DR. DANIEL MARC HOOPER (Orcid ID : 0000-0002-4198-4320)

Article type : Special Issue

Sex chromosome inversions enforce reproductive isolation across an avian hybrid zone

[1,2]Daniel M. Hooper (DMH), [3]Simon C. Griffith (SCG), & [4]Trevor D. Price (TDP)

[1]Cornell Lab of Ornithology, Cornell University, Ithaca, NY 14850, USA
[2]Committe on Evolutionary Biology, University of Chicago, Chicago, IL 60637, USA
[3]Department of Biological Sciences, Macquarie University, Sydney, NSW, Australia
[4]Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA

Corresponding author:
Daniel M. Hooper: dmh356@cornell.edu

Last modified: August 28, 2018

Abstract

Across hybrid zones, the sex chromosomes are often more strongly differentiated than the autosomes. This is regularly attributed to the greater frequency of reproductive incompatibilities accumulating on sex chromosomes and their exposure in the heterogametic sex. Working within an avian hybrid zone, we explore the possibility that chromosome inversions differentially accumulate on the Z chromosome compared to the autosomes and thereby contribute to Z chromosome differentiation. We analyze the northern Australian hybrid zone between two subspecies of the long-tailed finch (*Poephila acuticauda*), first described based on differences in bill color, using reduced representation genomic sequencing for 293 individuals over a 1530 km transect. Autosomal differentiation between subspecies is minimal. In contrast, 75% of the Z chromosome is highly differentiated and shows a steep genomic cline which is displaced 350 km to the west of the cline in bill color. Differentiation is associated with two or more putative chromosomal inversions, each predominating in one subspecies. If inversions reduce recombination between hybrid incompatibilities, they are selectively favored and should therefore accumulate in hybrid zones. We argue that this predisposes inversions to differentially accumulate on the Z chromosome. One genomic region affecting bill color is on the Z, but the main candidates are on chromosome 8. This, and the displacement of the bill color and Z chromosome cline centers suggest that bill color has not strongly contributed to inversion accumulation. Based on cline width, however, the Z chromosome and bill color both contribute to reproductive isolation established between this pair of subspecies.

Keywords: Chromosome inversion, hybridization, sex chromosome, speciation, Estrildidae

Introduction

Genomic evidence across a wide variety of taxa indicates that hybridization between different lineages occurs regularly and leads to some amount of introgression (Payseur and Rieseberg, 2016). Hence, the development of reproductive isolation during speciation may often be slowed by the homogenizing influence of gene flow. During the speciation process, some genomic regions become resistant to gene flow before others and can result in so-called 'islands of differentiation' (Harrison, 1986; 1990; Wu, 2001; Feder et al., 2012). Coincident with this heterogeneity, chromosomal inversions and the sex chromosomes are two regions of the genome that are over-represented in surveys of genomic differentiation between hybridizing taxa (Payseur and Rieseberg, 2016; Muirhead and Presgraves, 2016). Chromosomal inversions commonly suppress recombination when paired against the uninverted chromosome. Pairing difficulties during meiosis reduce the frequency of cross overs and, when they do occur, cross overs may often result in the deleterious production of duplications, deletions, or aneuploid gametes. In hybrid zones, a recombination suppressor functions as an effective barrier to gene flow and may accumulate other incompatibilities over time (Navarro and Barton 2003; Kirkpatrick 2010). The sex chromosomes, in turn, tend to accumulate and expose deleterious gene combinations in hybrids faster than the autosomes do and so are expected to play a disproportionately large role relative to their size early in the speciation process (Charlesworth et al., 1987; Coyne and Orr, 2004, chs. 7, 8; Presgraves, 2008; 2010). In this paper we investigate the contribution of inversions and sex chromosomes to patterns of genomic differentiation across an avian hybrid zone. We review each in turn, focusing on avian studies, and the special case of ZW sex

74 chromosome systems applicable to birds, in which the female is the heterogametic sex
75 (Irwin, 2018).
76
77 Recombination suppressors may become established as a result of selection
78 against hybrids, and hence increase in hybrid zones (Charlesworth and Charlesworth
79 1973; 1979; Kirkpatrick and Barton, 2006; Dagilis and Kirkpatrick, 2016; Ortiz-
80 Barrientos et al., 2016). This selection arises because recombination in F1 hybrids
81 results in a higher fraction of gametes that carry combinations of alleles with potentially
82 deleterious effects (Kirkpatrick and Barton, 2006; Dagilis and Kirkpatrick, 2016). In this
83 way, contact between hybridizing species can increase the rate at which reproductive
84 isolation is achieved, rather than acting as a homogenizing force retarding
85 differentiation. In passerine birds, the rate of fixation for large pericentric inversions—
86 those inversions that include the centromere—scales with the proportion of closely
87 related species that are found in sympatry (Hooper and Price, 2015; 2017). Further,
88 sister species with overlapping ranges are more likely to differ by chromosome
89 inversions than those in allopatry (Hooper and Price, 2017). These findings are
90 consistent with a model in which gene flow between incipient species following
91 secondary contact establishes a selective advantage for inversions to increase in
92 frequency due to their ability to suppress local recombination (Kirkpatrick and Barton,
93 2006; Dagilis and Kirkpatrick, 2016). While evidence is lacking for birds, chromosome
94 inversions often contain genes affecting both pre-zygotic and post-zygotic isolation in a
95 diverse array of naturally hybridizing taxa including monkeyflowers (Lowry and Willis,
96 2010; Fishman et al. 2013), flies (Noor et al., 2001; Brown et al., 2004; McGaugh and
97 Noor, 2012) and fish (Jones et al., 2012; Kirubakaran et al., 2016). This could be because

98 the inversions themselves were favored as they hold together linked blocks of genes that
99 affect hybrid fitness and/or because they subsequently accumulated incompatibilities
100 that were unable to pass from one taxon to the other (Navarro and Barton, 2003;
101 Hoffmann and Rieseberg, 2008; Kirkpatrick, 2010; Fuller et al., 2018).
102
103 Sex chromosomes have long been recognized to play a disproportionately large role
104 in the speciation process (Charlesworth et al., 1987; Coyne and Orr, 2004, chs. 7, 8;
105 Presgraves, 2008; 2010). In birds, net genetic differentiation between species or
106 population pairs is almost always observed to be greater on the sex chromosomes than
107 on the autosomes (Irwin, 2018). This is also true across avian hybrid zones (Price, 2008;
108 Carling and Brumfield, 2008; Irwin, 2018). Explanations hinge on the presence of
109 differences between the sex chromosomes (Coyne and Orr, 2004; Presgraves, 2008;
110 2010; Payseur and Rieseberg, 2016; Muirhead and Presgraves, 2016). First, in an
111 isolation by distance model the smaller effective population size of the primary sex
112 chromosome (X or Z) is expected to lead to greater differentiation than the autosomes
113 (Fst 33% higher). Differentiation across hybrid zones is often much greater than this
114 (Mank et al., 2010; Ellegren 2013; Payseur and Rieseberg, 2016; Irwin, 2018). Second,
115 hybrids of the heterogametic sex (i.e. XY or ZW) suffer a greater loss of fitness in the
116 form of hybrid sterility or inviability than the homogametic sex (XX or ZZ; Haldane's
117 rule). Under the dominance theory of Haldane's rule, the heterogametic sex suffers from
118 the immediate exposure of any X-linked recessive incompatibilities (Charlesworth et al.,
119 1987; Turelli and Orr, 1995; 2000). Haldane's rule is exceptionally strong in birds
120 (Schilthuizen et al., 2011) and has been invoked to explain limited introgression of Z
121 linked alleles across avian hybrid zones (Carling and Brumfield, 2008; Storchova et al.,

2010; Gowen et al., 2014; Toews et al., 2015; Irwin, 2018). Third, sex chromosomes are enriched relative to the autosomes for loci involved in hybrid sterility and hybrid loss of fitness across a wide range of taxa, including *Drosophila*, mice, and butterflies (Coyne and Orr, 1989; Coyne, 1992; Masly and Presgraves, 2007; Presgraves, 2002; 2008; 2010). In birds, the Z chromosome is enriched for male-biased genes and is also the location of faster rates of female-biased gene expression divergence, at least in the Galloanserae (Ellegren, 2013; Dean et al., 2015).

Finally, theory suggests that ZW systems, in contrast to XY systems, should be especially susceptible to accumulating genes affecting hybrid sterility. In birds, the reduced level of recombination compared to the autosomes, due to there being only one copy in the heterogametic sex, makes the sex chromosomes particularly susceptible to meiotic drive. In ZW, but not XY, systems, meiotic drive is relatively easily achieved by passing one or other chromosome to the polar body (Rutkowska and Badyaev, 2008). In ZW systems too, mutations on the Z that cause females to differentially invest in sons (or on the W that cause females to differentially invest in daughters) are favored (Miller et al., 2006). This doesn't apply in XY systems, because the investing sex – the female – is homogametic. Both meiotic drive and differential investment result in deleterious consequences in hybrids that may be subsequently compensated by recombination modifiers or suppressors when favorable gene combinations are broken apart.

Plausibly, inversion differentiation and Z chromosome differentiation become associated. As differences accumulate on the sex chromosomes, they should be more likely to generate incompatibilities between species which in turn favors the increase of

146 inversions on these chromosomes in hybrid zones. In support of this, reviews of
147 cytological studies across passerine birds show a relatively high frequency of inversions
148 on the sex chromosomes, even when one might expect more on the autosomes, because
149 the larger autosomal population size results in higher mutational input (Hooper and
150 Price, 2015; 2017). Possibly then, the sex chromosomes accumulate genomic differences,
151 which then favors the spread of inversions, that in turn increases differentiation in a
152 feedback process. Even without this feedback cycle, the excessive presence of inversions
153 on the sex chromosomes may be an important contributor to sex chromosome
154 differentiation across hybrid zones.
155
156 Here, we take a genomic approach to examine the differentiation landscape
157 between hybridizing subspecies of the long-tailed finch (*Poephila acuticauda*, family
158 Estrildidae). Members of the Estrildidae frequently differ by large chromosome
159 inversions, especially between closely-related sympatric species (Hooper and Price,
160 2015; Stryjewski and Sorenson, 2017). We chose to study the long-tailed finch because
161 cytological studies have uncovered the presence of large segregating inversions on the Z
162 chromosome in this species (Christidis, 1986). Singhal et al. (2015) showed that the Z
163 chromosome is strongly differentiated between subspecies. We ask what might be
164 preserving this differentiation across their hybrid zone.
165
166 Classification of the two subspecies of *P. acuticauda* is based on bill color (yellow
167 in the western subspecies *acuticauda* and red in the eastern subspecies *hecki*; Fig. 1).
168 The geographic extent of bill color admixture between subspecies is restricted to a
169 narrow area approximately 150 km in width, or ~8% of their total range, containing

orange-billed individuals (Griffith and Hooper, 2017). The limited extent of bill color admixture observed in the wild indicates an appreciable degree of selection against intermediate and/or parental forms when they are rare (i.e., when individuals disperse through the zone; Griffith and Hooper, 2017). Hence, bill color is a trait whose genetic underpinning could contribute to differentiation of the sex chromosomes, if found on that chromosome, and to promote the spread of an inversion, if associated with at least one other nearby gene that also reduces hybrid fitness. While we focus in this paper on bill color, subspecies also differ in song (Zann, 1976) and sperm morphology and performance (Rowe et al., 2015) – traits known in birds to contribute to pre-mating and post-mating pre-zygotic isolation, respectively (Price, 2008, chs. 10, 14), and which would potentially contribute to inversion accumulation if their genetic underpinnings are linked to bill color differences.

The presence of chromosome inversions, together with the availability of traits germane to the development of reproductive isolation between these hybridizing subspecies, makes the long-tailed finch a robust system in which to examine the contribution of inversions to speciation and sex chromosome differentiation. We utilize reduced-representation genomic data from 293 finches drawn from 18 populations spanning nearly the entire longitudinal distribution of this species, plus previously published whole genome sequence data for 20 individuals (10 of each subspecies, Singhal et al., 2015) in order to: (1) assess the landscape of genomic differentiation between subspecies and the geographic extent of their genetic admixture, (2) search for evidence of the chromosomal inversions previously identified cytologically and evaluate

193 the degree to which they may contribute to reproductive isolation, and (3) identify
194 candidate genomic regions underlying variation in bill color.
195

196 Materials and Methods

197 *Study system and sampling*

198 We captured long-tailed finches in the dry season (between late August and October), in
199 three years (2009, 2010, and 2015). We captured 784 adult individuals in total across 18
200 sites that cover >80% of the longitudinal distribution of the species (1500 km of 1800
201 km; Figure 1). Griffith and Hooper (2017) report bill color, measured via spectral
202 reflectance, for 737 of these birds. Blood samples were taken from the brachial vein and
203 stored on EDTA-treated FTA cards or in 90% ethanol at -20°C. Long-tailed finch tissue
204 samples from an additional site (Koolpinyah; Fig. 1, pop. 14) and tissue and blood
205 materials from the black-throated finch (*Poephila cincta*), sister species to the long-
206 tailed finch, were sourced from the Australian National Wildlife Collection (ANWC).
207 Samples of the black-throated finch were included to root genomic analyses in the long-
208 tailed finch system. Details for sampled populations are presented in Tables 1 and S1 .
209

210 *Genomic library preparation, sequencing, and variant calling*

211 DNA was extracted using the Qiagen Gentra Puregene Tissue Kit (Qiagen) following the
212 manufacturer's instructions. We created genomic libraries following the MSG protocol
213 previously described in Andolfatto et al. (2011) and Schumer et al. (2014). A total of 351
214 individual samples were 100 bp paired-end sequenced across seven Illumina HiSeq
215 4000 lanes (*Poephila acuticauda*, N = 311 and *P. cincta*, N = 40; Tables 1 and S1). Raw
216 reads from seven pooled genomic libraries were first de-multiplexed by sample-specific

barcodes, trimmed for restriction enzyme overhang, and quality filtered to remove reads with more than seven bases below Q20 using ipyrad v.0.6.11 (http://github.com/dereneaton/ipyrad). Following de-multiplexing, individual reads were mapped to the zebra finch reference genome assembly taeGut1, otherwise known as WUGSC 3.2.4 (ftp://hgdownload.cse.ucsc.edu/goldenPath/taeGut1/chromosomes/), using BWAmem v.0.7.12 (Li, 2013). Sequence divergence between the long-tailed finch or black-throated finch and the zebra finch is low, about 1.6% (Singhal et al., 2015), and on average 98% of reads for both species successfully mapped to unique positions in the zebra finch reference assembly.

Following initial alignment, we called nucleotide and indel variants across the entire zebra finch reference assembly using the GATK v3.7 HaplotypeCaller (McKenna et al., 2010). We integrated variant data calculated individually across all samples into a combined VCF file using the CombineGVCF and GenotypeGVCF modules in GATK. SNPs were pruned from this initial variant call set via hard quality filters in GATK as follows: variant confidence by depth (QD < 2.0), strand bias (FS > 60.0), mapping quality across samples (MQ < 40.0), mapping quality of heterozygous variants (MQRankSum < -12.5), and distance from read ends for variant calls (ReadPosRankSum < -8.0). Moreover, we masked SNPs at sites that overlapped the repeat annotation for the zebra finch to avoid including error-prone variants and variants from genomic regions that are likely to be collapsed duplicates in a genome assembly, based on annotation from the RepeatMasker Repeat Library (Smit et al., 1996-2010; http://www.repeatmasker.org). Finally, this quality-filtered variant set was refined using vcftools v0.1.14 to exclude singletons (--mac 2), variants missing coverage in more

than 60% of samples (--max-missing 0.4), and samples with less than 25% of the remaining genotyped variable sites. The remaining set of 4,389,581 SNP variants included both long-tailed and black-throated finch samples (N = 330).

To obtain accurate sample sizes for clinal analyses we sexed all individuals by amplifying the sex-linked CHD gene using primers 2550F and 2718R (Fridolfsson and Ellegren, 1999) and visualized the PCR product upon a gel following electrophoresis. Samples were scored as male from the presence of a single PCR band (homozygous, ZZ) or as female from the presence of two bands of different size (heterozygous, ZW).

Population genetic structuring

We calculated Estimated Effective Migration Surfaces (EEMS; Petkova et al., 2016) between populations to assess the genomic context and geographic location of any historical barriers to migration between long-tailed finch subspecies. The EEMS approach utilizes a population genetic model to visualize departures from strict isolation by distance by comparing a genetic distance matrix to a geographic distance matrix using a geo-referenced set of genotyped individuals. In order to evaluate the magnitude of any migration barrier between long-tailed finch subspecies relative to that between species, we also included geo-referenced genotype data from the black-throated finch; the parapatric sister taxon to the long-tailed finch (Table S1; Figure 1). We used EEMS to analyze autosomal and Z chromosome data separately. We calculated pairwise genetic distance between individuals using the *bed2diffs_v1* function in EEMS with the entire set of quality-filtered SNPs for the autosomes (4,310,439 SNPs and 330 individuals), and Z chromosome (79,142 SNPs and 330 individuals), respectively. We performed

three independent chains of each EEMS run for 1 x 10^7 MCMC iterations (numMCMCIter), with a 5 x 10^5 burn-in (numBurnIter), iteratively between 100, 300, and 500 demes (nDemes); and with all other tuning parameters set to give acceptance ratios between 20-30%. Replicate runs for each dataset were combined, following individual assessment of EEMS convergence, and the fit of each model was assessed visually by comparing the relationship between fitted and observed dissimilarities between demes. EEMS also outputs estimates of genetic diversity across the range, which we present in Figure S1.

Geographic cline analyses

To further examine the geographic dimensions of genetic admixture between long-tailed finch subspecies we analyzed subsets of our variants comprising either ancestry-informative autosomal or Z chromosome SNPs. We defined ancestry-informative SNPs to be at or above an F_{ST} threshold of 0.85 between the most distant populations of subspecies *acuticauda* (populations 1 and 2, combined) and subspecies *hecki* (populations 18 and 19, combined). To remove inference inflation from close physical linkage these SNPs were randomly filtered so that no two markers were located within 10kb of each other. This resulted in subsets of 33 autosomal and 870 Z chromosome SNPs, respectively. As the proportion of ancestry informative (i.e. $F_{ST} > 0.85$) autosomal SNPs is five orders of magnitude lower than Z linked SNPs (38 of 1,919,450 autosomal SNPs against 1148 of 26,827 Z-linked SNPs; prior to SNP filtering for physical proximity < 10kb), and because no barrier to migration was observed on the autosomes using EEMS, we restricted downstream analyses in the main text using ancestry-informative markers to the Z chromosome (see Figs. 1 and S2 for results based on ancestry-

289 informative autosomal SNPs).

290

291 Focusing on the Z chromosome, we used the Bayesian clustering approach
292 implemented in STRUCTURE v2.3.4 (Pritchard et al., 2000; Falush et al., 2003) in
293 order to assign all individuals to genetic groups (Figure S2). We set the hypothesis about
294 the number of genetic clusters (*K*) equal to 2 as we sought to delineate the hybrid zone
295 between subspecies. Subsequent runs increasing *K* up to 4 did not result in a superior fit
296 to the data as determined by log likelihood comparison. We ran 10 iterations for each
297 expectation of *K* using the admixture model with correlated allele frequencies for
298 500,000 MCMC replicates and a burn-in period of 50,000 replicates. Results across
299 iterations of *K* were concatenated and summarized using the program CLUMPP
300 (Jakobsson and Rosenberg, 2007). Following Singhal and Moritz (2013), we generated a
301 hybrid index score for all samples based upon the probability (*Q*) each individual was
302 assigned by STRUCTURE to the eastern *hecki* subspecies. Accordingly, a hybrid index
303 was generated for each individual based on its Z chromosome *Q* score: HI_Z. Individuals
304 were classified as pure *hecki* if $Q \geq 0.9$, pure *acuticauda* if $Q \leq 0.1$, or admixed ancestry
305 if $0.1 < Q < 0.9$.

306

307 To estimate the geographic center and width of genomic admixture between long-
308 tailed finch subspecies, we fit the mean hybrid index value (HI_Z) and sample size for
309 each of the 18 populations with genetic data to a series of equilibrium geographic cline
310 models using the Metropolis-Hastings Markov chain Monte Carlo algorithm employed
311 in the R package *HZAR* (Derryberry et al., 2014). Distances between sampling sites were
312 estimated from their latitude and longitude in Google Earth and are cumulative from

313 the most western site (Population 1; Table 1). We ran 5 separate models that varied in
314 the number of cline shape parameters estimated. All models estimated cline center
315 (distance from sampling location 1, *c*) and width (1/maximum slope, *w*), but varied in
316 their estimation of different combinations of the exponential decay curve parameters δ
317 and τ (no tails, right tail only, left tail only, mirrored tails, or both exponential tails
318 estimated individually). Parameters delta (δ) and tau (τ) correspond to the distance from
319 the cline center to the exponential tail and the slope of that tail, respectively (Gay et al.,
320 2008; Derryberry et al., 2014). We evaluated support for alternative models using AIC
321 corrected for small sample size (AICc). The model with the lowest AICc score was
322 selected as the best-supported model.
323
324 In order to evaluate coincidence in the geographic location of genetic and bill
325 color admixture, we compared parameter values for the best-supported hybrid index
326 (HI_Z) cline to the previously inferred bill color cline (Griffith and Hooper, 2017). We
327 compared the two log-likelihood unit support intervals around cline center location for
328 genetic and bill color models and conservatively considered the centers of the genetic
329 and bill color clines to be significantly displaced if support intervals did not overlap.
330
331 *Genomic differentiation landscape*
332 Heterogeneity in genomic differentiation between taxa results from a complicated
333 combination of factors besides the influence of selection acting against hybrids. These
334 include incomplete lineage sorting, linked background selection, and recombination rate
335 variation (Clark, 1997; Cruickshank and Hahn, 2014; Burri et al., 2015; Payseur and
336 Rieseberg, 2016). For example, outlier regions of high differentiation may indicate a

resistance to gene flow produced by selection targeting hybrids but could also result from recombination rate variation, shared patterns of background selection, or selective sweeps before and after hybridization (e.g. Irwin et al., 2016). To control for these potentially confounding forces we built an empirical null model of differentiation across the genome in 50kb sliding windows with 25kb steps for two pairs of allopatric populations, one within each subspecies (following Vijay et al., 2016). Allopatric populations are assumed to be subject to similar forces of background selection and recombination rate variation as the hybrid zone. Using these as control comparisons, we determined outlier windows at the 99th percentile of the Z-transformed F_{ST} distribution (hereafter F_{ST}'), with differentiation represented in units of standard deviation relative to the mean. We evaluated outlier regions in our focal hybrid zone comparison (between individuals of *acuticauda* and *hecki* background from populations 4 and 5) in the same way. Outlier regions in our focal comparison that are also present in control comparisons are assumed to be a product of shared selection since the common ancestor of both subspecies. Accordingly, we subtracted the maximum value of F_{ST}' from orthologous windows in control comparisons from our focal comparison and focused attention on outliers at the 99th percentile for this so-called net genetic differentiation statistic (ΔF_{ST}'). Outlier windows in the focal comparison for F_{ST}' but not for ΔF_{ST}' are interpreted as differentiation regions resulting from processes shared across the entire species. Outlier regions in the focal comparison for both F_{ST}' and ΔF_{ST}', are considered to be resistant to gene flow, containing or linked to loci contributing to reproductive isolation between subspecies.

Linkage disequilibrium analyses

Chromosome inversion polymorphisms, by virtue of rarely recombining, leave characteristic patterns of long-range linkage disequilibrium across the chromosome, particularly near inversion breakpoints (Bansal et al., 2007; Stefansson et al., 2005; Knief et al., 2016). A finescale recombination map for the long-tailed finch has previously been described based on linkage disequilibrium patterns for SNPs called using whole-genome resequencing data for 20 individuals, 10 of each subspecies (Singhal et al., 2015). Just as in the closely-related zebra finch, broad-scale rates of recombination (cM/Mb) in the long-tailed finch were inferred to be elevated near chromosome telomeres and largely reduced to 'recombination deserts' across chromosome interiors (Backström et al., 2010; Singhal et al., 2015). More recent analyses of linkage disequilibrium patterns and chromosome-scale genomic substructure in the zebra finch suggest that the inferred recombination deserts on at least four chromosomes in this taxon are a byproduct of large chromosome inversion polymorphisms segregating at high frequency in both wild and captive populations (Knief et al., 2016; 2017; Kim et al., 2017). Indeed, the recombination desert on the zebra finch Z chromosome, with recombination rates estimated to be two orders of magnitude lower than on an equivalently sized autosome (Singhal et al., 2015), is likely the result of two polymorphic inversions that together encompass 87% of the length of the chromosome and segregate at high frequency, such that 68% of all males are heterokaryotypes (459 of 676 males, Kim et al., 2017; Knief et al., 2016; 2017). Here we make the assumption that strong linkage disequilibrium is likely a consequence of segregating inversions. We estimated linkage disequilibrium for the Z chromosome and the twenty-one autosomal chromosomes greater than 10 Mb in size using all 293 individuals together. Genotypic linkage disequilibrium (henceforth LD) is calculated as

385 the squared Pearson's correlation coefficient (r^2, calculated using PLINK v1.9; Purcell et
386 al., 2007), between all pairs of SNPs with a minor allele frequency above 10%, less than
387 50% missing data, and thinned to include a maximum of 1000 markers/chromosome
388 separated by at least 50kb.
389
390 To evaluate how linkage disequilibrium patterns for the set of Z-linked ancestry
391 informative markers changed across the long-tailed finch range, and in effect test for Z
392 chromosome inversion differences between subspecies, we estimated LD using the set of
393 870 SNPs above an F_{ST} threshold of 0.85 plus an additional 50 SNPs, randomly selected
394 after filtering for coverage and physical distance, from outside the region of high
395 differentiation between the subspecies, between coordinates 0-11.6 Mb and 68.7-72.2
396 Mb along the chromosome. Individuals with greater than 80% missing data were
397 excluded from analyses. We calculated pairwise LD using only SNPs with a minor allele
398 frequency above 10% within this set of quality and distance-filtered SNPs in turn for 1)
399 all individuals together (pops. 1-19, N = 247), 2) individuals identified as genotypically
400 admixed (pops. 3-5, N = 20), 3) individuals from subspecies *acuticauda* (pops. 1-5, N =
401 53), and 4) individuals from subspecies *hecki* (pops. 4-19, N = 169; Table 1).
402
403 Using whole genome sequence data available for the long-tailed finch (10 of
404 subspecies *acuticauda* and 10 of *hecki*) and the zebra finch (10 samples; ENA Study:
405 PRJEB10586; Singhal et al., 2015; Table S5), we calculated a suite of summary statistics
406 in order to independently evaluate signatures of evolutionary change on the Z
407 chromosome. Read data were mapped and SNP variants were called and filtered using
408 the bioinformatic pipeline described above. We first calculated an admixture-corrected

409 version of the population branch statistic (PBS; Yi et al., 2010; Huerta-Sánchez et al.,
410 2013). This statistic uses an outgroup taxon, in this case the zebra finch, to infer the
411 population containing the derived form of the SNP. This approach identifies which
412 regions of the genome show signatures of selection unique to each subspecies. We next
413 calculated Tajima's *D* (Tajima, 1989). The *D* statistic compares intermediate to low
414 frequency sites within a population: negative values indicate a surplus of rare frequency
415 variants with respect to a neutral model; a cluster of negative values would be consistent
416 with recent selection on a linked block. We also calculate Fay and Wu's *H* (Fay and Wu,
417 2000), which uses an outgroup (again the zebra finch) to polarize the frequency
418 spectrum: negative values indicate a surplus of high-frequency derived alleles as
419 expected if a selective sweep had recently occurred.
420
421 *Genome-wide mapping of bill color loci*
422 We take advantage of the finding that the center of Z chromosome genomic admixture
423 lies to the west of bill color admixture. Hence, association driven by geographical
424 structure in the genome and bill color can be minimized by mapping individuals drawn
425 only from populations to the east of the genomic hybrid zone (i.e. 192 individuals from
426 populations 5-18; Figure 1). We used a linear regression of bill color on SNP genotype,
427 examining each bi-allelic SNP in the quality-filtered SNP set with association
428 significance adjusted for multiple testing, as the Benjamini and Hochberg (1995) step-
429 up false discovery rate control (implemented in PLINK v1.9). We additionally used
430 PLINK to evaluate association significance adjusted for the genomic-control significance
431 in order to account for population structure. Sample phenotype was scored and
432 independently evaluated in three ways, following Griffith and Hooper (2017), as bill

chroma (S5, from 450-650nm), the individual loading for PC2 (from 300-650nm), and as a dichotomous variable (0 = pure *acuticauda* yellow or admixed orange, samples in the lower three quartiles of the bill chroma distribution, or 1 = pure *hecki* red, all remaining samples). Results scoring bill color as a dichotomous variable and as chroma were similar and we present results for chroma in the supplement only (Figure S5).

As a means to confirm these regions as important contributors to bill color, we searched for likely candidate genes, using the set of SNPs called from whole genome sequence data. We first used the population branch statistic, with the zebra finch as outgroup, in order to detect evidence of positive selection on bill color loci and polarize selection to subspecies. We calculated this statistic in 10kb sliding windows across chromosomes 8 and Z, where three candidate bill color loci were found to reside. One of our candidate regions encompasses a gene previously implicated in carotenoid-based color polymorphism in birds (*CYP2J19*; Mundy et al., 2016; Lopes et al., 2016; Twyman et al., 2018). We investigated this region (chr8: 24577711-24603991) encompassing both the A and B copy of this gene for fixed differences between subspecies using consensus sequences called for all 20 long-tailed finch samples with whole genome sequence data.

Results

Genomic library preparation, sequencing, and variant calling

Across the 311 long-tailed finch individuals sequenced in this study, approximately 1.42 billion 100bp reads were generated with an average of 4.56 million reads per sample (median of 3.15 million reads). From an initial variant call set of 24.55 million SNPs, filtering by quality, frequency, coverage, and masking across repeat regions, our final

quality-filtered dataset for long-tailed finch restricted analyses consisted of 3,565,420 SNPs (55,791 of which are located on chromosome Z). After filtering for missing data, 293 of 311 individuals sequenced were retained. We were able to confidently sex all 293 long-tailed finch samples used in analyses via amplification and electrophoresis of the CHD locus.

Population structure analyses

After filtering for quality and coverage, all but one of the 953 SNPs fixed between our most distant populations (populations 1 and 2 against 18 and 19) were located on the Z chromosome (99.9% of fixed differences). Averaged pairwise F_{ST} for the autosomes is low between all populations ($F_{ST} \approx 0.0$). Averaged pairwise F_{ST} for the Z chromosome fell into two highly differentiated ($F_{ST} \approx 0.4$) population pools. Based on the Z chromosome F_{ST} values we define subspecies *acuticauda*, populations 1-3 (N = 54); subspecies *hecki*, populations 6-19 (N = 197); and two putative hybrid zone populations of mixed ancestry, populations 4 and 5 (N = 42, Table 1).

Estimated effective migration surfaces from the complete sets of quality-filtered autosomal and the Z-linked SNPs are shown in Figure 2 with colors representing relative rates of migration on a $\log_{10}$ scale ranging from lower than average (orange) to higher than average (blue). Autosomal migration barriers are inferred between (1) the black-throated finch and the long-tailed finch, as expected, (2) across the unsampled drier region to the south of the long-tailed finch transect, (3) the area spanning bill color admixture, populations 12 and 15, and (4) between northern populations 11 and 14 (Figs. 1 and 2; Table 1). Autosomal migration barriers are also inferred within the range of the

481 black-throated finch but are not considered further in this paper. On the Z chromosome,
482 a prominent migration barrier is located between populations 4 and 5, running north to
483 south just west of the Western Australia – Northern Territory border. This is of greater
484 strength than the barrier between long-tailed and black-throated finch populations
485 (Figure 2b). Notably, *within* both long-tailed finch subspecies, estimated effective
486 migration rates are greater than expected by an isolation by distance model, implying
487 that natal dispersal distance *per se* is not the only limit to gene flow between subspecies.
488
489 *Geographic cline analyses*
490 The best-fitting cline model for the Z chromosome hybrid index (HI_Z, constructed from
491 870 ancestry informative SNPs), included estimates of cline width, center, mirrored
492 shape parameters for both left and right exponential tails, and fixed p_{min}/p_{max} at 0 and 1
493 (Table S2). This cline was centered 351 km (336 – 369, two log-likelihood support
494 limits) east of Mt. House (population 1) and is 115 km (86 – 166) in width and located
495 between populations 4 and 5 just across the Western Australia – Northern Territory
496 border (Figs. 1, 3, Table 1).
497
498 The center of bill color admixture is displaced 350km to the east of the center of
499 the Z chromosome hybrid index cline (Figure 1). The best-fitting cline model for bill
500 chroma has a center 702 km (698 – 708) east of population 1, falls between populations
501 12 and 13 and has a width of 146 km (Figure 1; Table S4; Griffith and Hooper, 2017). The
502 centers of the Z chromosome hybrid index and bill color clines are significantly
503 displaced (two log-likelihood support units do not overlap). The autosomal analysis is
504 based on only 33 ancestry informative SNPs, but the hybrid index shows a cline

505 intermediate between that of the bill color and the Z chromosome clines (Figure 1; Table
506 S3). Analysis of the SNP data using STRUCTURE gives entirely concordant results
507 (Figure S2).
508
509 *Genomic differentiation on the Z chromosome*
510 The Z chromosome is strongly differentiated and the autosomes much less so (Figure 3).
511 In this section we ask if particular regions of the Z are more strongly differentiated than
512 others. In order to control for the confounding influence of linked background selection,
513 we first assessed the landscape of genomic differentiation within subspecies using two
514 control comparisons (Figure 3a). Outlier windows (>99% percentile) for F_{ST}' were often
515 held in common between control comparisons, consistent with shared patterns of
516 background selection acting in each subspecies (Figure 3a). In both control
517 comparisons distantly located from the hybrid zone, the Z chromosome is enriched for
518 outlier windows for F_{ST}' relative to the autosomes (49.8% and 46.3% of outlier windows
519 are Z-linked, respectively). By contrast, F_{ST}' outlier windows are 98.8% Z-linked in the
520 comparison contrasting populations adjacent to the center of the hybrid zone. The Z
521 chromosome remains highly differentiated in our focal comparison even after
522 controlling for shared linked selection via ΔF_{ST}' (Figure 3b). 99.3% of the ΔF_{ST}' outlier
523 windows reside in a region that spans approximately 78% of the Z chromosome: 57 Mb
524 from approximately 11.5-68.8 Mb. Together this suggests that a large region of the Z
525 chromosome is resistant to gene flow and hence contains genes involved in reproductive
526 isolation (Figure 3b). These genes likely form a linked non-recombining block (see
527 below).
528

In Figure 4c we plot the population branch statistic between subspecies (populations 1 and 2 against 16), evaluated against the zebra finch for the Z chromosome. This shows no differentiation along the first 10 Mb of the chromosome and an asymmetric pattern of evolution across the rest of the chromosome. Hence the landscape of Z chromosome differentiation between subspecies is best explained as a result of differential selection acting on Z-linked loci and not a homogenous byproduct of sex chromosome demography in general.

Linkage disequilibrium analyses

Analysis of disequilibrium patterns in all long-tailed finch individuals combined reveals a large block of high LD on the Z chromosome, spanning a total of ~57.1 Mb containing 542 genes (from 11.6-68.7 Mb; Figure 4a). This could potentially reflect geographical structure. However, most of the LD pattern observed in our full sample set remained after restricting analysis to 20 individuals identified as admixed by STRUCTURE (populations 3-5, Figure 4b, center panel; Fig. S2), but now a mosaic distribution of blocks of long-range linkage appears, with perhaps four LD blocks, at least two of which are in LD with each other (Fig. 4b). Restricting analysis to samples drawn from the same subspecies reveals two putative Z chromosome LD blocks, one in each subspecies (between SNPs located from 11.6-62.3 Mb in *acuticauda* and from 10.8-21.4 Mb and between this region and SNPs located around 68.4 Mb in *hecki,* Fig. 4b). No autosome has long-range LD patterns comparable in magnitude or size to that observed on the Z chromosome (Figure S3).

Genome-wide mapping of bill color loci

We first performed association mapping using 3,565,420 quality-filtered genome-wide SNPs in 192 individuals from populations 5-18, with bill color scored in two ways (as a dichotomous color, or as PC2 from the reflectance curve). Two chromosomes contain SNPs that showed strong associations with bill color: chromosome 8 and the Z chromosome (Figure 5). On chromosome 8 the most compelling region contained a SNP located at 24.68 Mb. A second region contained 36 SNPs located between 4.99-5.27 Mb. Genes lying nearby that have been identified from other studies as influencing color are indicated in Figure 5. Further analyses identify the *CYP2J19* gene as an especially strong candidate on chromosome 8 (Fig. S6). A single region on the Z chromosome shows a strong association with bill color variation (Figure 5). This region contains 21 SNPs located between 59.45-60.22 Mb that were significant when bill color is scored as a dichotomous trait, but not for PC2. It spans nine protein coding genes in the zebra finch genome assembly. Moreover, SNPs within each of these three genomic regions remained significantly associated with bill color scored as a dichotomous trait after correcting both for genomic-inflation from population structure and adjusting for multiple tests.

Discussion

A large fraction of genomic differences between hybridizing species associates with chromosomal inversions (Hoffmann and Rieseberg, 2008; Payseur and Rieseberg, 2016) and the sex chromosomes (Presgraves, 2008; 2010; Muirhead and Presgraves, 2016; Toews et al., 2016; Irwin, 2018). This is true in the long-tailed finch hybrid zone, but in this case all the putative inversions we have identified reside on the Z chromosome. The Z chromosome is resistant to gene flow across ¾ of its length and is

likely to contain genes that affect reproductive isolation between the subspecies. Indeed, with the exception of the displaced cline in bill color, and autosomal variants likely associated with bill color (subspecies barrier in Fig. 2a), the only clear barrier to effective migration between subspecies is on the Z chromosome, where it is greater in strength than that between the long-tailed finch and the black-throated finch (Fig. 2b). These findings raise the following questions. What is the evidence for inversions, and why are they associated with the Z chromosome? Can we detect traits associated with reproductive isolation that may contribute to the establishment of inversions on the Z? If both bill color and genes on the Z chromosome affect reproductive isolation, why are their cline centers discordant? We consider each in turn.

If inversion differences are fixed between taxa, detecting them relies on the presence of large co-segregating blocks in late generation hybrids where recombination has had plenty of opportunity to break down any LD induced by migration. Here, we have only been able to study 20 admixed individuals, seven of which are likely to be F1s and the rest early backcrosses (Fig. S2). Despite this, we see evidence of recombination within them, with some areas of reduced LD when compared to the entire subspecies (Fig. 4). The remaining blocks may be regions where recombination has not broken down parental combinations in the samples we studied or reflect genuine regions of suppressed recombination. Evidence supports the idea that they do in fact reflect more than one inversion. First, maintaining a large block of differentiation in the face of extensive hybridization would require many incompatibilities, which would seem to reduce the fitness of hybrids more than has been observed in captive populations, where F1s readily are able to backcross (S.C.G *pers. obs.*). Second, two lines of direct evidence

support the presence of inversions. Previous cytological work has described the occurrence of at least three Z chromosome types, differentiated by pericentric and paracentric inversions, segregating within a mixed-ancestry captive population (Christidis 1986). We also observe LD patterns on the Z chromosome within each of the two subspecies, where we expect recombination to have destroyed large blocks (Fig. 4b). It is additionally difficult to explain correlated patterns of reduced genetic variation in *hecki* across this chromosome (Fig. S1) but nowhere else in the genome, unless it consists of a large non-recombining block. We therefore suggest that these LD patterns are consistent with two partially overlapping chromosomal inversion, as we describe in the next two paragraphs.

In subspecies *hecki*, we found LD between SNPs concentrated between 10.8 and 21.4 Mb, an area encompassing 68 genes, as well as between this 10.6 Mb large region and SNPs located around 68.4 Mb. An identical LD signature is present in the 10 whole-genome sequenced *hecki* individuals from pop. 15 (Fig. S4; Table S5). We suggest two explanations for this pattern. The first is that an inversion, 47-58 Mb in size, has been segregating in this subspecies for a sufficiently long time that double-crossovers and gene conversion have reduced LD except between the inversion breakpoints (Andolfatto et al., 2001; Stefansson et al., 2005; Bansal et al., 2007; Knief et al., 2016). The other is that the two regions are in close physical proximity to each other, and make up an inversion 11-12 Mb in size, and only appear to be displaced due to rearrangements between the long-tailed finch and the zebra finch reference. The SNPs producing this LD signal are ancestral and at low frequency. They were all found to be heterozygous in a single male individual, suggesting that the *hecki* inversion is near fixation at a frequency

of 0.93 in this population (i.e. 14 of 15 Z copies; Table S5). In the regions of high LD, we observe low values for Tajima's *D* in subspecies *hecki*. This is itself indicative of an excess of low frequency variants (Fig. 4c). The pattern is expected if an inversion had previously arisen, swept to high frequency, and the mutations that accumulated subsequently have remained at low frequency. In line with this expectation for a partial sweep, genetic diversity across the Z chromosome is substantially lower in *hecki* than in *acuticauda* and the black-throated finch (Fig. S1).

In subspecies *acuticauda*, we found evidence of long-range LD between SNPs concentrated around 11.6 and 63.0 Mb, but not for the majority of SNPs between these regions. There also appears to be LD between SNPs spanning the region from 68.2 and 68.9 Mb. A similar Z chromosome LD pattern was found in our analysis of 10 whole-genome sequenced *acuticauda* individuals from pops. 1 and 3 (Fig. S4; Table S5). Again, this pattern may be the result of a large inversion whose breakpoints are located at these coordinates, or physical displacement of LD blocks resulting from zebra finch reference mapping, or some combination. If a large inversion, it would encompass some 544 genes. We observe low values for Tajima's *D* and particularly low values for Fay and Wu's *H*, a result of an excess of high-frequency derived variants relative to intermediate frequency ones, in subspecies *acuticauda* between 11.6-21.4 Mb and around 68.4 Mb. This low *H* signature is potentially evidence of positive selection acting on the inversion to increase it, with the associated derived SNPs having risen to high frequency.

Our reliance on the zebra finch reference genome for mapping means that inferences about the exact dimensions and gene content of inversions in the long-tailed

finch must be taken with considerable caution, especially as the zebra finch and long-tailed finch Z chromosomes themselves differ by inversions (Hooper and Price, 2015; Singhal et al., 2015). Indeed, the mosaic distribution of SNPs in high LD within subspecies, and admixed individuals in particular, is suggestive of several fixed rearrangement differences between the zebra finch reference and any long-tailed finch Z chromosome (Fig. 4b). Together, however, it appears that 75% of the Z chromosome carriers one or more inversions, holding genes in tight linkage, which do not pass through the hybrid zone. The Z chromosome may preferentially accumulate inversions because of a greater structural mutation rate than the autosomes, and/or stronger selection pressures. If two alleles cause deleterious effects in hybrids, an inversion that reduces recombination between them gains a selective advantage (Kirkpatrick and Barton, 2006; Dagilis and Kirkpatrick, 2016; Ortiz-Barrientos et al., 2016). Hence, stronger selection is expected if the Z chromosome carries more deleterious alleles in hybrids than the autosomes. This may occur for multiple reasons, including exposure in the heterogametic sex and greater accumulation of sexually antagonistic alleles.

If such selection is the mechanism for inversion increase, it should eventually be possible to identify the traits and genes involved. One obvious set of candidates are those associated with bill color, which defines the textbook difference between the subspecies. The steep cline in bill color suggests that intermediates are selected against (Griffith and Hooper, 2015). We found three genomic regions that co-segregate with bill color (Figure 5). While one of these three regions is on the Z chromosome, the strongest candidates are on chromosome 8. One excellent candidate on chromosome 8 is associated with the SNP at 24.68 Mb, just 70kb downstream of the gene *CYP2J19*, a

member of the cytochrome P450 gene family recently shown to be associated with both the production and perception of red coloration in birds: the *yellowbeak* mutation in zebra finch (Mundy et al., 2016) and red plumage in hybrid red-factor canaries (Lopes et al., 2016). We describe color associated genes on chromosome 8 further in the supplemental material (see Fig. S7).

The third region co-segregating with bill color is on the Z chromosome and comprises 21 SNPs between 59.45-60.22Mb, a region that spans eight genes, two of which are candidates of interest: *TTC39B* and *BNC2*. The gene *TTC39B* functions in lipoprotein metabolism and regulation (Crooke et al., 2012; Koseki et al., 2014). Carotenoids are highly hydrophobic and require lipoprotein partnership in order to be transported to their tissue of deposition, so this gene could play a role in differential bill color expression by interfering in the movement of carotenoids (Sakudoh et al., 2013; Toews et al., 2017). The gene *BNC2* encodes a highly conserved zinc-finger protein and has previously been associated with variation in the saturation of skin color in humans, color stripe patterning in zebrafish, and coat color in mice (Jacobs et al., 2013; 2015; Lang et al., 2009; Smyth et al., 2006). What relationship *BNC2* might have with variation in carotenoid-based coloration, however, is less obvious as its previous associations appear to be primarily related to melanogenic coloration. In summary, the region co-segregating with bill color on the Z chromosome contains two reasonable candidates. Nevertheless, given much of the variation in bill color is autosomal, and displaced from the Z chromosome cline, it appears unlikely that genes associated with bill color are strong drivers of the inversions, although they may contribute if linked to other genes affecting hybrid fitness.

Bill color is not sexually dimorphic in the long-tailed finch (van Rooij and Griffith, 2012), but other traits, such as sperm, are sex-limited and are hence possible candidates for driving differentiation on the sex chromosomes. Notably, males of subspecies *acuticauda* have longer sperm with narrower head widths and a lower midpiece:flagellum ratio than males of subspecies *hecki* (Rowe et al., 2015). Across animals, sperm morphology evolution tends to co-evolve with and track the evolution of the female reproductive environment (Kleven et al., 2009; Higginson et al., 2012) and the strength of postmating prezygotic reproductive isolation increases between populations as these traits diverge (Miller and Pitnick, 2002). An inversion that suppressed recombination between genes involved in sperm morphology, the female reproductive environment, and/or other factors affecting hybrid fitness should be favored. Two recent studies in the zebra finch show that inversion polymorphisms on the Z chromosome are acting as a supergene for alternative sperm morphologies that result in substantial fitness variation across males (Kim et al., 2017; Knief et al., 2017). A similar situation may be at play in the long-tailed finch, with inversions on the Z chromosome involved in the observed sperm variation and nascent reproductive isolation between subspecies. We are now in the process of assessing this possibility.

As reproductive isolation develops between hybridizing forms, a point should be reached at which the genomes 'coalesce' and multiple traits affecting reproductive isolation become associated to complete the barrier to gene flow (Feder et al., 2012; 2013; Flaxman et al., 2013). Before that time, however, it is common to see different parts of the genome showing clines at different geographical locations and for some

parts of the genome to remain highly differentiated whereas other parts are homogenized (Brumfield et al., 2001; Baldassarre et al., 2014; Price and Hooper, 2016; reviewed in Harrison and Larson, 2014). In our case, genes underlying bill color in the long-tailed finch are largely decoupled from the Z chromosome. Both the clines for color and the Z chromosome are narrow, suggesting selection against hybrids (and against parental forms that move across each zone) is operating for both the Z chromosome and bill color. However, the centers of the clines are displaced by 350 km. While widespread, such displacement is surprising, given we expect clines to become trapped in regions of low hybrid fitness (Barton and Hewitt, 1985; Harrison, 1993; Harrison and Larson, 2014). Dissociation could be the result of differential selection on color, or the Z chromosome, or some combination (see Fig. 1).

Gene trees for the *CYP2J19B* recover *acuticauda* as a monophyletic clade (Fig. S7). Alongside the results of the population branch statistic (Fig. 5b), and analysis of the *CYP2J19B* amino acid sequence (Fig. S6), this suggests that the yellow bill form has arisen and been selectively favored, and these same selective forces resulted in movement east beyond the cline set in the Z chromosome (Fig. 1). For example, if individuals of both long-tailed finch subspecies favor yellower bills or yellow billed individuals are more aggressive and successful in securing mates, sexual selection for this trait could facilitate the adaptive introgression of "yellow alleles" out of the zone of initial secondary contact. However, based on the exponential shape parameters of the best-fitting cline model, Griffith and Hooper (2017) suggested that the red form was currently moving into western populations. This remains one possible alternative.

744 It is also worth considering the possibility that, although secondary contact
745 occurred where the bill color cline center is currently located, selection has caused the Z
746 chromosome cline to spread further to the west. Why would this occur? One possible
747 explanation is that the female reproductive tract of the western subspecies *acuticauda* is
748 more compatible with the shorter sperm from the eastern subspecies *hecki* than the
749 tract of *hecki* is compatible with the longer sperm of *acuticauda* (Rowe et al., 2015). If
750 genes affecting these traits are on the Z, then the Z chromosome should become
751 displaced from the bill color cline, even if both traits cause deleterious effects in the
752 hybrids and dispersal is completely symmetric. While Z chromosome movement may
753 appear less likely than bill color movement, we note that Z differentiation appears to be
754 largely a single non-recombining block, and the lack of differentiation of the autosomes
755 implies that there has been massive introgression of most of the genome between
756 subspecies. A model of gene flow of most of the genome has been proposed to explain
757 low genetic differentiation in the dark-eyed junco (*Junco hyemalis*) despite distinctive
758 color differences between subspecies (Price and Hooper, 2016). Alternatively, a possible
759 scenario is that the Z chromosome and bill color genes have been introduced from a now
760 extinct or undocumented population, with one spreading further than the other, while
761 the rest of the species' autosomal genomic variation remains essentially undifferentiated
762 across the range (Price and Hooper, 2016; Tuttle et al., 2016).
763
764 In conclusion, we find strong differentiation on the Z chromosome separating two
765 hybridizing subspecies. As noted by others, this is likely a consequence of the increased
766 expression of genes that drive reproductive isolation on this chromosome (Carling and
767 Brumfield 2008; Muirhead and Presgraves, 2016; Irwin 2018), but we also find the

likely presence of large inversions that, by creating linked blocks of genes, should limit gene flow at many other loci besides those causing reproductive incompatibilities. We suggest that the presence of inversions on the Z at a greater frequency than naively expected is because recombination suppressors between reproductive incompatibilities are generally expected to be favored. The contribution of inversions to large differentiation of the Z across hybrid zones may be more widespread than is currently appreciated.

Acknowledgments

We are grateful to the management teams of all stations listed in Table 1 for allowing us to work on their properties. Peri Bolton, Terry Burke, Kyle Kostrzewa, Dylan Meyer, Sarah Pryke, Nina Svedin, and Chris Wiley assisted with fieldwork in 2009, 2010, and 2015. Leo Joseph at CSIRO and Stanley Tang provided tissue and blood samples for black-throated finches used in this study, respectively. Kristen Wacker performed PCR sexing for long-tailed and black-throated finches. Molly Schumer and Julie Peng provided guidance with MSG protocol and sequencing optimization. B. Harr gave critical comments on the manuscript. This work was supported by Doctoral Dissertation Improvement Grant 1601323 to T.D.P. and D.M.H., National Geographic Young Explorer's Grant 9270-13 to D.M.H., funds from the Pritzker Laboratory for Molecular Systematics and Evolution at the Field Museum of Natural History to D.M.H., and Australian Research Council Grant DP0881019 to S.C.G.

Data Accessibility

Binary Alignment Map (BAM) files of genomic data from long-tailed finches and black-throated finches used in this study are available at www.ebi.ac.uk/ena/data/view/PRJEB28333.

Author Contributions

D.M.H. conceived the study, collected samples, designed and carried out molecular laboratory work, analyzed the data, and drafted the manuscript; S.C.G. conceived the study, procured funding, and collected samples; T.D.P. conceived the study, procured funding, and helped draft the manuscript. All authors gave final approval for publication.

Figures

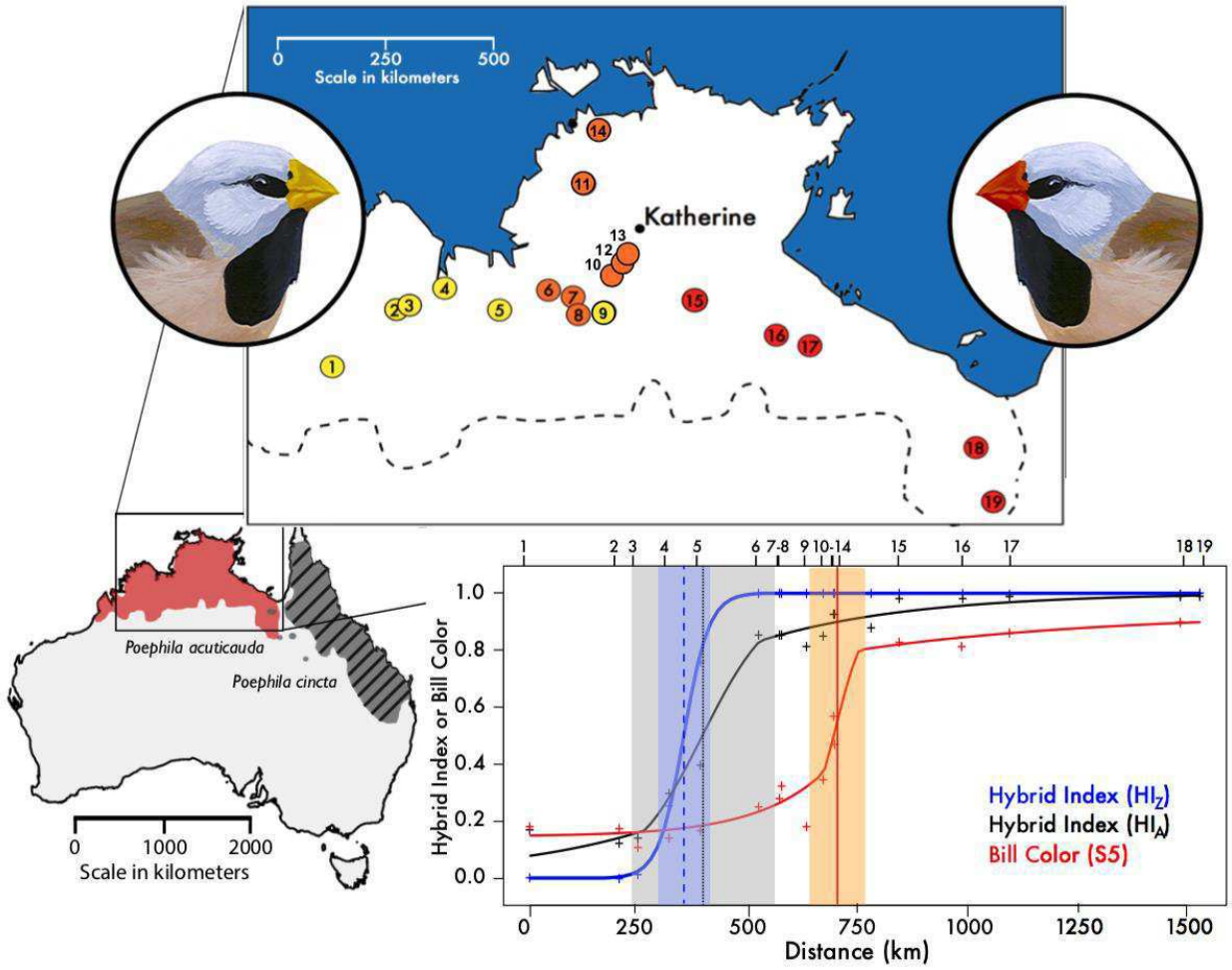

Figure 1 | Range maps for the long-tailed finch (*Poephila acuticauda*) and black-throated finch (*P. cincta*), population sampling, and best-fit hybrid index and bill color clines. Clockwise from bottom-left: Geographic range of the long-tailed finch (red) and black-throated finch (shaded black and gray), map inset with geographic locations of the 19 populations sampled in this study (see Table 1). Lower right, maximum likelihood HI_Z (blue), HI_A (black), and bill color (red) clines as estimated from HZAR (Derryberry et al., 2014). Distance reported in kilometers (km) from westernmost site (population 1, Mt. House) with greater values for hybrid index (Q) and bill chroma (S5) corresponding to a higher proportion of genetic ancestry from subspecies *hecki* and redder bills, respectively. The inferred centers and widths of the HI_Z, HI_A, and bill color clines are represented by vertical lines and color boxes in light blue, gray, and orange. Note that the autosomal cline fit was estimated using an autosomal hybrid index (HI_A) comprising only 33 SNPs and should be interpreted with due caution.

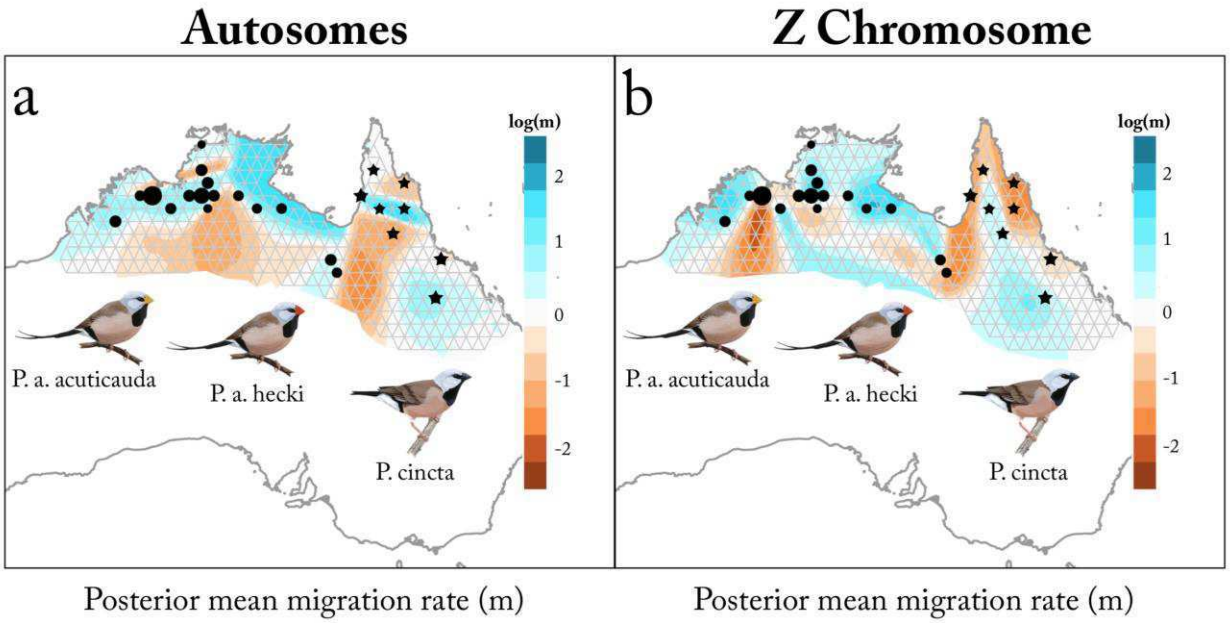

Figure 2 | Estimated effective migration surfaces (EEMS) for the autosomes (a) and Z chromosome (b) for the long-tailed finch and the black-throated finch combined. Sampled long-tailed finch and black-throated finch populations are represented by circles and stars, respectively. Migration rates (m) are color contoured on a $\log_{10}$ scale with cooler and warmer colors indicating migration rates greater than or less than the average expected under isolation-by-distance, respectively. A value of -1, for instance, indicates an effective migration rate tenfold slower than average.

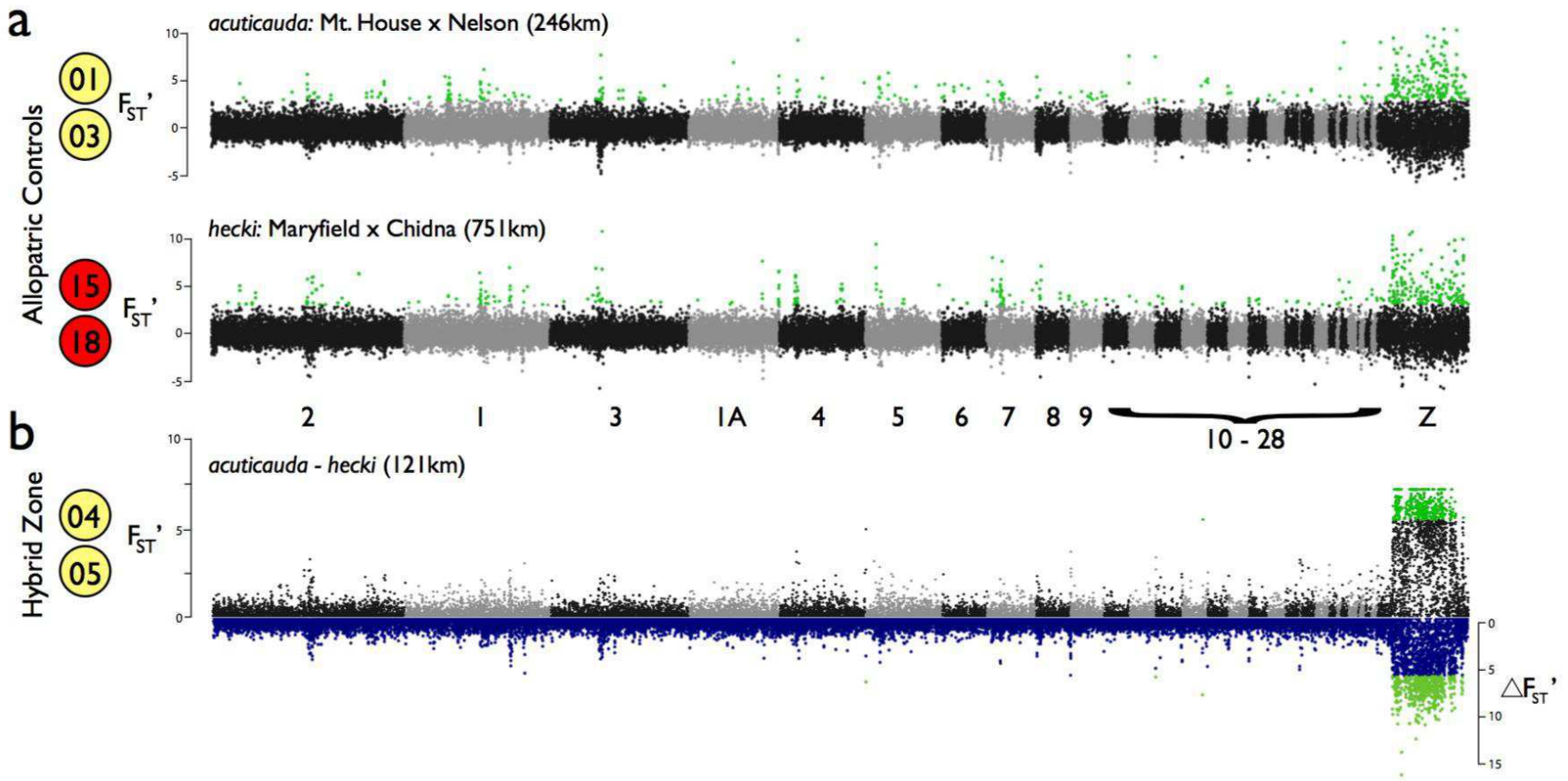

Figure 3 | Genomic differentiation (F_{ST}) between populations of the long-tailed finch. **a)** standardized genetic differentiation F_{ST}' calculated in 50kb sliding windows across the genome (colors alternate black and gray between chromosomes) between individuals from two control population comparisons of the same subspecies and bill color (see Table 1). The y-axis shows standard deviations of a normal distribution. Geographic distance between populations is given in kilometers. **b)** standardized genetic differentiation F_{ST}' (black and grey, positive axis) and net genetic differentiation ΔF_{ST}' (blue, mirrored to the negative axis) in 50kb sliding windows across the genome for our hybrid zone focal

comparison of $HI_Z < 0.1$ against $HI_Z > 0.9$ using individuals from populations 4 and 5. Genomic regions of extreme differentiation (>99th percentile) are shown in green for both F_{ST}' and ΔF_{ST}'.

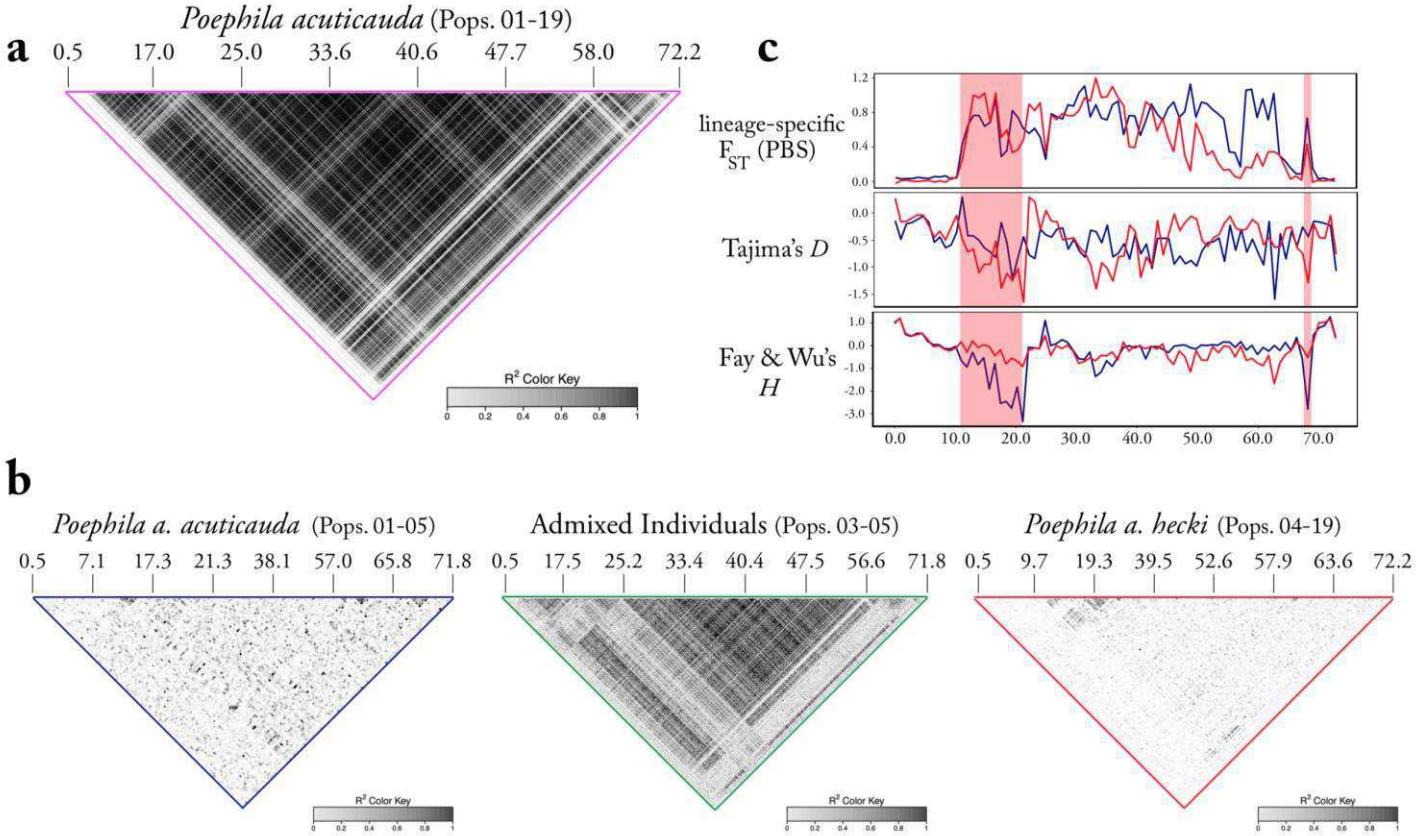

Figure 4 | Linkage disequilibrium (LD) and site-frequency spectrum summary statistics for the Z chromosome. **a)** pairwise LD patterns for all 293 long-tailed finches used in this study. Marker positions in Mb are given above all LD plots. **b)** pairwise LD calculated, from left to right, for individuals assigned, from left to right, by STRUCTURE to subspecies *acuticauda* (N = 53), admixed (N = 20), and subspecies *hecki* (N = 169). **c)** Summary statistics (PBS, Tajima's *D*, and Fay and Wu's *H*) calculated in 10kb sliding windows using 20 whole-genome sequenced long-tailed finches, 10 of each subspecies, and 10 zebra finches, used as outgroup. Results for subspecies *acuticauda* are shown in blue and results for

subspecies *hecki* are shown in red. The region of the putative inversion segregating in subspecies *hecki* is showcased with a pink box.

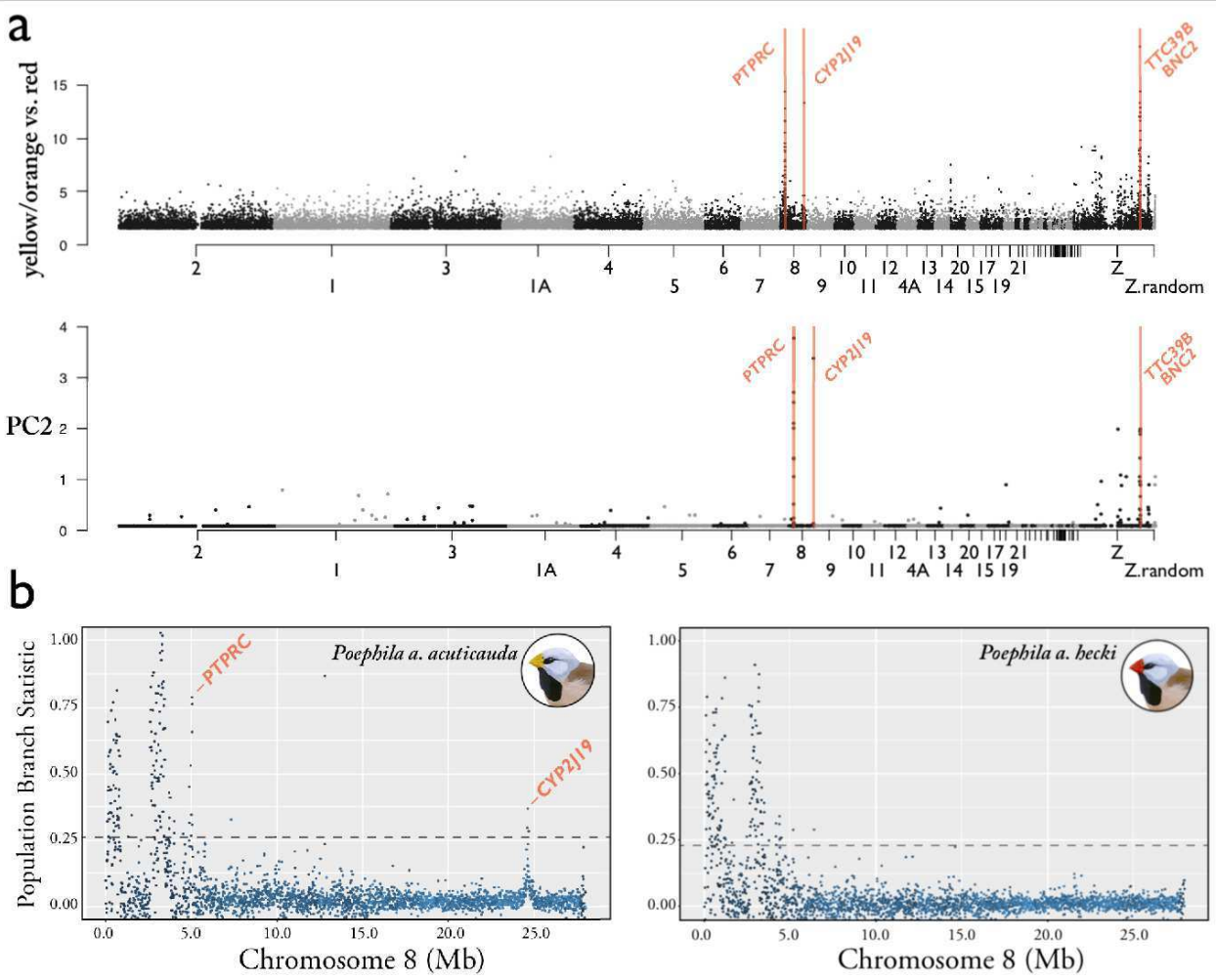

Figure 5 | a) Genome wide association mapping bill color in the long-tailed finch. Mapping performed using 192 genotyped long-tailed finches with associated bill color measure from populations 5-19 with Z chromosome characteristic of subspecies *hecki* (i.e. Q > 0.9). Chromosomes alternate in color between black and gray in order of descending size. Data points represent false-discovery rate-adjusted *P* values for a total of 500k SNPs on the y-axis (Benjamini and Hochberg, 1995). Bill color was evaluated across individuals in two ways: as a dichotomous variable (0 = pure *acuticauda* yellow and admixed orange or 1 = pure *hecki* red) and as PC2 from a principal components analysis of reflectance variation between 300-750 nm (31.3% variation). See Griffith and Hooper (2017) for details of scoring color. Vertical orange bars highlight three regions on chromosome 8 and Z that are enriched for SNPs significantly associated with bill color in all three comparisons – candidate genes in these regions are labeled. **b)** results for population branch statistic analyses for chromosome 8 in subspecies *acuticauda* (left) and *hecki* (right). Population branch statistic peaks found exclusively in subspecies *acuticauda* are indicative of recent selection acting at our two candidate bill color loci.

867 # Tables

868

No.	Population	Latitude	Longitude	Distance (km)	Ancestry	Bill Chroma (S5)	Hybrid Index (HIz)	Bill Color *N*	Genetic *N*
1	Mt. House	17.1°S	125.6°E	0	*acuticauda*	0.18 ± 0.05	0.005	89	19
2	Ellenbrae	16.0°S	127.1°E	205	*acuticauda*	0.18 ± 0.04	0.004	12	14
3	Nelson	15.8°S	127.5°E	246	*acuticauda*	0.11 ± 0.05	0.025	48	21
4	Wyndham	15.6°S	128.1°E	317	Admixed	0.15 ± 0.08	0.276	79	27
5	Newry	16.1°S	129.1°E	389	Admixed	0.17 ± 0.05	0.746	16	15
6	Auvergne	15.6°S	130.3°E	523	*hecki*	0.25 ± 0.08	0.999	92	20
7	Coolibah	15.8°S	130.8°E	571	*hecki*	0.28 ± 0.06	1.000	68	18
8	Kidman Springs	16.1°S	130.9°E	574	*hecki*	0.32 ± 0.05	0.999	18	17
9	Moolooloo	16.1°S	131.4°E	632	*hecki*	0.18 ± 0.05	1.000	6	8
10	Willeroo	15.3°S	131.6°E	670	*hecki*	0.35 ± 0.08	1.000	33	18
11	Tipperary	13.6°S	131.0°E	695	*hecki*	0.57 ± 0.08	0.999	55	19
12	Stapleton – Mathison	15.1°S	131.8°E	697	*hecki*	0.47 ± 0.12	0.999	59	19
13	Stapleton – Bullyard	14.9°S	131.9°E	710	*hecki*	0.46 ± 0.09	N/A	6	0
14	Koolpinyah	12.7°S	131.2°E	779	*hecki*	N/A	1.000	0	6
15	Maryfield	15.8°S	133.4°E	844	*hecki*	0.83 ± 0.05	1.000	27	15
16	October Creek	16.6°S	134.9°E	988	*hecki*	0.82 ± 0.05	1.000	27	13
17	McArthur River	16.5°S	135.9°E	1096	*hecki*	0.86 ± 0.05	0.997	32	14
18	Chidna	19.4°S	139.4°E	1486	*hecki*	0.90 ± 0.04	0.999	70	17
19	Leichardt	20.3°S	139.7°E	1530	*hecki*	N/A	0.999	0	13
	Total							**737**	**293**

869

Table 1 | Sampling transect across the range of the long-tailed finch. Sampling locations (from west to east, with distance from Mt. House), and sample sizes (individuals with spectrometric data and genetic data), as well as average and standard deviation for bill chroma (S_5), between 450-650nm, and hybrid index (HI_Z), respectively.

References

Andolfatto, P., Depaulis, F., & Navarro, A. (2001). Inversion polymorphisms and nucleotide variability in Drosophila. *Genetics Research*, *77*(1), 1.

Andolfatto, P., Davison, D., Erezyilmaz, D., Hu, T. T., Mast, J., Sunayama-Morita, T., & Stern, D. L. (2011). Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome research*, *21*(4), 610-617.

Backström, N., Forstmeier, W., Schielzeth, H., Mellenius, H., Nam, K., Bolund, E., ... & Ellegren, H. (2010). The recombination landscape of the zebra finch Taeniopygia guttata genome. *Genome research*, *20*(4), 485-495.

Baldassarre, D. T., White, T. A., Karubian, J., & Webster, M. S. (2014). Genomic and morphological analysis of a semipermeable avian hybrid zone suggests asymmetrical introgression of a sexual signal. *Evolution*, *68*(9), 2644-2657.

Bansal, V., Bashir, A., & Bafna, V. (2007). Evidence for large inversion polymorphisms in the human genome from HapMap data. *Genome research*, *17*(2), 219-230.

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual review of Ecology and Systematics*, *16*(1), 113-148.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.

Brown, K. M., Burk, L. M., Henagan, L. M., & Noor, M. A. (2004). A test of the chromosomal rearrangement model of speciation in Drosophila pseudoobscura. *Evolution*, *58*(8), 1856-1860.

Brumfield, R. T., Jernigan, R. W., McDonald, D. B., & Braun, M. J. (2001). Evolutionary implications of divergent clines in an avian (Manacus: Aves) hybrid zone. *Evolution*, *55*(10), 2070-2087.

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., ... & Hogner, S. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. *Genome research*, *25*(11), 1656-1665.

Carling, M. D., & Brumfield, R. T. (2008). Haldane's rule in an avian system: using cline theory and divergence population genetics to test for differential introgression of mitochondrial, autosomal, and sex- linked loci across the passerina bunting hybrid zone. *Evolution*, *62*(10), 2600-2615.

Charlesworth, B., & Charlesworth, D. (1973). Selection of new inversions in multi-locus genetic systems. *Genetics Research*, *21*(2), 167-183.

Charlesworth, D., & Charlesworth, B. (1979). Selection on recombination in clines. *Genetics*, *91*(3), 581-589.

Charlesworth, B., Coyne, J. A., & Barton, N. H. (1987). The relative rates of evolution of sex chromosomes and autosomes. *The American Naturalist*, *130*(1), 113-146.

Christidis, L. (1986). Chromosomal evolution within the family Estrildidae (Aves) I. The Poephilae. *Genetica*, *71*(2), 81-97.

Clark, A. G. (1997). Neutral behavior of shared polymorphism. *Proceedings of the National Academy of Sciences*, *94*(15), 7730-7734.

Coyne, J. A. (1992). Genetics and speciation. *Nature*, *355*(6360), 511.

Coyne, J. A., & Orr, H. A. (1989). Patterns of speciation in Drosophila. *Evolution*, *43*(2), 362-381.

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sunderland, MA.

Crooke, R. M., Graham, M., & Bell, T. A. (2012). *U.S. Patent Application No. 13/498,055*.

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

Dagilis, A. J., & Kirkpatrick, M. (2016). Prezygotic isolation, mating preferences, and the evolution of chromosomal inversions. *Evolution*, *70*(7), 1465-1472.

Dean, R., Harrison, P. W., Wright, A. E., Zimmer, F., & Mank, J. E. (2015). Positive selection underlies faster-Z evolution of gene expression in birds. *Molecular biology and evolution*, *32*(10), 2646-2656.

Derryberry, E. P., Derryberry, G. E., Maley, J. M., & Brumfield, R. T. (2014). HZAR: hybrid zone analysis using an R software package. *Molecular ecology resources*, *14*(3), 652-663.

Ellegren, H. (2013). The evolutionary genomics of birds. *Annual review of ecology, evolution, and systematics*, *44*, 239-259.

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, *164*(4), 1567-1587.

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

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, *28*(7), 342-350.

Feder, J. L., Flaxman, S. M., Egan, S. P., Comeault, A. A., & Nosil, P. (2013). Geographic mode of speciation and genomic divergence. *Annual Review of Ecology, Evolution, and Systematics*, *44*, 73-97.

Fishman, L., Stathos, A., Beardsley, P. M., Williams, C. F., & Hill, J. P. (2013). Chromosomal rearrangements and the genetics of reproductive barriers in Mimulus (monkey flowers). *Evolution*, *67*(9), 2547-2560.

Flaxman, S. M., Feder, J. L., & Nosil, P. (2013). Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. *Evolution*, *67*(9), 2577-2591.

Fridolfsson, A. K., & Ellegren, H. (1999). A simple and universal method for molecular sexing of non-ratite birds. *Journal of avian biology*, 116-121.

Fuller, Z., Leonard, C., Young, R., Schaeffer, S., & Phadnis, N. (2017). The role of chromosomal inversions in speciation. *bioRxiv*, 211771.

Gay, L., Crochet, P. A., Bell, D. A., & Lenormand, T. (2008). Comparing clines on molecular and phenotypic traits in hybrid zones: a window on tension zone models. *Evolution*, *62*(11), 2789-2806.

Gowen, F. C., Maley, J. M., Cicero, C., Peterson, A. T., Faircloth, B. C., Warr, T. C., & McCormack, J. E. (2014). Speciation in Western Scrub-Jays, Haldane's rule, and genetic clines in secondary contact. *BMC evolutionary biology*, *14*(1), 135.

Griffith, S. C., & Hooper, D. M. (2017). Geographical variation in bill colour in the Long-tailed Finch: evidence for a narrow zone of admixture between sub-species. *Emu-Austral Ornithology*, *117*(2), 141-150.

Harrison, R. G. (1986). Pattern and process in a narrow hybrid zone. *Heredity*, *56*(3), 337.

Harrison, R. G. (1990). Hybrid zones: Windows on evolutionary process. *Oxford surveys in evolutionary biology*, *7*, 69-128.

Harrison, R. G. (Ed.). (1993). *Hybrid zones and the evolutionary process*. Oxford University Press on Demand.

Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, *105*(S1), 795-809.

Higginson, D. M., Miller, K. B., Segraves, K. A., & Pitnick, S. (2012). Female reproductive tract form drives the evolution of complex sperm morphology. *Proceedings of the National Academy of Sciences*, *109*(12), 4538-4543.

Hoffmann, A. A., & Rieseberg, L. H. (2008). Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation?. *Annual review of ecology, evolution, and systematics*, *39*, 21-42.

Hooper, D. M., & Price, T. D. (2015). Rates of karyotypic evolution in Estrildid finches differ between island and continental clades. *Evolution*, *69*(4), 890-903.

Hooper, D. M., & Price, T. D. (2017). Chromosomal inversion differences correlate with range overlap in passerine birds. Nature Ecology & Evolution 1: 1526–1534.

Huerta-Sánchez, E., DeGiorgio, M., Pagani, L., Tarekegn, A., Ekong, R., Antao, T., ... & Weale, M. E. (2013). Genetic signatures reveal high-altitude adaptation in a set of Ethiopian populations. *Molecular biology and evolution*, *30*(8), 1877-1888.

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

Irwin, D. E. (2018). Sex chromosomes and speciation in birds and other ZW systems. *Molecular ecology*.

Jacobs, L. C., Wollstein, A., Lao, O., Hofman, A., Klaver, C. C., Uitterlinden, A. G., ... & Liu, F. (2013). Comprehensive candidate gene study highlights UGT1A and BNC2 as new genes determining continuous skin color variation in Europeans. *Human genetics*, *132*(2), 147-158.

Jacobs, L. C., Hamer, M. A., Gunn, D. A., Deelen, J., Lall, J. S., Van Heemst, D., ... & Beekman, M. (2015). A genome-wide association study identifies the skin color genes IRF4, MC1R, ASIP, and BNC2 influencing facial pigmented spots. *Journal of Investigative Dermatology*, *135*(7), 1735-1742.

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*(14), 1801-1806.

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., ... & Birney, E. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*(7392), 55.

Kim, K. W., Bennison, C., Hemmings, N., Brookes, L., Hurley, L. L., Griffith, S. C., ... & Slate, J. (2017). A sex-linked supergene controls sperm morphology and swimming speed in a songbird. *Nature ecology & evolution*, *1*(8), 1168.

Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, *173*(1), 419-434.

Kirkpatrick, M. (2010). How and why chromosome inversions evolve. *PLoS biology*, *8*(9), e1000501.

Kirubakaran, T. G., Grove, H., Kent, M. P., Sandve, S. R., Baranski, M., Nome, T., ... & Sonesson, A. (2016). Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Molecular ecology*, *25*(10), 2130-2143.

Kleven, O., Fossøy, F., Laskemoen, T., Robertson, R. J., Rudolfsen, G., & Lifjeld, J. T. (2009). Comparative evidence for the evolution of sperm swimming speed by sperm competition and female sperm storage duration in passerine birds. *Evolution*, *63*(9), 2466-2473.

Knief, U., Hemmrich-Stanisak, G., Wittig, M., Franke, A., Griffith, S. C., Kempenaers, B., & Forstmeier, W. (2016). Fitness consequences of polymorphic inversions in the zebra finch genome. *Genome biology*, *17*(1), 199.

Knief, U., Forstmeier, W., Pei, Y., Ihle, M., Wang, D., Martin, K., ... & Albrecht, T. (2017). A sex-chromosome inversion causes strong overdominance for sperm traits that affect siring success. *Nature ecology & evolution*, *1*(8), 1177.

Koseki, M., Hsieh, J., Yakushiji, E., Welch, C., Iqbal, J., Hussain, M. M., ... & Tall, A. R. (2014). TTC39B deficiency promotes HDL production and impairs non-hdl absorption in small intestine. *Atherosclerosis*, *235*(2), e104.

Küpper, C., Stocks, M., Risse, J. E., dos Remedios, N., Farrell, L. L., McRae, S. B., ... & Kitaysky, A. S. (2016). A supergene determines highly divergent male reproductive morphs in the ruff. *Nature genetics*, *48*(1), 79.

Lamichhaney, S., Fan, G., Widemo, F., Gunnarsson, U., Thalmann, D. S., Hoeppner, M. P., ... & Chen, W. (2016). Structural genomic changes underlie alternative reproductive strategies in the ruff (Philomachus pugnax). *Nature Genetics*, *48*(1), 84.

Lang, M. R., Patterson, L. B., Gordon, T. N., Johnson, S. L., & Parichy, D. M. (2009). Basonuclin-2 requirements for zebrafish adult pigment pattern development and female fertility. *PLoS genetics*, *5*(11), e1000744.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Lopes, R. J., Johnson, J. D., Toomey, M. B., Ferreira, M. S., Araujo, P. M., Melo-Ferreira, J., ... & Carneiro, M. (2016). Genetic basis for red coloration in birds. *Current Biology*, *26*(11), 1427-1434.

Lowry, D. B., & Willis, J. H. (2010). A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS biology*, *8*(9), e1000500.

Mank, J. E., Vicoso, B., Berlin, S., & Charlesworth, B. (2010). Effective population size and the Faster- X Effect: empirical results and their interpretation. *Evolution*, *64*(3), 663-674.

Masly, J. P., & Presgraves, D. C. (2007). High-resolution genome-wide dissection of the two rules of speciation in Drosophila. *PLoS biology*, *5*(9), e243.

McGaugh, S. E., & Noor, M. A. (2012). Genomic impacts of chromosomal inversions in parapatric Drosophila species. *Phil. Trans. R. Soc. B*, *367*(1587), 422-429.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, *20*(9), 1297-1303.

Miller, G. T., & Pitnick, S. (2002). Sperm-female coevolution in Drosophila. *Science*, *298*(5596), 1230-1233.

Miller, P. M., Gavrilets, S., & Rice, W. R. (2006). Sexual conflict via maternal-effect genes in ZW species. *Science*, *312*(5770), 73-73.

Mundy, N. I., Stapley, J., Bennison, C., Tucker, R., Twyman, H., Kim, K. W., ... & Slate, J. (2016). Red carotenoid coloration in the zebra finch is controlled by a cytochrome P450 gene cluster. *Current Biology*, *26*(11), 1435-1440.

Muirhead, C. A., & Presgraves, D. C. (2016). Hybrid incompatibilities, local adaptation, and the genomic distribution of natural introgression between species. *The American Naturalist*, *187*(2), 249-261.

Navarro, A., & Barton, N. H. (2003). Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution*, *57*(3), 447-459.

Noor, M. A., Grams, K. L., Bertucci, L. A., & Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences*, *98*(21), 12084-12088.

Ortiz-Barrientos, D., Engelstädter, J., & Rieseberg, L. H. (2016). Recombination rate evolution and the origin of species. *Trends in ecology & evolution*, *31*(3), 226-236.

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular ecology*, *25*(11), 2337-2360.

Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature genetics*, *48*(1), 94.

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

Presgraves, D. C. (2008). Sex chromosomes and speciation in Drosophila. *Trends in Genetics*, *24*(7), 336-343.

Presgraves, D. C. (2010). The molecular evolutionary basis of species formation. *Nature Reviews Genetics*, *11*(3), 175.

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

Price, T. D., & Hooper, D. M. (2016). The Potential Role of Parapatric and Alloparapatric Divergence in Junco Speciation. *Snowbird: Integrative Biology and Evolutionary Diversity in the Junco*, 199.

Pritchard, J. K., Stephens, M., Rosenberg, N. A., & Donnelly, P. (2000). Association mapping in structured populations. *The American Journal of Human Genetics*, *67*(1), 170-181.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559-575.

Rowe, M., Griffith, S. C., Hofgaard, A., & Lifjeld, J. T. (2015). Subspecific variation in sperm morphology and performance in the Long-tailed Finch (Poephila acuticauda). *Avian Research*, *6*(1), 23.

Rutkowska, J., & Badyaev, A. V. (2008). Meiotic drive and sex determination: molecular and cytological mechanisms of sex ratio adjustment in birds. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1497), 1675-1686.

Sakudoh, T., Kuwazaki, S., Iizuka, T., Narukawa, J., Yamamoto, K., Uchino, K., ... & Tsuchida, K. (2013). CD36 homolog divergence is responsible for the selectivity of carotenoid species migration to the silk gland of the silkworm Bombyx mori. *Journal of lipid research*, *54*(2), 482-495.

Schilthuizen, M., Giesbers, M. C. W. G., & Beukeboom, L. W. (2011). Haldane's rule in the 21st century. *Heredity*, *107*(2), 95.

Schumer, M., Cui, R., Powell, D. L., Dresner, R., Rosenthal, G. G., & Andolfatto, P. (2014). High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. *Elife*, *3*.

Singhal, S., & Moritz, C. (2013). Reproductive isolation between phylogeographic lineages scales with divergence. *Proc. R. Soc. B*, *280*(1772), 20132246.

Singhal, S., Leffler, E. M., Sannareddy, K., Turner, I., Venn, O., Hooper, D. M., ... & Griffith, S. C. (2015). Stable recombination hotspots in birds. *Science*, *350*(6263), 928-932.

Smit, AFA, Hubley, R & Green, P. *RepeatMasker Open-3.0*. 1996-2010 <http://www.repeatmasker.org>.

Smyth, I. M., Wilming, L., Lee, A. W., Taylor, M. S., Gautier, P., Barlow, K., ... & Pelan, S. (2006). Genomic anatomy of the Tyrp1 (brown) deletion complex. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3704-3709.

Stefansson, H., Helgason, A., Thorleifsson, G., Steinthorsdottir, V., Masson, G., Barnard, J., ... & Desnica, N. (2005). A common inversion under selection in Europeans. *Nature genetics*, *37*(2), 129.

Storchova, R., Reif, J., & Nachman, M. W. (2010). Female heterogamety and speciation: reduced introgression of the Z chromosome between two species of nightingales. *Evolution*, 64, 456-471.

Stryjewski, K. F., & Sorenson, M. D. (2017). Mosaic genome evolution in a recent and rapid avian radiation. *Nature ecology & evolution*, *1*(12), 1912.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585-595.

Toews, D. P., Campagna, L., Taylor, S. A., Balakrishnan, C. N., Baldassarre, D. T., Deane-Coe, P. E., ... & Mason, N. A. (2015). Genomic approaches to understanding population divergence and speciation in birds. *The Auk*, *133*(1), 13-30.

Toews, D. P., Hofmeister, N. R., & Taylor, S. A. (2017). The evolution and genetics of carotenoid processing in animals. *Trends in Genetics*, *33*(3), 171-182.

Turelli, M., & Orr, H. A. (1995). The dominance theory of Haldane's rule. *Genetics*, *140*(1), 389-402.

Turelli, M., & Orr, H. A. (2000). Dominance, epistasis and the genetics of postzygotic isolation. *Genetics*, *154*(4), 1663-1679.

Tuttle, E. M., Bergland, A. O., Korody, M. L., Brewer, M. S., Newhouse, D. J., Minx, P., ... & Gonser, R. A. (2016). Divergence and functional degradation of a sex chromosome-like supergene. *Current Biology*, *26*(3), 344-350.

Twyman, H., Andersson, S., & Mundy, N. I. (2018). Evolution of CYP2J19, a gene involved in colour vision and red coloration in birds: positive selection in the face of conservation and pleiotropy. *BMC evolutionary biology*, *18*(1), 22.

Van Rooij, E. P., & Griffith, S. C. (2012). No evidence of assortative mating on the basis of putative ornamental traits in Long- tailed Finches Poephila acuticauda. *Ibis*, *154*(3), 444-451.

Vijay, N., Bossu, C. M., Poelstra, J. W., Weissensteiner, M. H., Suh, A., Kryukov, A. P., & Wolf, J. B. (2016). Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature communications*, *7*, 13195.

Wu, C. I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, *14*(6), 851-865.

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X. P., Pool, J. E., ... & Zheng, H. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, *329*(5987), 75-78.

Zann, R. A. (1976). Variation in the songs of three species of estrildine grassfinches. *Emu*, *76*(3), 97-108.