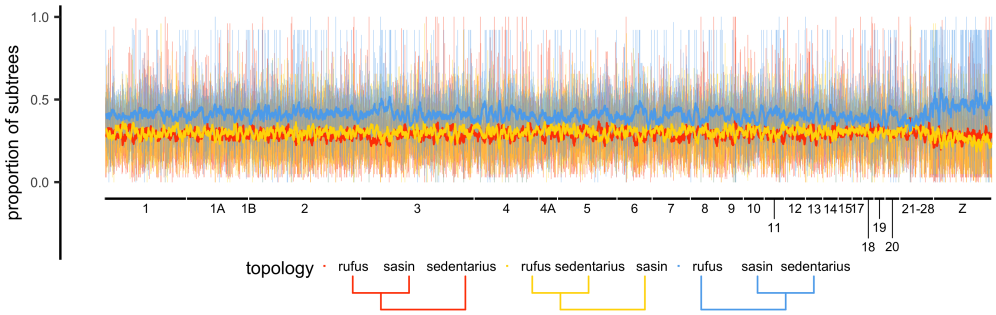

Figure 2: Local tree support across the genome. Neighbor-joining trees were constructed from 50kb windows. Thin lines show the proportion of subtrees supporting each topology, and heavy lines are a 1-megabase rolling average. Contigs of the *C. anna* reference genome are ordered by their approximate position in the *T. guttata* genome.

3.4 Demographic Models

SMC++ (Terhorst et al., 2017) was used to estimate demographic histories from whole-genome sequence data. SMC++ implements a simple demographic model without migration between populations, but takes advantage of more of the information in whole-genome data than standard SFS based approaches like $\delta a \delta i$ by analyzing the distribution of variation across the genome rather than working from a single genome-wide estimate of the SFS. Because the whole genome data was unphased and we lack a detailed map of inaccessible regions of the *C. anna* reference genome, contiguous stretches of homozygosity greater than 30kbp were masked and inferences of population size in the last 2,000 years were dropped from smc++ run. This reduces power to infer events in the recent past, but should minimize false signals of recent bottlenecks (Terhorst et al., 2017). I ran all contigs $>$ 1 x 10^6bp, covering a total of 8.5 x 10^8bp, and constructed composite datasets using 4-8 different individuals as "designated individuals" (see Terhorst et al. 2017). Uncertainty in SMC++ analyses was estimated by creating bootstrapped datasets (sampling over contigs of the *C. anna* reference genome) and re-fitting models to 10 bootstrap replicates per population. I assumed a mutation rate of 4.6×10^{-9} substitutions/bp/generation and a generation time of one year, following a recent study based on deep sequencing of a pedigreed population of flycatchers (Smeds et al., 2016).

3.5 Introgression

The D statistic (Green et al., 2010) was used to test for introgression across all species, assuming a true topology of (*calliope*,(*rufus*,(*sasin sasin*, *sasin sedentarius*))), with *C. anna* used as an outgroup for all tests. *D* was calculated for each combination of individuals conforming to the tree above using the "doAbbababba"

function in ANGSD (6,113 comparisons total). Significance was assessed by block jacknife over 1mbp windows. Results were summarized by estimating the range of D values and associated bootstrap Z scores for each unique subspecies-level topology. A significance cutoff of Z = 4.31 was applied, equivalent to a p value of 0.05 after correcting for multiple comparisons. As a conservative test of introgression at the populatin level, I considered a population pair to be introgressed if the median Z score across all combinations of individuals in a given topology was greater than the cutoff.

3.6 Genome Scans

Within-population genetic diversity (π) and Tajima's D was calculated in 50,000bp non-overlapping windows for each species using the "doThetaStat" program in ANGSD. Pairwise relative divergence (F_{st}) was estimated with the "realSFS" program for three interspecific comparisons involving *S. rufus*: (1) *rufus* x *calliope*, (2) *rufus* x *sasin sedentarius*, and (3) *rufus* x *sasin sasin*; as well as for two intraspecific comparisons of populations showing behavioral or morphological variation: (4) *sasin sasin* x *sasin sedentarius* and (5) east x west migrating *rufus*. All of the above analyses were limited to only male samples (7 breeding *S. rufus*, 13 migrant *S. rufus*, 3 *S. calliope*, 4 *S. sasin sedentarius*, and 2 *S. sasin sasin*) to ensure that results were not biased by low coverage on the Z chromosome or incorrect alignment of variants from the W chromosome in females.

To estimate the position of *C. anna* contigs on avian chromosomes and facilitate comparisons with previous studies, each contig in the Korlach et al. 2017 *C. anna* genome was aligned to the most recent chromosome-scale assembly of the Zebra Finch (T̂aenopygia guttata; Warren et al. 2010) with the program 'MUMmer' (Delcher et al. 2003) and contigs were arranged by the start position of the longest matching stretch over 10,000 base pairs in the *T. guttata* genome. Contigs without matches in the *T. guttata* genome or with over 10,000bp of matching sequence to multiple chromosomes were binned into a separate "NA" category. Because the reference individual in Korlach et al. 2017 is a male, the W chromosome is not included in this analysis. All windowed analyses were conducted at the contig rather than the chromosome level.

The top and bottom 0.5% of windows for each summary statistic were identified as outliers. Linear regression was used to test for correlations in F_{st} across different pairwise comparisons, and for per-window nucleotide diversity (π) across populations and species. After initial results suggested that differentiation was concentrated on the Z chromosome, I also estimated mean values of all statistics on the Z and the autosomes (excluding "NA" contigs).

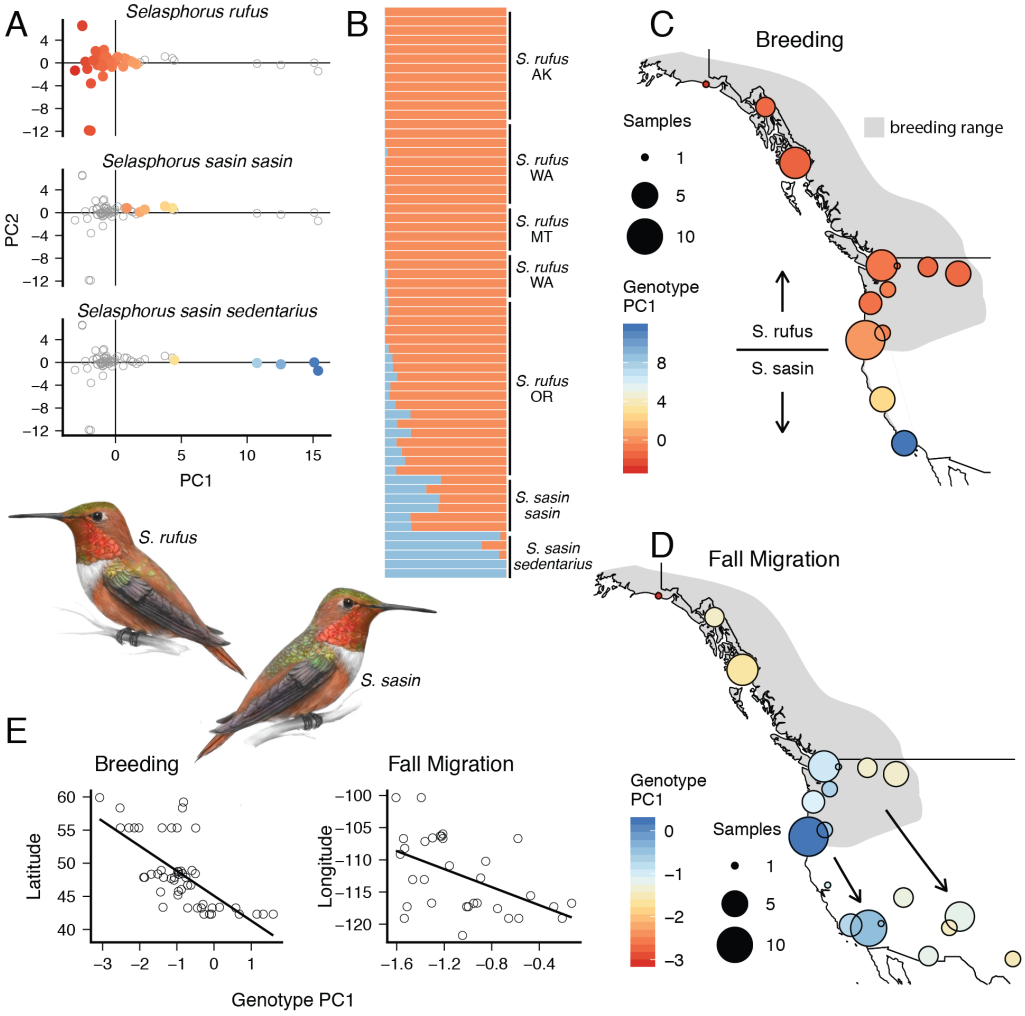

Figure 3: Spatial structure of *S. rufus* and *S. sasin* described from ddRAD data. A: PCA of *S. rufus* and *S. sasin* samples. B: `Structure` results for breeding samples at k=2. C: Map of breeding *S. rufus* and *S. sasin* samples with colors scaled by the mean genotype PC1 in each locality, showing a strong north-south cline concordant with known species ranges. Breeding ranges are shown in grey (note *S. sasin* is restricted to a narrow coastal band in California). D: Subset map including only breeding and fall-migrant *S. rufus*, showing a weak north-south cline across the breeding range and corresponding east-west cline during fall migration. E: Correlations of genotype PC1 with breeding latitude and fall longitude. Two highly admixed California migrants with PC1>1 were removed from the right panel. Illustrations show adult male *S. rufus* and *S. sasin*.

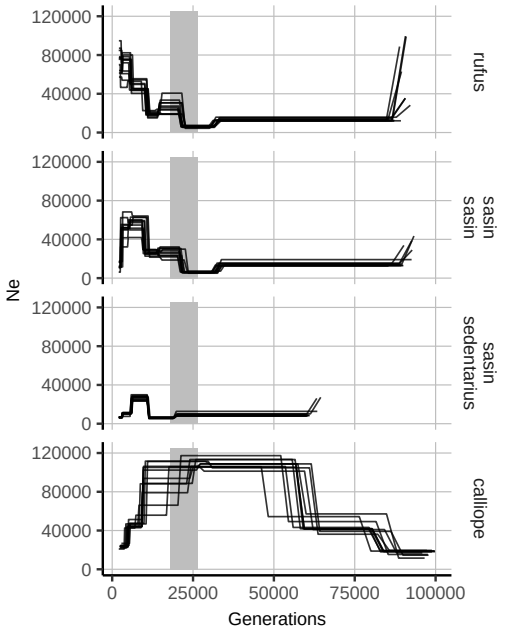

Figure 4: Estimates of population size histories from smc++ assuming a mutation rate of $4.6 \times 10^{-9}/bp/year$ (Smeds et al., 2016). Black lines show bootstrap replicates sampled over contigs, and the grey bar indicates the timing of the last glacial maximum (LGM).

4 Results

5 Sequence Data

For ddRAD data we recovered an average of 750,112 reads per sample, of which an average of 639,841 passed filters. An average of 62% of reads per sample aligned to a single location in the Korlach et al. (2017) *C. anna* genome assembly in a single location and were retained for downstream inference. After clustering and paralog filters we retained a total of 14,616 SNPs from 6,575 putatively unlinked genomic regions sequenced in at least 90% of individuals, with an average of 14.9x coverage. Though the proportion of reads aligning to the reference was relatively low, we found that an alternate de-novo clustering assembly with pyRAD returned very similar numbers of SNPs and produced the same patterns in a PCA; suggesting that our assembly is sufficiently high quality to describe spatial structure in the group.

The whole-genome resequencing data returned an average of 3.02×10^7 reads per sample. After joining paired-end reads we succesfully aligned an average of 88% of reads to the Korlach et al. (2017) *C. anna* reference assembly. Following

removal of optical and PCR duplicates we retained an average of 25,788,414 reads per sample. This yields an expected coverage of 4.07x across the ≈ 950,000,000bp *C. anna* reference assembly. For the called SNP alignment produced with GATK and filtered to retain only sites with no missing data, we retained a total of 5,688,922 variants with a mean sequencing depth of 4.35x. Given a genome size of 1GB and Hiseq 3000 throughput of ≈ 87.5GB per lane with 150-bp paired-end reads, our realized coverage values are close to our expected coverage of 4.375 given 20 samples per lane.

5.1 Phylogenetics and Population Structure

Phylogenetic and population tree inference on both concatenated ddRAD data and static whole-genome SNP calls found that *S. calliope* is sister to a combined clade of *S. rufus* and *S. sasin* (Figure 1), consistent with a recent study examining variation at six sanger-sequenced nuclear loci (Licona-Vera & Ornelas 2017). Though *S. calliope* is monophyletic, *S. rufus* and *S. sasin* form a paraphyletic grade in concatenated analysis. This is consistent with the TWISST (Martin and Van Belleghem, 2017) analysis of local tree support across the genome, which found that the expected species tree lumping the two *S. sasin* subspecies is the most well-supported tree in 54.6% of windows across most of the genome (Figure 2). Though local tree support varies considerably across the genome, contigs aligning to the Z chromosome stand out in more strongly supporting a topology lumping the two *sasin* subspecies (65.2% of windows on the Z). The topology inferred by treemix was consistent with all previous analyses, and also infers a migration edge from *S. rufus* to *S. sasin sasin*. This is consistent with their parapatric ranges and lack of monophyly in concatenated trees.

PCA's and Genotype clustering in *structure* both indicate that two-population models best describe the data, and suggest that most *S. sasin sasin* are admixed with *S. rufus* (Figure 3). *Structure* also infers 1-7% *S. sasin* ancestry in all breeding individuals from southern Oregon and 4 out of 16 migrants from California. However, the exact ancestry proportions in putative hybrids were highly sensitive to variation in sample size across runs so should be interpreted with some caution. In PC space the primary axis of differentiation is between *S. sasin sedentarius* and *S. rufus*, with *S. sasin sasin* and Oregon *S. rufus* falling in intermediate locations along PC1. Within *S. rufus*, both genotype PC1 and the fraction of inferred *S. sasin* ancestry in *structure* results decline with latitude (PC1: $p<0.01$, $R^2=0.48$). Across migrant populations of *S. rufus* PC1 is negatively correlated with longitude ($p=02$, $R^2=0.24$), suggesting that California migrants come primarily from southern and western areas of the breeding range while eastern migrants come from the rest of the range.

Mantel tests found that genetic and geographic distance were not correlated either across the full breeding range of *S. rufus* ($p=0.77$) or among migrants ($p=0.71$). However, genetic and geographic distance were correlated across breeding samples from Oregon and Washington ($p=0.017$), likely reflecting a consistent signal of *S.*

sasin ancestry in Oregon birds.

5.2 Demography

Single-population demographic histories from smc++ inferred a population bottleneck between 20 and 30kya in *S. rufus* and *S. sasin sasin*, possibly reflecting the impacts of the last glacial maximum (LGM) c. 18-26.5kya (Figure 4). Both *S. rufus* and *S. sasin* then grow to populations between 50,000 and 100,000 by 10kya, with *S. sasin* crashing in the recent past and *S. rufus* continuing to grow. *S. sasin sedentarius* is inferred to have experienced a small reduction in population size near the end of the last glacial maximum (LGM) followed by a peak and then decline after 7kya. *S. calliope* grows to reach a peak population prior to the LGM and gradually declines after c. 20kya (similar to several of the avian demographic trajectories estimated from whole-genome data in Nadachowska-Brzyska et al. (2015)). Note that these results should be interpreted cautiously – the exact effects of the rough run-of-homozygosity length mask used here are not well known, and our low coverage makes estimation of the true sequence length searched difficult to estimate. In addition, many processes other than population size change will cause variation in the coalescence rate through time, including varying migration rates and changing population structure (Mazet et al., 2016).

5.3 Introgression

D tests found evidence of significant introgression between *S. rufus* and *S. sasin* (Figure 5, Supplementary Table S1). Most introgression appears to be from *rufus* into *sasin*, because D scores were much higher for tests with two *sasin* in the P1/P2 positions than for those with two *rufus* in the P1/P2 positions. All combinations of individuals tested on topologies with different subspecies of *S. sasin* in the P1/P2 positions returned significant results, reflecting a consistent pattern of excess *S. rufus* ancestry in *S. sasin sasin*.

Lower levels of introgression were also observed from *S. calliope* into both *S. sasin* and *S. rufus* (13-31% of tests), and from *S. sasin* into *S. rufus* (16-18% of tests). However these tests may simply reflect high levels of introgression between *rufus* and *sasin*. For example, when testing the topology ((*sasin*, *sasin*),*calliope*), some *sasin* may have received ancestral alleles shared with *S. calliope* as a result of gene flow from *rufus* and consequently have an unusually high number of discordant sites. The lack of any significant gene flow into *calliope* supports this hypothesis, though it could also be interpreted as evidence of unidirectional gene flow.

5.4 Genome Scans

Regions with the highest relative differentiation (F_{st}) in interspecific comparisons across northern *Selasphorus* are concentrated on the Z chromosome, but these regions

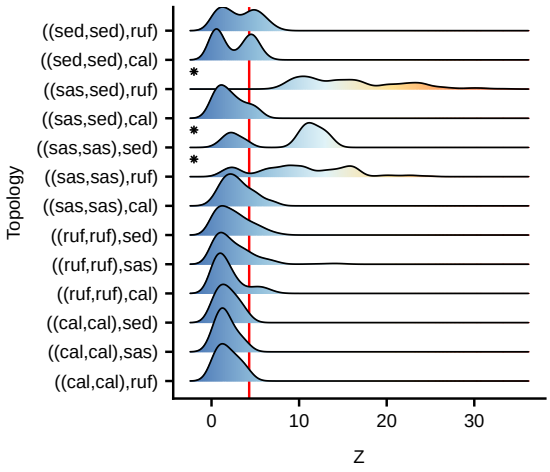

Figure 5: Z scores for "ABBA-BABA" introgression tests. High values indicate significant introgression. Distributions are over combinations of individuals corresponding to each topology. Toplogies are listed as ((P1,P2),P3), with abbreviations cal=*calliope*, ruf=*rufus*, sas=*sasin sasin*, and sed=*sasin sedentarius*. Asterisks indicate topologies with a median Z score over 4.31 (p<0.05 after correcting for 6,113 comparisons), and can be interpreted as trees for which a randomly sampled set of individuals would show significant signal of introgression.

have much lower than average within-population diversity (π) (Figure 6). F_{st} is not elevated on the Z chromosome in intraspecific comparisons of either the two *S. sasin* subspecies, or between east and west-migrating *S. rufus*. Tajima's D is lower on the Z chromosome than the autosomes in all populations (Figure 7, Supplementary Figure S2), with the largest relative differences in *S. rufus* and *S. sasin sedentarius*.

Per-window nucleotide diversity is strongly correlated in all comparisons of *rufus* and *sasin* (R^2=0.82-0.89) and moderately correlated in comparisons with *S. calliope* (R^2=0.44-0.47; Figure 7). In *rufus* and *sasin*, peaks of within population diversity occur near the end of macrochromosomes (Figure 6), as was also observed in the Passenger Pigeon genome (Murray et al., 2017). The ratio of autosomal to Z-linked π ranges from 0.44 to 0.58 across species (Table 1) – lower than the 0.75 ratio predicted by population size effects alone (Table 1). F_{st} is weakly correlated across population pairs including *rufus* and *sasin* (R^2=0.04 - 0.29) but uncorrelated between east x west *rufus* and all other population pairs (Figure S1).

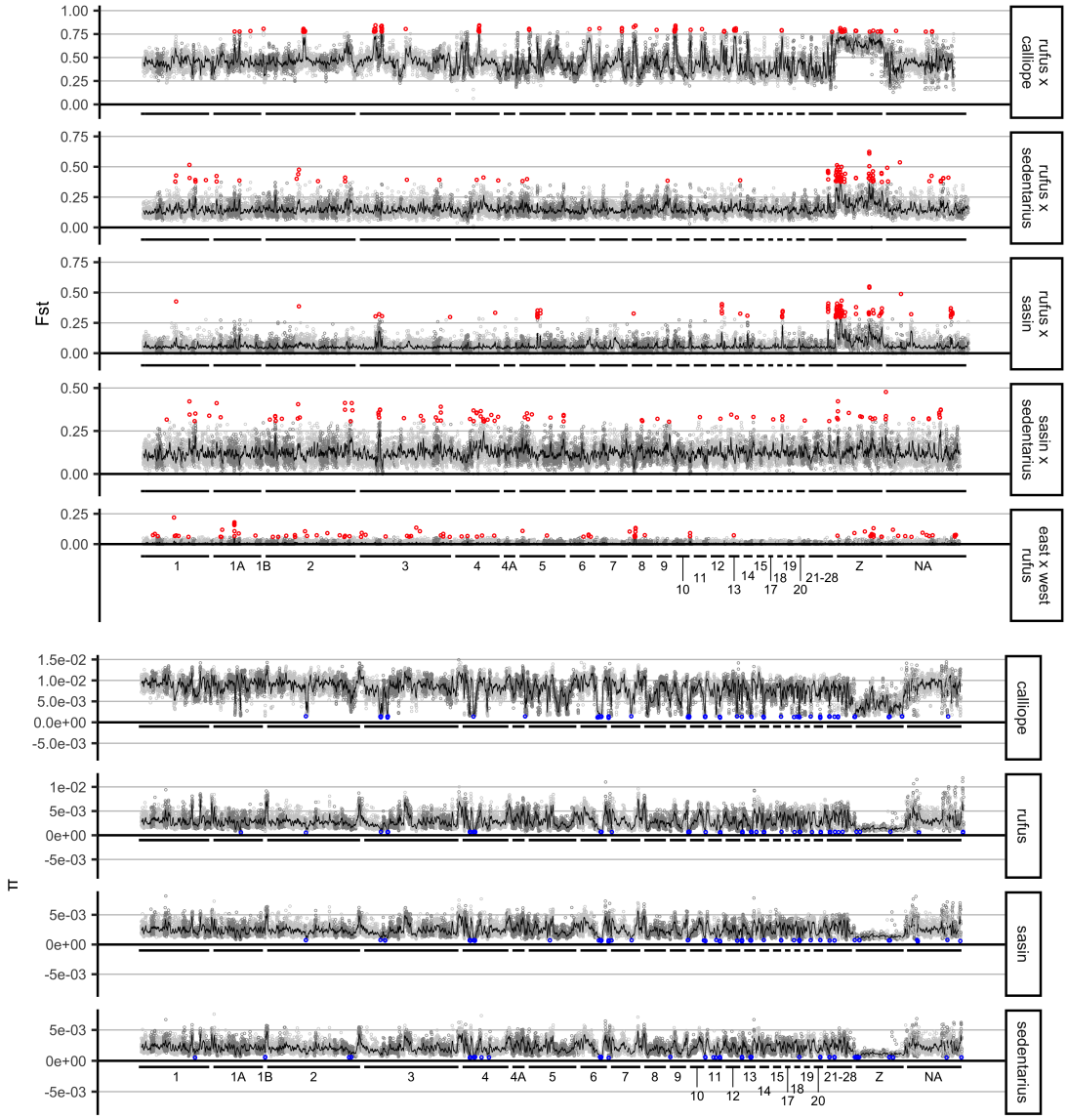

Figure 6: Relative divergence (F_{st}) and within-population diversity (π) in 50,000bp nonoverlapping windows. Red and blue circles are the top and bottom 0.5% of windows, the black line is a 1e6bp rolling average, and grey banding shows contigs of the Korlach et al. 2017 *C. anna* genome assembly. Contigs are ordered by their approximate position in the *T. guttata* genome.

Species	π_a	π_z	π_z/π_a
calliope	0.00857	0.00380	0.44400
rufus	0.00266	0.00136	0.50900
sasin sasin	0.00230	0.00133	0.57900
sasin sedentarius	0.00197	0.00108	0.54700

Table 1: Mean within–population diversity on autosomes (π_a) and the Z-chromosome (π_z), by species.

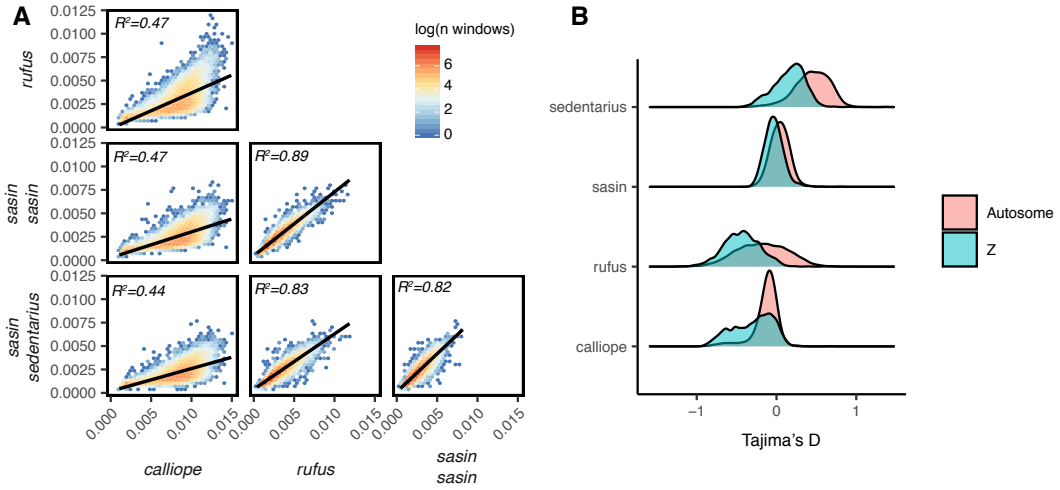

Figure 7: A: Correlation of nucleotide diversity (π) by genomic window. B: Distribution of Tajima's D on autosomes and the Z chromosome.

6 Discussion

6.1 Phylogeography and Spatial Structure

S. rufus and *S. sasin* diverged from *S. calliope* during the early Pleistocene (Licona-Vera and Ornelas, 2017) and their demographic trends begin to diverge in the late Pleistocene at the approximate time of the LGM (Figure 5). Introgression and clustering analyses indicate that *S. rufus* and *S. sasin* have continued to experience gene flow after their divergence, with most introgression limited to the migratory subspecies *S. sasin sasin*. Whole-genome demographic analyses infer a population bottleneck in both species at approximately 25,000 years in the past, possible reflecting the impacts of the last glacial maximum from 26 - 18kya. Both migratory populations then increase in population, which is consistent with an increase in available habitat in the Pacific Northwest after glacial retreat. Together these observations suggest that climate cycling in the late Pleistocene played a major role in

lineage diversification in the group.

For *S. rufus*, most of the modern breeding range was either glaciated (all of the Puget Sound lowlands and most high-altitude regions of the Pacific Northwest) or was characterized by a much drier, colder climate than modern times until glacial retreat beginning around 18ya (Hovan et al., 1991). The expansion in suitable habitats as the climate of the Pacific Northwest became warmer and wetter likely contributed to increasing populations of this species, which is reflected in estimates of rising populations after c. 20kya in single-population demographic models. Glacial retreat also corresponded with a significant drying and warming of the climate in California after 26 kya and a large rise in sea levels after 15kya (Herbert et al., 2001). In coastal California, sea level rise after the LGM shrunk the size of the coastal plain and decreased the land area of the Channel Islands by nearly 20%, with the most rapid change in land area occurring c. 3.5-5kya (Kinlan et al., 2005). Demographic models of *S. sasin sedentarius* indicate that effective population size has decreased over time, which may reflect the impacts of these biogeographic processes.

S. calliope shows a different demographic trajectory, with N_e peaking during the LGM and then gradually declining to the present day. This species currently occurs in the relatively drier habitats of the Great Basin, Central Valley, and the east slopes of the Cascade and Sierra Nevada mountains. In the south this region received significantly more precipitation during Pleistocene glacial periods (McGee et al., 2018), which may have expanded the area of suitable habitat and allowed larger population sizes in the past.

The migratory *S. sasin sasin* appears to be a hybrid taxon, including ancestry from both *S. rufus* and *S. sasin*. The subspecies likely diverged from *S. sasin sedentarius* during or after the onset of glacial retreat and subsequently began hybridizing with *S. rufus*. This has resulted in a highly admixed genome in which the "species tree" topology lumping the migratory and resident subspecies of *S. sasin* is most common in only 54.6% of genomic windows. Even within windows supporting the species tree, the average proportion of subtrees supporting this topology is just 49%. Genetic evidence of hybridization described here is consistent with a recent study of behavior and morphology showing clines in both stretching across the putative range boundary in northern California and southern Oregon (Myers et al., 2019).

For extant populations of *S. rufus* the dominant spatial pattern of genetic variation is caused by hybridization with *S. sasin* to the south. This creates a north-south cline across the breeding range which is maintained as a weak east-west cline during migration (Figure 3). Most fall migrants in the Rockies breed in the northern and eastern areas of the breeding range, while most migrants in the Sierra Nevada breed in Oregon and southern Washington. Assortative mating across breeding populations may increase fitness during migration by passing on combinations of alleles that direct juveniles either west to the Sierra Nevada or east to the Rockies while avoiding the relatively poor stopover habitats of the Columbia Plateau and Great Basin. However, the differentiation I observed across migratory populations is much

less distinct than that found in species with true migratory divides such as Swainson's Thrush (*Catharus ustulatus*; (Delmore and Irwin, 2014)) or Painted Buntings (*Passerina ciris*), (Battey et al., 2018)), possibly because *S. rufus* is a much younger species than either of these taxa, which diverged from their closest ancestors in the early Pleistocene and likely maintained structured populations through multiple glacial cycles.

6.2 Linked Selection and Sex Chromosome Differentiation

Relative differentiation across species of northern *Selasphorus* is concentrated on the Z chromosome, and this region is also enriched for sites supporting a topology that lumps the migratory and resident subspecies of *S. sasin* (Figure 6, Figure 2). In addition, both within-population diversity and Tajima's D are reduced on the Z chromosome relative to the autosomes in all populations (Figure 7). These observations are analogous to the "fast X" effects documented in many species with X/Y sex chromosome systems (Vicoso and Charlesworth (2006); Meisel and Connallon (2013)).

In part these patterns are caused by demographic effects. The Z chromosome has 3/4 the population size of the autosomes so we expect drift to create a Z:autosome diversity ratio of ≈ 0.75 at equilibrium. The smaller population size of sex chromosomes also causes them to equilibrate more quickly to changing demographic conditions (Pool and Nielsen, 2007), such that recent bottlenecks will depress the Z:autosome diversity ratio while population growth will increase it (Pool and Nielsen, 2007; Van Belleghem et al., 2018). Last, higher variance in male relative to female reproductive success could lead to additional drops in Z-linked N_e, with theory supporting a minimum Z:autosome diversity ratio of 0.56 (Charlesworth, 2001).

Here we estimate Z:autosome diversity ratios ranging from 0.44 to 0.58 (Table 1) – at the extreme low end of those caused by hemizygosity and variance in male reproductive success. Though these values could be generated by recent population bottlenecks (Pool and Nielsen, 2007), we instead inferred recent population growth in *S. rufus* and *S. sasin* (Figure 4); suggesting that demographic effects are unlikely to explain all of the patterns we observed.

An alternate explanation is that background selection and recurrent selection against hybrid ancestry on the Z chromosome explain the observations reported here. Background selection is expected to decrease sex-chromosome relative to autosomal diversity, and this effect is strongest when maladaptive alleles are recessive and sexual selection is strong (Charlesworth, 1994). *Selasphorus* hummingbirds are sexually dimorphic lekking species in which males conduct display flights and "sing" for females by generating noise through vibration of tail and wing feathers during dives (Clark, 2014; Myers et al., 2019), suggesting that sexual selection does occur in these species. In addition, birds have slightly male-biased mutation rates (Carmichael et al., 2000), which should increase the rate of new mutations occurring on the Z relative to autosomes and further increase the effects of background

selection.

Background selection could then lead to depressed Z chromosome diversity, and this in turn could cause elevated F_{st} (Cruickshank and Hahn, 2014). This process would also explain the reduced Tajima's D observed on the Z relative to the autosomes in all species, and is consistent with the strong correlation in per-window nucleotide diversity we observed across the full genome in all populations (Figure 6). Specifically, we expect the effects of background selection on linked neutral regions to be strongest when recombination rate is low (Charlesworth, 1994), and because birds have a relatively conserved recombination landscape (Kawakami et al., 2017; Singhal et al., 2015) this would lead to a highly correlated landscape of genomic diversity.

However, two pieces of evidence suggests that "divergence hitchhiking" or "barrier loci" (Ravinet et al., 2017) on the Z also play a role in the elevated *rufus* x *sasin* differentiation seen here. First, we observe that F_{st} is elevated in between-species but not within-species comparisons (Figure 6), suggesting that within-population processes are not driving all of this trend. And second, local tree reconstruction shows that the Z chromosome is specifically enriched for a topology lumping *S. sasin sasin* and *S. sasin sedentarius* rather than simply being a global F_{st} outlier. Though not conclusive, these patterns would be generated by a scenario in which loci on the Z chromosome were locally adapted within each species and generated selection against individuals with hybrid Z-chromosome ancestry (possibly through the extensive sexual selection on plumage and behavior documented in Myers et al. (2019)).

Though I did not attempt to identify specific variants or genes under selection in this study, it is notable that males are much more morphologically differentiated than females across all species studied here – to the degree that females and juveniles often cannot be distinguished in the field. This is similar to the case of several other species in which Z-linked genes involved in plumage differentiation have recently been reported as F_{st} outliers in several other avian species complexes (Toews et al., 2016; Saether et al., 2007; Campagna et al., 2017); suggesting that this may be a recurrent feature of speciation in birds.

6.3 Systematics

S. rufus and *S. sasin* are often indistinguishable in the field, appear to hybridize frequently (Myers et al., 2019), have heavily introgressed genomes (Figure 5), and diverged much more recently than any other pair of Hummingbird species in North America (McGuire et al., 2014; Licona-Vera and Ornelas, 2017). Further, all *S. sasin sasin* ddRAD samples were inferred to have the majority of their ancestry from *S. rufus* despite being caught at least 500 kilometers south of the range boundary, and autosomal F_{st} is lower between *S. rufus* and *S. sasin sasin* than between the two subspecies of *S. sasin*. Are they really different species, and if so, should *S. sasin sasin* be seen as the migratory subspecies of Allen's Hummingbird or the southernmost

population of Rufous Hummingbird?

Under strict versions of biological (Mayr, 1948) or phylogenetic (Donoghue, 1985; Baum and Shaw, 1995) species criteria, *S. rufus* and *S. sasin* are not species because they interbreed and are not reciprocally monophyletic over the vast majority of the genome (Figure 2). Under the criteria used by the US Fish and Wildlife Service when delimiting Distinct Population Segments as "species" for the purposes of the Endangered Species Act (Waples and Gaggiotti, 2006), which emphasizes demographic independence and behavioral or morphological differentiation, they are certainly species. As with most cases of divergence with gene flow, whether or not we recognize populations as species depends on the criteria we use rather than the underlying concept of populations differentiating over time, and reasonable arguments can be made advocating for the recognition of "species" at any stage of divergence (De Queiroz, 2007).

More concretely, this analysis suggests that *S. rufus* and *S. sasin* as currently recognized are separate lineages displaying some degree of assortative mating and likely capable of responding to different selective regimes in the face of gene flow. First, assortative mating is reflected in the much greater divergence between Oregon *S. rufus* and the parapatric *S. sasin sasin* than is observed between any pair of breeding or migratory populations within *S. rufus*, and by the inference of at least two genotype clusters in *structure* results and PCA's. Second, breeding behaviors are reported to be differentiated between species (Healy and Calder, 2006), with *S. rufus* and *S. sasin* each performing characteristic flight patterns while displaying for females at leks and putative hybrids conducting intermediate displays (Myers et al., 2019).

Third, relative differentiation of the sex chromosomes is elevated between *rufus* and both *sasin* subspecies, but not within either species. The reduction in genetic diversity and Tajima's D on this chromosome suggests that it is under selection in both species, and the elevated differentiation (relative to autosomes) in interspecific but not intraspecific comparisons suggests that the Z chromosome specifically is resistant to gene flow across species. Similar to cases of Blue and Golden-Winged Warblers (Toews et al., 2016) and Ficedula flycatchers (Saether et al., 2007), loci associated with sexually selected plumage or behavioral traits may be concentrated on the Z in *Selasphorus* hummingbirds, and elevated relative differentiation may reflect disruptive selection and indicate that combinations of traits important in reproductive isolation are maintained in different species. Whether this level of differentiation and selective coherence will eventually lead to reproductive isolation is unknown, but at present these populations appear to be evolving on separate evolutionary trajectories.

7 Conclusion

S. rufus and *S. sasin* diverged in the late Pleistocene and subsequently experienced

different demographic histories, with both migratory populations increasing in population and hybridizing in northern California and southern Oregon while sedentary birds on the Channel Islands and southern California coast maintained low populations through the Holocene. Relative genetic differentiation across northern *Selasphorus* is concentrated on the Z chromosome, which is also the least diverse region of the genome in all species. The drop in diversity observed on the Z chromosome is greater than that expected from population size and mutation rate effects alone, suggesting that increased variance in male reproductive success and linked selection explains much of the reduction in diversity. Though within-population diversity is strongly correlated across species, relative divergence (F_{st}) is weakly correlated in different taxonomic comparisons. Together these results suggest that linked selection is a prominent driver of variation in genetic differentiation across the genome of northern *Selasphorus* hummingbirds, and that the Z chromosome is a likely site of genes underlying behavioral and morphological variation in the group.

8 Data

Demultiplexed sequence reads for the WGS data are available at the sequence read archive (https://www.ncbi.nlm.nih.gov/sra) under project PRJNA576990. ddRAD data, specimen information, and analysis scripts are available on the Dryad data repository: `https://doi.org/10.5061/dryad.s1rn8pk3s`. Please contact the corresponding author at cjbattey@gmail.com for access to any other intermediate files generated during sequence assembly or analysis.

9 Acknowledgements

Tissue samples analyzed in this study were contributed by the Burke Museum of Natural History, Louisiana State University Museum of Natural Sciences, University of New Mexico Museum of Southwestern Biology, University of California Berkeley Museum of Vertebrate Zoology, and University of Alaska Museum of the North. Many thanks to the curators, staff, and volunteers at these institutions, who made this work possible. Ethan Linck, Adam Leaché, Andrew Kern, Lauren Buckley, and John Klicka provided helpful comments on early versions of this manuscript. Members of Klicka Lab at the University of Washington from 2013 - 2018 all contributed to the development and refinement of ideas and analyses presented here – thanks to Ethan Linck, Cooper French, Dave Slager, John Klicka, and Rob Bryson. Thoughtful comments from two anonymous reviewers significantly improved the final paper. This project was supported by grants from the University of Washington Dept. of Biology and NSF grant DEB1600945.

10 Supplementary Figures and Tables

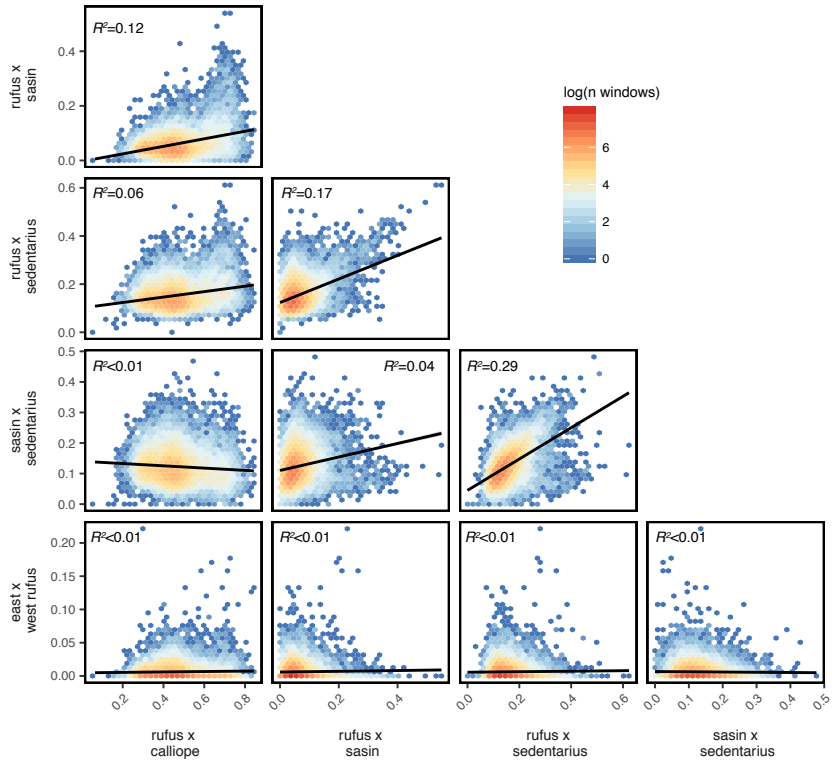

Figure S1: Correlation of relative divergence (F_{st}) by genomic window across population pairs.

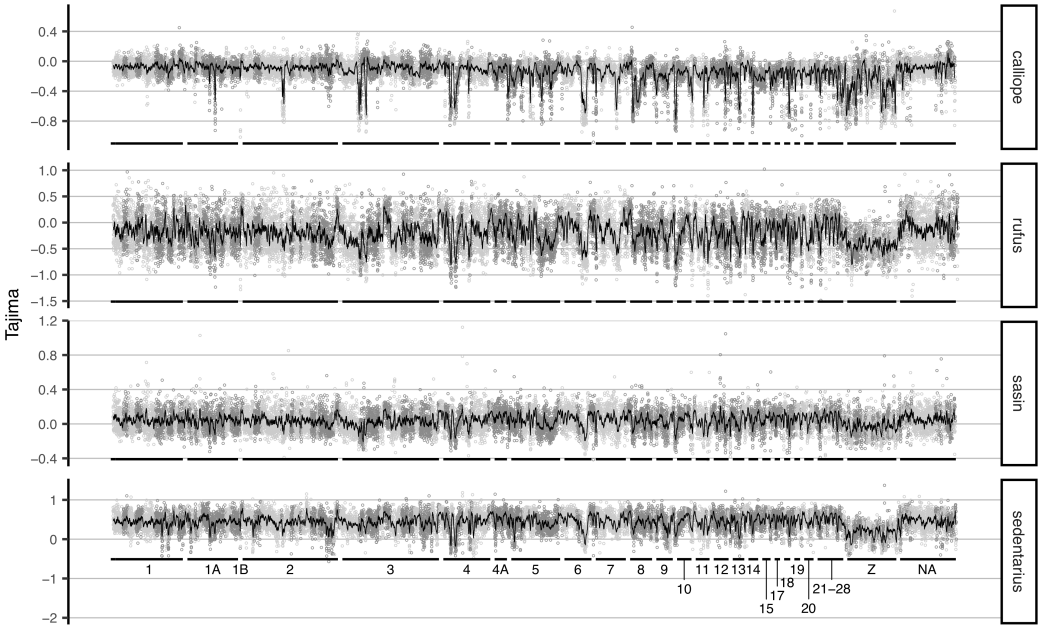

Figure S2: Tajima's D in 50,000bp nonoverlapping windows.

Topology	median(D)	median(Z)	Proportion Significant
((cal,cal),ruf)	2.15×10^{-3}	1.60	0.01
((cal,cal),sas)	1.64×10^{-3}	1.37	0
((cal,cal),sed)	2.19×10^{-3}	1.59	0
((ruf,ruf),cal)	2.47×10^{-3}	1.30	0.13
((ruf,ruf),sas)	3.38×10^{-3}	1.84	0.16
((ruf,ruf),sed)	4.24×10^{-3}	2.12	0.18
((sas,sas),cal)	4.93×10^{-3}	2.76	0.21
((sas,sas),ruf)	2.26×10^{-2}	9.85	0.82
((sas,sas),sed)	2.72×10^{-2}	10.7	0.67
((sas,sed),cal)	3.82×10^{-3}	1.83	0.14
((sas,sed),ruf)	3.22×10^{-2}	14.8	1
((sed,sed),cal)	4.55×10^{-3}	1.99	0.31
((sed,sed),ruf)	5.9×10^{-3}	2.94	0.38

Table S1: Summary of D test results. D values and counts of ABBA and BABA sites are medians; Z scores are median absolute values. Proportion significant is the proportion of all individual combinations under a given topology with a abs(Z) $>$ 4.3.

SampleID	*Selasphorus* sp.	Museum	Prep Number	Museum Number	State	Latitude	Longitude
AK1	*rufus*	UAM	RWD24097	7468	AK	55.312	-131.570
AK10	*rufus*	UAM	ABJ651	15394	AK	55.331	-131.619
AK11	*rufus*	UAM	UAMX178	9977	AK	55.331	-131.619
AK12	*rufus*	UAM	ABJ103	13100	AK	55.331	-131.619
AK2	*rufus*	UAM	JJW1573	30194	AK	59.904	-141.326
AK3	*rufus*	UAM	RWD24014	7469	AK	58.3449	-134.555
AK4	*rufus*	UAM	UAMX2175	13891	AK	55.331	-131.619
AK5	*rufus*	UAM	UAMX2174	13896	AK	55.331	-131.619
AK6	*rufus*	UAM	UAMX2173	13895	AK	55.331	-131.619
AK7	*rufus*	UAM	UAMX4263	22246	AK	59.232	-135.4642
AK8	*rufus*	UAM	JJW877	27373	AK	58.344	-134.555
AK9	*rufus*	UAM	JJW1158	20383	AK	55.331	-131.619
AZ1	*rufus*	UWBM	RBB794	87082	AZ	32.222	-110.926
AZ2	*rufus*	UWBM	CEC500	119790	AZ	31.380	-110.228
CA1	*rufus*	UWBM	RBB479	86827	CA	38.535	-121.754
CA10	*rufus*	LSU	SWC7652	B-30336	CA	34.820	-119.097
CA11	*rufus*	LSU	SWC7653	B-30337	CA	34.820	-119.097
CA12	*rufus*	LSU	DLD6758	B-30514	CA	34.172	-116.725
CA13	*rufus*	LSU	EAC7798	B-41970	CA	34.172	-116.725
CA14	*rufus*	LSU	EAC7823	B-41998	CA	34.172	-116.725
CA15	*rufus*	LSU	EAC9383	B-51775	CA	34.223	-116.750
CA16	*rufus*	LSU	EAC9412	B-51804	CA	34.223	-116.750
CA17	*rufus*	LSU	EAC9413	B-51805	CA	34.223	-116.750
CA2	*rufus*	LSU	EAC3105	B-19499	CA	35.008	-117.284
CA3	*rufus*	LSU		B-19500	CA	35.009	-117.284
CA4	*rufus*	LSU	EAC310?	B-19514	CA	35.008	-117.284
CA5	*rufus*	LSU	EAC3108	B-19515	CA	35.008	-117.284
CA6	*rufus*	LSU	EAC7470	B-24253	CA	34.602	-117.579
CA7	*rufus*	LSU	DLD4794	B-25054	CA	34.971	-115.571
CA8	*rufus*	LSU	SWC7647	B-30331	CA	34.819	-119.097
CA9	*rufus*	LSU	SWC7651	B-30335	CA	34.819	-119.097
MT1	*rufus*	UWBM	CJB081	pending	MT	47.862	-113.841
MT2	*rufus*	UWBM	CJB082	pending	MT	47.862	-113.841
MT3	*rufus*	UWBM	CJB083	pending	MT	47.862	-113.841
MT4	*rufus*	UWBM	CJB085	pending	MT	47.862	-113.841
MT5	*rufus*	UWBM	CJB087	pending	MT	47.862	-113.841

Table S2: ddRAD specimen information - part 1.

SampleID	*Selasphorus* sp.	Museum	Prep Number	Museum Number	State	Latitude	Longitude
NM1	*rufus*	MSB	116031	23795	NM	35.768	-106.692
NM10	*rufus*	MSB	221743	40230	NM	31.91	-109.14
NM2	*rufus*	MSB	142268	25453	NM	35.875	-106.328
NM3	*rufus*	MSB	169609	26896	NM	35.151	-108.211
NM4	*rufus*	MSB	170948	29562	NM	35.889	-106.287
NM5	*rufus*	MSB	170952	29566	NM	35.89	-105.98
NM6	*rufus*	MSB	173138	29844	NM	33.993	-107.144
NM7	*rufus*	MSB	174757	30838	NM	35.77	-106.69
NM8	*rufus*	MSB	174808	30908	NM	35.090	-106.592
NM9	*rufus*	MSB	218283	39413	NM	35.13	-106.68
OR1	*rufus*	UWBM	TNL183	91462	OR	44.0537	-121.313
OR10	*rufus*	UWBM	CSW6067	64429	OR	43.345	-122.091
OR11	*rufus*	UWBM	CEC685	120043	OR	46.024	-123.911
OR12	*rufus*	UWBM	BTS06279	100539	OR	42.303	-123.781
OR13	*rufus*	UWBM	BTS06275	100534	OR	43.245	-124.12
OR14	*rufus*	UWBM	BTS06280	100540	OR	42.304	-123.781
OR15	*rufus*	UWBM	BTS06281	100541	OR	43.245	-124.12
OR16	*rufus*	UWBM	BTS06282	100542	OR	42.303	-123.781
OR17	*rufus*	UWBM	BTS06286	100546	OR	42.303	-123.781
OR18	*rufus*	UWBM	CJB088	pending	OR	43.245	-124.045
OR19	*rufus*	UWBM	CJB089	pending	OR	42.303	-123 46.9
OR2	*rufus*	UWBM	SMB101	64552	OR	43.345	-122.09
OR20	*rufus*	UWBM	CJB090	pending	OR	42.303	-123 46.9
OR3	*rufus*	UWBM	JK06 824	112781	OR	43.78	-124.015
OR4	*rufus*	UWBM	GKD595	79617	OR	45.659	-122.863
OR5	*rufus*	UWBM	GHL034	86050	OR	45.758	-122.88
OR6	*rufus*	UWBM	BTS06277	100537	OR	42.304	-123.781
OR7	*rufus*	UWBM	BTS06278	100538	OR	43.245	-124.12
OR8	*rufus*	UWBM	BTS06276	100535	OR	43.245	-124.12
OR9	*rufus*	UWBM	CEC686	119823	OR	45.2023	-123.9629
S.cal1	*calliope*	UWBM	CSW7612	90538	WA		
S.cal2	*calliope*	UWBM	CEC602	119809	NM	35.082	-106.817
S.cal3	*calliope*	UWBM	CJB091	pending	WA	45.125	-116.479
S.cal4	*calliope*	MVZ		183560	CA	37.53286	-118.157
S.cal5	*calliope*	MVZ		182181	CA	40.565523	-120.756
S.cal6	*calliope*	MVZ		182180	CA	40.6661	-120.837
S.cal7	*calliope*	MVZ		182361	CA	40.34428	-121.433
S.cal8	*calliope*	MVZ		183708	CA	39.433192	-120.261

Table S3: ddRAD specimen information - part 2.

SampleID	*Selasphorus* sp.	Museum	Prep Number	Museum Number	State	Latitude	Longitude
S.sasin1	*sasin sasin*	UWBM	GSB 054	80061	CA	37.919	-122.696
S.sasin10	*sasin sasin*	MVZ		180045	CA	37.866	-122.152
S.sasin11	*sasin sedentarius*	MVZ		183552	CA	33.996	-119.725
S.sasin2	*sasin sedentarius*	MVZ		183554	CA	33.996	-119.725
S.sasin3	*sasin sasin*	MVZ		180487	CA	37.972	-122.012
S.sasin4	*sasin sasin*	MVZ		182025	CA	37.681	-121.756
S.sasin5	*sasin sedentarius*	MVZ		183549	CA	33.996	-119.725
S.sasin6	*sasin sedentarius*	MVZ		183551	CA	33.996	-119.725
S.sasin7	*sasin sasin*	MVZ		183714	CA	37.889	-122.319
S.sasin8	*sasin sedentarius*	MVZ		183550	CA	33.996	-119.725
S.sasin9	*sasin sasin*	MVZ		183713	CA	37.890	-122.318
TX1	*rufus*	LSU		B-91622	TX	31.550	-100.327
TX2	*rufus*	LSU		B-91623	TX	31.550	-100.327
UT1	*rufus*	UWBM	JK09 597	114234	UT	36.678	-113.0612
UT2	*rufus*	UWBM	JK04 576	111649	UT	37.678	-113.0612
UT3	*rufus*	UWBM	JK00 304	99309	UT	37.842	-112.828
WA12	*rufus*	UWBM		62641	WA	48.8	-121.92
WA21	*rufus*	UWBM		57305	WA	48.907	-121.659
WA26	*rufus*	UWBM		57307	WA	48.907	-121.659
WA27	*rufus*	UWBM		84224	WA	47.642	-122.542
WA28	*rufus*	UWBM		99531	WA	48.71	-122.442
WA29	*rufus*	UWBM		99530	WA	47.69	-122.565
WA35	*rufus*	UWBM		54064	WA	48.37	-117.19
WA36	*rufus*	UWBM		54058	WA	48.34	-117.14
WA40	*rufus*	UWBM		118223	WA	48.42	-120.503
WA43	*rufus*	UWBM		72972	WA	46.686	-121.523
WA44	*rufus*	UWBM		72973	WA	46.686	-121.523
WA45	*rufus*	UWBM		119827	WA	48.425	-122.288
WA46	*rufus*	UWBM		119658	WA	47.447	-122.459
WA49	*rufus*	UWBM		120045	WA	48.741	-122.474
WA62	*rufus*	UWBM	CJB092	pending	WA	48.34	-117.14

Table S4: ddRAD specimen information - part 3.

SampleID	*Selasphorus* sp.	Museum	Museum Number	Latitude	Longitude	Sex
WA35	*rufus*	UWBM	64429	43.35	-122.09	male
WA36	*rufus*	UWBM	64552	43.35	-122.09	male
OR2	*rufus*	UWBM	54058	48.34	-117.14	male
OR10	*rufus*	UWBM	54064	48.37	-117.19	male
WA22	*rufus*	UWBM	86002	47.69	-122.32	male
WA31	*rufus*	UWBM	86027	47.80	-122.66	male
AK1	*rufus*	UAM	7468	55.31	-131.57	female
AK3	*rufus*	UAM	7469	58.34	-134.56	male
S.sasin4	*sasin sasin*	MVZ	182025	37.68	-121.76	female
S.sasin7	*sasin sasin*	MVZ	183714	37.89	-122.32	male
S.sasin9	*sasin sasin*	MVZ	183713	37.89	-122.32	male
S.sasin10	*sasin sasin*	MVZ	180045	37.87	-122.15	female
S.sasin2	*sasin sedentarius*	MVZ	183554	34.00	-119.73	male
S.sasin5	*sasin sedentarius*	MVZ	183549	34.00	-119.73	male
S.sasin6	*sasin sedentarius*	MVZ	183551	34.00	-119.73	male
S.sasin11	*sasin sedentarius*	MVZ	183552	34.00	-119.73	male
S.cal1	*calliope*	UWBM	90538	47.01	-120.66	female
S.cal3	*calliope*	UWBM	pending	45.12	-116.48	female
S.cal5	*calliope*	MVZ	182181	40.57	-120.76	male
S.cal7	*calliope*	MVZ	182361	40.34	-121.43	female
UT1	*rufus*	UWBM	114234	36.68	-113.06	male
UT2	*rufus*	UWBM	111649	37.68	-113.06	male
UT3	*rufus*	UWBM	99309	37.84	-112.83	male
AZ1	*rufus*	UWBM	87082	32.22	-110.93	female
AZ2	*rufus*	UWBM	119790	31.38	-110.23	male
NM1	*rufus*	MSB	23795	35.77	-106.69	male
NM2	*rufus*	MSB	25453	35.88	-106.33	female
NM3	*rufus*	MSB	26896	35.15	-108.21	male
NM4	*rufus*	MSB	29562	35.89	-106.29	male
NM5	*rufus*	MSB	29566	35.89	-105.98	male
NM6	*rufus*	MSB	29844	33.99	-107.14	male
NM7	*rufus*	MSB	30838	35.77	-106.69	male
NM8	*rufus*	MSB	30908	35.09	-106.59	male
CA10	*rufus*	LSU	B-30336	34.82	-119.10	female
CA19	*rufus*	MVZ	182183	40.57	-120.76	male
CA21	*rufus*	MVZ	183710	39.43	-120.26	male
CA22	*rufus*	MVZ	183712	39.43	-120.26	male
CA5	*rufus*	LSU	B-19515	35.01	-117.28	male
CA6	*rufus*	LSU	B-24253	34.60	-117.58	male
CA33	*rufus*	MVZ	182363	40.37	-121.54	female
CA15	*rufus*	LSU	B-51775	34.22	-116.75	female
S.cal2	*calliope*	UWBM	119809	35.08	-106.82	male
S.cal4	*calliope*	MVZ	183560	37.53	-118.16	female
S.cal6	*calliope*	MVZ	182180	40.67	-120.84	male

Table S5: Whole-genome sequencing specimen data.

References

C J Battey, Ethan B Linck, Kevin L Epperly, Cooper French, David L Slager, Paul W Sykes, Jr, and John Klicka. A migratory divide in the painted bunting (passerina ciris). *Am. Nat.*, 191(2):259–268, February 2018.

David A Baum and Kerry L Shaw. Genealogical perspectives on the species problem. *Experimental and molecular approaches to plant biosystematics*, 53(289-303):123–124, 1995.

Richard A Becker, Allan R Wilks, Ray Brownrigg, and Thomas P Minka. maps: Draw geographical maps. *R package version*, 2, 2013.

Andrea J Betancourt, Yuseob Kim, and H Allen Orr. A pseudohitchhiking model of x vs. autosomal diversity. *Genetics*, 168(4):2261–2269, 2004.

Leonardo Campagna, Márcio Repenning, Luís Fábio Silveira, Carla Suertegaray Fontana, Pablo L Tubaro, and Irby J Lovette. Repeated divergent selection on pigmentation genes in a rapid finch radiation. *Science advances*, 3(5):e1602404, 2017.

Ariane N Carmichael, Anna-Karin Fridolfsson, Joy Halverson, and Hans Ellegren. Male-biased mutation rates revealed from z and w chromosome-linked atp synthase α-subunit (atp5a1) sequences in birds. *Journal of Molecular Evolution*, 50(5):443–447, 2000.

Julian Catchen, Paul A Hohenlohe, Susan Bassham, Angel Amores, and William A Cresko. Stacks: an analysis tool set for population genomics. *Mol. Ecol.*, 22(11): 3124–3140, June 2013.

B Charlesworth. The effect of life-history and mode of inheritance on neutral genetic variability. *Genet. Res.*, 77(2):153–166, April 2001.

Brian Charlesworth. The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genetics Research*, 63(3):213–227, 1994.

Christopher James Clark. Harmonic hopping, and both punctuated and gradual evolution of acoustic characters in selasphorus hummingbird tail-feathers. *PLoS One*, 9(4):e93829, April 2014.

Christopher James Clark. Wing, tail, and vocal contributions to the complex acoustic signals of courting Calliope hummingbirds. *Current Zoology*, 57(2): 187–196, 11 2015. ISSN 2396-9814. doi: 10.1093/czoolo/57.2.187. URL `https://doi.org/10.1093/czoolo/57.2.187`.

Brandon S Cooper, Alisa Sedghifar, W Thurston Nash, Aaron A Comeault, and Daniel R Matute. A maladaptive combination of traits contributes to the maintenance of a drosophila hybrid zone. *Curr. Biol.*, 0(0), August 2018.

Tami E Cruickshank and Matthew W Hahn. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.*, 23(13):3133–3157, July 2014.

Petr Danecek, Adam Auton, Goncalo Abecasis, Cornelis A Albers, Eric Banks, Mark A DePristo, Robert E Handsaker, Gerton Lunter, Gabor T Marth, Stephen T Sherry, Gilean McVean, Richard Durbin, and 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics*, 27(15): 2156–2158, August 2011.

Kevin De Queiroz. Species concepts and species delimitation. *Syst. Biol.*, 56(6): 879–886, December 2007.

Kira E Delmore and Darren E Irwin. Hybrid songbirds employ intermediate routes in a migratory divide. *Ecol. Lett.*, 17(10):1211–1218, October 2014.

Kira E Delmore, Juan S Lugo Ramos, Benjamin M Van Doren, Max Lundberg, Staffan Bensch, Darren E Irwin, and Miriam Liedvogel. Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds. *Evolution Letters*, 2(2):76–87, April 2018.

Michael J Donoghue. A critique of the biological species concept and recommendations for a phylogenetic alternative. *Bryologist*, pages 172–181, 1985.

Dent A Earl and Bridgett M vonHoldt. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the evanno method. *Conserv. Genet. Resour.*, 4(2):359–361, June 2012.

Deren A R Eaton. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, 30(13):1844–1849, July 2014.

Hans Ellegren, Linnea Smeds, Reto Burri, Pall I Olason, Niclas Backström, Takeshi Kawakami, Axel Künstner, Hannu Mäkinen, Krystyna Nadachowska-Brzyska, Anna Qvarnström, et al. The genomic landscape of species divergence in ficedula flycatchers. *Nature*, 491(7426):756, 2012.

G Evanno, S Regnaut, and J Goudet. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.*, 14(8):2611–2620, July 2005.

Michael C Fontaine, James B Pease, Aaron Steele, Robert M Waterhouse, Daniel E Neafsey, Igor V Sharakhov, Xiaofang Jiang, Andrew B Hall, Flaminia Catteruccia,

Evdoxia Kakani, Sara N Mitchell, Yi-Chieh Wu, Hilary A Smith, R Rebecca Love, Mara K Lawniczak, Michel A Slotman, Scott J Emrich, Matthew W Hahn, and Nora J Besansky. Mosquito genomics. extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, 347(6217):1258524, January 2015.

Richard E Green, Johannes Krause, Adrian W Briggs, Tomislav Maricic, Udo Stenzel, Martin Kircher, Nick Patterson, Heng Li, Weiwei Zhai, Markus Hsi-Yang Fritz, Nancy F Hansen, Eric Y Durand, Anna-Sapfo Malaspinas, Jeffrey D Jensen, Tomas Marques-Bonet, Can Alkan, Kay Prüfer, Matthias Meyer, Hernán A Burbano, Jeffrey M Good, Rigo Schultz, Ayinuer Aximu-Petri, Anne Butthof, Barbara Höber, Barbara Höffner, Madlen Siegemund, Antje Weihmann, Chad Nusbaum, Eric S Lander, Carsten Russ, Nathaniel Novod, Jason Affourtit, Michael Egholm, Christine Verna, Pavao Rudan, Dejana Brajkovic, Željko Kucan, Ivan Gušic, Vladimir B Doronichev, Liubov V Golovanova, Carles Lalueza-Fox, Marco de la Rasilla, Javier Fortea, Antonio Rosas, Ralf W Schmitz, Philip L F Johnson, Evan E Eichler, Daniel Falush, Ewan Birney, James C Mullikin, Montgomery Slatkin, Rasmus Nielsen, Janet Kelso, Michael Lachmann, David Reich, and Svante Pääbo. A draft sequence of the neandertal genome. *Science*, 328(5979):710–722, May 2010.

S Healy and W A Calder. Rufous hummingbird - introduction | birds of north america online. `https://birdsna.org/Species-Account/bna/species/rufhum`, 2006. Accessed: 2018-8-7.

Heliconius Genome Consortium. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, 487(7405):94–98, July 2012.

T D Herbert, J D Schuffert, D Andreasen, L Heusser, M Lyle, A Mix, A C Ravelo, L D Stott, and J C Herguera. Collapse of the california current during glacial maxima linked to climate change on land. *Science*, 293(5527):71–76, July 2001.

Geoffrey E Hill, Robert R. Sargent, and Martha B. Sargent. Recent change in the winter distribution of rufous hummingbirds. *Auk*, 115(1):240–245, 1998.

S A Hovan, D K Rea, and N G Pisias. Late pleistocene continental climate and oceanic variability recorded in northwest pacific sediments. *Paleoceanography*, 6 (3):349–370, June 1991.

Darren E Irwin. Sex chromosomes and speciation in birds and other ZW systems. *Mol. Ecol.*, February 2018.

Thibaut Jombart. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11):1403–1405, June 2008.

Takeshi Kawakami, Linnéa Smeds, Niclas Backström, Arild Husby, Anna Qvarnström, Carina F Mugal, Pall Olason, and Hans Ellegren. A high-density linkage

map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Molecular ecology*, 23(16):4035–4058, 2014.

Takeshi Kawakami, Carina F Mugal, Alexander Suh, Alexander Nater, Reto Burri, Linnea Smeds, and Hans Ellegren. Whole-genome patterns of linkage disequilibrium across flycatcher populations clarify the causes and consequences of fine-scale recombination rate variation in birds. *Molecular ecology*, 26(16):4158–4172, 2017.

Andrew D Kern and Matthew W Hahn. The neutral theory in light of natural selection. *Mol. Biol. Evol.*, 35(6):1366–1371, June 2018.

M Kimura. Evolutionary rate at the molecular level. *Nature*, 217(5129):624–626, February 1968.

Brian P Kinlan, Michael H Graham, and Jon M Erlandson. Late-Quaternary changes in the size and shape of the california channel islands: implications for marine subsidies to terrestrial communities. In *Proceedings of the California Islands Symposium*, volume 6, pages 119–130. academia.edu, 2005.

Jonas Korlach, Gregory Gedman, Sarah B Kingan, Chen-Shan Chin, Jason T Howard, Jean-Nicolas Audet, Lindsey Cantin, and Erich D Jarvis. De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. *Gigascience*, 6(10):1–16, October 2017.

Thorfinn Sand Korneliussen, Anders Albrechtsen, and Rasmus Nielsen. ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, 15:356, November 2014.

Vikas Kumar, Fritjof Lammers, Tobias Bidon, Markus Pfenninger, Lydia Kolter, Maria A Nilsson, and Axel Janke. The evolutionary history of bears is characterized by gene flow across species. *Sci. Rep.*, 7:46487, April 2017.

Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with bowtie 2. *Nat. Methods*, 9(4):357–359, March 2012.

Daniel J Lawson, Lucy Van Dorp, and Daniel Falush. A tutorial on how not to over-interpret structure and admixture bar plots. *Nature Communications*, 9(1): 3258, 2018.

Yuyini Licona-Vera and Juan Francisco Ornelas. The conquering of north america: dated phylogenetic and biogeographic inference of migratory behavior in bee hummingbirds. *BMC Evol. Biol.*, 17(1):126, June 2017.

E B Linck and C J Battey. Minor allele frequency thresholds strongly affect population structure inference with genomic datasets. *bioRxiv*, 2017.

Stinus Lindgreen. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res. Notes*, 5:337, July 2012.

Simon H Martin and Steven M Van Belleghem. Exploring evolutionary relationships across the genome using topology weighting. *Genetics*, 206(1):429–438, 2017.

E Mayr. The bearing of the new systematics on genetical problems; the nature of species. *Adv. Genet.*, 3b(2):205–237, 1948.

Olivier Mazet, Willy Rodríguez, Simona Grusea, Simon Boitard, and Lounès Chikhi. On the importance of being structured: instantaneous coalescence rates and human evolution—lessons for ancestral population size inference? *Heredity*, 116(4): 362, 2016.

D McGee, E Moreno-Chamarro, J Marshall, and ED Galbraith. Western us lake expansions during heinrich stadials linked to pacific hadley circulation. *Science advances*, 4(11):eaav0118, 2018.

Jimmy A McGuire, Christopher C Witt, J V Remsen, Jr, Ammon Corl, Daniel L Rabosky, Douglas L Altshuler, and Robert Dudley. Molecular phylogenetics and the diversification of hummingbirds. *Curr. Biol.*, 24(8):910–916, April 2014.

Aaron McKenna, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, David Altshuler, Stacey Gabriel, Mark Daly, and Mark A DePristo. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, 20(9): 1297–1303, September 2010.

Richard P Meisel and Tim Connallon. The faster-x effect: integrating theory and data. *Trends in Genetics*, 29(9):537–544, 2013.

Gemma G R Murray, André E R Soares, Ben J Novak, Nathan K Schaefer, James A Cahill, Allan J Baker, John R Demboski, Andrew Doll, Rute R Da Fonseca, Tara L Fulton, M Thomas P Gilbert, Peter D Heintzman, Brandon Letts, George McIntosh, Brendan L O'Connell, Mark Peck, Marie-Lorraine Pipes, Edward S Rice, Kathryn M Santos, A Gregory Sohrweide, Samuel H Vohr, Russell B Corbett-Detig, Richard E Green, and Beth Shapiro. Natural selection shaped the rise and fall of passenger pigeon genomic diversity. *Science*, 358(6365):951–954, November 2017.

Brian M Myers, David T Rankin, Kevin J Burns, and Christopher J Clark. Behavioral and morphological evidence of an allen's× rufous hummingbird (selasphorus sasin× s. rufus) hybrid zone in southern oregon and northern california. *The Auk*, 2019.

Krystyna Nadachowska-Brzyska, Cai Li, Linnea Smeds, Guojie Zhang, and Hans Ellegren. Temporal dynamics of avian populations during pleistocene revealed by Whole-Genome sequences. *Curr. Biol.*, 25(10):1375–1380, May 2015.

Patrik Nosil. Speciation with gene flow could be common. *Mol. Ecol.*, 17(9):2103–2106, May 2008.

Patrik Nosil, Daniel J Funk, and Daniel Ortiz-Barrientos. Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.*, 18(3):375–402, February 2009.

Emmanuel Paradis, Julien Claude, and Korbinian Strimmer. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20(2):289–290, January 2004.

Brant K Peterson, Jesse N Weber, Emily H Kay, Heidi S Fisher, and Hopi E Hoekstra. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, 7(5):e37135, May 2012.

Albert B Phillimore, C David L Orme, Gavin H Thomas, Tim M Blackburn, Peter M Bennett, Kevin J Gaston, and Ian P F Owens. Sympatric speciation in birds is rare: insights from range data and simulations. *Am. Nat.*, 171(5):646–657, May 2008.

Joseph K Pickrell and Jonathan K Pritchard. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.*, 8(11):e1002967, November 2012.

John E Pool and Rasmus Nielsen. Population size changes reshape genomic patterns of diversity. *Evolution: International Journal of Organic Evolution*, 61(12):3001–3006, 2007.

J K Pritchard, M Stephens, and P Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959, June 2000.

Peter Pyle, Steve N G Howell, and Others. *Identification guide to North American birds*. sidalc.net, 1997.

Mark Ravinet, R Faria, RK Butlin, J Galindo, N Bierne, M Rafajlović, MAF Noor, B Mehlig, and AM Westram. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of evolutionary biology*, 30(8): 1450–1477, 2017.

Stein A Saether, Glenn-Peter Saetre, Thomas Borge, Chris Wiley, Nina Svedin, Gunilla Andersson, Thor Veen, Jon Haavie, Maria R Servedio, Stanislav Bures, Miroslav Král, Mårten B Hjernquist, Lars Gustafsson, Johan Träff, and Anna

Qvarnström. Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. *Science*, 318(5847):95–97, October 2007.

Molly Schumer, Chenling Xu, Daniel L Powell, Arun Durvasula, Laurits Skov, Chris Holland, John C Blazier, Sriram Sankararaman, Peter Andolfatto, Gil G Rosenthal, and Molly Przeworski. Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, 360(6389):656–660, May 2018.

Sonal Singhal, Ellen M Leffler, Keerthi Sannareddy, Isaac Turner, Oliver Venn, Daniel M Hooper, Alva I Strand, Qiye Li, Brian Raney, Christopher N Balakrishnan, Simon C Griffith, Gil McVean, and Molly Przeworski. Stable recombination hotspots in birds. *Science*, 350(6263):928–932, November 2015.

Viviane Slon, Fabrizio Mafessoni, Benjamin Vernot, Cesare de Filippo, Steffi Grote, Bence Viola, Mateja Hajdinjak, Stéphane Peyrégne, Sarah Nagel, Samantha Brown, et al. The genome of the offspring of a neanderthal mother and a denisovan father. *Nature*, 561(7721):113, 2018.

Linnéa Smeds, Anna Qvarnström, and Hans Ellegren. Direct estimate of the rate of germline mutation in a bird. *Genome Res.*, 26(9):1211–1218, September 2016.

Alexandros Stamatakis. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9):1312–1313, May 2014.

R Core Team and Others. R: A language and environment for statistical computing. 2013.

Jonathan Terhorst, John A Kamm, and Yun S Song. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.*, 49 (2):303–309, February 2017.

David P L Toews, Scott A Taylor, Rachel Vallender, Alan Brelsford, Bronwyn G Butcher, Philipp W Messer, and Irby J Lovette. Plumage genes and little else distinguish the genomes of hybridizing warblers. *Curr. Biol.*, 26(17):2313–2318, September 2016.

Steven M Van Belleghem, Margarita Baquero, Riccardo Papa, Camilo Salazar, W Owen McMillan, Brian A Counterman, Chris D Jiggins, and Simon H Martin. Patterns of z chromosome divergence among heliconius species highlight the importance of historical demography. *Molecular ecology*, 27(19):3852–3872, 2018.

Beatriz Vicoso and Brian Charlesworth. Evolution on the x chromosome: unusual patterns and processes. *Nature Reviews Genetics*, 7(8):645, 2006.

Robin S Waples and Oscar Gaggiotti. INVITED REVIEW: What is a population? an empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Mol. Ecol.*, 15(6):1419–1439, 2006.

Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer, June 2016.

Claus O Wilke. cowplot: streamlined plot theme and plot annotations for 'ggplot2'. *R package version 0. 6*, 2, 2016.