

# The Genomic Landscape of Divergence Across the Speciation Continuum in Island-Colonising Silvereyes (*Zosterops lateralis*)

Ashley T. Sendell-Price,<sup>\*,1</sup> Kristen C. Ruegg,<sup>\*,1,\*</sup> Eric C. Anderson,<sup>§</sup> Claudio S. Quilodrán,<sup>\*</sup> Benjamin M. Van Doren,<sup>\*</sup> Vinh L. Underwood,<sup>†</sup> Tim Coulson,<sup>\*</sup> and Sonya M. Clegg<sup>\*,\*\*</sup>

<sup>\*</sup>Department of Zoology, University of Oxford, United Kingdom, <sup>†</sup>Center for Tropical Research, Institute of the Environment and Sustainability, University of California Los Angeles, California, <sup>‡</sup>Department of Biology, Colorado State University, Fort Collins, Colorado, <sup>§</sup>Fisheries Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, Santa Cruz, California, and <sup>\*\*</sup>Environmental Futures Research Institute, Griffith University, Queensland, Australia  
ORCID IDs: 0000-0002-1227-8929 (A.T.S.-P.); 0000-0001-5579-941X (K.C.R.); 0000-0001-7197-9154 (C.S.Q.); 0000-0002-7355-6005 (B.M.V.D.); 0000-0002-2526-1851 (V.L.U.); 0000-0001-9371-9003 (T.C.); 0000-0002-3092-3864 (S.M.C.)

**ABSTRACT** Inferring the evolutionary dynamics at play during the process of speciation by analyzing the genomic landscape of divergence is a major pursuit in population genomics. However, empirical assessments of genomic landscapes under varying evolutionary scenarios that are known *a priori* are few, thereby limiting our ability to achieve this goal. Here we combine RAD-sequencing and individual-based simulations to evaluate the genomic landscape of divergence in the silvereye (*Zosterops lateralis*). Using pairwise comparisons that differ in divergence timeframe and the presence or absence of gene flow, we document how genomic patterns accumulate along the speciation continuum. In contrast to previous predictions, our results provide limited support for the idea that divergence accumulates around loci under divergent selection or that genomic islands widen with time. While a small number of genomic islands were found in populations diverging with and without gene flow, in few cases were SNPs putatively under selection tightly associated with genomic islands. The transition from localized to genome-wide levels of divergence was captured using individual-based simulations that considered only neutral processes. Our results challenge the ubiquity of existing verbal models that explain the accumulation of genomic differences across the speciation continuum and instead support the idea that divergence both within and outside of genomic islands is important during the speciation process.

## KEYWORDS

Genomic heterogeneity  
Gene flow  
Genomic islands  
Genomic valleys  
Simulated divergence  
RAD sequencing  
Genomic landscape

Darwin described the process of speciation as a continuum in which differentiation between populations accumulates, first giving rise to well-marked ‘varieties’ or ‘races,’ and in some instances culminating in the formation of new species (Darwin 1859), a view that today is

widely accepted (Mallet 2008; Nosil 2012; Powell *et al.* 2013). Numerous evolutionary processes can promote, stall or reverse trajectory along the speciation continuum, including gene flow, selection, recombination, mutation and drift (Ricklefs and Bermingham 2007; Ravinet *et al.* 2017; Kearns *et al.* 2018). Understanding how these processes interact at different stages and how this interaction shapes the speciation process remains a central but challenging goal in evolutionary biology (Marie Curie SPECIATION Network *et al.* 2012; Ravinet *et al.* 2017). Examining divergence at the level of the genome is a potentially powerful way to address this challenge (Nosil and Feder 2012; Seehausen *et al.* 2014).

During the divergence process, some regions of the genome diverge rapidly and others more slowly (Seehausen *et al.* 2014). This variation in the tempo of divergence across the genome results in a heterogeneous genomic landscape (Nosil *et al.* 2009; Andrew and

Copyright © 2020 Sendell-Price *et al.*

doi: <https://doi.org/10.1534/g3.120.401352>

Manuscript received May 3, 2020; accepted for publication July 6, 2020; published Early Online July 13, 2020.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material available at figshare: <https://doi.org/10.25387/g3.12482000>.

<sup>1</sup>Corresponding author: Department of Zoology, University of Oxford, Zoology Research and Administration Building, 11a Mansfield Road, Oxford, OX1 3SZ, United Kingdom. E-mail: [ashley.sendell-price@zoo.ox.ac.uk](mailto:ashley.sendell-price@zoo.ox.ac.uk)

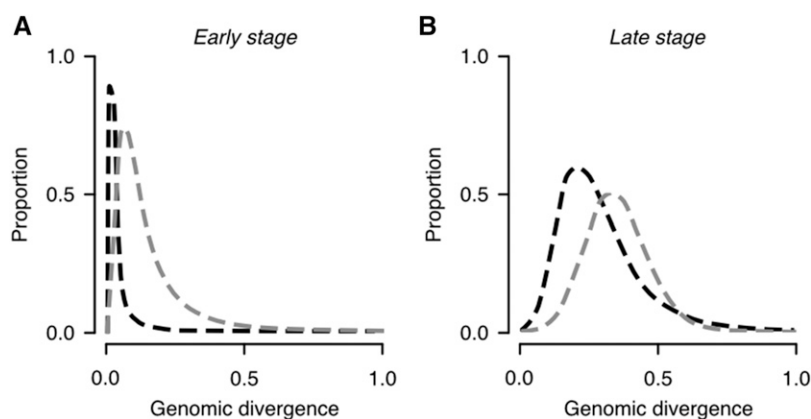
Rieseberg 2013; Feder *et al.* 2013; Seehausen *et al.* 2014; Feulner *et al.* 2015; Ravinet *et al.* 2017). During the initial stages of the speciation continuum, divergence is expected to be localized to regions of the genome where loci are under strong divergent selection, forming peaks of divergence often referred to as ‘genomic islands of divergence’ (Feder *et al.* 2013; Seehausen *et al.* 2014). As populations move along the speciation continuum, genomic islands are predicted to widen as linkage disequilibrium facilitates divergence of neutral and weakly selected loci via divergence hitchhiking (Nosil *et al.* 2009; Feder and Nosil 2010). The growth of genomic islands of divergence forms the basis of the divergence hitchhiking model of speciation (Via 2012), which until recently was the prevailing mechanism through which genome-wide levels of divergence were thought to be achieved (Smadja *et al.* 2008; Via and West 2008). However, the conditions under which divergence hitchhiking can generate large regions of differentiation are limited, requiring small effective population sizes, low rates of migration, and strong selection (Feder and Nosil 2010).

Using closely related taxa at different stages of divergence as a proxy for the speciation continuum, a number of studies have provided evidence that the pattern of genomic divergence accumulates in a way consistent with the mechanisms described above (Martin *et al.* 2013; Feulner *et al.* 2015; Supple *et al.* 2015; Vijay *et al.* 2016). However, the pace at which divergence accumulates could be accelerated or hindered by other evolutionary processes. For example, gene flow could slow the rate at which divergence accumulates by having a homogenizing effect at neutrally evolving or weakly selected loci (Nosil *et al.* 2017). In contrast, in genetic isolation, divergence proceeds via selection and/or drift, unfettered by this homogenizing effect. As such, these different modes of divergence are expected to affect the distribution of genetic divergence values in predictable ways, accounting for the stage of the speciation continuum (Feder *et al.* 2012) (see Figure 1). To date, comparisons of genomic divergence between races of *Heliconius* butterflies distributed in allopatry *vs.* those in parapatry (Martin *et al.* 2013) provide the strongest empirical basis for the likely role of gene flow in shaping the genomic landscape. However, studies are needed with comparisons matched for divergence timeframe and with rates of gene flow inferred from more than geographic distribution alone.

Genomic valleys – highly conserved regions where differentiation falls far below background levels – are also a common feature of the genomic landscape. While these have been identified in a range of species (Hofer *et al.* 2012; Roesti *et al.* 2012; Wang *et al.* 2016; Van Doren *et al.* 2017), their role in shaping the genomic landscape has

been overlooked in comparison to the focus on genomic islands. Genomic valleys may play an important role in maintaining genomic heterogeneity by slowing the approach to genome-wide divergence. Genomic valleys may occur because of alleles favored in both populations (parallel adaptation) (Nielsen 2005; Roesti *et al.* 2012), or they may result from purifying selection in which newly arising deleterious mutations are selected against in order to preserve biological function (Cvijović *et al.* 2018). Alternatively, genomic valleys may occur due to incomplete lineage sorting at neutral loci, where diverging populations share alleles for some time, manifesting as regions of below background level divergence (Stölting *et al.* 2013). Should parallel adaptation be important in their formation, we would expect the position of genomic valleys to closely correspond to the position of loci under parallel selection. Further, as parallel selection leaves a distinct signature of reduced diversity, where genomic islands are the product of parallel selection, between population diversity ( $d_{xy}$ ) would be expected to be reduced (Roesti *et al.* 2014). A similar pattern of reduced  $F_{st}$  and reduced  $d_{xy}$  is expected under purifying selection (Cvijović *et al.* 2018). Where genomic valleys are the product of incomplete lineage sorting at neutral loci, they are expected to be more numerous during the early stage of divergence as shared alleles become less numerous over time. Under these proposed mechanisms, a reasonable expectation of the temporal dynamics of genomic valleys is that they decrease in size over longer divergence timeframes, as genomic valleys are broken down by recombination and loosely linked/neutral loci diverge (Charlesworth 2006). This expectation is yet to be tested empirically.

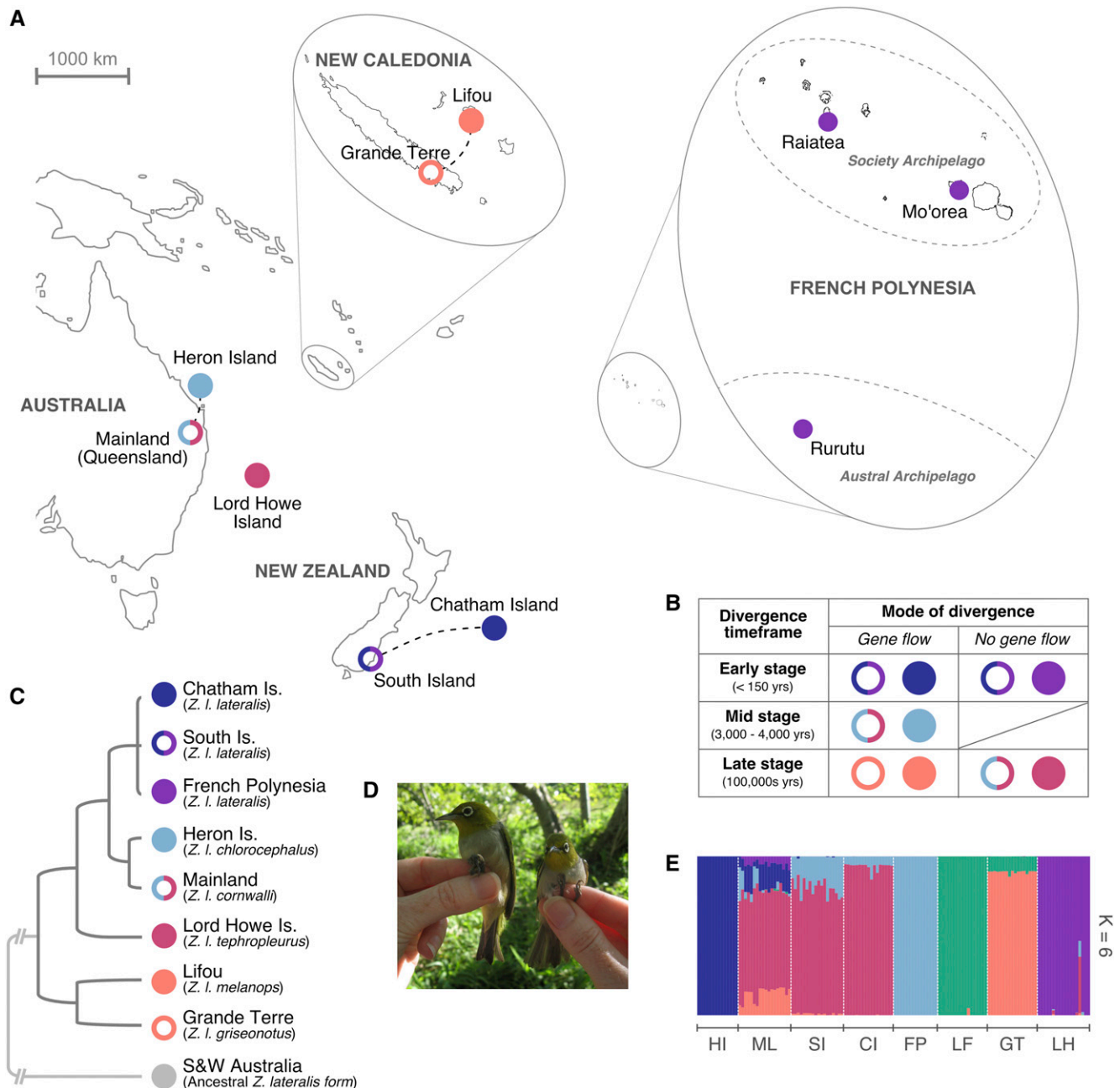
A more nuanced understanding of how genome-wide divergence develops from the heterogeneous genomic landscape requires simultaneous consideration of the stage along the speciation continuum and the gene flow context of divergence for systems that have well characterized population histories. Members of the silveryeye subspecies complex (*Zosterops lateralis*) of Australia and southwest Pacific islands offer an exceptional opportunity to explore patterns of divergence across different timescales and gene flow contexts. The species has repeatedly colonized islands in the region since its origin on the Australian mainland (Mees 1969) across timeframes from decades to hundreds of thousands of years. This provides a spectrum of divergence timeframes that can be used as a proxy for the speciation continuum, capturing incipient phenotypic divergence through to highly divergent subspecies (Clegg *et al.* 2002a). Importantly, silveryeye populations have diverged under different gene flow scenarios, with some populations diverging with gene flow and others



**Figure 1** Hypothetical distributions of genetic differentiation expected for divergence with and without gene flow. (A) During the early stages of the speciation continuum divergence is expected to be limited to few loci, resulting in a highly skewed distribution of divergence values. This expectation will be most extreme for populations diverging with gene flow (black lines) than for populations diverging in genetic isolation (gray lines) as the homogenizing effect of gene flow limits divergence to only loci under strong divergent selection. (B) During later stages of the speciation continuum skew is expected to break down as divergence accumulates, but less so in populations diverging with gene flow (black lines) than those diverging without (gray lines) as occasional gene flow may reduce levels of divergence at positions of the genome even as reproductive isolation is approached.

diverging in genetic isolation (Mees 1969). Silveryeyes also show a repeated pattern of phenotypic change on islands toward increased body and bill size (Mees 1969; Clegg *et al.* 2002b; Clegg *et al.* 2008). While we do not have direct evidence that body size acts as a reproductively isolating mechanism, empirical and modeling evidence supports that increased size is the result of directional natural selection that is strongest in the early generations following colonization of novel environments (Clegg *et al.* 2002b; Clegg *et al.* 2008).

There are several hypotheses to explain the selective advantage of larger body size, including the positive relationship between successful intraspecific aggressive encounters and body size in high density insular populations (Robinson-Wolrath and Owens 2003), and competitive displacement in the presence of closely related species. While we do not know the genetic architecture of this repeated phenotypic pattern, several studies have found genes of large effect for body/bill traits (Bosse *et al.* 2017; Chaves *et al.* 2016; Cornetti *et al.* 2015;



**Figure 2** Silveryeye populations used in the study. (A) Sampling locations of *Zosterops lateralis* across the south Pacific (dashed lines indicate gene-flow between populations); (B) Population divergence timeframes (known from historical records, or inferred from island ages and genetic divergence dates) and modes of divergence (gene-flow or no gene-flow); (C) evolutionary relationships between *Z. lateralis* subspecies and populations based on mitochondrial DNA (Black, R., unpublished results), SNP data (Sendell-Price, A.T., unpublished results) and known colonization histories (subspecies indicated in brackets); (D) *Z. l. chlorocephalus* from Heron Island (left) are up to 40% larger than *Z. l. cornwalli* from the Australian mainland (right). Image by Nick Clark; (E) Maximum likelihood estimation of individual ancestries calculated with ADMIXTURE (Alexander *et al.* 2009), based on 6,822 LD-filtered SNPs.

Lamichhane *et al.* 2015; vonHoldt *et al.* 2018) and we anticipate that genomic islands should form around loci underlying this phenotypic change across comparisons, allowing the dynamics of individual genomic islands to be studied.

Here we use empirical analysis of Restriction-site Associated DNA Sequencing (RAD-Seq) data from silvereyes and individual-based simulation models to address the following questions: 1) How does genomic divergence accumulate across the genome over time? 2) How is the accumulation of divergence affected by gene flow? 3) Does the position of genomic islands of divergence and valleys of similarity correspond to the location of loci under divergent selection or parallel/purifying selection, respectively? 4) Do genomic islands widen, and valleys narrow, over longer divergence timeframes? and 5) Are the features of genomic valleys consistent with proposed mechanisms that could contribute to their formation?

## MATERIALS & METHODS

### Sample collection

Blood samples were collected from eight silvereye populations across the south Pacific, capturing a range of divergence levels and gene flow variation within this species, and providing five diverging population pairs for comparison. Population pairs varied in their divergence timeframes (early stage: <150 years, mid stage: 3,000–4,000 years, and late stage: 100,000s years) and mode of divergence (gene flow or no gene flow) (Figure 2). The populations used in this study provide a rare example of where gene flow scenarios and divergence timescales are either known *a priori* or where a strong case can be made to infer the likely gene flow scenario. This is based on a combination of historical records, contemporary bird movements, geographic proximity, and phylogenetic inferences as outlined below:

**South Island vs. Chatham Island:** The Chatham Island silvereye population was colonized from South Island (New Zealand) during the 1850s. This timing is historically documented (Mees 1969; Clegg *et al.* 2002a) and reflected in population level phylogenies (R. Black, unpublished results). In the mid 1800s, South Island silvereyes were undergoing a population expansion which resulted in the colonization of other islands in the region (notably: North Island of New Zealand and Norfolk Island) (Mees 1969; Clegg *et al.* 2002a). Given that population expansion was taking place within this region for several decades after the colonization of Chatham Island, these population are likely to have experienced gene flow. Therefore, this comparison is characterized as Early Stage – Gene Flow.

**South Island vs. French Polynesia:** The French Polynesia silvereye population is the product of a documented human-mediated introduction of South Island silvereyes (*Z. l. lateralis*) to Tahiti in 1937 (Monnet *et al.* 1993; Thibault and Cibois 2017). Given that French Polynesia is located well beyond the natural distribution limit for this species and that the French Polynesian population is thought to be the product of a single and recent introduction event (Sendell-Price *et al.* 2020), this comparison is characterized as Early Stage – No Gene Flow.

**Mainland (Australia) vs. Heron Island:** The Heron Island silvereye population was established from the mainland silvereye subspecies (*Z. l. cornwalli*) between 3,000–4,000 years ago. This divergence timeframe is supported by geological records which suggest that Heron Island has been vegetated (and therefore habitable) for a maximum of 4,000 years (Degnan and Moritz 1992; Hopley 1982).

Given that *Z. l. cornwalli* is regularly observed on Heron Island, this comparison is characterized as Mid Stage – Gene flow.

**Grande Terre vs. Lifou:** Based on mitochondrial DNA divergence estimates, silvereyes on the islands of Grande Terre (*Z. l. griseonotus*) and Lifou (*Z. l. melanops*) diverged in the past few hundred-thousand years (Black, R., unpublished results). Given the silver-eye's ability to cross water boundaries and the geographic proximity of these islands (200km apart), this comparison is characterized as Late Stage – Gene Flow.

**Mainland (Queensland) vs. Lord Howe Island:** Based on phylogenetic analyses, *Z. l. tephropleurus* on Lord Howe Island is derived from the mainland subspecies *Z. l. cornwalli* in the past few hundred-thousand years (A. T. Sendell-Price, unpublished results). Given the geographic isolation of Lord Howe Island, this comparison is characterized as Late Stage – No Gene Flow.

For each of the population comparisons used in this study the diverging populations show body size differentiation. *Z. l. lateralis* from Chatham Island and French Polynesia are larger than those from their South Island source (Clegg *et al.* 2002a; Sendell-Price *et al.* 2020), *Z. l. chlorocephalus* from Heron Island is approximately 40% larger than its mainland ancestor *Z. l. cornwalli* (Clegg *et al.* 2008), *Z. l. melanops* from Lifou is larger in many traits than its Grand Terre ancestor, *Z. l. griseonotus* (Black, R., unpublished results) and *Z. l. tephropleurus* from Lord Howe Island is substantially larger than its mainland ancestor (Clegg *et al.* 2002a; Clegg and Phillimore 2010).

Sampling locations and collection dates are provided in Table 1. To provide large enough sample sizes for analysis, the French Polynesian population included samples taken from multiple islands. This is justified based on an admixture analysis which grouped these samples together into a single cluster (Figure 2E). All birds were caught using mist nets or traps and 20–40  $\mu$ l of blood collected from the brachial wing vein was stored in  $\sim$ 0.5 ml of lysis buffer (0.01M Tris-HCl; 0.01M NaCl; 0.01M EDTA; 1% n-lauroylsarcosine, pH 8.0) (Seutin *et al.* 1991).

### RAD-PE sequencing and bioinformatics

Restriction-site associated DNA paired-end (RAD-PE) sequencing was used to assess the genomic landscape of divergence among silvereye populations. This reduced representation sequencing method has the potential to identify many hundreds of thousands of Single Nucleotide Polymorphisms (SNPs) distributed throughout the genome (Davey and Blaxter 2010). Genomic DNA in blood samples was extracted from between 20 and 23 individuals per population (or set of populations in the case of French Polynesia) (n = 168) using QIAGEN DNeasy blood and tissue extraction kits following manufacturer protocols. RAD-PE libraries were then prepared using restriction enzyme *Sbf*I-HF (New-England Biolabs Inc., Beverly MA, USA) following the protocol by Ali *et al.* (2016). See supplementary material for a comprehensive description of library preparation. Libraries were sequenced on three Illumina HiSeq4000 lanes (Illumina, San Diego, CA, USA) at the UC Davis Genome Center using paired-end 150-bp sequence reads.

Quality of sequencing reads was checked visually using FASTQC (Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). The *process\_radtags* script included in the STACKS version 1.4 software pipeline (Catchen *et al.* 2013) was used to assign sequence reads to individuals. In addition, reads containing uncalled bases and/or bases of low quality were discarded in this step using default quality thresholds (an average Phred score of 10 in sliding

■ **Table 1** *Z. lateralis* sampling information. No. Sequenced = number of individuals including in RAD-Seq libraries; No. Filtered = number of individuals retained after quality filtering

Subspecies	Pop. code	Sampling location	Lat / Long	Collection date	No. Sequenced	No. Filtered
<i>chlorocephalus</i>	HI	Heron Island	-23.44 / 151.91	2015	20	14
<i>cornwalli</i>	ML	Mainland Australia	-28.18 / 153.45	2013	20	18
<i>griseonotus</i>	GT	Grande Terre	-21.78 / 166.02	2014	20	17
<i>lateralis</i>	CI	Chatham Island	-44.58 / 176.32	1997	22	17
<i>lateralis</i>	SI	South Island	-45.53 / 170.30	1997	23	18
<i>lateralis</i>	FP	Mo'orea	-17.51 / -149.91	2005	17	10
<i>lateralis</i>	FP	Raiatea	-16.82 / -151.46	2010	3	2
<i>lateralis</i>	FP	Rurutu	-22.48 / -151.34	2012	3	3
<i>melanops</i>	LF	Lifou	-20.89 / 167.25	2014	20	17
<i>tephropleurus</i>	LH	Lord Howe Island	-31.33 / 159.05	1998	20	18

windows of 15% of the length of the read). Sequences with possible adapter contamination and/or missing the *SbfI*-HF restriction site were also discarded. Following this, reads were filtered for PCR duplicates using the STACKS *clone\_filter* script. The remaining reads were then mapped to the *Zosterops lateralis melanops* genome assembly version 1 (NCBI Assembly GCA001281735.1, (Cornetti *et al.* 2015)) with BOWTIE2 version 2.2.6 (Langmead and Salzberg 2012) using end-to-end alignment and default settings (allowing for a maximum of two mismatches in the seed (-n 2)). Genotypes were then called using the *HaplotypeCaller* and *GenotypeGVCFs* tools from the Genome Analysis Toolkit (GATK) nightly build version 2016-12-05-ga159770 (McKenna *et al.* 2010). To output both variant and non-variant sites, *HaplotypeCaller* was ran using `-output_mode 'EMIT_ALL_CONFIDENT_SITES'` and *GenotypeGVCFs* ran using mode `'-includeNonVariantSites'`. The resulting output was then filtered to remove indels and only include sites where the minor allele count was  $\leq 2$ ; minimum genotype quality = 30; minimum depth = 8; and sites were called in at least 70% of individuals. Following this the outputted VCF file was further filtered to remove individuals missing > 30% of sites.

As the *Z. l. melanops* genome is only assembled to the scaffold level (Cornetti *et al.* 2015), *Z. l. melanops* scaffolds were mapped to chromosomes of the *Taeniopygia guttata* genome assembly version 3.2.4 (NCBI Assembly GCA\_000151805.2) using Satsuma Synteny (Grabherr *et al.* 2010). Output from Satsuma Synteny was then used to assign scaffolds to chromosomes and determine order, location, and orientation using custom R scripts from Van Doren *et al.* (2017). Custom scripts from Sendell-Price *et al.* (2020) were then used to reorder the GATK outputted VCF file accordingly and remove SNPs where chromosomal positions could not be determined. Because synteny is high in birds (Ellegren 2010), 96.8% of the *Zosterops* scaffolds were assigned to assembled chromosomes and arranged in the presumed correct order and orientation. As inversions and other chromosomal rearrangements are known to occur in birds (Backström *et al.* 2010), it may be possible that a small percentage of scaffolds were ordered or oriented incorrectly. However, a small number of misplaced scaffolds is not expected to impact our overall interpretation of empirical patterns.

Finally, Wright's (1951) fixation index ( $F_{ST}$  – a measure of relative divergence) and Nei's (1987) measure of absolute divergence ( $d_{xy}$ ) were calculated in non-overlapping windows of 5kb, 50kb and 100kb for each population comparison, using Python scripts developed by Martin *et al.* (2013) ([https://github.com/simonhmartin/genomics\\_general](https://github.com/simonhmartin/genomics_general)).  $F_{ST}$  and  $d_{xy}$  were outputted only for windows that contained at least 10 sites. For all downstream analyses that used windowed statistics we filtered our dataset so that windows containing missing

data were removed, *i.e.*, only those windows containing  $F_{ST}/d_{xy}$  for all pairwise comparisons were retained.

### Population genomic analyses

We assessed the shift from localized divergence at few loci toward more genome-wide levels of divergence by comparing the third moment (skewness) of  $F_{ST}$  distributions between population comparisons. In line with expectations outlined in Figure 1, skewness provides a useful metric to describe a population comparisons' stage of transition from localized (high positive skew) to genome-wide levels of divergence (less positive skew). Note: as  $F_{ST}$  is bounded between 0 and 1, distributions are expected to become negatively skewed when populations have been diverging over very long time-scales as most loci approach fixation. We tested for significant differences in the skewness of empirical distributions using a randomization test. For each pairwise comparison of diverging populations (*e.g.*, South Island vs. Chatham Island compared to South Island vs. French Polynesia) we first calculated a test statistic (the absolute difference in distributional skew of observed  $F_{ST}$  values) and compared this to test statistics calculated for 10,000 randomized distributions. Randomized distributions were produced by assigning observed  $F_{ST}$  values to comparisons at random without replacement. *P*-values were then calculated as the percentile of the distribution of randomized test statistics that the observed test statistic lied on.

R scripts, based on those used in Van Doren *et al.* (2017), were used to identify highly diverged regions (genomic islands of divergence) and regions of low divergence (genomic valleys of similarity) occurring across the genomic landscapes of the diverging populations. First, for each population comparison, a kernel-based smoothing algorithm was applied to windowed  $F_{ST}$  values (box density with bandwidth of 20), and the smoothed line compared to 10,000 smoothed lines obtained after permuting the order of the windows (see Ruegg *et al.* (2014)). Genomic islands were identified as any location where the observed smoothed line was greater than the most extreme value from the permutation distribution (see Van Doren *et al.* (2017)). In contrast, genomic valleys were identified as any region where the observed smoothed line was lower than the lowest value of the permuted distribution. We merged outlier regions separated by ten windows or fewer. For each genomic island and genomic valley, we calculated mean  $F_{ST}$  and  $d_{xy}$ . Wilcoxon signed-rank tests were used to determine if  $d_{xy}$  within genomic islands and genomic valleys differed significantly from chromosomal background levels. To assess the effect of time and gene flow on the development of patterns of heterogeneity the number and size of identified genomic islands and valleys were compared across timescales and gene flow contexts using Kruskal-Wallis tests. In the main text we

present results of island/valley detection using 50kb windows, however island/valley detection was also conducted using 5kb and 100kb windows – these results are presented in the supplementary material.

### Identifying outlier SNPs under selection

To identify genomic regions that may be under selection, we scanned for outlier loci using *PCAdapt*, a principal components-based method of outlier detection with a low rate of false-positive detection (Luu *et al.* 2017). *PCAdapt* requires selection of  $K$  principal components, based on inspection of a scree plot, in which  $K$  is the number of PCs with eigenvalues that depart from a straight line. *PCAdapt* then computes a test statistic based on Mahalanobis distance and controls for inflation of test statistics and false discovery rate (FDR). Outlier SNPs were identified using the following settings for all population comparisons:  $K = 2$ ,  $\text{MinMAF} = 0.2$ , and  $\text{FDR} = 0.01$ . As *PCAdapt* has been shown to be sensitive to linkage-disequilibrium (LD) (Lotterhos *et al.* 2019), prior to conducting outlier detection we filtered SNPs so that those with very high LD ( $r^2 > 0.8$ ) were removed. LD-filtering was conducted using the ‘-indep-pairwise 1000 kb 1 0.8’ command in PLINK (Purcell *et al.* 2007).

### Candidate gene analysis

To determine if genomic islands of divergence occurred around genes known to be associated with body and beak size differences in birds, we compiled a list of candidate genes based upon a literature review of gene expression and association studies. The resulting candidate gene list was restricted to only those genes where (1) locations within the zebra finch genome are known; and (2) genes were mapped to directly by SNPs sequenced as part of this study or genes were located within the 50kb windows used when summarizing divergence statistics. Positions of candidate genes were then compared to the position of identified genomic islands of divergence. The position of candidate genes was also compared to the position of outlier SNPs. Genes were determined to fall within genomic islands, if island bounds (start and end positions) overlapped gene start and end positions (this includes 5' UTRs and introns). Likewise, outlier SNPs were determined as within candidate genes if they fell within 10kb of the gene start/end positions.

### Simulations of divergence

We performed individual-based simulations where genomic divergence could only occur via neutral processes and compared model predictions with empirically observed genomic patterns. The simulations consequently provide a neutral model. If such a model can generate patterns observed in nature, then it is not necessary to invoke selection as the underlying driver of observed patterns. In other words, the simulation modeling allowed us to address whether the transition from localized to more genome-wide divergence could be explained by drift alone, and also allowed us to explore the fine scale effects of gene flow and recombination by simulating divergence under various levels of gene flow and recombination (both of which were not possible with empirical comparisons alone).

The simulations of divergence are based on the framework available in the *R* package “glads” (genomic landscape of divergence simulations) (Quilodrán *et al.* 2020). This package allows individual-based simulation for demographic, genetic, and genomic data forward in time. The model considers two populations that may or may not evolve independently according to different levels of migration between them, allowing a level of gene flow. There are three hierarchical levels: genotypes, phenotypes and demographic rates. All individuals are characterized by sex and genetic identity.

Each population started with individuals with randomly drawn genotypes across 46,853 biallelic SNPs on chromosome 5. Each genotype was determined by sampling with replacement from the distribution of all observed genotyped birds at each SNP. Individuals were then assigned a sex assuming a 50% sex ratio and assigned to a population to produce two populations, each with an initial population size of 400 individuals (similar to the Heron Island breeding population size (Kikkawa and Wilson 1983)).

Each population was iterated forward on a per-generation time step. At each generation, individuals were assigned a number of offspring independent of genotype. Fitness (the number of offspring) was determined by randomly drawing from a Poisson distribution with a mean ( $\lambda$ ) that varied with population density ( $N$ ). We thus defined mean population fitness as the average number of offspring produced. We included density-dependence ( $\sigma_{dem}$ ) in this function to avoid exponential growth. The final fitness of individuals at a given time was computed as  $\text{Poisson}(\lambda) - N\sigma_{dem}$ . The density dependence ( $\sigma_{dem}$ ) kept the population within empirically observed bounds for the Heron Island population (100-300 breeding pairs, see Table S1). The mean population fitness ( $\lambda$ ) was obtained by multiplying the average number of fledglings of observed clutches by the mean survival during the summer, autumn and winter of young individuals along with the generation time (Table S1). These parameter values were obtained from literature (Kikkawa and Wilson 1983, Brook and Kikkawa 1998, Clegg *et al.* 2002b, Sandvig *et al.* 2017). Male fitness was then scaled such that the total sum of male fitness equalled the total sum of female fitness. Mating pairs were formed by combining male and female parents as a function of their ranked fitness values. Because the fitness value of individuals is obtained from a random distribution, this process is equivalent to a random mating pair within populations. This approach assumes monogamous reproduction which is empirically observed (Robertson *et al.* 2001).

To determine offspring genotype, we first identified crossover points along the parental genomes by drawing random values via a Poisson process. This procedure considers the physical position of each SNP in the chromosome, within which crossovers are obtained from a Poisson distribution along the length of the chromosome. Recombination is thus a homogeneous Poisson process where the rate is defined by the expected per base pair recombination rate (e.g.,  $1e-08$ , corresponding to 1 cM/Mb). Four different recombination rates were simulated: 2 cM/Mb, 3 cM/Mb, 6 cM/Mb, 10 cM/Mb and 19 cM/Mb (translated to crossovers per base pair in our simulations). This corresponds to the range of values known from the collared flycatcher (*Ficedula albicollis*) genome (Kawakami *et al.* 2014) (Table S1). Flycatcher recombination rates were used, as unlike the zebra finch genome, the flycatcher is not thought to contain large recombination deserts (Backström *et al.* 2010). Copies of chromosome 5 were then segregated from the parental genomes with recombination occurring at each crossover. Migration rate ( $m$ ) was set to zero in simulations of no-gene-flow scenarios. In divergence-with-gene-flow simulations, offspring either stayed in the same population or dispersed, with simulations repeated for values of  $m$  of 0.0001, 0.001 and 0.01. Simulations were run for a maximum of 2,000 generations (assuming a generation time of three years (Clegg *et al.* 2008)). More information about the individual based framework is available in Quilodrán *et al.* (2020). *R* code is available at GitHub: <https://github.com/eriqande/glads>.

### Data availability

Raw sequencing reads have been submitted to the National Center for Biotechnology Information (NCBI; <https://www.ncbi.nlm.nih.gov/>) under accession number PRJNA489169. Supplementary material, VCF files

## RESULTS

### RAD-PE sequencing and bioinformatics

Overall, RAD-PE sequencing resulted in 1,149,472,877 paired-end reads. Post-quality filtering (removal of indels and only including sites where: the minor allele count was  $\leq 2$ ; minimum genotype quality = 30; minimum depth = 8; and genotypes were called in at least 70% of individuals) reads covered a total of 15,872,783 sites (of which 788,169 were biallelic). Of the 168 samples sequenced, 134 were retained after removing individuals where  $\geq 30\%$  of data were missing. The number of individuals retained per location ranged from 15 to 18 (Table 1).

### Overall levels of divergence

Mean  $F_{ST}$  was lowest for populations in the earliest stages of divergence (South Island vs. Chatham Island and South Island vs. French Polynesia) and increased over longer divergence timeframes, with populations at later stages of divergence (Grande Terre vs. Lifou and Mainland vs. Lord Howe Island) having the highest mean  $F_{ST}$  (Figure 3A). Mean  $F_{ST}$  was also consistently lower for populations diverging with gene flow when compared to non-gene flow scenarios matched for divergence timeframe (*i.e.*, South Island vs. Chatham Island compared to South Island vs. French Polynesia) demonstrating how the homogenizing nature of gene flow impedes divergence. Compared to autosomes, chromosome Z exhibited higher levels of divergence for all population comparisons (Figure 3A). This difference was significant for South Island vs. Chatham Island and Mainland vs. Heron Island comparisons (Mann-Whitney: all  $P$  values  $< 0.05$ ).

### Accumulation of genomic divergence across the genome

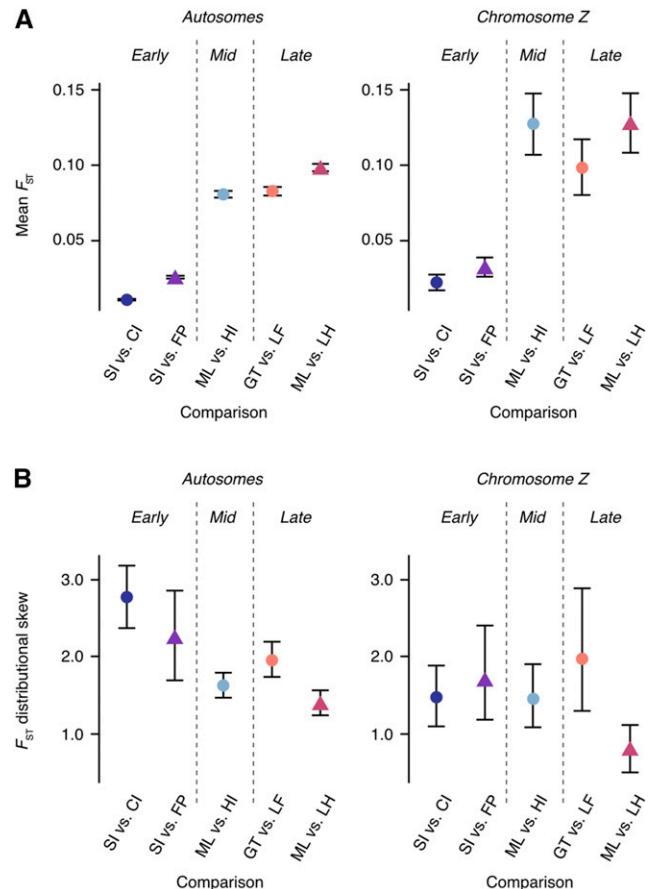
In the early stage of the speciation continuum (South Island vs. Chatham Island and South Island vs. French Polynesia comparisons), the distributions of  $F_{ST}$  values calculated in 50kb windows for autosomes were highly skewed toward large values of  $F_{ST}$  (skew = 2.69 and 2.07 respectively) (Figure 3B) and characterized by extreme L-shaped distributions with divergence limited to few 50kb windows (Figure 4B). Over longer divergence timeframes distributional skew was reduced, with the lowest levels of skew (skew = 1.41) observed for the late stage – no gene flow comparison (Mainland vs. Lord Howe Island) (Figure 3B). The accumulation of genome-wide divergence over longer divergence timeframes is clearly visible in Figure 4A. When matched for divergence timeframe, the  $F_{ST}$  distributions of populations diverging under the influence of gene flow showed higher levels of skew (more localized divergence) than populations diverging in the absence of gene flow – this difference was significant for the two late-stage comparisons ( $P < 0.001$ ). Although not all pairwise differences were significant, recently diverging populations in most cases had significantly higher skew for autosomes than late stage diverging populations (Table 2). When treating each autosomal chromosome as an independent unit, more chromosomes showed significantly reduced  $F_{ST}$  distributional skew than increased skew when diverging under gene flow, during both early and late stage divergence (Table S2).

Distributional skew of chromosome Z was lowest for the Mainland vs. Lord Howe Island (late stage - no gene flow) comparison, meaning that divergence was more widespread for this comparison (Figure 4A). This difference in skew was significant when compared to all other diverging population comparisons (all  $P$ -values  $< 0.05$ ,

Table 2). Unlike for autosomes, there were no significant differences in chromosome Z distributional skew between South Island vs. Chatham Island, South Island vs. French Polynesia, Mainland vs. Heron Island, and Grande Terre vs. Lifou comparisons (Table 2).

### Dynamics of genomic islands and genomic valleys

Genomic islands were identified in all population comparisons, including populations diverging with gene flow (South Island vs. Chatham Island, Mainland vs. Heron Island, and Grande Terre vs. Lifou) as well as populations diverging in the absence of gene flow (South Island vs. French Polynesia and Mainland vs. Lord Howe Island) (Figure 4A). Genomic island size – although variable (mean size ranged from 350 kb to 575 kb, see Table 3 and Figure 5) – did not differ significantly between population comparisons (Kruskal-Wallis:  $\chi^2_4 = 0.44$ ,  $P = 0.98$ ). This finding was robust to different window sizes, with no significant difference in genomic island size observed when identifying genomic islands using a 5kb window (Kruskal-Wallis:  $\chi^2_4 = 7.897$ ,  $P = 0.095$ ) or 100kb window (Kruskal-Wallis:  $\chi^2_4 = 1.43$ ,  $P = 0.84$ ) (see Figure S1). We identified one region of the genome in which a genomic island of divergence occurred at the same location across multiple population comparisons. This genomic island, which was located on chromosome 5 and occurred in both



**Figure 3** Mean  $F_{ST}$  and distributional skew of  $F_{ST}$  values. (A) Mean  $F_{ST}$  and (B) distributional skew of  $F_{ST}$  values calculated in 50kb non-overlapping windows. Calculated for autosomes and chromosome Z separately for each population comparison. Populations diverging with gene flow are indicated by circles and populations diverging in isolation by triangles. 95% confidence intervals obtained via bootstrapping over 500kb windows.

late stage comparisons, was largest for the Mainland vs. Lord Howe Island comparison (Late stage – No gene flow) (Table 4). In most instances, absolute divergence ( $d_{xy}$ ) within individual genomic islands was not significantly different from background chromosomal levels. However, genomic islands with significantly elevated  $d_{xy}$  were identified in small numbers, as were genomic islands with below background level  $d_{xy}$  (Figure 4A).

Genomic valleys were identified in all diverging population comparisons (Figure 4A). Their size – although variable (mean size ranged from 250 kb to 1,142 kb, see Table 3 and Figure 5) – did not differ significantly between population comparisons (Kruskal-Wallis:  $\chi^2_4 = 2.45$ ,  $P = 0.65$ ). Although significant differences in genomic valley size were observed when identifying genomic valleys using a 5kb window (Kruskal-Wallis:  $\chi^2_4 = 14.95$ ,  $P < 0.01$ ) and 100kb window (Kruskal-Wallis:  $\chi^2_4 = 9.61$ ,  $P < 0.05$ ) (see Figure S1), Dunn’s post-hoc test indicated that this difference was limited to few pairwise comparisons and followed no specific pattern in terms of gene flow scenario or divergence timeframe (Table S3). We identified two regions of the genome (located on chromosomes 1 and 7) in which a genomic valley occurred across multiple population comparisons (Table 4). Both shared genomic valleys were identified in the late stage comparisons and were largest for the comparison thought to be diverging in the absence of gene flow (Mainland vs. Lord Howe Island). In most instances, absolute divergence ( $d_{xy}$ ) within individual genomic valleys was not significantly different from background chromosomal levels.

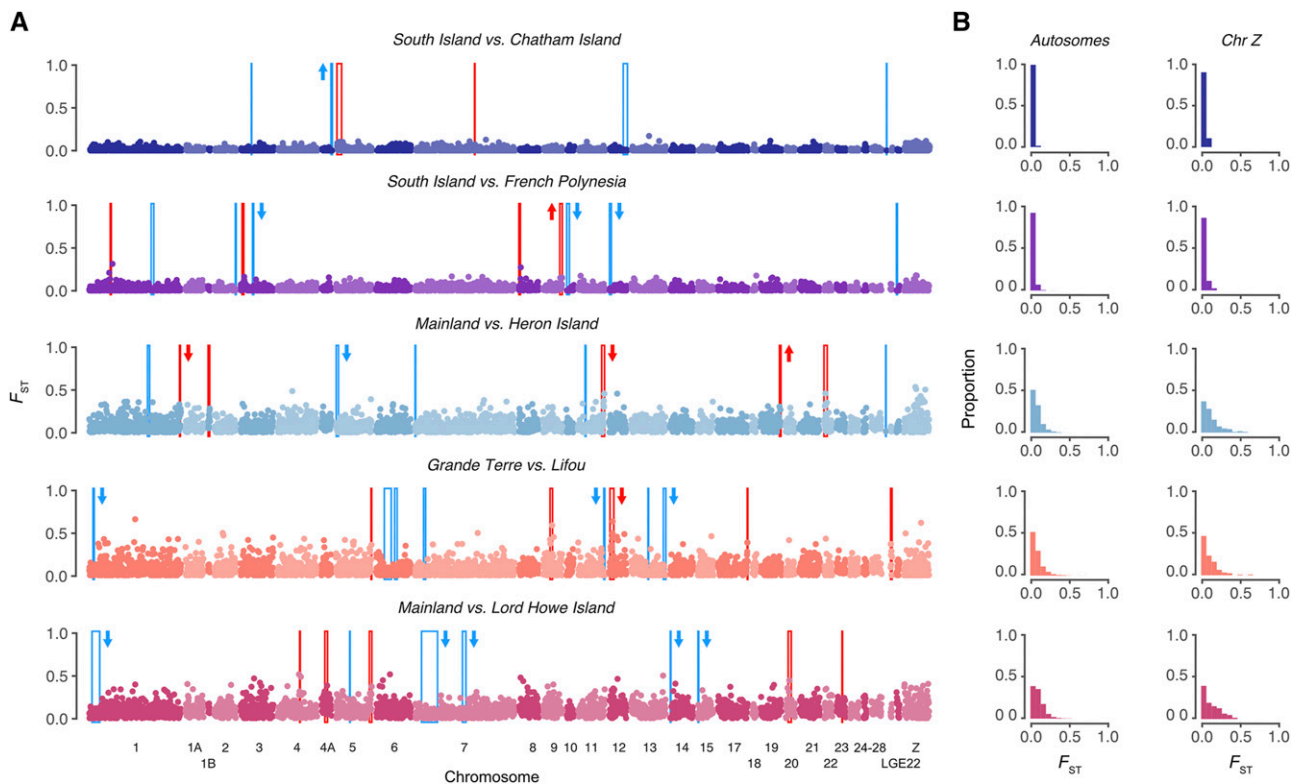
However, genomic valleys with significantly decreased  $d_{xy}$  were also frequently identified (Figure 4A). A single genomic valley with significantly elevated  $d_{xy}$  was identified for the South Island vs. Chatham Island comparison (Early stage – Gene flow).

### Correspondence of genomic islands and valleys to the position of outlier SNPs

*PCAdapt* identified between 14 and 235 outlier SNPs in each population comparison (Table 5 and Figure 6). For the South Island vs. Chatham Island, Mainland vs. Heron Island, and Grande Terre vs. Lifou comparisons the position of four outlier SNPs fell within genomic islands of divergence. A single outlier SNP was identified as occurring within a genomic valley for the Mainland vs. Lord Howe Island comparison.

### Candidate gene analysis

A review of the literature identified 22 candidate genes associated with body and/or beak size differences in birds, for which we had sequence data or occurred within the 50kb windows used when summarizing divergence statistics (Table 6). The location of candidate genes did not correspond to the position of any genomic islands or genomic valleys identified. However, for the South Island vs. French Polynesia (Early stage – No gene flow) and Mainland vs. Heron Island (Mid stage – Gene flow) comparisons an outlier SNP was located within *BMPRIA* (Figure 6 and Table 6). This was the only candidate gene to contain outlier SNPs.



**Figure 4** Distributions of divergence across the genome for diverging silveryeye populations. (A) Pairwise  $F_{ST}$  across the genome for each population comparison calculated in non-overlapping 50kb windows. Regions of elevated differentiation (genomic islands) are highlighted in red and regions with low differentiation (genomic valleys) are highlighted in blue. Arrows beside genomic islands/valleys indicate if  $d_{xy}$  within these regions was significantly elevated (upwards pointing arrows) or decreased (downward pointing arrows) compared to chromosomal background levels. Significance determined using Wilcoxon signed-rank tests. Chromosomes are numbered according to the zebra finch nomenclature. (B) Distributions of windowed  $F_{ST}$  values calculated in 50kb windows for each comparison. Distributions are shown for Autosomes and Chromosome Z separately.



■ **Table 2** P-values for pairwise comparisons of  $F_{ST}$  distributional skew tested using a randomization test. Below diagonal = autosomes only; above diagonal = chromosome Z only

	SI vs. CI	SI vs. FP	ML vs. HI	GT vs. LF	ML vs. LH
SI vs. CI	—	0.688	0.967	0.732	0.040
SI vs. FP	0.447	—	0.593	0.885	0.005
ML vs. HI	<0.001	<0.001	—	0.226	0.031
GT vs. LF	<0.001	0.225	0.062	—	0.004
ML vs. LH	<0.001	<0.001	0.060	<0.001	—

### Simulations of neutral divergence

Simulations of neutral divergence, using an individual based model, were able to capture the transition from localized to more genome-wide levels of divergence, in which  $F_{ST}$  distributions were characterized by low values of  $F_{ST}$  during the early stages of divergence, shifting toward higher  $F_{ST}$  values over longer divergence timeframes (Figure 7). Simulating divergence under various levels of gene flow ( $m = 0$ ,  $m = 0.0001$ ,  $m = 0.001$ ,  $m = 0.01$ ) showed that the accumulation of divergence slowed under increasing levels of gene flow (Figure 8).

### DISCUSSION

We take advantage of an exceptional natural system where gene flow scenarios and divergence timescales are either known *a priori* or can be reasonably presumed, to reveal key insights into how divergence accumulates at the genomic level. Our results provide an empirical characterization of the change in the distribution of genomic divergence using a proxy for the speciation continuum. The pattern of decreasing skew in the distribution of genomic divergence is consistent with the intuitive explanation of how heterogeneity originates, and divergence accumulates, *i.e.*, divergence at a small number of selected loci followed by divergence hitchhiking. However closer inspection of the genomic landscape provides equivocal support for this particular mechanism. Genomic islands were rarely associated with SNPs putatively under selection and genomic islands did not widen as expected under the divergence hitchhiking model of speciation. Genomic valleys of similarity were also a common feature of the heterogeneous genomic landscape, but we did not find clear evidence to attribute these to selective processes. Furthermore, by simulating divergence under a neutral model we were able to capture the transition from localized to more genome-wide levels of divergence. The transition from localized divergence (highly skewed distribution) to more genome-wide levels of divergence (a more even distribution of genetic differentiation) therefore appears to occur largely outside of the context of genomic islands and is not dependent of selection, raising questions about the divergence hitchhiking model of speciation, where divergence is initiated through selection of a

small number of genes (presumably of large effect), genomic islands form around them, and divergence hitchhiking completes genome-wide divergence.

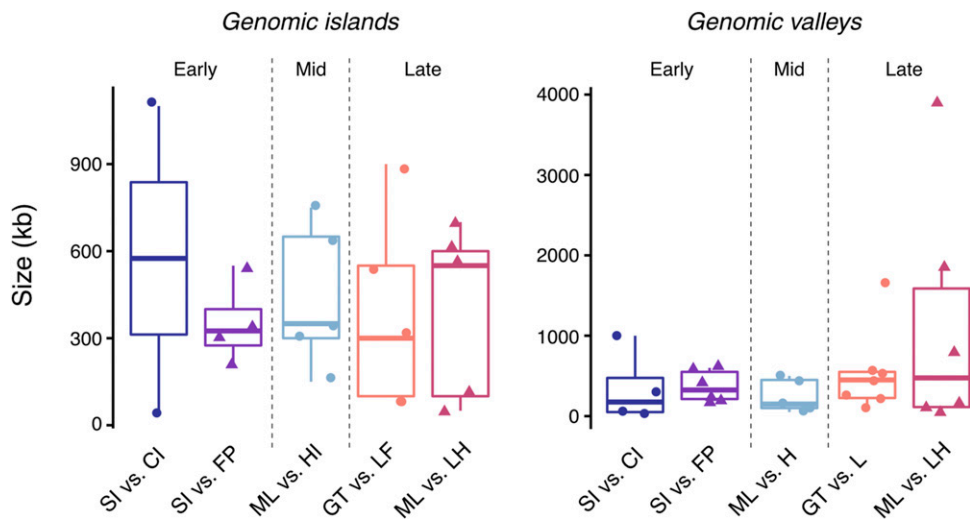
### The accumulation of divergence across the genome

Our results demonstrate how divergence accumulates across the genome as populations move along the speciation continuum. *Zosterops lateralis* populations at the earliest stages of divergence followed extreme L-shaped  $F_{ST}$  frequency distributions where divergence was limited to few loci, with a reduction of skew in  $F_{ST}$  values as divergence became more wide-spread at later stages of the speciation continuum. This shift in distribution occurred in both the presence and absence of gene flow. We are aware of only two other studies where similar comparisons could be made: *Heliconius* butterflies (Martin *et al.* 2013) and *Ficedula* flycatchers (Burri *et al.* 2015). Unlike Burri *et al.* (2015), we did not observe a shift toward an extreme left-skewed distribution where the majority of SNPs reach fixation. This is likely due to the inclusion of comparisons between flycatcher species with divergence times of millions of years, whereas our within-species divergence timeframes were too short for an extreme left-skewed distribution to be reached.

On the whole, the observed distributions of genomic differences across the genome fit with expectations for how genomic divergence accumulates across the speciation continuum (Nosil and Feder 2012; Seehausen *et al.* 2014). However, our simulations demonstrated that the shift from localized to more genome-wide levels of divergence could be achieved assuming a neutral model of molecular evolution. This finding was robust to varying levels of recombination, which is interesting given the emphasis on linkage disequilibrium in driving localized and ultimately genome-wide divergence (Seehausen *et al.* 2014). One important difference, however, was that our simulations suggest that the transition from localized to more genome-wide levels of divergence occurs at a faster rate than is observed empirically - under allopatry and low-level gene flow  $F_{ST}$  approached fixation by 2,000 generations. This suggests that some other mechanism, such as frequency-dependent selection, may be maintaining genetic variation at individual SNPs. However, our simulation framework only

■ **Table 3** Summary statistics of genomic islands and genomic valleys identified for each population comparison

Divergence timeframe	Mode of divergence	Comparison	No. identified	Mean size (kb)	Max size (kb)	Mean $F_{ST}$	Mean $d_{xy}$
<b>Genomic islands</b>							
Early stage	Gene flow	SI vs. CI	2	575	1,100	0.031	0.003
Early stage	No-gene flow	SI vs. FP	4	350	550	0.038	0.005
Mid stage	Gene flow	ML vs. HI	5	440	750	0.147	0.005
Late stage	Gene flow	GT vs. LF	5	390	900	0.233	0.002
Late stage	No-gene flow	ML vs. LH	5	400	700	0.153	0.006
<b>Genomic valleys</b>							
Early stage	Gene flow	SI vs. CI	4	350	1,000	0.001	0.005
Early stage	No-gene flow	SI vs. FP	6	367	600	0.005	0.003
Mid stage	Gene flow	ML vs. HI	5	250	500	0.040	0.004
Late stage	Gene flow	GT vs. LF	7	536	1,650	0.019	0.002
Late stage	No-gene flow	ML vs. LH	6	1,142	3,900	0.059	0.003



**Figure 5** Size (kb) of genomic islands and genomic valleys identified for each population comparison. Populations diverging with gene flow are indicated by circles and populations diverging in isolation by triangles.

considered diverging population consisting of 400 individuals each, whereas in reality many of the diverging populations have much larger population sizes. As the potential for drift is greatest when population size is small (Wright 1931), this discrepancy in the pace of divergence may reflect differences in population size not captured in our simulation framework.

When matched for divergence timeframe, the pace at which divergence accumulated across the genome was moderated by gene flow. This was manifested as higher distributional skew of autosomal  $F_{ST}$  values for silvereye populations thought to be diverging under gene flow compared to populations thought to be diverging in genetic isolation. Martin *et al.* (2013) found that genomic divergence was more widespread for races and species of *Heliconius* butterflies diverging in geographic allopatry than for those diverging in parapatry and sympatry, suggesting an underlying role for gene flow in slowing the accumulation of genomic divergence. We provide a further empirical characterization of this difference in the accumulation of genomic divergence. While we can be certain that the Mainland and Heron Island populations have been diverging under the effects of gene flow (the mainland subspecies is frequently observed on Heron Island) and that the South Island and French Polynesian populations have been diverging under genetic isolation (the French Polynesian population is the product of a single human-mediated introduction well beyond the natural distribution limit for this species) it is important to state that there is some uncertainty as to the gene flow status of the other comparisons used in our study. However, with that said, the moderating effect of gene flow was also apparent when simulating divergence.

Given that the silvereye population in French Polynesia was the product of a human introduction and therefore likely had lower effective founder size than natural colonisations involving large expanding populations of silvereyes (Sendell-Price *et al.* 2020), we

cannot rule out that the more wide-spread genomic divergence observed for the South Island vs. French Polynesia population comparison (when compared to other early stage comparison: South Island vs. Chatham Island) is the product of a potential severe population bottleneck during introduction. The potential for population founding to accelerate the accumulation of genomic differences has been highlighted previously (Ravinet *et al.* 2017), and in the silvereye system could be addressed through genomic comparisons of a number of other populations with varying intensities of population bottlenecks (Estoup and Clegg 2003).

Divergence of silvereye chromosome Z was elevated beyond that observed for autosomes, as seen in a number of other avian divergence studies (Storchová *et al.* 2010; Sætre and Sæther 2010; Ruegg *et al.* 2014; Oyler-McCance *et al.* 2015). Sex chromosomes are thought to contain a large number of genes related to sexual selection and reproductive isolation, and as such have been proposed to play a disproportionately large role in speciation (Dobzhansky 1974; Coyne 1985; Ellegren 2009; Charlesworth *et al.* 2017). Despite this, no genomic islands, and only a handful of significant outlier SNP, were identified within chromosome Z. Instead, the elevated levels of divergence may be partly explained by the lower effective population size of the Z chromosome (3/4 of an autosome), amplifying the effects of genetic drift (Mank *et al.* 2010).

### The dynamics of genomic islands of divergence and valleys of similarity

Our results run counter to several patterns expected from the divergence hitchhiking model of speciation. First, genomic islands of divergence were seen in similar frequency in populations diverging both with and without gene flow. This adds to a number of empirical studies in natural systems where genomic islands have been observed in populations that are, or are assumed to be, genetically isolated

**Table 4** Size of genomic islands and genomic valleys occurring at the same location across multiple population comparisons

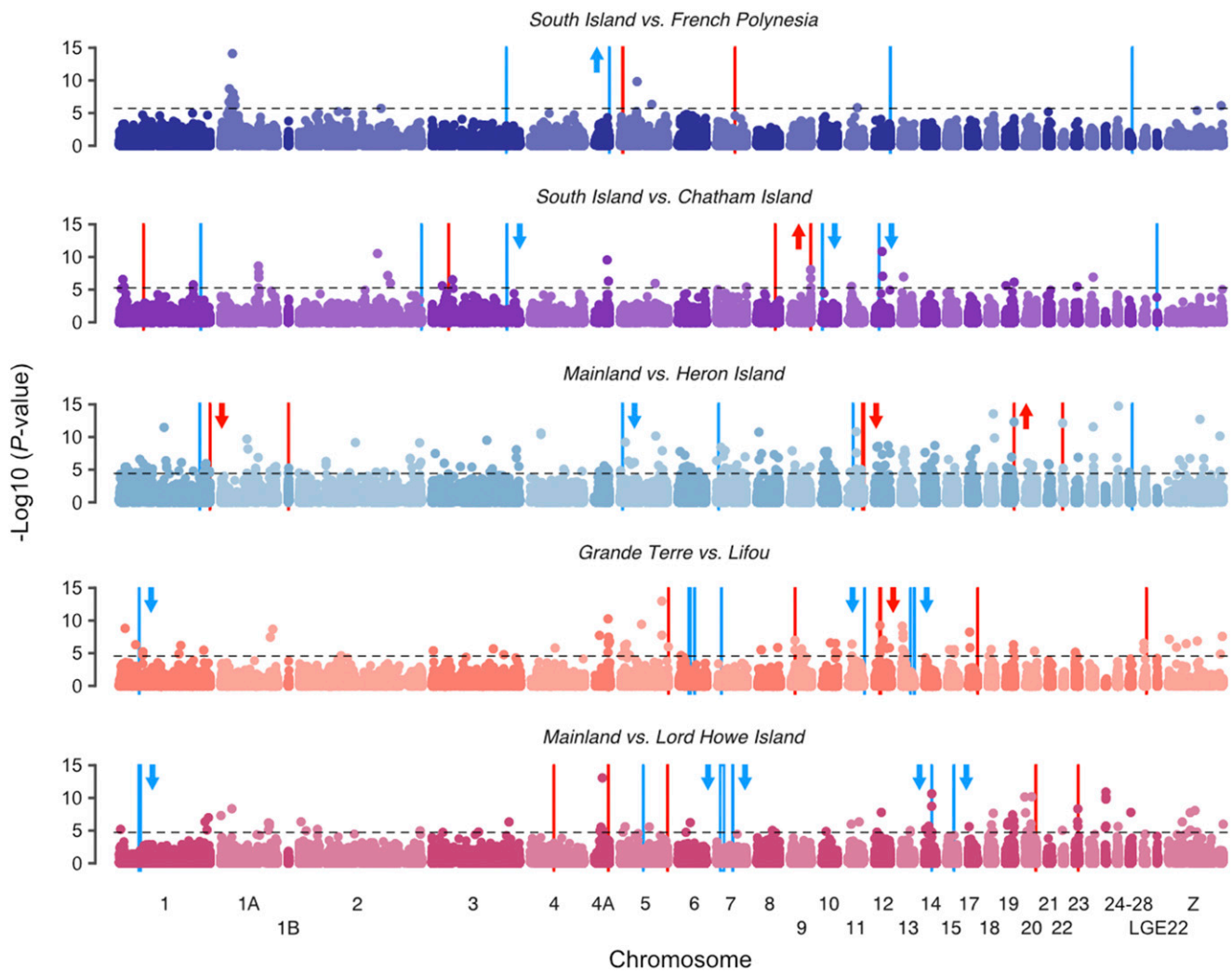
Divergence timeframe	Mode of divergence	Comparison	Genomic island size (kb)			Genomic valley size (kb)	
			5	1	7	1	7
Early stage	Gene flow	SI vs. CI	—	—	—	—	—
Early stage	No-gene flow	SI vs. FP	—	—	—	—	—
Mid stage	Gene flow	ML vs. HI	—	—	—	—	—
Late stage	Gene flow	GT vs. LF	100	250	450	—	—
Late stage	No-gene flow	ML vs. LH	550	1,850	3,900	—	—

■ **Table 5 Outlier SNP summaries.** Number of outlier SNPs identified using *PCAdapt* and the number located within genomic islands and genomic valleys per population comparison

Divergence timeframe	Mode of divergence	Comparison	No. outliers identified	No. outliers in genomic islands	No. outliers in genomic valleys
Early stage	Gene flow	SI vs. CI	14	0	0
Early stage	No-gene flow	SI vs. FP	28	4	0
Mid stage	Gene flow	ML vs. HI	235	4	0
Late stage	Gene flow	GT vs. LF	89	4	0
Late stage	No-gene flow	ML vs. LH	76	0	1

(Renaut *et al.* 2013; Marques *et al.* 2016; Han *et al.* 2017; Van Doren *et al.* 2017; Zhang *et al.* 2017). Second, genomic islands were not wider on average for older divergence timeframes, which does not align with the prediction that genomic islands expand with time as linkage disequilibrium facilitates divergence of neutral and weakly selected loci via divergence hitchhiking (Smadja *et al.* 2008; Via and West 2008; Via 2012). A small number of recent studies have also failed to find a positive association between divergence time and

average genomic island size (*Timema* stick insects (Riesch *et al.* 2017); Darwin's finches (Han *et al.* 2017); and Yellow-rumped warblers (Toews *et al.* 2016)). Taken together, growth of genomic islands may not be pervasive in the generation of genome-wide divergence, with the accumulation of divergence in our example instead taking place largely outside of the context of genomic islands. However, given that we have identified relatively few genomic islands we may lack the statistical power needed to detect differences in genomic island size



**Figure 6** Manhattan plot of negative  $\log_{10}(P\text{-values})$  estimated using *PCAdapt*. Dashed line indicates  $P$ -value threshold above which SNPs are determined outliers (FDR = 0.01). 50kb windows with elevated differentiation (genomic islands) are highlighted in red and regions with low differentiation (genomic valleys) are highlighted in blue. Genomic islands and genomic valleys are based on  $F_{ST}$  and locations are the same as those shown in Figure 4. Arrows beside genomic islands/valleys indicate if  $d_{xy}$  within these regions was significantly elevated (upwards pointing arrows) or decreased (downward pointing arrows) compared to chromosomal background levels. Significance determined using Wilcoxon signed-rank tests. Chromosomes are numbered according to the zebra finch nomenclature.

■ **Table 6** Patterns of differentiation in genes linked to body/bill size differentiation in birds

Gene	Chromosome	Source	Associated with bill or body size?	Within genomic island?	Mapped to by outlier SNPs?
ALX1	1A	<sup>a</sup>	Bill	N	N
ALX4	5	<sup>b</sup>	Bill	N	N
BMP4	5	<sup>c, d</sup>	Bill	N	N
BMPR1A	6	<sup>b</sup>	Bill	N	Y (SI vs. FP and ML vs. HI)
CACNA1G	18	<sup>e</sup>	Bill	N	N
COL17A1	6	<sup>f</sup>	Body	N	N
COL4A5	4A	<sup>b</sup>	Bill	N	N
COL6A3	7	<sup>f</sup>	Body	N	N
FGF10	Z	<sup>a</sup>	Bill	N	N
FOXC1	2	<sup>a</sup>	Bill	N	N
GSC	5	<sup>a</sup>	Bill	N	N
HMGA2	1A	<sup>g</sup>	Bill	N	N
IGF1	1A	<sup>h</sup>	Bill	N	N
INHBA	2	<sup>b</sup>	Bill	N	N
ITPR2	1A	<sup>f</sup>	Body	N	N
ITPR3	26	<sup>f</sup>	Body	N	N
NELL1	5	<sup>b</sup>	Bill	N	N
RDH14	3	<sup>a</sup>	Bill	N	N
SATB2	7	<sup>b</sup>	Bill	N	N
SIX2	3	<sup>b</sup>	Bill	N	N
TRPS1	2	<sup>b</sup>	Bill	N	N
VPS13B	2	<sup>a, b</sup>	Bill	N	N

<sup>a</sup>Lamichhaney *et al.* (2015).

<sup>b</sup>Bosse *et al.* (2017).

<sup>c</sup>Abzhanov *et al.* (2004).

<sup>d</sup>Campàs *et al.* (2010).

<sup>e</sup>Cheng *et al.* (2017).

<sup>f</sup>Cornetti *et al.* (2015).

<sup>g</sup>Lamichhaney *et al.* (2016).

<sup>h</sup>vonHoldt *et al.* (2018).

across different divergence timeframes and gene flow scenarios. We therefore caution readers against over-interpreting this primarily negative result.

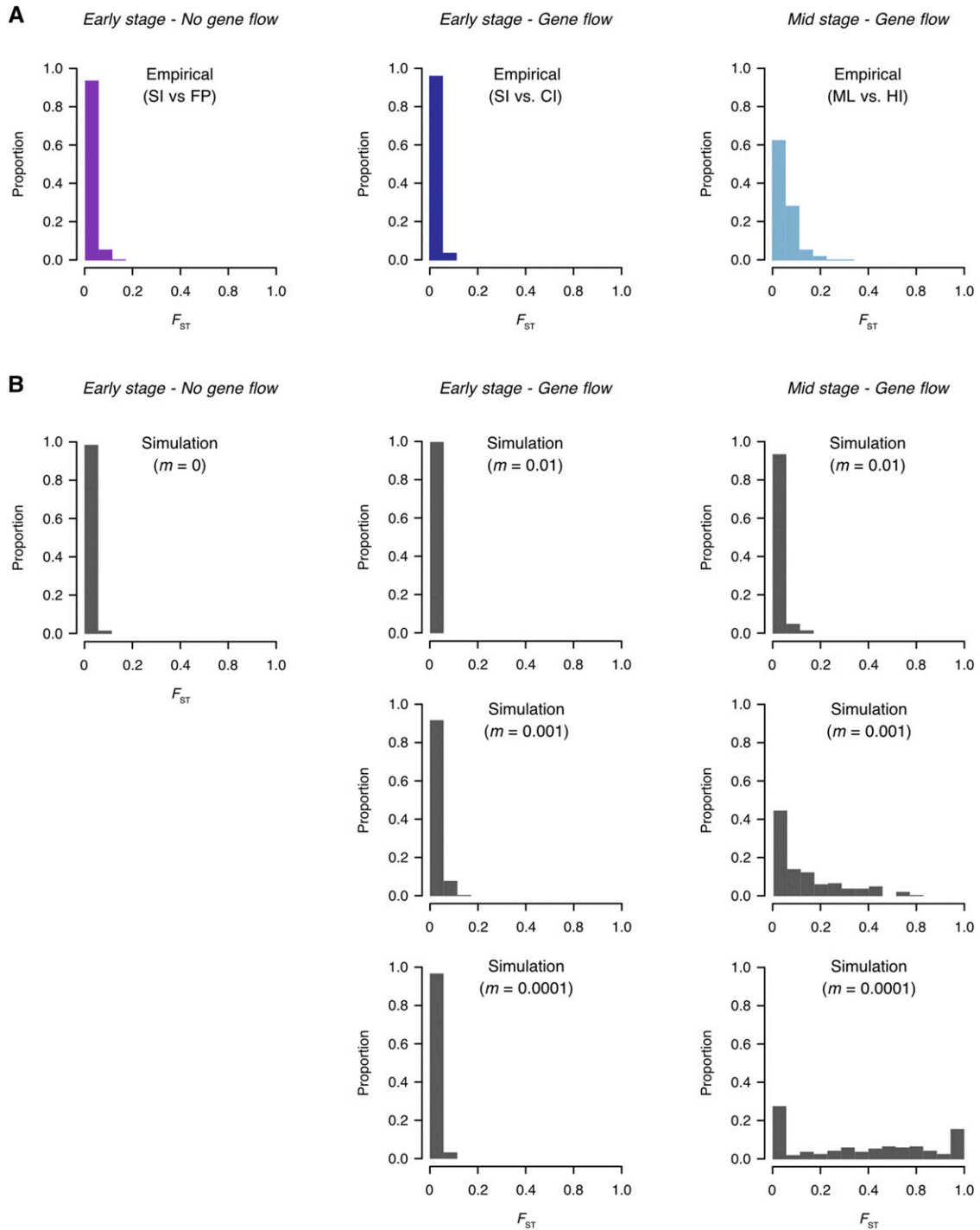
Associating average genomic island size with divergence time may be too coarse an approach, and comparisons of genomic islands at specific genomic locations may be a better method to understand how time and gene flow shape growth of genomic islands. In our case, only one genomic island was located at overlapping genomic positions across population comparisons, however this genomic island did not occur across different divergence timeframes. Whole genome analysis would likely provide more shared genomic islands, but currently we lack support for the idea that expansion of genomic islands of divergence at specific locations over time occurs.

Contrary to the expectation that genomic islands of divergence contain loci under strong divergent selection (Wu 2001), we find that genomic islands regularly occur in regions of the *Z. lateralis* genome where outlier SNPs putatively under directional selection were not identified. A similar result was reported in diverging Swainson's thrush *Catharus ustulatus* subspecies (Ruegg *et al.* 2014). Hence, it appears that genomic islands of divergence are not always seeded by directional selection, but whether this is frequently the case requires a broader number of empirical examples. The formation of genomic islands without obvious evidence of directional selection highlights the need for alternative explanations for their formation (Turner *et al.* 2005; Hohenlohe *et al.* 2010; Feder *et al.* 2013). One such explanation is that genomic islands of divergence may arise in regions of low recombination such as centromeres (Carneiro *et al.* 2009; Renaut *et al.* 2013; Burri *et al.* 2015). However, lack of recombination and karyotype maps for silvereyes at present prevents formal testing of this hypothesis here. In addition to variation in recombination rate,

other genomic features that may contribute to the formation of genomic islands include variation in gene density and variation in mutation rate along the genome (Ravinet *et al.* 2017). Whole genome sequencing would help to address the contribution of these genomic features.

A comparison of absolute and relative measures of divergence (*e.g.*,  $d_{xy}$  and  $F_{ST}$  respectively) can provide further insight into the formation of genomic islands (Nachman and Payseur 2012; Han *et al.* 2017; Delmore *et al.* 2018). Unlike  $F_{ST}$ , which is consistently elevated around loci under directional selection,  $d_{xy}$  may be elevated, reduced or unchanged. Under divergence with gene flow, genomic islands with elevated measures of both relative and absolute divergence are expected to form around regions where gene flow is disadvantageous, such as around regions containing variants involved in local adaptation or reproductive isolation (Han *et al.* 2017). We observed such a pattern in only one of the three comparisons diverging with gene flow. In the absence of gene flow, selective sweeps, which convert between-population variation into fixed differences, are expected to produce regions of elevated  $F_{ST}$  but not  $d_{xy}$  (Nachman and Payseur 2012; Cruickshank and Hahn 2014). Such islands were frequently observed across all comparisons, suggesting that genomic islands arise frequently within regions where gene flow is absent. Genomic islands (based on  $F_{ST}$ ) but with reduced  $d_{xy}$ , were also frequently observed. Such a pattern is thought to be caused by background selection and recurrent selective sweeps, both of which result in locally reduced levels of genetic variation (Seehausen *et al.* 2014; Wolf and Ellegren 2016; Ma *et al.* 2017).

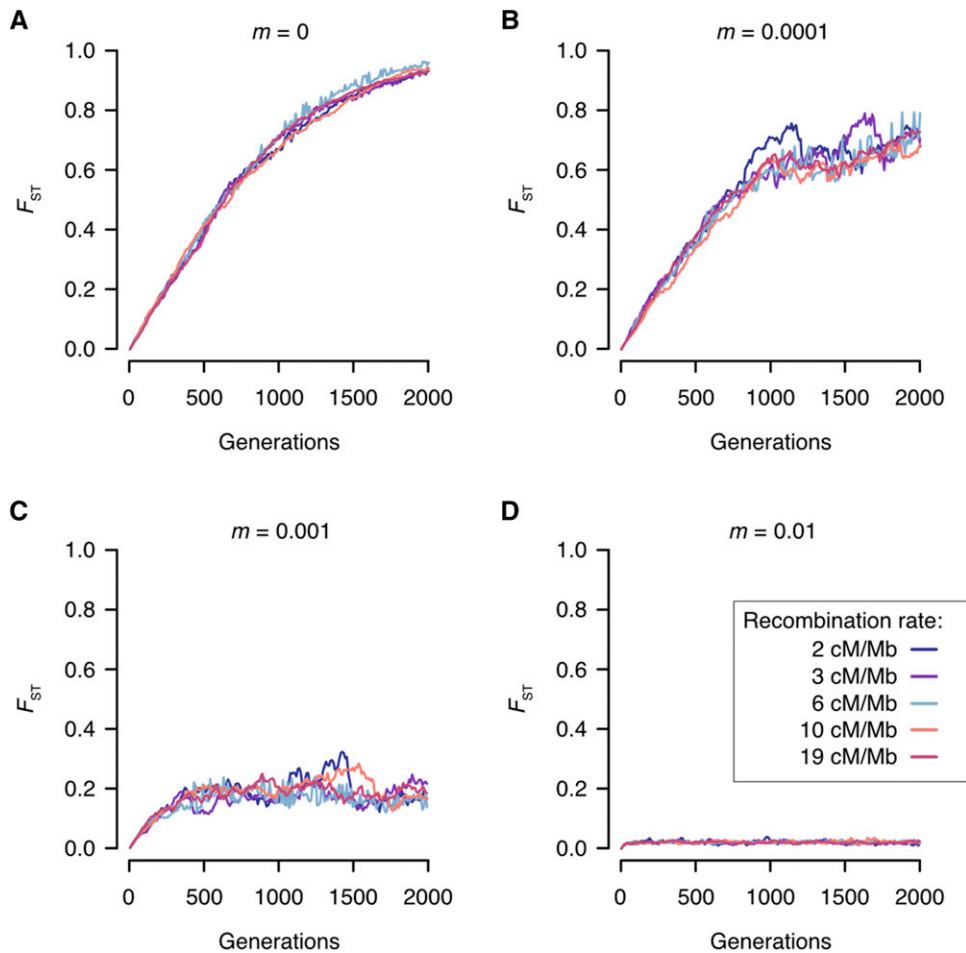
Genomic valleys play an important role in shaping the genomic landscape across the speciation continuum, occurring across all population comparisons. In particular, their occurrence during the



**Figure 7** Comparison of observed and simulated  $F_{ST}$  distributions. (A) Frequency distributions of pairwise differentiation ( $F_{ST}$ ) for empirical data (chromosome 5 only). (B) Frequency distributions of pairwise differentiation ( $F_{ST}$ ) for simulated data. The  $F_{ST}$  values are calculated in 500kb windows. Simulations were conducted using a recombination rate of 3 cM/Mb (the distance between loci for which the expected average number of intervening chromosomal crossovers in a single generation is 0.01). Simulations of divergence with gene flow were conducted under different migration rates ( $m = 0.01$ ,  $m = 0.001$ , and  $m = 0.0001$ ). Simulation timeframes matched that for each empirical comparison (Early stage - no gene flow: 27 generations; Early stage - gene flow: 47 generations; Mid stage - gene flow: 1,000 generations).

late stage of divergence may aid maintenance of heterogeneity by slowing the approach to genome-wide divergence at some genomic regions. Should genomic valleys remain present in the very advanced stages of divergence - *i.e.*, comparisons at the species level - these regions likely contribute to the formation of the tail of

extreme values that characterizes the left skewed distributions observed during the late stage of divergence such as those observed between diverging *Ficedula* flycatcher species by Burri *et al.* (2015). Comparisons between *Zosterops* species are needed to confirm this expectation.



**Figure 8** Expected  $F_{ST}$  values for 2,000 generations of simulated neutral divergence. (A) During divergence without gene flow ( $m = 0$ )  $F_{ST}$  rapidly increases, achieving fixation at 2,000 generations under all recombination rates; (B) During divergence with low levels of migration ( $m = 0.0001$ )  $F_{ST}$  increases rapidly, approaching fixation at 2,000 generations for all recombination rates; (C) During divergence with intermediate levels of migration ( $m = 0.001$ )  $F_{ST}$  increases rapidly during the initial  $\sim 300$  generations, after which it continues to fluctuate around 0.2 with some variation between recombination rates; and (D) During divergence with high levels of gene flow ( $m = 0.01$ ) low levels of  $F_{ST}$  are maintained across all 2,000 generations of divergence.

A comparison of the nature of genomic valleys across the speciation continuum and gene flow contexts allows us to comment on mechanisms proposed to generate genomic valleys. Although genomic valleys frequently showed a pattern of reduced  $F_{ST}$  and  $d_{xy}$ , as would be expected if genomic valleys form at loci under parallel selection (Roesli *et al.* 2014) or purifying selection (Cvijović *et al.* 2018), in few cases did the position of an outlier SNP correspond to the position of a genomic valleys. As such, at present we have limited evidence to confirm that genomic valleys frequently arise due to effects of parallel or purifying selection. Second, as the size of genomic valleys was found to be stable over time, we do not find support that genomic valleys are solely the product of incomplete lineage sorting at neutral loci, as under this mechanism genomic valleys would be expected to become smaller as shared ancestral variation breaks down over time (Stölting *et al.* 2013).

### Repeated evolutionary patterns

While we found no evidence that candidate genes thought to be associated with body and/or bill size differences in passerines were concentrated within genomic islands, we did find evidence that one such gene may be implicated in the repeated evolutionary pattern seen in island silvereyes. Bone morphogenetic protein receptor, type 1A (*BMPRIA*) was identified as containing outlier SNPs potentially under selection in both the South Island vs. French Polynesia and Mainland vs. Heron Island comparisons. This gene is associated with the ontology term “palate development” [Gene Ontology (GO) 0060021] and has been previously identified as a candidate gene underlying bill length variation in great tits (*Parus major*) (Bosse *et al.*

2017). Interestingly, both French Polynesia and Heron Island populations have increased bill sizes (and overall body size) when compared to their respective source populations (Clegg *et al.* 2010; Sendell-Price *et al.* 2020). A further gene within the Bone Morphogenetic Protein (BMP) pathway, *BMP4*, has been strongly associated with bill shape variation in Darwin’s finches (genus: *Geospiza*) (Abzhanov *et al.* 2004).

### CONCLUSION

While the goal of identifying unambiguous signatures of particular evolutionary processes in patterns of genomic divergence remains elusive, the addition of empirical studies of well-characterized systems provides valuable insight into the nature of divergence across the speciation continuum. In the silvereye system, gene flow is clearly important in shaping genome-wide divergence. While these results provide an excellent baseline for inferring the role of gene flow in other silvereye populations, equivalent calibrations would likely be needed on a species-by-species basis for the role of gene flow in a population of unknown history to be assessed. Genome-wide divergence in silvereyes does not hinge on the formation and growth of genomic islands. This is at odds with the intuitive and until recently, frequently invoked verbal model describing the accumulation of genomic divergence during the speciation continuum. Instead, differences are spread in a remarkably even way across the genome. The joint empirical and theoretical approach we present offers a potentially powerful tool to test a range of hypotheses about the mechanisms that underlie genomic speciation.

## ACKNOWLEDGMENTS

This work was funded by a grant from the John Fell Oxford University Press Research Fund to T.C. and S.M.C., with additional support from an anonymous donation to K.C.R. and a Natural Environment Research Council (NERC) studentship awarded to A.T.S.P. C.S.Q. acknowledges support from the Swiss National Science Foundation (P2GEP3\_168973 and P400PB\_183930). Sample collection was conducted with the permission of the governments of Australia, French Polynesia, New Caledonia, New Zealand, and Vanuatu. The sample collection used in this manuscript covers an 18-year period and we are thankful to the many people who facilitated the fieldwork including help with logistics, field assistance, land access and accommodation. We thank the Museum d'Histoire Naturelle de Genève, Alice Cibois, Nick Clark, Rob Fleischer, and Ally Phillimore for providing additional samples. The authors would like to thank Bruce Robertson, Fiona Robertson, Ari Robertson, and Erik Sandvig for assisting with Heron Island sample collection. The authors would like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work. We thank anonymous reviewers who provided helpful comments on earlier drafts of this manuscript.

## LITERATURE CITED

- Abzhanov, A., M. Protas, B. R. Grant, P. R. Grant, and C. J. Tabin, 2004 Bmp4 and Morphological Variation of Beaks in Darwin's Finches. *Science* 305: 1462–1465. <https://doi.org/10.1126/science.1098095>
- Alexander D.H., J. Novembre, and K. Lange, 2009 Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19: 1655–1664
- Ali, O. A., S. M. O. Rourke, S. J. Amish, M. H. Meek, G. Luikart *et al.*, 2016 RAD Capture (Rapture): Flexible and efficient sequence-based genotyping. *Genetics* 202: 389–400. <https://doi.org/10.1534/genetics.115.183665>
- Andrew, R. L., and L. H. Rieseberg, 2013 Divergence is focused on few genomic regions early in speciation: Incipient speciation of sunflower ecotypes. *Evolution*. 67: 2468–2482. <https://doi.org/10.1111/evo.12106>
- Backström, N., W. Forstmeier, H. Schielzeth, H. Mellenius, K. Nam *et al.*, 2010 The recombination landscape of the zebra finch *Taeniopygia guttata* genome. *Genome Res.* 20: 485–495. <https://doi.org/10.1101/gr.101410.109>
- Bosse, M., L. G. Spurgin, V. N. Laine, E. F. Cole, J. A. Firth *et al.*, 2017 Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science* 358: 365–368. <https://doi.org/10.1126/science.aal3298>
- Brook, B. W., and J. Kikkawa, 1998 Examining threats faced by island birds: A population viability analysis on the Capricorn silveryeye using long-term data. *J. Appl. Ecol.* 35: 491–503. <https://doi.org/10.1046/j.1365-2664.1998.3540491.x>
- Burri, R., A. Nater, T. Kawakami, and C. F. Mugal, 2015 Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Res.* 25: 1656–1665. <https://doi.org/10.1101/gr.196485.115>
- Marie Curie SPECIATION Network; Butlin, R., A. Debelle, C. Kerth, R. R. Snook, L. W. Beukeboom *et al.*, 2012 What do we need to know about speciation? *Trends Ecol. Evol.* 27: 27–39. <https://doi.org/10.1016/j.tree.2011.09.002>
- Campàs, O., R. Mallarino, A. Herrel, A. Abzhanov, and M. P. Brenner, 2010 Scaling and shear transformations capture beak shape variation in Darwin's finches. *Proc. Natl. Acad. Sci. USA* 107: 3356–3360. <https://doi.org/10.1073/pnas.0911575107>
- Carneiro, M., N. Ferrand, and M. W. Nachman, 2009 Recombination and speciation: Loci near centromeres are more differentiated than loci near telomeres between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics* 181: 593–606. <https://doi.org/10.1534/genetics.108.096826>
- Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko, 2013 Stacks: An analysis tool set for population genomics. *Mol. Ecol.* 22: 3124–3140. <https://doi.org/10.1111/mec.12354>
- Charlesworth, D., 2006 Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet.* 2: e64. <https://doi.org/10.1371/journal.pgen.0020064>
- Charlesworth, B., J. A. Coyne, and N. H. Barton, 2017 The relative rates of evolution of sex chromosomes and autosomes. *Am. Nat.* 130: 113–146. <https://doi.org/10.1086/284701>
- Chaves, J. A., E. A. Cooper, A. P. Hendry, J. Podos, L. F. De León *et al.*, 2016 Genomic variation at the tips of the adaptive radiation of Darwin's finches. *Mol. Ecol.* 25: 5282–5295. <https://doi.org/10.1111/mec.13743>
- Cheng, Y., B. Gao, H. Wang, N. Han, S. Shao *et al.*, 2017 Evolution of beak morphology in the ground tit revealed by comparative transcriptomics. *Front. Zool.* 14: 58. <https://doi.org/10.1186/s12983-017-0245-6>
- Clegg, S. M., S. M. Degnan, J. Kikkawa, C. Moritz, A. Estoup *et al.*, 2002a Genetic consequences of sequential founder events by an island-colonising bird. *Proc. Natl. Acad. Sci. USA* 99: 8127–8132. <https://doi.org/10.1073/pnas.102583399>
- Clegg, S. M., S. M. Degnan, C. Moritz, A. Estoup, J. Kikkawa *et al.*, 2002b Microevolution in island forms: The roles of drift and directional selection in morphological divergence of a passerine bird. *Evolution* 56: 2090–2099. <https://doi.org/10.1111/j.0014-3820.2002.tb00134.x>
- Clegg, S. M., F. D. Frentiu, J. Kikkawa, G. Tavecchia, and I. P. F. Owens, 2008 4000 Years of phenotypic change in an island bird: Heterogeneity of selection over three microevolutionary timescales. *Evolution* 62: 2393–2410. <https://doi.org/10.1111/j.1558-5646.2008.00437.x>
- Cornetti, L., L. M. Valente, L. T. Dunning, X. Quan, R. A. Black *et al.*, 2015 The genome of the “great speciator” provides insights into bird diversification. *Genome Biol. Evol.* 7: 2680–2691. <https://doi.org/10.1093/gbe/evv168>
- Coyne, J. A., 1985 The genetic basis of Haldane's rule. *Nature* 314: 736–738. <https://doi.org/10.1038/314736a0>
- Cruikshank, T. E., and M. W. Hahn, 2014 Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* 23: 3133–3157. <https://doi.org/10.1111/mec.12796>
- Cvijović, I., B. H. Good, and M. M. Desai, 2018 The effect of strong purifying selection on genetic diversity. *Genetics* 209: 1235–1278. <https://doi.org/10.1534/genetics.118.301058>
- Darwin, C., 1859 *On the origin of species by means of natural selection*, Murray, London.
- Davey, J. L., and M. W. Blaxter, 2010 RADseq: Next-generation population genetics. *Brief. Funct. Genomics* 9: 416–423. <https://doi.org/10.1093/bfpg/elq031>
- Degnan, S. M., and C. Moritz, 1992 Phylogeography of mitochondrial DNA in two species of white-eyes in Australia. *Auk* 109: 800–811. <https://doi.org/10.1023/1023074088154>
- Delmore, K. E., J. S. Lugo Ramos, B. M. Van Doren, M. Lundberg, S. Bensch *et al.*, 2018 Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds. *Evol. Lett.* 2: 76–87. <https://doi.org/10.1002/evl3.46>
- Dobzhansky, T., 1974 Genetic analysis of hybrid sterility within the species *Drosophila pseudoobscura*. *Hereditas* 77: 81–88. <https://doi.org/10.1111/j.1601-5223.1974.tb01356.x>
- Ellegren, H., 2009 Genomic evidence for a large-Z effect. *Proc. Biol. Sci.* 276: 361–366. <https://doi.org/10.1098/rspb.2008.1135>
- Ellegren, H., 2010 Evolutionary stasis: The stable chromosomes of birds. *Trends Ecol. Evol.* 25: 283–291. <https://doi.org/10.1016/j.tree.2009.12.004>
- Estoup, A., and S. M. Clegg, 2003 Bayesian inferences on the recent island colonisation history by the bird *Zosterops lateralis lateralis*. *Mol. Ecol.* 12: 657–674. <https://doi.org/10.1046/j.1365-294X.2003.01761.x>
- Feder, J. L., S. P. Egan, and P. Nosil, 2012 The genomics of speciation-with-gene-flow. *Trends Genet.* 28: 342–350. <https://doi.org/10.1016/j.tig.2012.03.009>
- Feder, J. L., S. M. Flaxman, S. P. Egan, A. A. Comeault, and P. Nosil, 2013 Geographic mode of speciation and genomic divergence. *Annu. Rev. Ecol. Evol. Syst.* 44: 73–97. <https://doi.org/10.1146/annurev-ecolsys-110512-135825>

- Feder, J. L., and P. Nosil, 2010 The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*. 64: 1729–1747. <https://doi.org/10.1111/j.1558-5646.2009.00943.x>
- Feulner, P. G. D., F. J. J. Chain, M. Panchal, Y. Huang, C. Eizaguirre *et al.*, 2015 Genomics of divergence along a continuum of parapatric population differentiation. *PLoS Genet*. 11: e1004966. <https://doi.org/10.1371/journal.pgen.1004966>
- Flaxman, S. M., J. L. Feder, and P. Nosil, 2013 Genetic hitchhiking and the dynamic build-up of genomic divergence during speciation with gene flow. *Evolution*. 67: 2577–2591. <https://doi.org/10.1111/evo.12055>
- Flaxman, S. M., A. C. Wacholder, J. L. Feder, and P. Nosil, 2014 Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Mol. Ecol*. 23: 4074–4088. <https://doi.org/10.1111/mec.12750>
- Grabherr, M. G. P., M. Russell, E. Meyer, J. Mauceli, F. Alföldi *et al.*, 2010 Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* 26: 1145–1151. <https://doi.org/10.1093/bioinformatics/btq102>
- Han, F., S. Lamichhaney, R. B. Grant, P. R. Grant, L. Andersson *et al.*, 2017 Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Res*. 27: 1004–1015. <https://doi.org/10.1101/gr.212522.116>
- Hofer, T., M. Foll, and L. Excoffier, 2012 Evolutionary forces shaping genomic islands of population differentiation in humans. *BMC Genomics* 13: 107. <https://doi.org/10.1186/1471-2164-13-107>
- Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson *et al.*, 2010 Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet*. 6: e1000862. <https://doi.org/10.1371/journal.pgen.1000862>
- Hopley, D., 1982 *The geomorphology of the Great Barrier Reef: Quaternary development of coral reefs*, John Wiley and Sons, New York.
- Kawakami, T., L. Smeds, N. Backström, A. Husby, A. Qvarnström *et al.*, 2014 A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Mol. Ecol*. 23: 4035–4058. <https://doi.org/10.1111/mec.12810>
- Kearns, A. M., M. Restani, I. Szabo, A. Schröder-Nielsen, J. A. Kim *et al.*, 2018 Genomic evidence of speciation reversal in ravens. *Nat. Commun*. 9: 906. <https://doi.org/10.1038/s41467-018-03294-w>
- Kikkawa, J., and J. M. Wilson, 1983 Breeding and dominance among the Heron Island silvereyes *Zosterops lateralis chlorocephala*. *Emu* 83: 181–198. <https://doi.org/10.1071/MU9830181>
- Lamichhaney, S., J. Berglund, M. S. Almén, K. Maqbool, M. Grabherr *et al.*, 2015 Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* 518: 371–375. <https://doi.org/10.1038/nature14181>
- Lamichhaney, S., F. Han, J. Berglund, C. Wang, M. S. Almén *et al.*, 2016 A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science* 352: 470–474. <https://doi.org/10.1126/science.aad8786>
- Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9: 357–359. <https://doi.org/10.1038/nmeth.1923>
- Lotterhos, K. E., 2019 The effect of neutral recombination variation on genome scans for selection. *G3 (Bethesda)* 9: 1851–1867.
- Luu, K., E. Bazin, and M. G. B. Blum, 2017 pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Mol. Ecol. Resour*. 17: 67–77. <https://doi.org/10.1111/1755-0998.12592>
- Ma, T., K. Wang, Q. Hu, Z. Xi, D. Wan *et al.*, 2017 Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. *Proc. Natl. Acad. Sci. USA* 115: E236–E243. <https://doi.org/10.1073/pnas.1713288114>
- Mallet, J., 2008 Hybridization, ecological races and the nature of species: Empirical evidence for the ease of speciation. *Philos. Trans. R. Soc. B*. 363: 2971–2986. <https://doi.org/10.1098/rstb.2008.0081>
- Mank, J. E., K. Nam, and H. Ellegren, 2010 Faster-Z evolution is predominantly due to genetic drift. *Mol. Biol. Evol*. 27: 661–670. <https://doi.org/10.1093/molbev/msp282>
- Marques, D. A., K. Lucek, J. I. Meier, S. Mwaiko, C. E. Wagner *et al.*, 2016 Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLoS Genet*. 12: e1005887. <https://doi.org/10.1371/journal.pgen.1005887>
- Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, C. Salazar, J. R. Walters *et al.*, 2013 Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res*. 23: 1817–1828. <https://doi.org/10.1101/gr.159426.113>
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, and K. Cibulskis, 2010 The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 20: 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Mees, G. F., 1969 A systematic review of the Indo-Australian Zosteropidae. *Zool. Verh*. 102: 1–390.
- Monnet, C., J. C. Thibault, and A. Varney, 1993 Stability and changes during the twentieth century in the breeding landbirds of Tahiti (Polynesia). *Bird Conserv. Int*. 3: 261–280. <https://doi.org/10.1017/S0959279090002550>
- Nachman, M. W., and B. A. Payseur, 2012 Recombination rate variation and speciation: Theoretical predictions and empirical results from rabbits and mice. *Philos. Trans. R. Soc. B*. 367: 409–421. <https://doi.org/10.1098/rstb.2011.0249>
- Nei, M., 1987 *Molecular Evolutionary Genetics*, Columbia University Press, New York. <https://doi.org/10.7312/nei-92038>
- Nielsen, R., 2005 Molecular signatures of natural selection. *Annu. Rev. Genet*. 39: 197–218. <https://doi.org/10.1146/annurev.genet.39.073003.112420>
- Nosil, P., 2012 *Ecological Speciation*, Oxford University Press, Oxford. <https://doi.org/10.1093/acprof:osobl/9780199587100.001.0001>
- Nosil, P., and J. L. Feder, 2012 Genomic divergence during speciation: Causes and consequences. *Philos. Trans. R. Soc. B*. 367: 332–342. <https://doi.org/10.1098/rstb.2011.0263>
- Nosil, P., J. L. Feder, S. M. Flaxman, and Z. Gompert, 2017 Tipping points in the dynamics of speciation. *Nat. Ecol. Evol*. 1: 1. <https://doi.org/10.1038/s41559-016-0001>
- Nosil, P., D. J. Funk, and D. Ortiz-Barrientos, 2009 Divergent selection and heterogeneous genomic divergence. *Mol. Ecol*. 18: 375–402. <https://doi.org/10.1111/j.1365-294X.2008.03946.x>
- Oyler-McCance, S. J., R. S. Cornman, K. L. Jones, and J. A. Fike, 2015 Z chromosome divergence, polymorphism and relative effective population size in a genus of lekking birds. *Heredity* 115: 452–459. <https://doi.org/10.1038/hdy.2015.46>
- Powell, T. H. Q., G. R. Hood, M. O. Murphy, J. S. Heilveil, S. H. Berlocher *et al.*, 2013 Genetic divergence along the speciation continuum: The transition from host race to species in rhaagoletis (diptera: *Tephritidae*). *Evolution*. 67: 2561–2576. <https://doi.org/10.1111/evo.12209>
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira *et al.*, 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet*. 81: 559–575. <https://doi.org/10.1086/519795>
- Quilodrán, C. S., K. Ruegg, A. T. Sendell-Price, E. C. Anderson, T. Coulson *et al.*, 2020 The multiple population genetic and demographic routes to islands of genomic divergence. *Methods Ecol. Evol*. 11: 6–21. <https://doi.org/10.1111/2041-210X.13324>
- Ravinet, M., R. Faria, R. K. Butlin, J. Galindo, N. Bierne *et al.*, 2017 Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *J. Evol. Biol*. 30: 1450–1477. <https://doi.org/10.1111/jeb.13047>
- Renaut, S., C. J. Grassa, S. Yeaman, B. T. Moyers, Z. Lai *et al.*, 2013 Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nat. Commun*. 4: 1827. <https://doi.org/10.1038/ncomms2833>
- Ricklefs, R. E., and E. Bermingham, 2007 The causes of evolutionary radiations in archipelagoes: Passerine birds in the lesser Antilles. *Am. Nat*. 169: 285–297. <https://doi.org/10.1086/510730>
- Riesch, R., M. Muschick, D. Lindtke, R. Villoutreix, A. A. Comeault *et al.*, 2017 Transitions between phases of genomic differentiation during stick-insect speciation. *Nat. Ecol. Evol*. 1: 82. <https://doi.org/10.1038/s41559-017-0082>
- Robertson, B. C., S. M. Degnan, J. Kikkawa, and C. C. Moritz, 2001 Genetic monogamy in the absence of paternity guards: The Capricorn silvereye, *Zosterops lateralis chlorocephalus*, on Heron Island. *Behav. Ecol. Evol*. 12: 666–673. <https://doi.org/10.1093/beheco/12.6.666>



- Robinson-Wolrath, S. I., and I. P. F. Owens, 2003 Large size in an island-dwelling bird: Intraspecific competition and the dominance hypothesis. *J. Evol. Biol.* 16: 1106–1114. <https://doi.org/10.1046/j.1420-9101.2003.00615.x>
- Roesti, M., S. Gavrillets, A. P. Hendry, W. Salzburger, and D. Berner, 2014 The genomic signature of parallel adaptation from shared genetic variation. *Mol. Ecol.* 23: 3944–3956. <https://doi.org/10.1111/mec.12720>
- Roesti, M., A. P. Hendry, W. Salzburger, and D. Berner, 2012 Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Mol. Ecol.* 21: 2852–2862. <https://doi.org/10.1111/j.1365-294X.2012.05509.x>
- Ruegg, K., E. C. Anderson, J. Boone, J. Pouls, and T. B. Smith, 2014 A role for migration-linked genes and genomic islands in divergence of a songbird. *Mol. Ecol.* 23: 4757–4769. <https://doi.org/10.1111/mec.12842>
- Sætre, G. P., and S. A. Sæther, 2010 Ecology and genetics of speciation in *Ficedula flycatchers*. *Mol. Ecol.* 19: 1091–1106. <https://doi.org/10.1111/j.1365-294X.2010.04568.x>
- Sandvig, E. M., T. Coulson, J. Kikkawa, and S. M. Clegg, 2017 The influence of climatic variation and density on the survival of an insular passerine *Zosterops lateralis*. *PLoS One* 12: e0176360. <https://doi.org/10.1371/journal.pone.0176360>
- Schilling, P. M., P. S. Mullen, M. Kronforst, J. R. Safran, P. Nosil *et al.*, 2018 Transitions from single- to multi-locus processes during speciation with gene flow. *Genes (Basel)* 9: 274. <https://doi.org/10.3390/genes9060274>
- Seehausen, O., R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman *et al.*, 2014 Genomics and the origin of species. *Nat. Rev. Genet.* 15: 176–192. <https://doi.org/10.1038/nrg3644>
- Sendell-Price, A. T., K. C. Ruegg, and S. M. Clegg, 2020 Rapid morphological divergence following a human-mediated introduction: The role of drift and directional selection. *Heredity* 124: 535–549. <https://doi.org/10.1038/s41437-020-0298-8>
- Seutin, G., B. White, and P. T. Boag, 1991 Preservation of avian blood and tissue samples for DNA analyses. *Can. J. Zool.* 69: 82–90. <https://doi.org/10.1139/z91-013>
- Smadja, C., J. Galindo, and R. Butlin, 2008 Hitching a lift on the road to speciation. *Mol. Ecol.* 17: 4177–4180. <https://doi.org/10.1111/j.1365-294X.2008.03917.x>
- Stölting, K. N., R. Nipper, D. Lindtke, C. Caseys, S. Waeber *et al.*, 2013 Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Mol. Ecol.* 22: 842–855. <https://doi.org/10.1111/mec.12011>
- Storchová, R., J. Reif, and M. W. Nachman, 2010 Female heterogamety and speciation: Reduced introgression of the Z chromosome between two species of nightingales. *Evolution.* 64: 456–471. <https://doi.org/10.1111/j.1558-5646.2009.00841.x>
- Supple, M. A., R. Papa, H. M. Hines, W. O. McMillan, and B. A. Counterman, 2015 Divergence with gene flow across a speciation continuum of *Heliconius* butterflies. *BMC Evol. Biol.* 15: 204. <https://doi.org/10.1186/s12862-015-0486-y>
- Thibault, J. C., and A. Cibois, 2017 *Birds of eastern Polynesia: A biogeographic atlas*, Lynx Publishing, Ontario.
- Toews, D. P. L., A. Brelsford, C. Grossen, B. Milá, and D. E. Irwin, 2016 Genomic variation across the yellow-rumped warbler species complex. *Auk* 133: 698–717. <https://doi.org/10.1642/AUK-16-61.1>
- Turner, T. L., M. W. Hahn, and S. V. Nuzhdin, 2005 Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol.* 3: e285. <https://doi.org/10.1371/journal.pbio.0030285>
- Van Doren, B. M., L. Campagna, B. Helm, J. C. Illera, I. J. Lovette *et al.*, 2017 Correlated patterns of genetic diversity and differentiation across an avian family. *Mol. Ecol.* 26: 3982–3997. <https://doi.org/10.1111/mec.14083>
- Via, S., 2012 Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philos. Trans. R. Soc. B.* 367: 451–460. <https://doi.org/10.1098/rstb.2011.0260>
- Via, S., and J. West, 2008 The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol. Ecol.* 17: 4334–4345. <https://doi.org/10.1111/j.1365-294X.2008.03921.x>
- Vijay, N., C. M. Bossu, J. W. Poelstra, M. H. Weissensteiner, A. Suh *et al.*, 2016 Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nat. Commun.* 7: 13195. <https://doi.org/10.1038/ncomms13195>
- vonHoldt, B. M., R. Y. Kartzinel, C. D. Huber, V. L. Underwood, Y. Zhen *et al.*, 2018 Growth factor gene IGF1 is associated with bill size in the black-bellied seedcracker *Pyrenestes ostrinus*. *Nat. Commun.* 9: 4855. <https://doi.org/10.1038/s41467-018-07374-9>
- Wang, J., N. R. Street, D. G. Scofield, and P. K. Ingvarsson, 2016 Variation in linked selection and recombination drive genomic divergence during allopatric speciation of European and American aspens. *Mol. Biol. Evol.* 33: 1754–1767. <https://doi.org/10.1093/molbev/msw051>
- Wolf, J. B. W., and H. Ellegren, 2016 Making sense of genomic islands of differentiation in light of speciation. *Nat. Rev. Genet.* 18: 87–100. <https://doi.org/10.1038/nrg.2016.133>
- Wright, S., 1931 Evolution in Mendelian populations. *Genetics* 16: 97–159.
- Wright, S., 1951 The genetical structure of populations. *Ann. Eugen.* 15: 323–354. <https://doi.org/10.1111/j.1469-1809.1949.tb02451.x>
- Wu, C. I., 2001 The genic view of the process of speciation. *J. Evol. Biol.* 14: 851–865. <https://doi.org/10.1046/j.1420-9101.2001.00335.x>
- Zhang, D., G. Song, B. Gao, Y. Cheng, Y. Qu *et al.*, 2017 Genomic differentiation and patterns of gene flow between two long-tailed tit species (*Aegithalos*). *Mol. Ecol.* 26: 6654–6665. <https://doi.org/10.1111/mec.14383>

Communicating editor: A. Sethuraman