

RESEARCH ARTICLE

Open Access



# Genomic differentiation across the speciation continuum in three hummingbird species pairs

Elisa C. Henderson\*  and Alan Brelsford

## Abstract

**Background:** The study of speciation has expanded with the increasing availability and affordability of high-resolution genomic data. How the genome evolves throughout the process of divergence and which regions of the genome are responsible for causing and maintaining that divergence have been central questions in recent work. Here, we use three pairs of species from the recently diverged bee hummingbird clade to investigate differences in the genome at different stages of speciation, using divergence times as a proxy for the speciation continuum.

**Results:** Population measures of relative differentiation between hybridizing species reveal that different chromosome types diverge at different stages of speciation. Using  $F_{ST}$  as our relative measure of differentiation we found that the sex chromosome shows signs of divergence early in speciation. Next, small autosomes (microchromosomes) accumulate highly diverged genomic regions, while the large autosomes (macrochromosomes) accumulate genomic regions of divergence at a later stage of speciation.

**Conclusions:** Our finding that genomic windows of elevated  $F_{ST}$  accumulate on small autosomes earlier in speciation than on larger autosomes is counter to the prediction that  $F_{ST}$  increases with size of chromosome (i.e. with decreased recombination rate), and is not represented when weighted average  $F_{ST}$  per chromosome is compared with chromosome size. The results of this study suggest that multiple chromosome characteristics such as recombination rate and gene density combine to influence the genomic locations of signatures of divergence.

**Keywords:** Speciation genomics, Speciation continuum, Divergence, Polymorphism, Hummingbird,  $F_{ST}$ ,  $d_{xy}$

## Background

A fundamental goal in evolutionary biology is to understand how the process of speciation occurs. The increasing availability of population genomic data has led to a new understanding of speciation beyond the classical categorization of sympatric, allopatric, and parapatric modes [1]. The recently developed field of “speciation genomics” has revealed that speciation with gene flow, a phenomenon that was once thought to be highly

unlikely [2], is common [3–7], including between extant and extinct taxa (reviewed in [8]). These revelations suggest that the individual is not the unit of isolation, and that there must be regions of isolation within the genome maintaining species boundaries. These sites are called barrier loci: specific regions of the genome that contribute to barriers to gene flow between populations [9]. This research has revealed that there are many genomic regions important to speciation, not just one or a few regions of large effect. For example, Martin et al. [10] found evidence of reduced introgression between two species of *Heliconius* butterflies at many regions of the genome, suggesting that many barrier loci of

\* Correspondence: [ehend004@ucr.edu](mailto:ehend004@ucr.edu)

Department of Evolution, Ecology, and Organismal Biology, University of California Riverside, 2710 Life Science Bldg, Riverside, CA 92521, USA



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

relatively small effect are responsible for maintaining species boundaries. Ellison et al. [11] found evidence that many genes of small effect are responsible for the divergence of sexual behaviors between species of *Lau-pala* crickets. Overall, investigating the genomic landscapes of differentiation between hybridizing species using modern genomics techniques will enhance our understanding of speciation [1, 12].

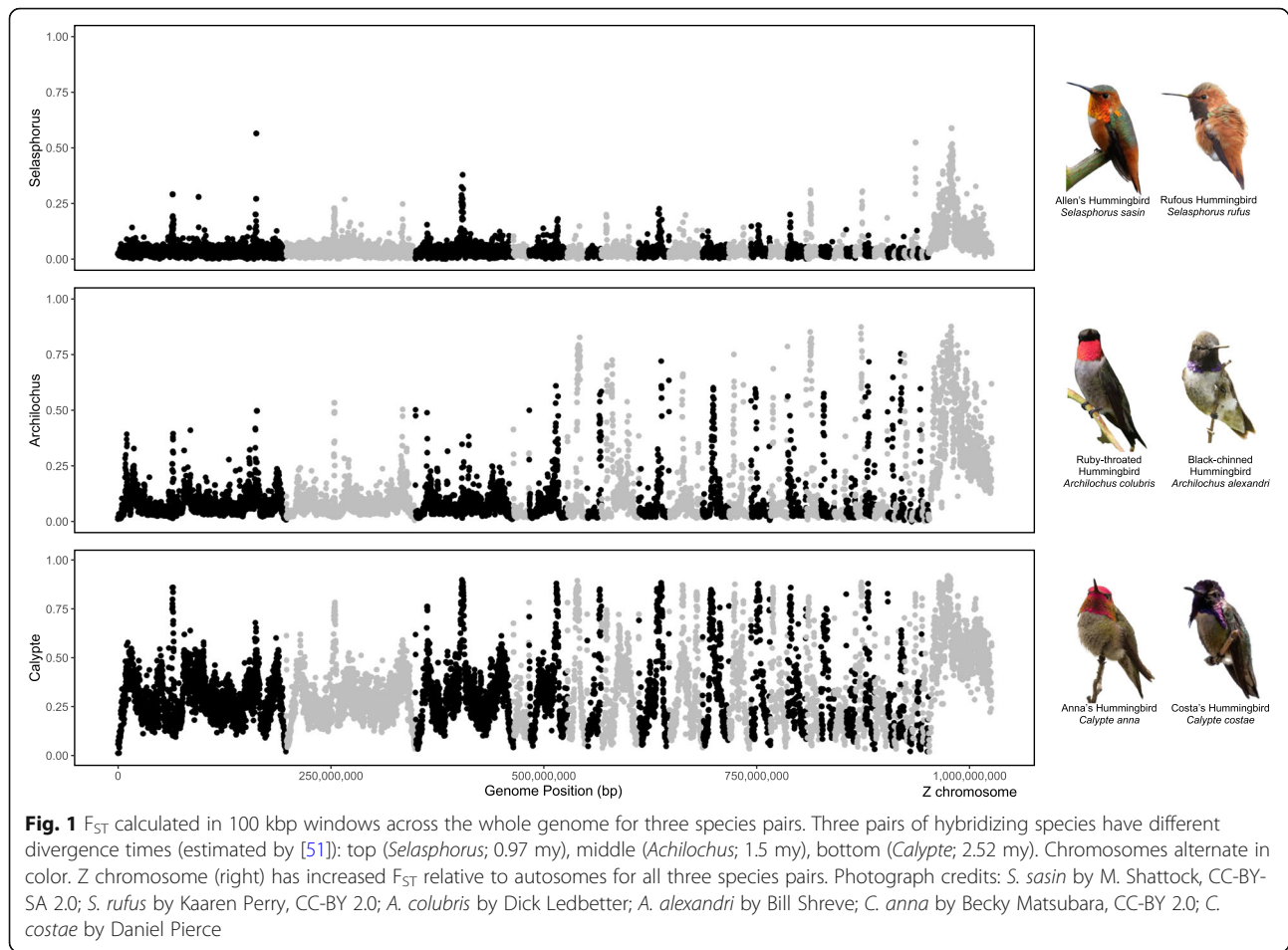
The study of genomic divergence was initially focused on the population statistic,  $F_{ST}$ , which is the standard statistical measure for genetic divergence between two populations and is based on between-population variance in allele frequencies compared to within-population variance [13]. It was observed early on that  $F_{ST}$  is variable among loci [14]. Studies have since shown that the variation in  $F_{ST}$  creates a heterogeneous landscape across the genome. This pattern has been observed in several diverse taxon pairs, including *Heliconius* butterflies [15], mussels [16], warblers [17, 18], and *Ficedula* flycatchers [19]. Originally, it was thought that peaks of differentiation were indicative of reproductively isolating genes [20], and were termed “genomic islands of speciation” [21]. Further scrutiny revealed that equating  $F_{ST}$  peaks to “islands of speciation” (i.e. barrier loci) is premature, and additional investigation is required to determine the true cause of  $F_{ST}$  heterogeneity [9, 22–24].

Though genome-wide average  $F_{ST}$  increases as divergence time increases [25, 26] it is clear that  $F_{ST}$  peaks relative to the rest of the genome are not necessarily indicative of barrier loci or reduced gene flow, and a variety of processes may lead to these “outlier regions”. Recombination rate variation across the genome predicts much of the variation in nucleotide diversity [27] and  $F_{ST}$  [28–30]. Reduced recombination rate, genetic drift, local adaptation, and other evolutionary processes may reduce local nucleotide diversity within at least one species in the pair causing peaks in  $F_{ST}$  that are not indicative of locally reduced gene flow [22, 23]. In contrast to  $F_{ST}$ , an absolute measure of divergence between populations ( $d_{xy}$ ) is not inflated by reduced within-population nucleotide diversity.  $F_{ST}$  peaks resulting from locally reduced gene flow are predicted to have elevated  $d_{xy}$ , while  $F_{ST}$  peaks resulting from low within-population diversity are not. Thus, comparing multiple statistics together may help elucidate the evolutionary mechanisms leading to the genomic patterns we find. For example, measuring nucleotide diversity (e.g.  $\pi$ ) across the genome can indicate specific regions of low diversity in one or both species that result in an  $F_{ST}$  peak that is not due to reduced gene flow, but rather due to species-specific selection at that locus. Calculating  $d_{xy}$  across the genome and locating  $F_{ST}$  peaks that are associated with elevated  $d_{xy}$  can help narrow down the potential causes of some islands of divergence (e.g. [18, 31–34]).

Though much progress has been made in characterizing the heterogeneity in divergence using these statistics, our understanding of how these patterns change over time is still limited. Because speciation is often a process with a duration of at least 1 million years [35], it is nearly impossible to investigate the different stages of speciation using only a single species pair. To help alleviate this problem, some studies have used independent pairs of closely related species that have different divergence times as a proxy for the different stages of speciation (e.g. [15, 18, 26, 28, 31, 36–40]). This can provide valuable insight into the genomic process of speciation over time.

Variation in chromosome type may be important to consider when investigating the speciation process. For example, sex chromosomes play a disproportionate role in reproductive isolation relative to the autosomes [41]. Greater differentiation on sex chromosomes relative to autosomes has been broadly identified in both male-heterogametic (XY/XX) and female-heterogametic (ZW/ZZ) taxa [42]. Proposed reasons for this include the large X-effect, reduced effective population size, and reduced recombination rate of X or Z chromosomes [42]. Chromosome size variation across autosomes may also contribute to the process of speciation. All birds and many species of reptiles have a largely conserved karyotype made up of large chromosomes (macrochromosomes) and small chromosomes (microchromosomes) [43]. Often, the karyotype contains up to 8 large chromosomes that are on average an order of magnitude larger than the average microchromosome [44]. Microchromosomes have a higher recombination rate, gene density, and GC content relative to macrochromosomes [43]. Low recombination rates are predicted to lead to reduced genetic diversity due to hitchhiking [45] and background selection [46, 47]. Consistent with this prediction, in birds larger chromosomes tend to have lower diversity [48, 49] and, in at least one case, higher  $F_{ST}$  between lineages [50]. Whether microchromosomes play a different role in speciation than macrochromosomes, however, is still unknown.

In this study, we investigate 1) how genomic signatures of divergence change as speciation proceeds, and 2) the differences between micro-, macro-, and Z chromosomes, how those differences compare across the speciation continuum, and what that tells us about the importance of different chromosome types in speciation. We use three pairs of hybridizing species from the bee hummingbird clade: *Calypte anna* and *C. costae* (Anna’s and Costa’s hummingbirds), *Archilochus alexandri* and *A. colubris* (Black-chinned and Ruby-throated hummingbirds), *Selasphorus sasin* and *S. rufus* (Allen’s and Rufous hummingbirds; Fig. 1). Previous studies have used multiple pairs of species with different divergence times as a proxy for the speciation continuum (e.g. [15, 18, 26, 28, 31, 36–40]).



However, recent studies on the phylogenetic relationships among these hummingbirds have produced conflicting estimates of relative divergence times. According to McGuire et al. [51] the species pairs *S. sasin*/*S. rufus* and *A. alexandri*/*A. colubris* diverged more recently than *C. anna*/*C. costae*, but numerical estimates of divergence dates are not provided. Abrahamczyk and Renner [52] is the only study to our knowledge to provide numerical estimates for the three species pairs included in this study: *C. anna*/*C. costae* estimated at 2.52 million years ago (mya), *A. alexandri*/*A. colubris* 1.5 mya, and *S. sasin*/*S. rufus* 0.97 mya. Licona-Vera and Ornelas [53] used improved within-species sampling (previous studies included only a single representative of four [51] or six [52] of our focal species), and did not recover monophyletic groups for *A. alexandri*, *S. sasin*, or *S. rufus*. The latter study also estimated an older divergence date for a node within *A. alexandri* than the node separating *C. anna* and *C. costae*, albeit with overlapping 95% HPD intervals. Because of the lack of consensus in the literature on divergence times and phylogenetic relationships among the three focal species pairs of this study, we chose to use extent of reproductive isolation (i.e., frequency of hybridization), rather

than divergence time, as a proxy for the speciation continuum.

Our species pairs are especially suitable for this study, as *A. alexandri*/*A. colubris* and *S. sasin*/*S. rufus* are, to our knowledge, the only two hummingbird species pairs in the US with a quantified extent of hybridization. *A. alexandri* and *A. colubris* have a narrow sympatric range in southwestern Oklahoma, and an estimated 9.3% of adult males are F1 hybrids [54]. *S. sasin* and *S. rufus* have a broad hybrid zone in northern California and southern Oregon where hybrids outnumber parental species and there is clinal variation in species diagnostic traits [55]. *C. anna* and *C. costae* inhabit a broad sympatric range in parts of California, Nevada, Arizona, and Baja California. Though the frequency of hybridization has not been quantified, hybrids are occasionally observed [[56], pers. obs.], and this is the only species pair in the United States listed as “extensive natural hybridization reported” by [57]. Species pairs will hereafter be referred to by genus (*Calypte*, *Archilochus*, *Selasphorus*).

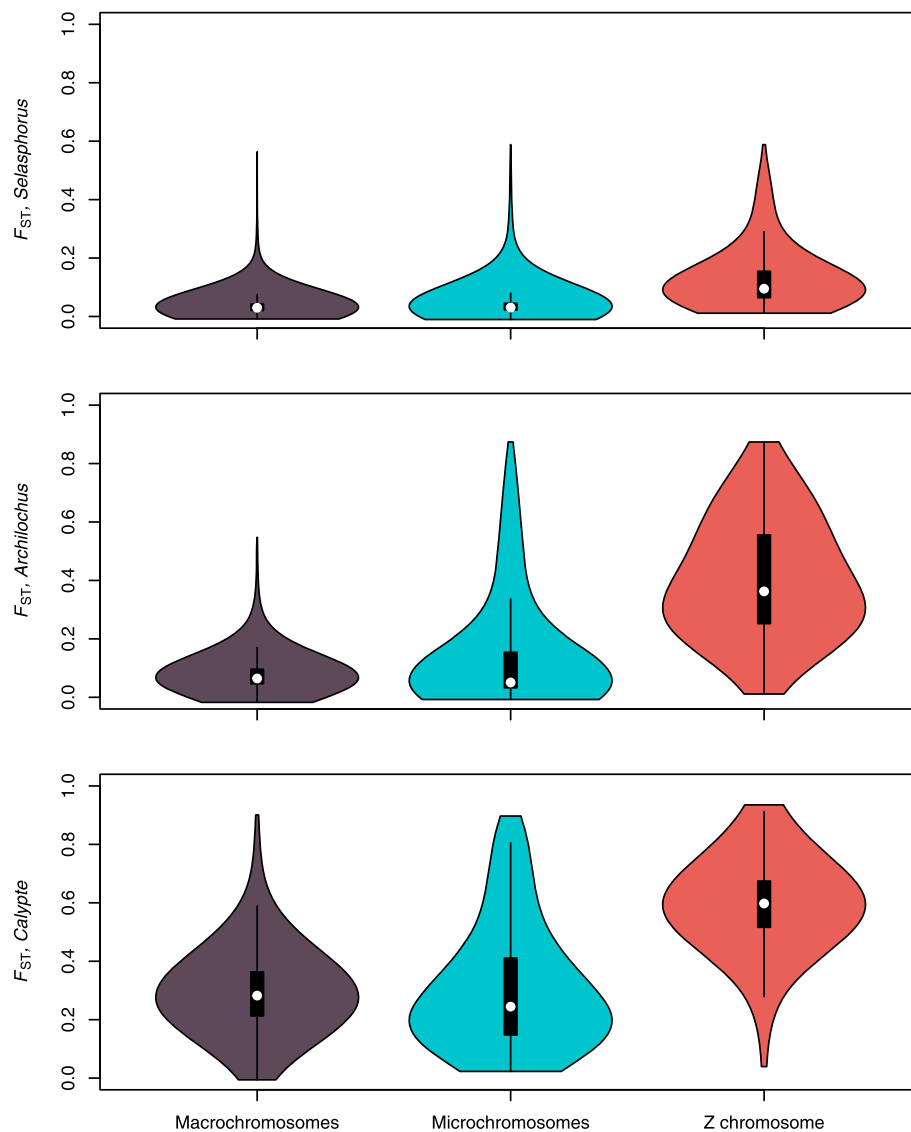
We use these three independent but closely related species pairs as a proxy for the speciation continuum.

First, we compare patterns of genomic differentiation and diversity across these three levels of reproductive isolation. Second, we compare these patterns across three different chromosome types: microchromosomes, macrochromosomes, and the sex chromosome, and observe how these comparisons differ across the three different levels of reproductive isolation. We found that speciation seems to progress at different rates based on chromosome type, with the sex chromosome diverging first, the microchromosomes diverging next, and divergence only appearing on the macrochromosomes in late stages of reproductive isolation.

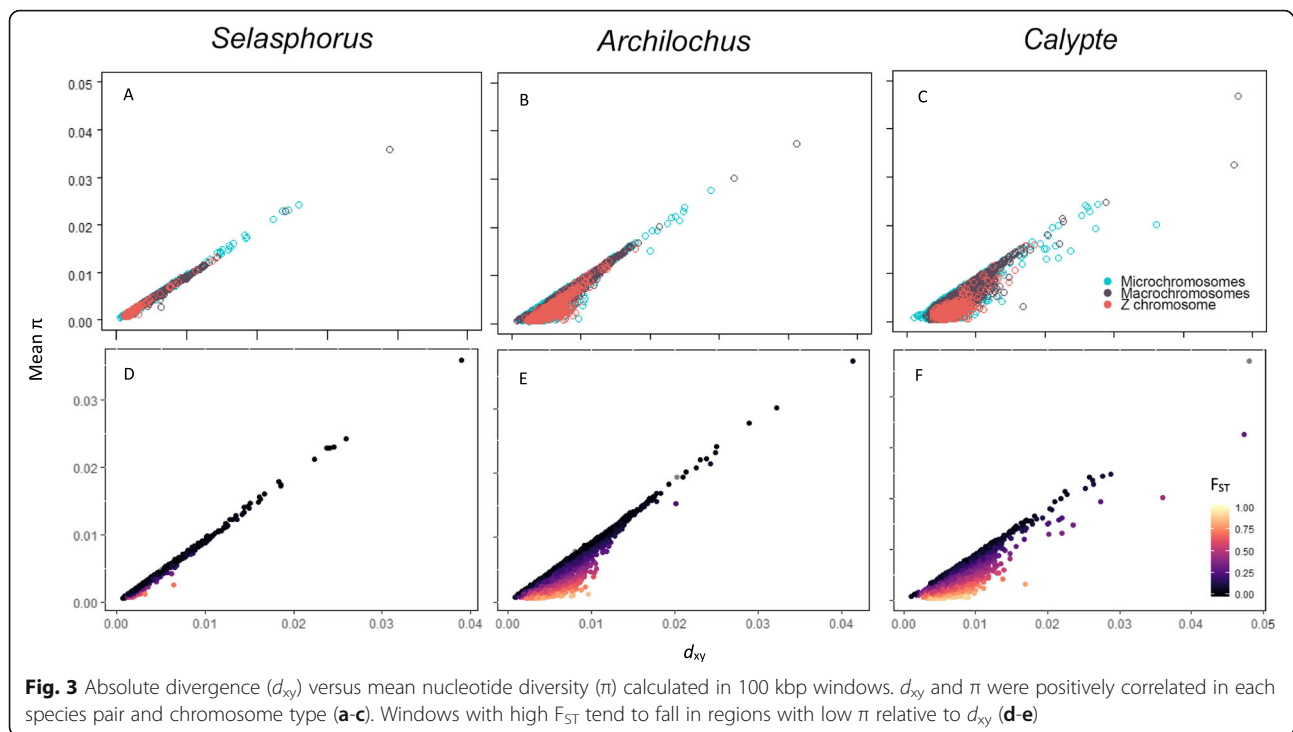
## Results

Our least reproductively-isolated species pair, *Selasphorus* had the lowest average  $F_{ST}$  (0.041) and fewer

overall  $F_{ST}$  peaks than either of the other species pairs (Fig. 1). Our most reproductively-isolated species pair, *Calypste* had the highest average  $F_{ST}$  (*Calypste*,  $F_{ST} = 0.323$ ; *Archilochus*,  $F_{ST} = 0.112$ ), though *Calypste* and *Archilochus* had a qualitatively similar number of  $F_{ST}$  peaks. For all three pairs,  $F_{ST}$  was higher on the Z chromosome than on the autosomes, with a significant overrepresentation of high- $F_{ST}$  windows on the Z chromosome (*Calypste*,  $\chi^2 = 40.975$ ,  $P = 1.27 \times 10^{-9}$ ; *Archilochus*,  $\chi^2 = 292.91$ ,  $P = 2.2 \times 10^{-16}$ ; *Selasphorus*,  $\chi^2 = 674.01$ ,  $P = 2.2 \times 10^{-16}$ ; Supplemental Table 1), consistent with the findings of Battey [58] in the *Selasphorus* species pair and Elgvin et al. in *Passer* sparrows [59]. When comparing  $F_{ST}$  across chromosome types, *Selasphorus*, the least-reproductively isolated species pair only had noticeably elevated  $F_{ST}$  on the Z chromosome, and



**Fig. 2** Density of windows with different  $F_{ST}$  values separated by chromosome type



not on autosomes, and no significant difference in number of elevated  $F_{ST}$  peaks on microchromosomes versus macrochromosomes ( $\chi^2 = 0.09577$ ,  $P = 0.757$ ). The next species pair, *Archilochus* showed elevated  $F_{ST}$  on the Z chromosome and significantly more elevated  $F_{ST}$  windows on the microchromosomes relative to the macrochromosomes ( $\chi^2 = 48.998$ ,  $P = 2.56 \times 10^{-12}$ ). The most reproductively-isolated species pair, *Calypte* had windows with elevated  $F_{ST}$  throughout the entire genome, including the macrochromosomes (Fig. 2), though the microchromosomes had significantly more elevated  $F_{ST}$  windows than the macrochromosomes ( $\chi^2 = 22.759$ ,  $P = 1.27 \times 10^{-9}$ ). Genes found within the top 1% of  $F_{ST}$  peaks are listed in Supplemental Table 2, although we emphasize that not all high- $F_{ST}$  regions should be interpreted as barrier loci, and this table likely includes many genes that have no involvement in reproductive barriers.

For each species pair,  $F_{ST}$  and  $d_{xy}$  were negatively correlated (Supplemental Fig. 1). The pattern held true across all three chromosome types, but the strength of the positive correlation did not vary consistently (Supplemental Fig. 1). Within species pairs, there was always a positive correlation between  $d_{xy}$  and mean  $\pi$  for all three chromosome types (Fig. 3). The correlation was strongest for the least reproductively-isolated species pair (*Selasphorus*; Fig. 3a). The correlation was weaker for the next species pair (*Archilochus*), with increased  $d_{xy}$  relative to  $\pi$  appearing only on the Z chromosome (Fig. 3b). Increased  $d_{xy}$  relative to  $\pi$  appeared on all chromosome types in the most reproductively-isolated

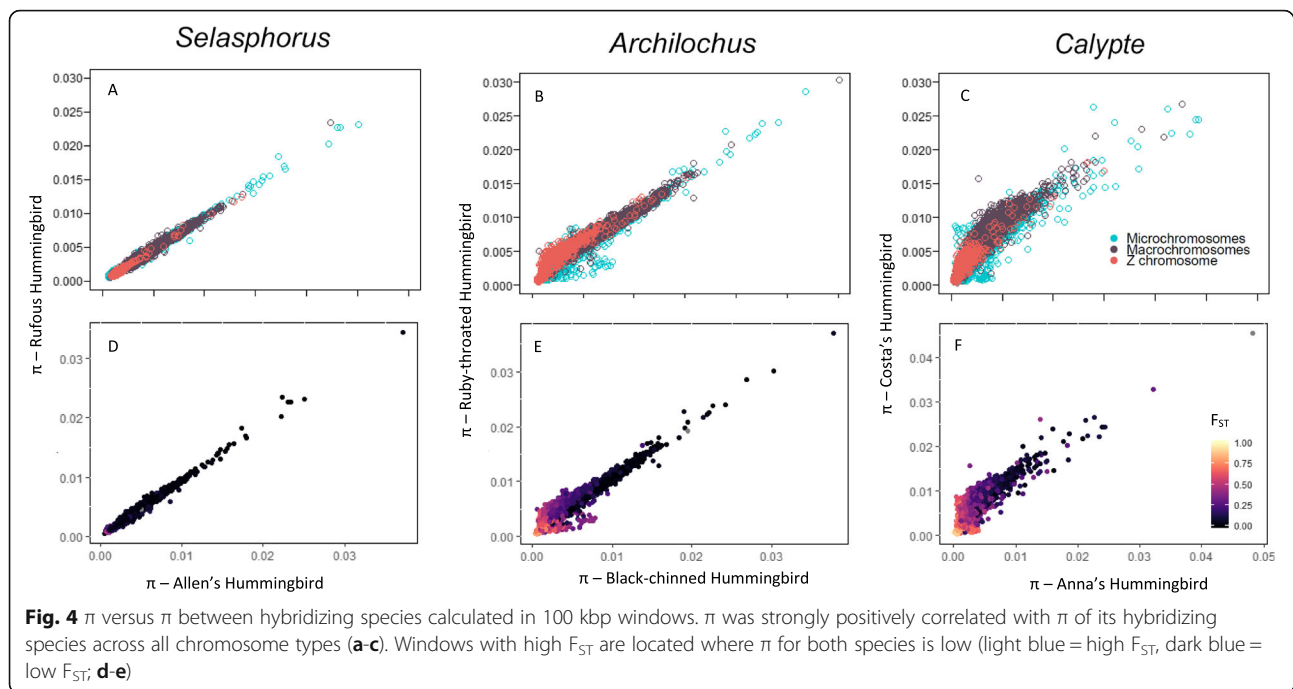
species pair (*Calypte*; Fig. 3c). Windows with elevated  $F_{ST}$  appeared in regions with higher  $d_{xy}$  relative to  $\pi$ , or where  $\pi$  was especially low (Fig. 3d–f).  $\pi$  was strongly positively correlated between hybridizing species (Fig. 4a–c). Elevated  $F_{ST}$  appeared mostly in windows where both species in a pair had low  $\pi$  (Fig. 4d–f).

Across species pairs,  $F_{ST}$  from one pair was always positively correlated with  $F_{ST}$  from either of the other species pairs (Fig. 5a–c; Table 1), indicating that patterns of  $F_{ST}$  peaks and valleys in one species pair can partially predict patterns in another species pair. The pattern was true for all chromosome types, but the strength in correlation did not vary consistently (Table 1; Supplemental Fig. 2).  $F_{ST}$  increased with chromosome size for the most reproductively-isolated species pair (*Calypte*; Fig. 6), with a weaker positive correlation for the less reproductively-isolated species pairs (*Selasphorus* and *Archilochus*; Fig. 6).

## Discussion

In this study we compared the genomic differentiation and diversity of three pairs of closely related hummingbirds at different stages of divergence (*Selasphorus*, extensive hybridization [55]; *Archilochus*, moderate hybridization [54], *Calypte*, rare hybridization [56, 57]);, allowing us to investigate changes in the genome as divergence progresses. Our estimates of  $F_{ST}$  are consistent with one previously published estimate of divergence times [52] for these three species pairs (Fig. 1), suggesting that both average levels of  $F_{ST}$  and number of  $F_{ST}$



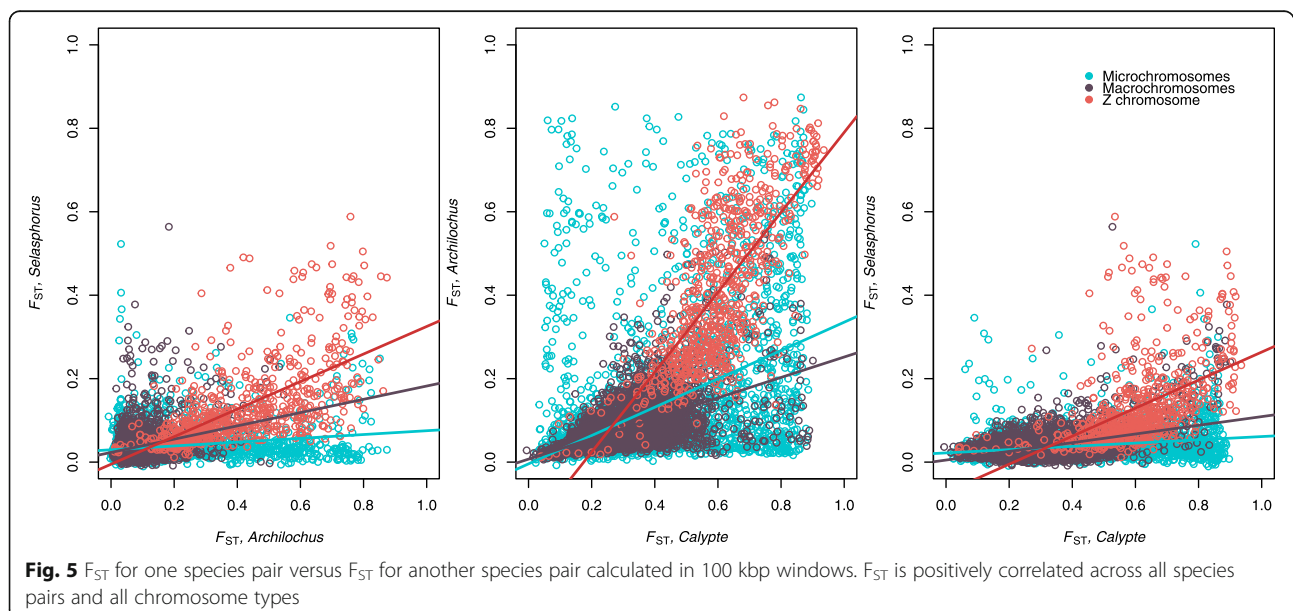


peaks increase as extent of reproductive isolation and divergence time increase.

Chromosome-wide average  $F_{ST}$  increased with chromosome size for our most reproductively-isolated species pair (*Calypte*; Fig. 6) as is theoretically expected given the correlation between chromosome size and recombination rate in birds [22, 60] and empirically tested in at least one other species [50]. Finding significance of this pattern in *Calypte* and not in the other species pairs suggests that it may be a pattern that appears late in

speciation. However, we cannot exclude that differences in natural history among the species pairs contributes to our results. For example, the *Calypte* species are the only hummingbirds in this study that are not long-distance migrants (sedentary populations of *S. sasin* exist, but were not sequenced). How natural history characteristics such as migration affect genomic patterns of speciation is unclear.

When looking at  $F_{ST}$  in windows across the genome, we did not find the expected pattern of increased  $F_{ST}$

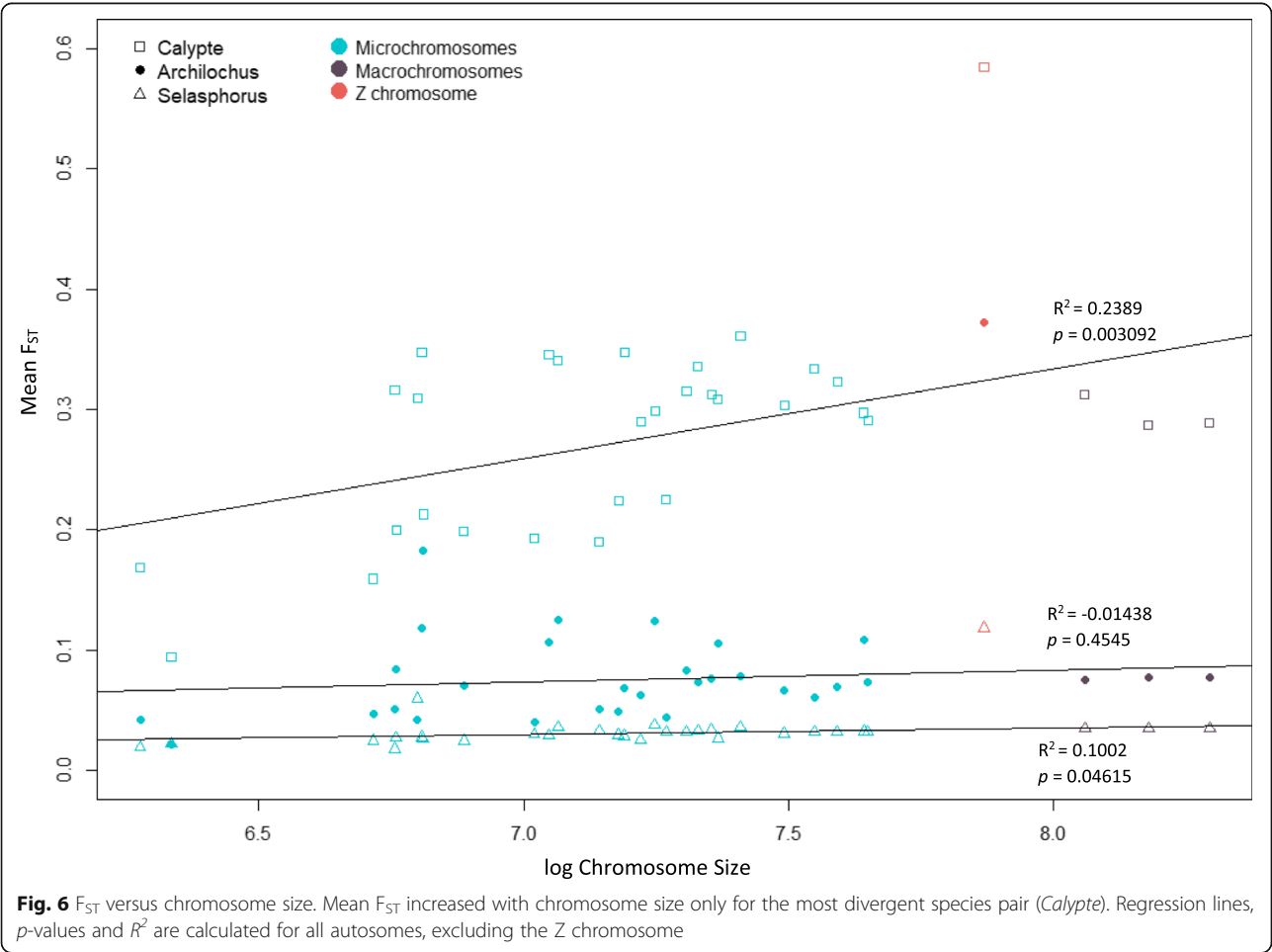


**Table 1**  $F_{ST}$  across species pairs is positively correlated for all chromosome types

| Genus 1 (x)        | Genus 2 (y)        | Chrom type | R-squared |
|--------------------|--------------------|------------|-----------|
| <i>Archilochus</i> | <i>Selasphorus</i> | Macro      | 0.0863347 |
|                    |                    | Micro      | 0.0522726 |
|                    |                    | Z          | 0.4069244 |
| <i>Calypte</i>     | <i>Archilochus</i> | Macro      | 0.3053203 |
|                    |                    | Micro      | 0.2404795 |
|                    |                    | Z          | 0.5296113 |
| <i>Calypte</i>     | <i>Selasphorus</i> | Macro      | 0.1883183 |
|                    |                    | Micro      | 0.0801977 |
|                    |                    | Z          | 0.2432459 |

with increased chromosome size. Windows with high  $F_{ST}$  accumulated on the microchromosomes earlier in speciation than on the macrochromosomes (Fig. 2). This pattern is unexpected given the reduced recombination rate on macrochromosomes relative to microchromosomes, and it was not predictable from the correlation between  $F_{ST}$  and chromosome size.  $F_{ST}$  is expected to

increase in regions of the genome that have reduced rates of recombination because the inheritance of linked loci results in locally reduced diversity within species over time. Likewise,  $F_{ST}$  is expected to decrease where recombination rate is high, because nucleotide diversity will increase in these regions over time. These predictions bring to light the peculiarity of our result that  $F_{ST}$  peaks seemed to accumulate on microchromosomes, where recombination rate is high, earlier in the divergence process than macrochromosomes, where recombination rate is lower. Average  $F_{ST}$  did not increase on small chromosomes relative to large chromosomes, but rather 100 kbp windows with elevated  $F_{ST}$  were more common on microchromosomes than on macrochromosomes in early stages of speciation. The early accumulation of  $F_{ST}$  peaks on microchromosomes may be due to a combination of characteristics of these small chromosomes. For example, higher gene density on microchromosomes may provide more targets for positive selection to act on, and higher recombination rate may increase the efficiency of selection in fixing beneficial mutations by reducing Hill-Robertson interference. If the rate of adaptive evolution is higher on



microchromosomes, this could lead to an earlier accumulation of  $F_{ST}$  peaks either as a direct result of within-species selective sweeps, or because some of these selective sweeps result in reproductive barrier loci. High linkage on macrochromosomes results in reduced diversity which leads to the accumulation of  $F_{ST}$  peaks on macrochromosomes observed at later stages of speciation.

Elevated  $F_{ST}$  windows were especially common on the Z chromosome in all three species, as expected based on prior work in hummingbirds [58] and other taxa [42]. The large X-effect, or the observation that sex chromosomes play a disproportionate role in speciation, is thought to be the overarching cause of the commonly found pattern of elevated  $F_{ST}$  on sex chromosomes relative to autosomes. Hypothesized reasons for the large X-effect include ploidy difference between sexes, faster evolution of the sex chromosomes and higher density of hybrid incompatibility loci on the sex chromosomes (reviewed in [42]). Using our data, we cannot distinguish between these processes nor determine whether the sexual dimorphism observed in hummingbirds plays a strong role in this genetic pattern.

Absolute divergence ( $d_{xy}$ ) and nucleotide diversity ( $\pi$ ) were strongly positively correlated and relative divergence ( $F_{ST}$ ) was negatively correlated with  $d_{xy}$  in all three species pairs. This suggests a pattern of selection before divergence wherein pre-speciation selection causes regions of reduced  $d_{xy}$  and low diversity ( $\pi$ ). Repeated linked selection at these regions before and after speciation can cause locally elevated relative divergence ( $F_{ST}$ ) despite reduced absolute divergence ( $d_{xy}$ ) [23]. Alternatively, this pattern could result from a global cross-species selective sweep after divergence (“sweep-before-differentiation;” [37]) at loci with elevated  $F_{ST}$  (e.g. [61]). This pattern is consistent with some previous findings in other species pairs of birds [62, 63]. However, given that the species in this study are the result of a recent rapid radiation and have recent common ancestors, a measure of absolute divergence might be unreliable for determining the evolutionary history of species pairs with such recent divergence times.  $d_{xy}$  measures the nucleotide differences that have accumulated since the divergence of the two focal species, but also reflects ancestral polymorphism that was present before divergence. Therefore, species pairs that have not been diverging for very long (including all three focal pairs of this study) are expected to have nucleotide diversity that is at least partially representative of ancestral polymorphism. Thus,  $d_{xy}$  that strongly correlates with  $\pi$  for these species pairs may indicate that much of the polymorphism in the ancestor of each species pair is retained in the extant populations [22].

The strong correlation between  $\pi$  and  $d_{xy}$  was present in all three species pairs, but the relationship weakened

as extent of reproductive isolation increased, indicating that absolute divergence increases relative to levels of within-species diversity over the course of the speciation process.  $F_{ST}$  peaks appearing in windows that have higher  $d_{xy}$  relative to  $\pi$  is expected, given that  $F_{ST}$  is a measure of differentiation relative to within-species polymorphism (Fig. 3d-f).

Nucleotide diversity ( $\pi$ ) was strongly correlated between hybridizing species (Fig. 4a-c) and the relationship was stronger for species pairs with less reproductive isolation. In species pairs with greater reproductive isolation the correlation weakened, with some genomic windows showing reduced  $\pi$  in one but not both species (Fig. 4b-c). Elevated  $F_{ST}$  in windows where one species has low nucleotide diversity relative to its closest relative is expected to be caused by within-species selection, rather than divergent selection between the species. However, we did not find strong evidence of elevated  $F_{ST}$  being caused by reduced polymorphism in one species within a pair as  $F_{ST}$  peaks appeared primarily on windows where both species had reduced  $\pi$  (Fig. 4d-f). A positive correlation of  $F_{ST}$  across species pairs might indicate that the landscape is partially driven by genomic features such as local recombination rate that are conserved across a higher phylogenetic level [63].

Differences in the genetic signatures of speciation across species pairs may be attributable to the natural history and phenotypic differences among species. For example, differences in plumage color are weak between Allen’s and Rufous hummingbird (genus *Selasphorus*), and these species are commonly misidentified. By contrast, the species pairs in *Calypte* and *Archilochus* have distinct differences in male plumage color, especially in the gorget feathers. Additionally, mating displays, habitats and migration habits differ across the six species in this study. Anna’s and Costa’s hummingbirds (genus *Calypte*) have complicated and poorly known migration patterns, with variation across populations and movement that is likely driven by variation in availability of food [64, 65]. Black-chinned (genus *Archilochus*), Ruby-throated (genus *Archilochus*) and Rufous hummingbirds perform complete migrations [66–68], while Allen’s hummingbirds include both migratory and sedentary populations [69]. Demographic differences across the species, such as historical changes in population size and distribution could also affect genetic signals that we interpret as signatures of speciation. For example, Anna’s hummingbird has drastically increased its breeding range in the last 100 years, likely as a result of increased availability of food in the form of exotic plants and hummingbird feeders [70]. In Allen’s hummingbird, the non-migratory subspecies (*S. s. sedentarius*) has expanded its breeding range into mainland southern California from the Channel Islands [71]. While the breeding ranges for



other species used in this study seem to have remained unchanged over time, observations of birds wintering over an expanded range have been observed [63–65]. These differences across the species used in this study highlight that each species pair is subject to its own evolutionary trajectory leading to a unique speciation event. While this is a general caveat of using independent species pairs as a proxy for the speciation continuum, we believe that the differences we observe among chromosome types can inform the ongoing debate about the roles of selection and recombination in the genetics of speciation.

## Conclusions

In this study we found evidence for the earlier divergence of microchromosomes than macrochromosomes in speciation through comparing genomic differentiation and diversity across the speciation continuum. Our study is the first, to our knowledge, to compare genomic statistics across different categories of autosomes and across independent, closely related species pairs with different levels of reproductive isolation. The results of this study suggest that variation in chromosome size, or in associated characteristics such as recombination rate and gene density, plays an important role in determining the genomic landscape of divergence at different points along the speciation continuum.

## Methods

### Sampling, extractions and sequencing

We collected samples from populations of three species pairs of hummingbirds for a total of six species and 59 individuals (Supplemental Table 3): (1) Allen's and Rufous (*Selasphorus sasin*, 9 samples; *S. rufus*, 7 samples), (2) Anna's and Costa's (*Calypte anna*, 12 samples; *C. costae*, 12 samples), and (3) Black-chinned and Ruby-throated (*Archilochus alexandri*, 10 samples; *A. colubris*, 9 samples). All populations were collected from allopatric regions with the exception of the *Calypte* species pair which was collected from the sympatric range in Riverside, CA. *C. anna* and *C. costae* hybridize infrequently, but the sympatric sampling for this species pair may lead to an underestimate of genetic differentiation between these species.

Samples from *Calypte* and *Selasphorus rufus* populations were extracted from dried blood spots on filter paper using the Qiagen DNeasy extraction protocol. Samples from both *Archilochus* species were acquired from Dr. Chris Clark's collected samples stored at the Yale Peabody Museum of Natural History, and samples from *Selasphorus sasin* were provided by the California Academy of Sciences. Small pieces of tissue were then extracted using the DNeasy extraction protocol. All DNA concentrations were quantified using a Qubit fluorometer and then diluted to 4 ng/uL in preparation

for a modified Nextera Whole Genome Library prep protocol [[72]; see [Supplementary Material](#) for our modifications to that protocol].

### Alignment and SNP calling

We used the Burrow-wheeler aligner (BWA-mem [73];) to align the sequences to an Anna's Hummingbird reference genome [74, 75]. We called variants using Samtools mpileup (v1.8 [76];) and filtered nucleotide positions for missing data (20% per locus, --max-missing 0.8), minimum depth (--minDP 2), biallelic sites (--maxalleles 2), and removed indels using VCFtools (v1.15 [77];), retaining all positions passing these filters, including invariant sites.

### Population statistics

We used VCFtools to calculate allele frequency and Weir and Cockerham's [78] weighted  $F_{ST}$ . Allele frequency was calculated for each nucleotide position passing our depth and missingness filters and  $F_{ST}$  was calculated for each non-overlapping 100 kbp window across the genome. From allele frequency,  $\pi$  and  $d_{xy}$  were calculated for each SNP following [37]. The result per SNP was then averaged over 100 kbp windows to match the windows in which  $F_{ST}$  was calculated. Separate files for  $\pi$ ,  $d_{xy}$  and  $F_{ST}$  were combined such that any window with a missing value for any one of the statistics was excluded from the final file.

We classified chromosomes 1, 2 and 3 as macrochromosomes and chromosomes 4–33 as microchromosomes. Though size varies across all chromosomes, the first three average an order of magnitude larger than the rest of the autosomes, and combined account for nearly one half of the entire genome size. For comparisons across the different chromosome types after filtering and SNP-calling, we separated the genome into three parts: microchromosomes, macrochromosomes, and Z chromosome. For analyses on the Z chromosome we used only the male individuals for population statistics calculations. Following Elgvin et al. [59] we compared the distribution of high- $F_{ST}$  regions (100 kbp windows in the top 1% for each species pair) among macrochromosomes, microchromosomes, and the Z chromosome using a chi-squared test in R v3.4.3. We used the Bedtools v2.28 [79] intersect command with the “-loj” option to obtain a list of annotated genes overlapping windows in the top 1% of  $F_{ST}$  values for each species pair.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12862-020-01674-9>.

**Additional file 1: Supplemental Fig. 1.**  $d_{xy}$  vs  $F_{ST}$ .

**Additional file 2: Supplemental Fig. 2.**  $F_{ST}$  for one species pair versus  $F_{ST}$  for another species pair.

**Additional file 3: Supplemental Table 1.** Number of  $F_{ST}$  windows from the top 1% distributed across different chromosome types for each species pair.

**Additional file 4: Supplemental Table 2.** List of genes associated with top 1%  $F_{ST}$  windows determined using genome annotation of Anna's Hummingbird from Rhie et al. 2020 [74].

**Additional file 5: Supplemental Table 3.** Sex and location for each sample.

**Additional file 6.** Supplementary Methods: Modified Nextera whole-genome library prep protocol for low volumes.

## Acknowledgements

Chris Clark provided training, supervision, and permits for hummingbird capture, and provided DNA samples. David Rankin provided training and advice on hummingbird capture and handling. Chris Clark, Jessica Purcell, Daniel Pierce, German Lagunas-Robles, and Nadje Najjar provided feedback on analyses and writing. The Yale Peabody Museum of Natural History and California Academy of Sciences provided tissue samples. Brian Myers assisted with DNA extractions. We thank David Rankin and Tom Schallert for permission to catch hummingbirds on private property. We are grateful to the reviewers for helping to improve the manuscript.

## Authors' contributions

EH and AB designed the study, obtained samples and data, and analyzed data. EH wrote the manuscript with input from AB. Both authors read and approved the final manuscript.

## Funding

Sampling and sequencing of *Selasphorus sasin* and *S. rufus* samples used in this project was supported by National Science Foundation grant IOS-1656708.

## Availability of data and materials

Raw sequence data is available at NCBI SRA (<https://www.ncbi.nlm.nih.gov/sra>) under Accession Number PRJNA640148.

## Ethics approval and consent to participate

All samples in this study were acquired under authority of USFWS permit #MB087454. *Calypte* and *Selasphorus* samples were collected in compliance with IACUC protocol #20160039, USGS Bird Banding permit #23516, and California Department of Fish and Wildlife permit #SC-006598. *Selasphorus rufus* samples were collected under the Oregon Dept of Fish and Wildlife permit #103-14, *Archilochus colubris* samples were collected under the Connecticut Dept. of Environmental Protection permit #1013003, and *Archilochus alexandri* were collected under Big Bend National Park permit and Texas Parks and Wildlife permit #SPR-0310-032.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

Received: 8 May 2020 Accepted: 19 August 2020

Published online: 03 September 2020

## References

- Campbell RC, Poelstra JW, Yoder AD. What is speciation genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. *Biol J Linnean Soc.* 2018;124:561–83.
- Mayr E. *Animal species and evolution.* Cambridge, Mass: Harvard University Press; 1963.
- Mallet J. Hybridization as an invasion of the genome. *Trends Ecol Evol.* 2005; 20:229–37.
- Ottenburghs J, Megens HJ, Kraus RHS, van Hooft P, van Wieren SE, Crooijmans RPMA, et al. A history of hybrids? Genomic patterns of introgression in the True Geese. *BMC Evol Biol.* 2017;17:201.
- Rutherford S, Rossetto M, Bragg JG, McPherson H, Benson D, Bonser SP, et al. Speciation in the presence of gene flow: population genomics of closely related and diverging *Eucalyptus* species. *Heredity.* 2018;121:126–41.
- Wang X, Que P, Heckel G, Hu J, Zhang X, Chiang C-Y, et al. Genetic, phenotypic and ecological differentiation suggests incipient speciation in two *Charadrius* plovers along the Chinese coast. *BMC Evol Biol.* 2019;19:135.
- Servedio MR, Hermisson J. The evolution of partial reproductive isolation as an adaptive optimum. *Evolution.* 2020;74:4–14.
- Ottenburghs J. Ghost introgression: spooky gene flow in the distant past. *Bioessays.* 2020;42. <https://doi.org/10.1002/bies.202000012>.
- Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlović M, et al. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J Evol Biol.* 2017;30:1450–77.
- Martin SH, Davey JW, Salazar C, Jiggins CD. Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biol.* 2019; 17:e2006288.
- Ellison CK, Wiley C, Shaw KL. The genetics of speciation: genes of small effect underlie sexual isolation in the Hawaiian cricket *Laupala*. *J Evol Biol.* 2011;24:1110–9.
- Ottenburghs J, Kraus RHS, van Hooft P, van Wieren SE, Ydenberg RC, Prins HHT. Avian introgression in the genomic era. *Avian Res.* 2017;8:30.
- Wright S. The genetical structure of populations. *Ann Eugenics.* 1949;15: 323–54.
- Lewontin RC, Krakauer J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics.* 1973;74:175–95.
- Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, et al. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res.* 2013;23:1817–28.
- Gosset CC, Bierne N. Differential introgression from a sister species explains high  $F_{ST}$  outlier loci within a mussel species. *J Evol Biol.* 2013;26:14–26.
- Toews DPL, Taylor SA, Vallender R, Brelsford A, Butcher BG, Messer PW, et al. Plumage genes and little else distinguish the genomes of hybridizing warblers. *Curr Biol.* 2016;26:2313–8.
- Irwin DE. Sex chromosomes and speciation in birds and other ZW systems. *Mol Ecol.* 2018;27:3831–51.
- Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, et al. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Res.* 2015;25:1656–65. <https://doi.org/10.1101/gr.196485.115>.
- Wu C-L. Genes and speciation. *J Evol Biol.* 2001;14:889–91.
- Turner TL, Hahn MW, Nuzhdin SV. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol.* 2005;3:e285.
- Noor MAF, Bennett SM. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity.* 2009;103:439–44.
- Cruickshank TE, Hahn MW. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol.* 2014;23:3133–57.
- Wolf JBW, Ellegren H. Making sense of genomic islands of differentiation in light of speciation. *Nat Rev Genet.* 2017;18:87–100.
- Nielsen R, Mountain JL, Huelsenbeck JP, Slatkin M. Maximum-likelihood estimation of population divergence times and population phylogeny in models without mutation. *Evolution.* 1998;52:669–77.
- Miller KMW, Bracewell RR, Eison MB, Bachtrog D. Patterns of genome-wide diversity and population structure in the *Drosophila athabasca* species complex. *Molec Biol Evol.* 2017;43:1912–23.
- Dutoit L, Vijay N, Mugal CF, Bossu CM, Burri R, Wolf J, Ellegren H. Covariation in levels of nucleotide diversity in homologous regions of the avian genome long after completion of lineage sorting. *Proc Royal Soc B.* 2017;284:20162756.
- Burri R. Interpreting differentiation landscapes in the light of long-term linked selection: differentiation and long-term linked selection. *Evol Lett.* 2017;1:118–31.
- Jackson BC, Campos JL, Zeng K. The effects of purifying selection on patterns of genetic differentiation between *Drosophila melanogaster* populations. *Heredity.* 2015;114:163–74.
- Wang J, Street NR, Scofield DG, Ingvarsson PK. Variation in linked selection and recombination drive genomic divergence during allopatric speciation of European and American aspens. *Mol Biol Evol.* 2016;33:1754–67.
- Delmore KE, Lugo Ramos JS, Van Doren BM, Lundberg M, Bensch S, Irwin DE, et al. Comparative analysis examining patterns of genomic

- differentiation across multiple episodes of population divergence in birds. *Evol Lett.* 2018;2:76–87.
32. Talla V, Johansson A, Dincă V, Vila R, Friberg M, Wiklund C, et al. Lack of gene flow: narrow and dispersed differentiation islands in a triplet of *Leptidea* butterfly species. *Mol Ecol.* 2019;28:3756–70.
  33. Mandeville EG, Walters AW, Nordberg BJ, Higgins KH, Burckhardt JC, Wagner CE. Variable hybridization outcomes in trout are predicted by historical fish stocking and environmental context. *Mol Ecol.* 2019;28:3738–55.
  34. Papadopoulos AST, Igea J, Dunning LT, Osborne OG, Quan X, Pellicer J, et al. Ecological speciation in sympatric palms: 3. Genetic map reveals genomic islands underlying species divergence in *Howea*. *Evolution.* 2019;73:1986–95.
  35. Uyeda JC, Hansen TF, Arnold SJ, Pienaar J. The million-year wait for macroevolutionary bursts. *Proc Natl Acad Sci.* 2011;108:15908–13.
  36. Supple MA, Papa R, Hines HM, McMillan WO, Counterman BA. Divergence with gene flow across a speciation continuum of *Heliconius* butterflies. *BMC Evol Biol.* 2015;15:204.
  37. Irwin DE, Alcáide M, Delmore KE, Irwin JH, Owens GL. Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. *Mol Ecol.* 2016;25:4488–507.
  38. Christie K, Strauss SY. Along the speciation continuum: quantifying intrinsic and extrinsic isolating barriers across five million years of evolutionary divergence in California juncos: reproductive isolation and genetic distance. *Evolution.* 2018;72:1063–79.
  39. Dufresnes C, Mazepa G, Jablonski D, Oliveira RC, Wenseleers T, Shabanov DA, et al. Fifteen shades of green: the evolution of *Bufo* toads revisited. *Molec Phylogenet Evol.* 2019;141:106615.
  40. Zieliński P, Dudek K, Amtzen JW, Palomar G, Niedzicka M, Fijarczyk A, et al. Differential introgression across newt hybrid zones: evidence from replicated transects. *Mol Ecol.* 2019;28:4811–24.
  41. Coyne JA. Genetics and speciation. *Nature.* 1992;355:511–5.
  42. Presgraves DC. Evaluating genomic signatures of “the large X-effect” during complex speciation. *Mol Ecol.* 2018;27:3822–30.
  43. Burt DW. Origin and evolution of avian microchromosomes. *Cytogenet Genome Res.* 2002;96:97–112.
  44. Rodionov AV. Micro versus macro: a review of structure and functions of avian micro- and macrochromosomes. *Russ J Genet.* 1996;32:517–27.
  45. Smith JM, Haigh J. The hitch-hiking effect of a favourable gene. *Genet Res.* 1974;23:23–55.
  46. Charlesworth D, Morgon MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. *Genetics.* 1993;134:1289–303.
  47. Kim Y, Stephan W. Joint effects of genetic hitchhiking and background selection on neutral variation. *Genetics.* 2000;155:1415–27.
  48. Huynh LY, Maney DL, Thomas JW. Contrasting population genetic patterns within the white-throated sparrow genome (*Zonotrichia albicollis*). *BMC Genet.* 2010;11:96.
  49. Ellegren H. The evolutionary genomics of birds. *Annu Rev Ecol Syst.* 2013;44:239–59.
  50. Manthey JD, Robbins MB, Moyle RG. A genomic investigation of the putative contact zone between divergent Brown creeper (*Certhia americana*) lineages: chromosomal patterns of genetic differentiation. *Genome.* 2016;59:115–25.
  51. McGuire JA, Witt CC, Renssen JV, Corl A, Rabosky DL, Altshuler DL, et al. Molecular Phylogenetics and the diversification of hummingbirds. *Curr Biol.* 2014;24:910–6.
  52. Abrahamczyk S, Renner S. The temporal build-up of hummingbird/plant mutualisms in North America and temperate South America. *BMC Evol Biol.* 2015;15:104.
  53. Licona-Vera Y, Ornelas JF. The conquering of North America: dated phylogenetic and biogeographic inference of migratory behavior in bee hummingbirds. *BMC Evol Biol.* 2017;17:126.
  54. Judd ER, Butler CJ, Batchelder N. Hybridization between black-chinned (*Archilochus alexandri*) and ruby-throated (*A. colubris*) hummingbirds in Oklahoma. *Bulletin of the OK Ornith. Soc.* 2011;44:1–7.
  55. Myers BM, Rankin DT, Burns KJ, Clark CJ. Behavioral and morphological evidence of an Allen's × Rufous hummingbird (*Selasphorus sasin* × *S. rufus*) hybrid zone in southern Oregon and northern California. *Auk.* 2019;136:ukz049.
  56. Wells S, Bradley RA, Baptista LF. Hybridization in Calypte hummingbirds. *Auk.* 1978;95:537–49.
  57. McCarthy EM. Handbook of avian hybrids of the world. New York, NY: Oxford University Press; 2006. p. 2006.
  58. Battey CJ. Evidence of linked selection on the Z chromosome of hybridizing hummingbirds. *Evolution.* 2019;74:725–39.
  59. Elgvin TO, Trier CN, Torresen OK, Hagen IJ, Lien S, Nederbragt AJ. The genomic mosaicism of hybrid speciation. *Sci Adv.* 2017;3:e1602996.
  60. Hillier L, Miller W, Birney E, et al. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature.* 2004;432:695–716.
  61. Norris LC, Main BJ, Lee Y, Collier TC, Fofana A, Cornel AJ, Lanzaro GC, et al. Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. *PNAS.* 2015;112:815–20.
  62. Van Doren BM, Campagna L, Helm B, Illera JC, Lovette IJ, Liedvogel M. Correlated patterns of genetic diversity and differentiation across an avian family. *Mol Ecol.* 2017;26:3982–97.
  63. Vijay N, Weissensteiner M, Burri R, Kawakami T, Ellegren H, Wolf JBW. Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa. *Mol Ecol.* 2017;26:4284–95.
  64. Clark CJ, Russell SM. Anna's Hummingbird (*Calypte anna*), version 1.0. In Birds of the World (A. F. Poole, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA. 2020. <https://doi.org/10.2173/bow.annhum.01>.
  65. Baltosser WH, Scott PE. Costa's Hummingbird (*Calypte costae*), version 1.0. In Birds of the World (A. F. Poole and F. B. Gill, Editors). Cornell Lab of Ornithology, Ithaca, NY, USA. 2020. <https://doi.org/10.2173/bow.coshum.01>.
  66. Baltosser WH, Russell SM. Black-chinned Hummingbird (*Archilochus alexandri*), version 1.0. In Birds of the World (A. F. Poole and F. B. Gill, Editors). Cornell Lab of Ornithology, Ithaca, NY, USA. 2020. <https://doi.org/10.2173/bow.bkchum.01>.
  67. Weidensaul S, Robinson TR, Sargent RR, Sargent MB, Zenzal TJ. Ruby-throated Hummingbird (*Archilochus colubris*), version 1.0. In Birds of the World (P. G. Rodewald, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA. 2020. <https://doi.org/10.2173/bow.rthhum.01>.
  68. Healy S, Calder WA. Rufous Hummingbird (*Selasphorus rufus*), version 1.0. In Birds of the World (A. F. Poole, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA. 2020. <https://doi.org/10.2173/bow.rufhum.01>.
  69. Clark CJ, Mitchell DE. Allen's Hummingbird (*Selasphorus sasin*), version 1.0. In Birds of the World (A. F. Poole, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA. 2020. <https://doi.org/10.2173/bow.allhum.01>.
  70. Zimmerman DA. Range expansion of Anna's hummingbird. *Am Birds.* 1973;27:827–35.
  71. Clark CJ. EBird records show substantial growth of the Allen's hummingbird (*Selasphorus sasin sedentarius*) population in urban Southern California. *Condor.* 2018;119:122–30.
  72. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One.* 2015;10:e0128036.
  73. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013; arXiv:13033997.
  74. Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S. Towards complete and error-free genome assemblies of all vertebrate species. 2020. bioRxiv:2020.05.22.110833.
  75. Vertebrate Genomes Project. Genome Ark: Calypte anna. Individual 1 Alternate Assembly. 2018. [https://vgp.github.io/genomeark/Calypte\\_anna/](https://vgp.github.io/genomeark/Calypte_anna/). Accessed Oct 2018.
  76. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics.* 2011;27:2987–93.
  77. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. And 1000 genomes project analysis group. The variant call format and VCFtools. *Bioinformatics.* 2011;27:2156.
  78. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population-structure. *Evolution.* 1984;38:1358–70.
  79. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26:841–2.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.