

The Genomics of Diversification
in South Pacific Silvereyes
(*Zosterops lateralis*)



Ashley T. Sendell-Price

Linacre College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2019

This thesis is dedicated to
Leslie Kennedy
I miss your stories Grandad
This story is for you

x

“

*This silveryeye it's caught, it's caught by hidden
mist; its shiny little eyes, can see that opened fist.
This silveryeye wants to fly, this silveryeye wants to
soar; this silveryeye shall reveal its secrets, but not
spill a drop more.*

*This silveryeye shall give answers, its wings shall
give it lift; this silveryeye shall sing, it shall sing
out its hidden gift.*

— **Cael Sendell-Price**
2019

Declaration

I declare that this thesis was composed by myself and that the work contained herein is my own except where explicitly stated in the text. The work has not been submitted for any degree or professional qualification except as specified.

Ashley T. Sendell-Price

Acknowledgements

This thesis, like all science, was a collaboration, and as such there are many people to thank. First and foremost, I would like to thank my principle supervisor Sonya Clegg. I could not have asked for a kinder supervisor or friend. Thank you for your constant guidance and help with everything, for informing the way I approach evolutionary questions, and providing an encouraging example of how to be both a great scientist and a balanced individual. I look forward to our continued work together and hope we can again drink chateau d'cardboard on a tropical island. Second, I would like to thank my co-supervisor Ben Sheldon. Thank you, Ben, for supporting my application to Oxford all those years ago, and for including me in your group even when my research interests turned towards evolution on islands.

Pursuing a doctoral degree was a dream that would not have become reality had it not been for the financial support provided by the Natural Environment Research Council. Likewise, undertaking both genomics and tropical fieldwork (both expensive research pastimes) would not have happened without additional funding from the Genetics Society, Environmental Research DTP, and Linacre College.

I would like to thank Kristen Ruegg and Eric Anderson. Prior to commencing this DPhil my knowledge of bioinformatics and genomics was non-existent, by sharing their expertise in both of these fields they have made acquiring these skills a much less daunting challenge.

Thank you to everyone in the Department of Zoology, especially members of the Clegg, Sheldon and Coulson labs, whose camaraderie has helped me get to this point, in particular, Benjamin Van Doren, Freddy Hillemann, Allison Roth and Sara Keen who have been better able than most to sympathize with the highs and lows associated with completing a DPhil in zoology. I especially want to thank Em Dale, Will Smith, Michał Jezierski and Roddy Pracana - I am not sure how I would have completed this thesis without their comfort and support. Thank you.

My time at Oxford would certainly not have been the same without two of my closest friends, Rob Paton and Kaja Fenn, who have shared this journey with me from its very beginning. Evenings spent (interpretive) dancing together are among my most cherished memories of Oxford.

Next I would like to thank my family, who have provided me with unlimited support and have always encouraged me to chase my dreams. I love you all so very much Mum, Dad, Jess, Jordan and Milo.

Last, and most importantly, I would like to thank my husband, Cael Sendell-Price, whose companionship over the past decade has brought me great happiness and strength. Without your patience and love this journey would not have been possible. I promise that we will one day again sit atop Mt. Aorai and gaze across paradise. Rwy'n dy garu di.

Abstract

The genomics of diversification in south Pacific silvereyes (*Zosterops lateralis*)

DPhil thesis by Ashley T. Sendell-Price, Linacre College, submitted Michaelmas 2019

Speciation is a dynamic process, shaped by interactions among different evolutionary mechanisms including natural selection, genetic drift, gene flow, and mutation. Over 150 years since the publication of ‘On the Origin of Species’ a major challenge in evolutionary biology is to determine the relative importance of these different mechanisms in the speciation process. In the era of genomics, a step towards addressing this challenge is to study the progression of speciation at the level of the genome. Using island colonisations by the silvereye (*Zosterops lateralis*), in this thesis I investigate the roles of gene flow, selection and drift in shaping patterns of divergence at both the genomic and phenotypic level. In Chapter 2, using populations where divergence timeframes and gene flow histories are known *a priori*, I documented how genomic patterns accumulate as populations progress along the speciation continuum. This chapter challenges the ubiquity of existing verbal models that explain the accumulation of genomic differences via the growth of genomic islands and instead supports the idea that divergence both within and outside of genomic islands is important during the speciation process. In Chapter 3, using the recent sequential colonization of south Pacific islands by the Tasmanian silvereye subspecies I provide the first assessment of how patterns of genomic divergence are shaped by the stochastic effects of drift following population founding. By tracking divergence across the genome, I demonstrate that population founding has an idiosyncratic effect on the buildup of divergence. In Chapter 4, using the human-mediated introduction of the silvereye to French Polynesia, I assess the role of drift and selection in driving rapid morphological change following the establishment of new populations. Despite only 80 years of separation from their New Zealand ancestors, French Polynesian silvereyes displayed significant changes in body and bill size and shape, but the magnitude of these changes are not large enough, considering effective population sizes, to require directional natural selection to explain them; instead drift alone is a sufficient explanation at the phenotypic level. However, identification of outlier loci at genes previously identified as candidates for bill size and body shape, also suggests selective processes are operating within this population. As a whole, this body of work furthers understanding of how different evolutionary mechanisms shape patterns of divergence at the level of the genome and how divergence unfolds following the establishment of new populations.

Author Contributions

The following manuscripts have arisen from this thesis, and are presented in Chapters 2, 3 and 4. These are collaborative efforts, although in each case the majority of work is my own.

Chapter 2: Sendell-Price AT, Ruegg KC, Anderson EC, Quilodrán CS, Van Doren BM, Underwood VL, Coulson T & Clegg SM. The genomic landscape of divergence across the speciation continuum in island-colonising silvereyes. Under review at: *G3: Genes, Genomes, Genetics*.

SMC, KCR, ECA and TC conceived the project. SMC and KCR supervised the project. VLU conducted the laboratory work. Bioinformatics and analyses were primarily carried out by ATSP with assistance from SMC, KCR, ECA, BVD and TC. Simulations were conducted by CSQ with input from TC and SMC. The manuscript was written by ATSP, SMC, and KCR with input on individual sections by TC and CSQ, and editing by all authors.

Chapter 3: Sendell-Price AT, Ruegg KC, Robertson BC & Clegg SM. The idiosyncratic nature of genomic divergence following population founding. *In preparation*.

ATSP and SMC jointly conceived the project, KCR coordinated the genomic data collection, and SMC supervised the project. The manuscript was written by ATSP and SMC, with input from KCR and BCR.

Chapter 4: Sendell-Price AT, Ruegg KC & Clegg SM. (2020) Rapid morphological divergence following a human-mediated introduction: The role of drift and directional selection. *Heredity* **124**: 535–549.

ATSP and SMC jointly conceived the project and SMC supervised the project. Fieldwork was conducted by ATSP and SMC. Bioinformatics and analyses were carried out by ATSP with assistance from SMC and KCR. The manuscript was written by ATSP and SMC, with input from KCR.

Contents

Declaration	i
Acknowledgements	iii
Abstract	v
Author Contributions	vii
1 General Introduction	1
1.1 A brief introduction to evolutionary processes driving divergence . . .	1
1.2 The genomic landscape of divergence	2
1.3 Studying speciation below the species level	4
1.4 The silvereye (<i>Zosterops lateralis</i>)	5
1.5 Aims of thesis	8
1.6 References	10
2 The genomic landscape of divergence across the speciation continuum in island-colonising silvereyes	15
2.1 Introduction	15
2.2 Methods	19
2.2.1 Sample collection	19
2.2.2 RAD-PE sequencing and bioinformatics	22
2.2.3 Population genomic analyses	23
2.2.4 Identifying outlier SNPs under selection	24
2.2.5 Candidate gene analysis	25
2.2.6 Simulations of divergence	25
2.3 Results	27
2.3.1 RAD-PE sequencing and bioinformatics	27
2.3.2 Overall levels of divergence	27
2.3.3 Accumulation of genomic divergence across the genome . . .	27
2.3.4 Dynamics of genomic islands of divergence and valleys of similarity	29

2.3.5	Correspondence of genomic islands and valleys to the position of outlier SNPs	31
2.3.6	Candidate gene analysis	34
2.3.7	Simulations of neutral divergence	34
2.4	Discussion	37
2.4.1	The accumulation of divergence across the genome	38
2.4.2	The dynamics of genomic islands of divergence and valleys of similarity	40
2.4.3	Repeated evolutionary patterns	42
2.5	Conclusion	43
2.6	Acknowledgements	43
2.7	References	44
3	The idiosyncratic nature of genomic divergence following population founding	53
3.1	Introduction	53
3.2	Methods	56
3.2.1	Sample collection	56
3.2.2	Sequencing and bioinformatics	56
3.2.3	Population structure	58
3.2.4	Inferring demographic history	58
3.2.5	Population genomic analyses	60
3.2.6	Identification of highly diverged genomic regions	60
3.3	Results	61
3.3.1	RAD-PE sequencing and bioinformatics	61
3.3.2	Population structure	62
3.3.3	Demographic history	62
3.3.4	Genetic diversity and differentiation in recently colonised populations	63
3.3.5	Genetic diversity and differentiation in evolutionarily old island populations.	66
3.3.6	Genomic patterns of divergence and outlier loci	66
3.4	Discussion	67
3.4.1	The distribution of divergence across the genome	70
3.4.2	Loss of diversity following population founding	71
3.4.3	Genetic divergence and diversity compared to evolutionarily old populations	72
3.5	Conclusion	73
3.6	Acknowledgements	73

3.7	References	73
4	Rapid morphological divergence following a human-mediated introduction: The role of drift and directional selection	81
4.1	Introduction	83
4.2	Materials and methods	85
4.2.1	Sample collection	85
4.2.2	RAD-PE sequencing and SNP calling	85
4.2.3	Population structure	86
4.2.4	Demographic history inference	86
4.2.5	Morphological divergence	87
4.2.6	Identifying signatures of selection and candidate gene analysis	88
4.2.7	Genetic associations with morphological traits	88
4.3	Results	89
4.3.1	RAD-PE sequencing and bioinformatics	89
4.3.2	Population structure	89
4.3.3	Demographic history	89
4.3.4	Patterns and rates of morphological divergence	89
4.3.5	Signatures of selection and candidate genes	90
4.4	Discussion	91
4.4.1	Demographic history	91
4.4.2	The role of drift	91
4.4.3	Non-random founding population	92
4.4.4	The role of natural selection	92
4.4.5	Candidate genes	93
4.5	Conclusion	94
4.6	Data availability	94
4.7	Compliance with ethical standards	94
4.8	References	94
5	General Discussion	99
5.1	Summary of results	99
5.2	Future directions	102
5.3	References	104
A	Library preparation	107
B	Supplementary Material: Chapter 2	109
C	Supplementary Material: Chapter 4	113

D	Samples sequenced	117
E	The multiple population genetic and demographic routes to islands of genomic divergence	123

General Introduction

“

... according to my view, varieties are species in the process of formation, or are, as I have called them, incipient species.

— Charles Darwin
1859

1.1 A brief introduction to evolutionary processes driving divergence

Charles Darwin's 'On the Origin of Species' laid out the conceptual foundation for understanding how new species arise (Darwin, 1859). The answer to this 'mystery of mysteries' is that the vast diversity of life owes its existence to the process of speciation. Although Darwin was not aware of its underlying genetic basis, he proposed natural selection as the evolutionary mechanism driving differentiation (Darwin, 1859). By increasing the frequency of alleles that confer greater fitness, natural selection can cause population differentiation when populations experience different selective pressures (Hartl et al., 1997). Divergent selection can result in the formation of reproductive barriers between populations – i.e. reduced hybrid fitness – ultimately resulting in reproductive isolation and speciation (Orr, 1995; Schluter, 2001). Selection primarily acts on both standing genetic variation and new variation arising via mutations (Burke, 2012). Although divergent selection remains the prominent mechanism, where populations are adapting to the same selective pressure, reproductive barriers can also arise via mutation-order-speciation, in which incompatible mutations are fixed between populations, hybrid speciation or changes in ploidy (Coyne & Orr, 2004).

However, speciation does not necessarily need to involve selection. Random genetic drift alone can be sufficient to create large variations in allele frequencies between

populations, eventually leading to their differentiation into reproductively isolated species (Lande, 1976; Coyne & Orr, 2004). Although normally a gradual process, genetic drift can be accelerated during population founding and through population bottlenecks. Population founding by a small number of individuals that are not representative of the variation within the source population can cause the new population to differ considerably from the source population (Huxley et al., 1958; Barton & Charlesworth, 1984). Similarly, population bottlenecks can accelerate divergence through loss of genetic variation following a drastic decline in population size (Mayr, 1966; Carson, 1971).

As the reproductive barriers required for speciation to take place require the accumulation of genetic differences between populations, high rates of gene flow can reduce genetic differentiation between populations, thus hindering speciation (Bolnick & Nosil, 2007). Due to its homogenizing effect divergence with gene flow was thought to occur rarely, if at all, in nature (Mayr, 1942; Coyne, 1992). However, today divergence with gene flow is widely accepted (Turelli et al., 2001; Via, 2001). Despite its homogenising effect, gene flow can be overcome when natural selection is sufficiently strong or when gene flow is weaker than genetic drift.

Speciation is a dynamic process that is shaped by interactions among different evolutionary mechanisms: natural selection, genetic drift, gene flow, and mutation (Gavrilets, 2014). Over 150 years since the publication of ‘On the Origin of Species’ a major challenge in evolutionary biology is to determine the relative importance of these different evolutionary mechanisms in the speciation process (van den Berg & Weissing, 2015). In the era of genomics, a step towards addressing this challenge is to study the progression of speciation at the level of the genome.

1.2 The genomic landscape of divergence

The genomes of diverging taxa do not change uniformly – instead some regions of the genome diverge rapidly; whereas other regions diverge much more slowly (Wu, 2001; Nosil et al., 2009). Such differences in the speed of divergence result in the formation of a heterogeneous genomic landscape that is somewhat analogous to a mountainscape. Consisting of peaks of divergence where genetic differentiation accumulates and valleys of little to no differentiation (Feder et al., 2013; Seehausen et al., 2014).

Over the past decade a major challenge in speciation genomics has been to understand how different evolutionary processes such as gene flow, natural selection and

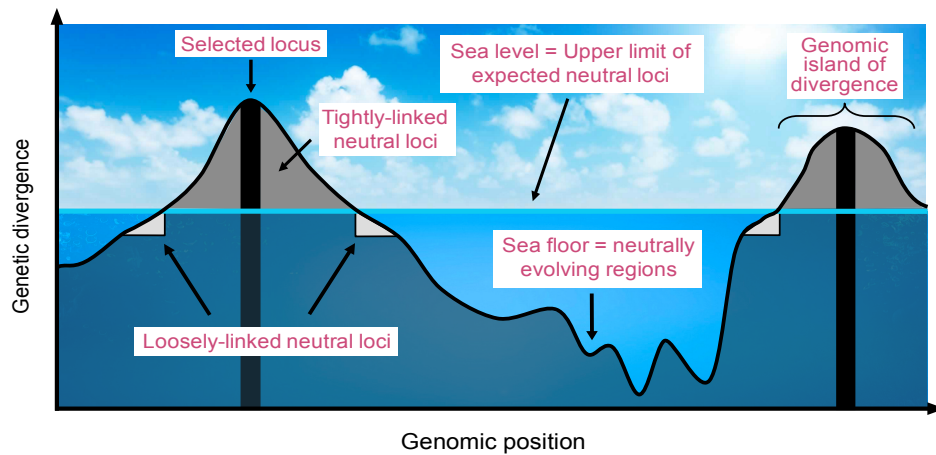


Figure 1.1. Schematic of the heterogeneous genomic landscape showing two genomic islands of divergence. Peaks above ‘sea level’ = ‘genomic islands’ - regions that exhibit significantly greater differentiation than expected under neutrality. Islands are composed of loci – both selected (black) and tightly linked (grey) – that are identified as outliers. Adapted from Nosil et al., 2009.

genetic drift shape the terrain of this heterogeneous genomic landscape (Ravinet et al., 2017). Identifying features of the genomic landscape that are diagnostic of these different processes would revolutionise evolutionary biology by allowing the processes involved during divergence between any two populations to be rapidly inferred without prior knowledge of their historical and contemporary occurrence; thereby freeing us from reliance on model organisms in artificial environments when addressing fundamental questions in this field.

A step towards using the genomic landscape to infer evolutionary processes was the use of ‘genomic islands of divergence’ as a signature of divergence with gene flow (Turner et al., 2005). By comparing the heterogeneous genetic differentiation observed across the genome to the topography of oceanic islands and the sea floor which connects them, genomic islands consist of highly differentiated regions of the genome that are surrounded by regions of little to no differentiation (Nosil et al., 2009). Composed of loci under strong divergent selection and those loci tightly linked to them (see Figure 1.1); these genomic islands are thought to be the product of selection operating over a small fraction of the genome, while the rest remains connected via gene flow (Turner et al., 2005; Hohenlohe et al., 2010; Feder et al., 2013). Genomic islands of divergence have been identified in a range of diverging taxa (e.g. house mouse (*Mus musculus*) (Harr, 2006); threespine

stickleback (*Gasterosteus aculeatus*) (Marques, 2016); Swainson's thrush (*Catharus ustulatus*) (Ruegg et al., 2014); European rabbit (*Oryctolagus cuniculus*) (Carneiro et al., 2014) and apple maggot (*Rhagoletis pomonella*) (Michel et al., 2010)) and for a time, their presence alone was interpreted as a signature of divergence with gene flow (Turner et al., 2005; Hohenlohe et al., 2010; Nadeau et al., 2012; Renaut et al., 2012; Renaut et al., 2013).

However, similar patterns of genomic heterogeneity can also arise through alternative evolutionary scenarios (Feder et al., 2013; Campagna et al., 2015). For example, in the absence of gene flow, adaptation may involve differential selection across the genome, much of which might occur post-speciation (Cruickshank & Hahn, 2014), set against a backdrop of maintained ancestral variation (incomplete lineage sorting) (Cruickshank & Hahn, 2014; Campagna et al., 2015). Such realisations have since tempered confidence in inferring evolutionary processes via the genomic landscape.

In light of the above realisations, a more detailed understanding of how evolutionary processes shape the genomic landscape is required before features of the genomic landscape can be used to correctly infer past processes. This can be achieved by studying the genomic landscape in well characterised systems where knowledge of divergence history is already available. Currently such studies have been limited to only a handful of systems. For example, comparing the genomic landscapes of *Heliconius* butterflies diverging under different geographic contexts (allopatry versus parapatry and sympatry), Martin et al. (2013) have shown that the accumulation of divergence across the genome is slowed by gene flow. Similarly by comparing diverging populations of *Ficedula* flycatchers, Burri et al. (2015) demonstrated that genome-wide variation in differentiation can be explained by variation in recombination rate. This study also demonstrated that the position of genomic islands can be highly repeatable. However, both of these studies are limited as they infer gene flow from geographic distribution alone, and many of the comparisons they make are not well matched for divergence timeframe.

1.3 Studying speciation below the species level

Studies of speciation have generally focused on the species level and above, rather than processes occurring at the population level (Barracough, 2010). Yet, because many models of speciation require some degree of initial population differentiation, elucidating the mechanisms that produce variation within species is of fundamental

importance (Foster, 1998). For example, in the two prevailing models of speciation ‘divergence-with-gene-flow’ and ‘divergence in isolation’ reproductive isolation is a by-product of genetic divergence between subpopulations (Coyne and Orr, 2004). If population differentiation is integral during the initial stages of speciation, a complete understanding of the most common mechanisms of speciation therefore requires that processes be examined prior to reproductive isolation. Despite this, studies of differentiated populations have rarely been incorporated into research on speciation. The first step towards resolving this is to identify and characterise species with high within-species population differentiation.

1.4 The silvereye (*Zosterops lateralis*)

The white-eyes (family: *Zosteropidae*) are a group of small, gregarious and arboreal, warbler-like passerines (Figure 1.2A) with a wide distribution (Figure 1.3). These species are among the fastest diversifying groups of terrestrial vertebrates, with approximately 120 recognised species, and have colonized more islands worldwide than any other group in the Order Aves (Moyle et al., 2009). Given their propensity to rapidly form new species across regions where other bird species show little or no diversification, the white-eyes have been dubbed ‘great speciators’ (Diamond et al., 1976). A term coined by Ernst Mayr and Jared Diamond following their visits to the Solomon Islands, during which they observed distinct white-eye species on each island visited; whereas most other bird species showed little variation from island to island (Mayr & Diamond, 2001). However, the white-eyes’ propensity to both colonise islands over large geographical distances yet rapidly differentiate at a local scale presents a paradox: How can these species be both highly dispersive, yet at the same time overcome gene flow from neighbouring populations? This question remains un-resolved.

A particularly interesting species of white-eye to study within-species diversification at both the phenotypic and genomic level is the silvereye (*Zosterops lateralis*) (Figure 1.2B). This species is one of the most widely distributed species of white-eye having naturally colonised numerous South Pacific islands and archipelagos since its origin on the Australian mainland, including New Caledonia, Vanuatu, Fiji, Tasmania, and New Zealand (Figure 1.4). The silvereye’s success at colonising islands is partly due to buffering from founder effects via large founding flocks and subsequent rapid population expansions (Clegg et al., 2002; Estoup and Clegg, 2003). Some island populations represent ancient colonisation events (occurring over timeframes of

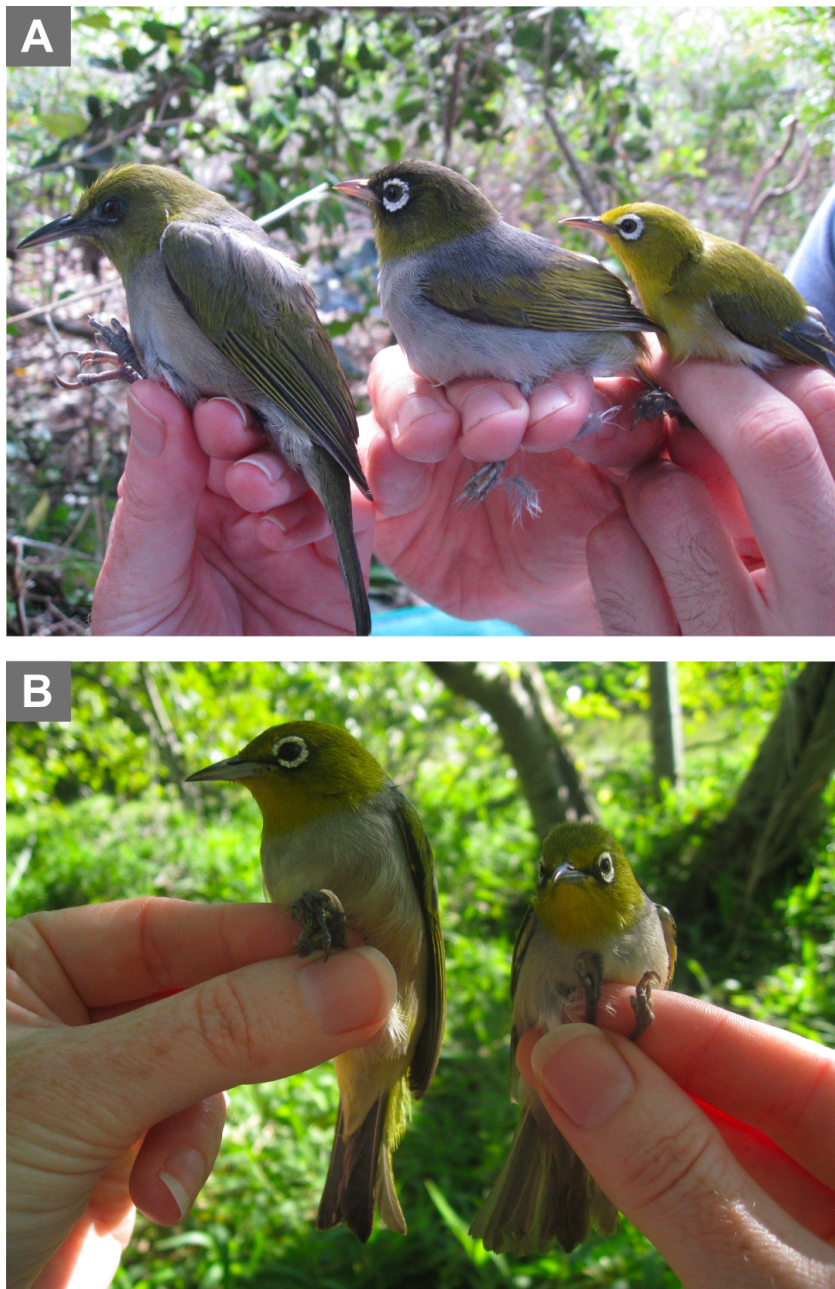


Figure 1.2. Examples of the phenotypic diversity seen within family Zosteropidae. (A) Three different species of white-eye caught on the island of Lifou, New Caledonia. From left to right: *Z. inornatus*, *Z. lateralis* (subspecies: *melanops*), and *Z. minutus*. Image by Sonya Clegg. (B) Within-species phenotypic diversity as seen between two *Zosterops lateralis* subspecies caught on Heron Island. The Heron Island silvereye (*Z. l. chlorocephalus*) (left) is up to 40% larger than (*Z. l. cornwalli* from the Australian mainland (right)). Image by Nick Clark.

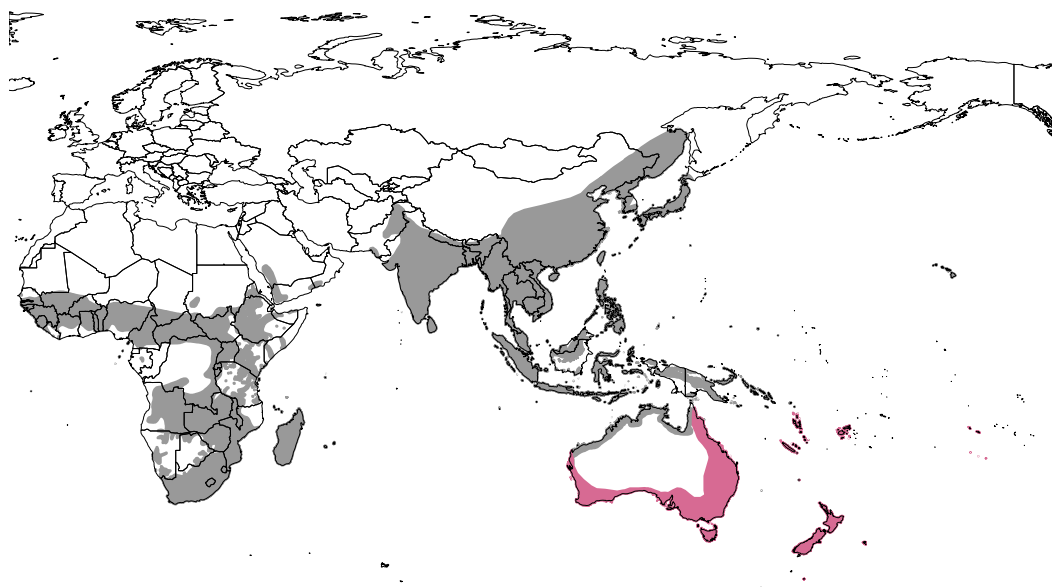


Figure 1.3. Distribution of *Zosterops lateralis* (pink) and other species of the family Zosteropidae (grey); based on data from BirdLife International and NatureServe (2015).

thousands and hundreds-of-thousands of years) where populations have a distinct phenotype and are considered distinct subspecies (for example *Z. l. chlorocephalus* is up to 40% larger than its mainland ancestor *Z. l. cornwalli*, and *Z. l. melanops* shows stark plumage differentiation from its ancestor *Z. l. griseonotus*) (Mees, 1969). In contrast more recently colonised populations represent incipient stages of genetic adaptation and population differentiation.

Silvereye populations across the south Pacific differ in the levels of current and historical gene flow. For example, reasoning based on geographic proximity, known population histories and contemporary bird movements, strongly supports divergence with gene flow in: 1) *Z. l. lateralis* from Chatham Island, which colonised from South Island, New Zealand during a natural population expansion during the 1850s (Mees, 1969; Clegg et al., 2002); 2) *Z. l. chlorocephalus* on Heron Island, which diverged in less than 4,000 years despite regular influxes of its mainland ancestor, *Z. l. cornwalli* (Clegg et al., 2008); and 3) *Z. l. melanops* on Lifou, New Caledonia which diverged in the past few hundred thousand years from the geographically proximate population of *Z. l. griseonotus* on New Caledonia's largest island, Grand Terre (Black, 2010). In contrast, a number of silvereeye populations have diverged in genetic isolation, these include: 1) *Z. l. lateralis* on islands of French Polynesia, which established following a human-mediated introduction in 1937, well beyond

the species natural eastern distribution limit (Monnet et al., 1993); and (2) *Z. l. tephropleurus* on the geographically isolated Lord Howe Island which is geographically far removed from its closest relative, *Z. l. cornwalli* on the Australian mainland (Black, 2010).

The variation within the silvereve subspecies complex provides a range of divergence timeframes and gene flow contexts; which in combination with the availability of the *Z. l. melanops* reference genome (Cornetti et al., 2015) makes the silvereve an ideal species in which to study the genomics of divergence, and the impact of different microevolutionary processes.

1.5 Aims of thesis

This thesis aims to advance understanding of how different evolutionary mechanisms shape patterns of divergence at both the genomic and phenotypic level and how divergence unfolds following the establishment of new populations.

In **Chapter 2**, using colonisations by the silvereve where divergence timeframes are known *a priori*, I document how genomic heterogeneity accumulates as divergence progresses. In contrast to previous predictions, results from this chapter provide limited support for the idea that divergence accumulates around loci under divergent selection or that genomic islands widen with time. This chapter therefore challenges the previously proposed and widely cited verbal model of speciation in which 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.

In **Chapter 3**, I use the historically recent, stepwise colonisation of south Pacific islands by the Tasmanian silvereve subspecies to provide the first genomic-level assessment of how patterns of divergence are shaped by the stochastic effects of population founding. By tracking the fate of highly diverged genomic regions between founding steps this chapter demonstrates that the effects of population founding are highly idiosyncratic. This chapter also provides a re-assessment of the effects of population founding on loss of overall levels of genetic diversity and levels of divergence.

In **Chapter 4**, I use the recent human-mediated introduction of the silvereve to French Polynesia, to assess the potential for rapid morphological divergence to arise following the establishment of a new population. In particular this chapter sets

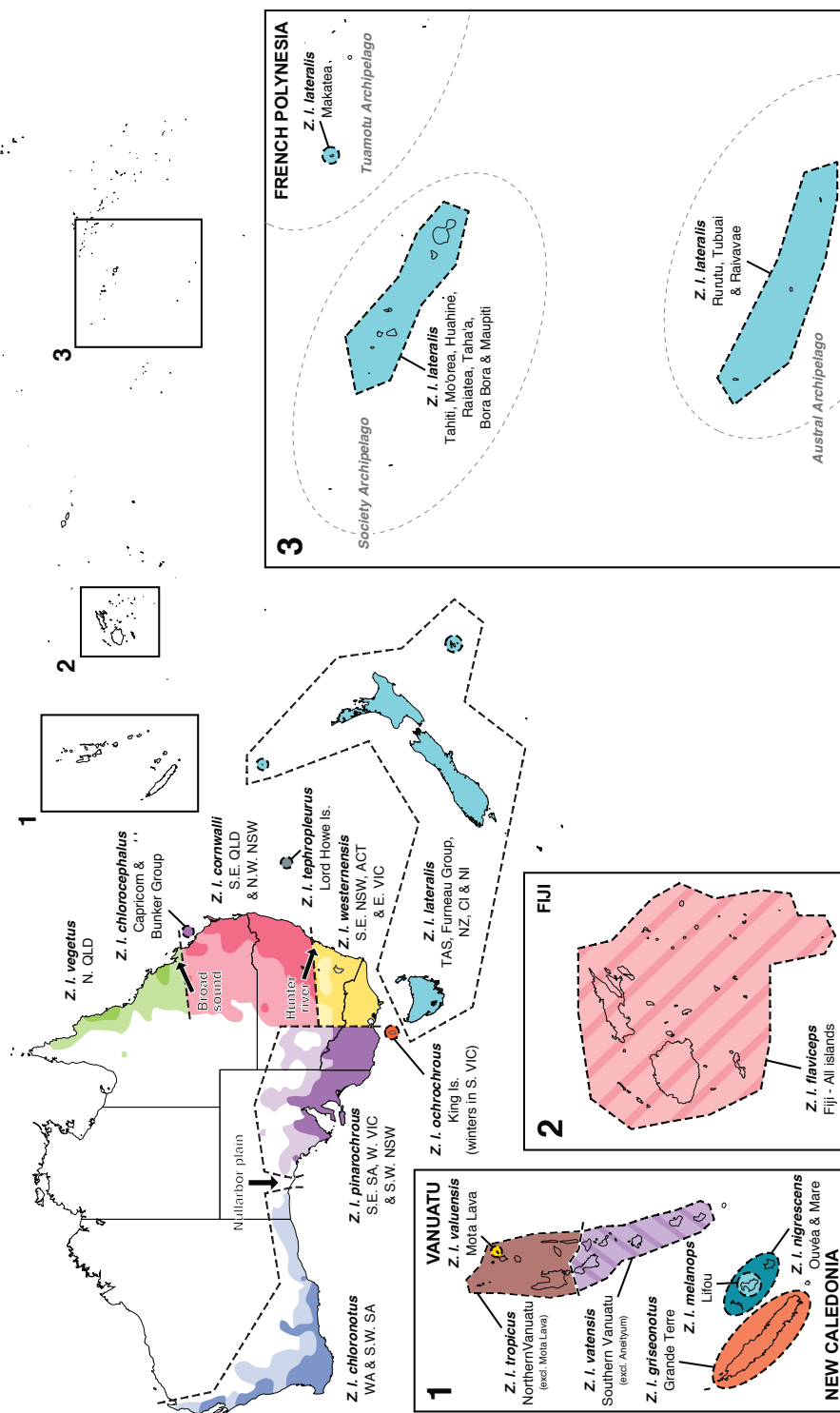


Figure 1.4. Distribution of currently recognised silvereye subspecies across the South Pacific. For mainland Australia pale shading indicates maximum extent of range and dark shading indicates where most often recorded. Dashed lines indicate subspecies boundaries.

out to test the idea that when populations are established by few individuals, the random sampling effect of drift can facilitate rapid phenotypic divergence in the absence of selective processes. Despite identifying signatures of selection operating in the French Polynesian population at the level of the genome, I demonstrate that where populations are established under founding extremes, rapid morphological shifts can be explained without the need to invoke selective processes.

Finally, in the **general discussion**, I summarise the main results of Chapters 2-4. I discuss broader implications for evolutionary theory, and future directions for speciation/population genomics.

1.6 References

- Barraclough TG** (2010). Evolving entities: Towards a unified framework for understanding diversity at the species and higher levels. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**: 1801–1813.
- Barton NH, Charlesworth B** (1984). Genetic revolutions, founder effects, and speciation. *Annual Review of Ecology and Systematics* **15**: 133–164.
- BirdLife International and NatureServe** (2015) *Bird species distribution maps of the world*. BirdLife International: Cambridge and NatureServe: Arlington.
- Black RA** (2010). *Phylogenetic and phenotypic divergence of an insular radiation of birds*. PhD Thesis: Imperial College London.
- Bolnick DI, Nosil P** (2007). Natural selection in populations subject to a migration load. *Evolution* **61**: 2229–2243.
- Burke MK** (2012). How does adaptation sweep through the genome? Insights from long-term selection experiments. *Proceedings of the Royal Society B: Biological Sciences* **279**: 5029–5038.
- Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, et al.** (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Research* **25**: 1656–1665.
- Campagna L, Gronau I, Silveira LF, Siepel A, Lovette IJ** (2015). Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation. *Molecular Ecology* **24**: 4238–51.

- Carneiro M, Albert FW, Afonso S, Pereira RJ, Burbano H, Campos R, et al.** (2014). The genomic architecture of population divergence between subspecies of the European Rabbit. *PLoS Genetics* **10**: e1003519.
- Carson HL** (1971). Speciation and the founder principle. *Stadler Genetics Symposia* **3**: 51-70.
- Clegg SM, Degnan SM, Kikkawa J, Moritz C, Estoup A, Owens IPF** (2002). Genetic consequences of sequential founder events by an island-colonizing bird. *PNAS* **99**: 8127–8132.
- Clegg SM, Frentiu FD, Kikkawa J, Tavecchia G, Owens IPF** (2008). 4000 Years of phenotypic change in an island bird: Heterogeneity of selection over three microevolutionary timescales. *Evolution* **62**: 2393–2410.
- Cornetti L, Valente LM, Dunning LT, Quan X, Black R A., Hébert O, et al.** (2015). The genome of the “great speciator” provides insights into bird diversification. *Genome Biology and Evolution* **7**: 2680–2691.
- Coyne JA** (1992). Genetics and speciation. *Nature* **355**: 511–515.
- Coyne JA, Orr HA** (2004). *Speciation*. Sinauer Associates: Sunderland, MA.
- Cruickshank TE, Hahn MW** (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* **23**: 3133–3157.
- Darwin C** (1859). *On the origin of species by means of natural selection*. Murray: London.
- Diamond JM, Gilpin ME, Mayr E** (1976). Species-distance relation for birds of the Solomon Archipelago, and the paradox of the great speciators. *PNAS* **73**: 2160–2164.
- Estoup A, Clegg SM** (2003). Bayesian inferences on the recent island colonization history by the bird *Zosterops lateralis lateralis*. *Molecular Ecology* **12**: 657–674.
- Feder JL, Flaxman SM, Egan SP, Comeault AA, Nosil P** (2013). Geographic mode of speciation and genomic divergence. *Annual Review of Ecology, Evolution, and Systematics* **44**: 73–97.
- Foster SA** (1998). Nested biological variation and speciation. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* **353**: 207–218.

- Gavrilets S** (2014). Models of speciation: Where are we now? *Journal of Heredity* **105**: 743–755.
- Harr B** (2006). Genomic islands of differentiation between house mouse subspecies. *Genome Research* **16**: 730–737.
- Hartl DL, Clark AG, Clark AG** (1997). *Principles of population genetics*. Sinauer associates: Sunderland.
- Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA** (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6**: e1000862.
- Huxley J, Hardy AC, Ford EB** (1958). *Evolution as a process*. George Allen and Unwin: London.
- Lande R** (1976). Natural Selection and Random Genetic Drift in Phenotypic Evolution. *Evolution* **30**: 314–334.
- Marques DA, Lucek KAY, Haesler MP, Feller AF, Meier JI, Wagner CE** (2016). Genomic landscape of early ecological speciation initiated by selection on nuptial colour. *Molecular Ecology* **26**: 7–24
- Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, et al.** (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research* **23**: 1817–1828.
- Mayr E** (1942). *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press: Cambridge.
- Mayr E** (1966). *Animal species and evolution*. Belknap Press of Harvard University Press: Cambridge
- Mayr E, Diamond JM** (2001). *The birds of northern Melanesia: Speciation, ecology and biogeography*. Oxford University Press: New York.
- Mees GF** (1969). A systematic review of the Indo-Australian Zosteropidae (Part III). *Zoologische Verhandelingen* **102**: 1–390.
- Michel AP, Sim S, Powell THQ, Taylor MS, Nosil P, Feder JL** (2010). Widespread genomic divergence during sympatric speciation. *PNAS* **107**: 9724–9729.
- Monnet C, Thibault JC, Varney A** (1993). Stability and changes during the twentieth century in the breeding landbirds of Tahiti (Polynesia). *Bird Conservation International* **3**: 261–280.

- Moyle RG, Filardi CE, Smith CE, Diamond J** (2009). Explosive Pleistocene diversification and hemispheric expansion of a ‘great speciator’. *PNAS* **106**: 1863–8.
- Nadeau NJ, Whibley A, Jones RT, Davey JW, Dasmahapatra KK, Baxter SW, et al.** (2012). Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* **367**: 343–53.
- Nosil P, Funk DJ, Ortiz-Barrientos D** (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* **18**: 375–402.
- Orr HA** (1995). The population genetics of speciation: The evolution of hybrid incompatibilities. *Genetics* **139**: 1805–1813.
- Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlović M, et al.** (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology* **30**: 1450–1477.
- Renaut S, Grassa CJ, Yeaman S, Moyers BT, Lai Z, Kane NC, et al.** (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications* **4**: 1827.
- Renaut S, Maillet N, Normandeau E, Sauvage C, Derome N, Rogers SM, et al.** (2012). Genome-wide patterns of divergence during speciation: The lake whitefish case study. *Philosophical Transactions of the Royal Society B* **367**: 354–363.
- Ruegg K, Anderson EC, Boone J, Pouls J, Smith TB** (2014). A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology* **23**: 4757–4769.
- Schluter D** (2001). Ecology and the origin of species. *Trends in Ecology and Evolution* **16**: 372–380.
- Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, et al.** (2014). Genomics and the origin of species. *Nature Reviews Genetics* **3**: 176–92.
- Turelli M, Barton NH, Coyne JA** (2001). *Theory and speciation*. *Trends in Ecology and Evolution* **16**: 330–343.
- Turner TL, Hahn MW, Nuzhdin S V.** (2005). Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology* **3**: 1572–1578.

- van den Berg P, Weissing FJ** (2015). The importance of mechanisms for the evolution of cooperation. *Proceedings of the Royal Society B: Biological Sciences* **282**: 20151382.
- Via S** (2001). Sympatric speciation in animals: The ugly duckling grows up. *Trends in Ecology and Evolution* **16**: 381–390.
- Wu CI** (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**: 851–865.

The genomic landscape of divergence across the speciation continuum in island-colonising silvereyes

Ashley T. Sendell-Price, Kristen C. Ruegg,
Eric C. Anderson, Claudio S. Quilodrán,
Benjamin M. Van Doren, Vinh Le Underwood,
Tim Coulson & Sonya M. Clegg

2.1 Introduction

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 & 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 (Butlin 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 & 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 & 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 localised 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 & 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 & 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 & Nosil, 2010). As such, alternative mechanisms such as the transition from single locus dynamics to multi-locus coupling (Flaxman et al., 2013; Flaxman et al., 2014; Butlin & Smadja, 2017; Schilling et al., 2018) have been proposed to explain the continued shift from localised to genome wide levels of divergence.

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 homogenising effect at neutrally evolving or weakly selected loci (Nosil et al., 2017). In contrast, under genetic isolation, divergence proceeds via selection and/or drift, unfettered by this homogenising 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 2.1). To date, comparisons of genomic divergence between races of *Heliconius* butterflies distributed in allopatry versus 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 of similarity – 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

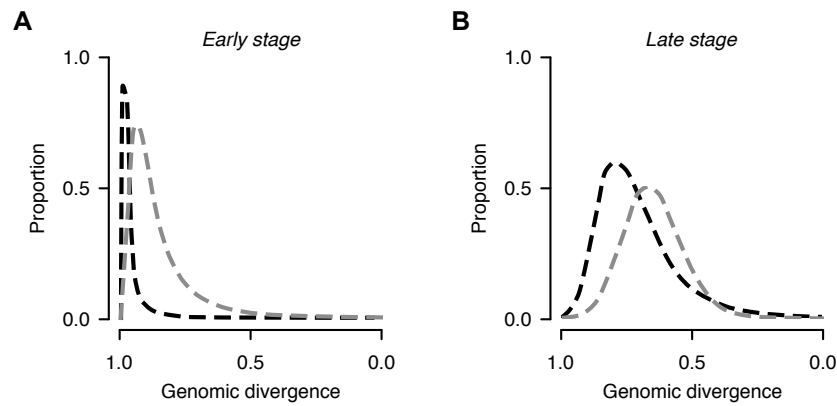


Figure 2.1. Hypothetical distributions of genetic differentiation expected for divergence with and without gene flow. (A) During the early stages of the speciation continuum divergence is expected to be limited to few loci, resulting in a highly skewed distribution of divergence values. This expectation will be most extreme for populations diverging with gene flow (black lines) than for populations diverging in genetic isolation (grey lines) as the homogenising 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 (grey lines) as occasional gene flow may reduce levels of divergence at positions of the genome even as reproductive isolation is approached.

by slowing the approach to genome-wide divergence. Genomic valleys may occur because of alleles favoured in both populations (parallel adaptation) (Nielsen, 2005; Roesti et al., 2012), or they may result from 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). 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 are lost 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, and the dynamics of genomic islands of divergence and genomic valleys of similarity, requires simultaneous consideration of the stage along the speciation continuum and the gene flow context of divergence for systems that have well characterised population histories. Members of the silverevee subspecies complex (*Zosterops lateralis*) of Australia and south Pacific islands offer an exceptional opportunity to explore patterns of divergence across different timescales and gene flow contexts. The species has repeatedly colonised 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, silverevee populations have diverged under different gene flow scenarios, with some populations diverging with gene flow and others diverging in genetic isolation (Mees, 1969). Further, as insular silverevees show a repeated pattern of phenotypic change towards increased body size and bill size, largely driven by directional natural selection (Mees, 1969; Clegg et al., 2002b; Clegg et al., 2008), genomic islands may form around loci underlying this phenotypic change across comparisons. While we do not know the genetic architecture of these traits, several studies have found genes of large effect for bill traits (Bosse et al., 2017; Chaves et al., 2016; Cornetti et al., 2015; Lamichhaney et al., 2015; Lamichhaney et al., 2016; vonHoldt et al., 2018).

Here we use empirical analysis of Restriction-site Associated DNA sequencing (RAD-Seq) data from silverevees 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 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?

2.2 Methods

2.2.1 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 2.2). The populations used in this study provide a rare example of where gene flow scenarios and divergence timescales are known *a priori* or where a very strong case can be made to infer the gene flow scenario. This is based on a combination of historical records, contemporary bird movements, and phylogenetic inferences as outlined below:

South Island vs. Chatham Island: The Chatham Island silvereye population was colonised 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 (Black, 2010). In the mid 1800s, South Island silvereyes were undergoing a population expansion which resulted in the colonisation of other islands in the region (notably: the North Island of New Zealand and Norfolk Island) (Mees, 1969; Clegg et al., 2002a). Given this recent colonisation date and that a population expansion was taking place within the region, this comparison is characterised as Early Stage – Gene Flow.

South Island vs. French Polynesia: The French Polynesian 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 & Cibois, 2017). Given that French Polynesia is located well beyond the natural distribution limit for this species and that the French Polynesian population is known to be the product of a single and recent introduction event, this comparison is characterised as Early Stage – No Gene Flow.

Mainland (Australia) vs. Heron Island: The Heron Island silvereye (*Z. l. chlorocephalus*) population was established from ancestral mainland subspecies (*Z. l. cornwalli*) between 3,000-4,000 years ago. This divergence timeframe is supported by geological records which suggest

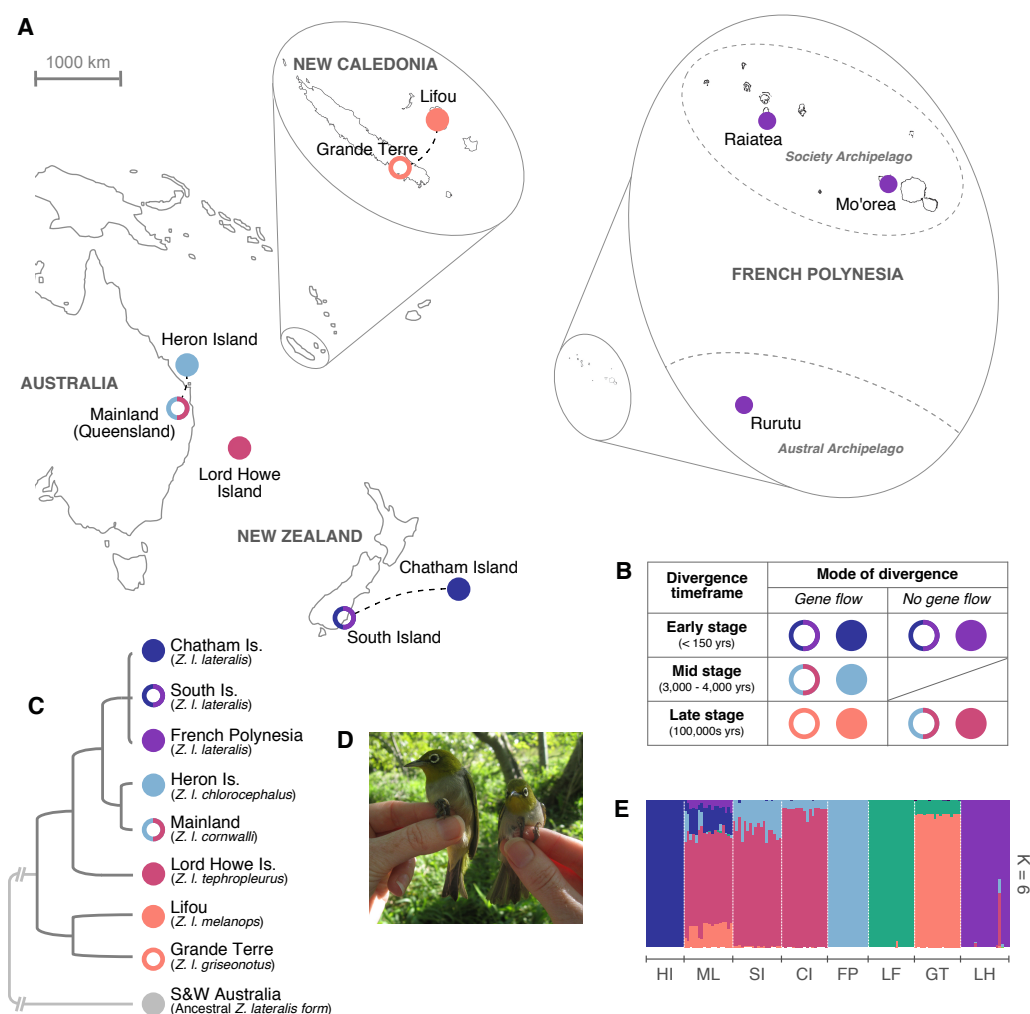


Figure 2.2. Silveryeye populations used in the study. (A) Sampling locations of *Zosterops lateralis* across the south Pacific (dashed lines indicate gene-flow between populations); (B) Population divergence timeframes (known from historical records, or inferred from island ages and genetic divergence 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 2010) and known colonisation 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.

Table 2.1. *Z. lateralis* sampling information. No. Filtered = number of individuals retained after quality filtering.

Subspecies	Pop. Code	Sampling location	Latitude / Longitude	Collection date	No. Sequenced	No. Filtered
<i>chlorocephalus</i>	HI	Heron Island	-23.44 / 151.91	2015	20	16
<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	19
<i>lateralis</i>	CI	Chatham Island	-44.58 / 176.32	1997	22	18
<i>lateralis</i>	SI	South Island	-45.53 / 170.30	1997	23	21
<i>lateralis</i>	FP	Mo'orea	-17.51 / -149.91	2005	17	11
<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	19
<i>tephropleurus</i>	LH	Lord Howe Island	-31.33 / 159.05	1998	20	19

that Heron Island has been vegetated (and therefore habitable) for a maximum of 4,000 years (Hopley, 1982; Degnan & Moritz, 1992). Given that the ancestral mainland subspecies is regularly observed on Heron Island, this comparison is characterised 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, 2010). Given the silvereye's ability to cross water boundaries and the geographic proximity of these islands (200km apart), this comparison is characterised as Late Stage – Gene Flow.

Mainland (Australia) vs. Lord Howe Island: Based on mtDNA phylogenetic analyses, *Z. l. tephropleurus* on Lord Howe Island derived from the mainland subspecies (*Z. l. cornwalli*) in the past few hundred-thousand years (Black 2010). Given the geographic isolation of Lord Howe Island, this comparison is characterised as Late Stage – No Gene Flow.

Sampling locations and collection dates are provided in Table 2.1. To provide large enough sample sizes for analysis, the French Polynesian population included samples taken from three islands. This is justified based on an admixture analysis which grouped these samples together into a single cluster (Figure 2.2E). All birds were caught using mist nets or traps and 20–40 μ l of blood collected from the brachial wing vein and stored in 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).

2.2.2 RAD-PE sequencing and bioinformatics

Restriction-site associated DNA paired-end (RAD-PE) sequencing was used to assess the genomic landscape of divergence among silveryeye 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 & Blaxter, 2010). Genomic DNA in blood samples was extracted from between 20 and 24 individuals per population (or set of populations in the case of French Polynesia) ($n = 168$) using QIAGEN DNeasy blood and tissue extraction kits (Qiagen, Hilden, Germany) 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 Appendix A 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 (Andrews 2010). 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 windows of 15% of the length of the read). Sequences with possible adapter contamination and/or missing the *Sbf*I 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 & 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 ≤ 2 ; minimum genotype quality = 30; minimum depth = 8; and sites were called in at least 70% of individuals. Following this the outputted Variant Call Format (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 ($N50 = 3.5$ Mb) (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 their 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 fixation index (F_{ST} - a measure of relative divergence) and Nei's measure of absolute divergence (d_{xy}) were calculated in non-overlapping windows of 50kb and 500kb for each population comparison, using Python scripts developed by Martin et al. (2013) (https://github.com/simonhmartin/genomics_general).

2.2.3 Population genomic analyses

We assessed the shift from localised divergence at few loci towards 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 2.1, skewness provides a useful metric to describe a population comparisons' stage of transition from localised (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 timescales as most loci approach fixation ($F_{ST} = 1$). We tested for significant differences in the skewness of empirical distributions using a randomisation 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 randomised distributions. Randomised 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 randomised 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 (bandwidth = 20) was applied to F_{ST} values calculated across the genome in 50kb non-overlapping windows, and the smoothed line compared to 10,000 smoothed lines obtained after permuting the order of the windows (see Ruegg et al. (2014)). We used a window size of 50kb for genomic island and genomic valley detection because it provided a sufficiently fine resolution across the genome while containing on average 1,093 sites per window. Further, windows containing <100 sites were removed prior to conducting analyses. 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. Adjacent outlier windows considered in the same genomic island or genomic valley were merged into larger divergent regions. 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 size of identified genomic islands and genomic valleys were compared across timescales and gene flow contexts using Kruskal-Wallis tests.

2.2.4 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 the 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, MinMAF = 0.2, and FDR = 0.01. 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 PLINK (Purcell et al., 2007).

2.2.5 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 summarising 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/valleys, if island/valley bounds (start and end positions) overlapped with 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.

2.2.6 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 modelling allowed us to address whether the transition from localised 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).

Each simulation considered two populations diverging with or without gene flow. Each population started with individuals with randomly drawn genotypes across 8,317 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 & Wilson, 1983)).

Each population was iterated forward on a per-generation time step. At each generation in the simulation, 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 that varied with population density. We thus defined mean population fitness as the average number of offspring produced. We included density-dependence in this function to avoid exponential growth. The density dependence kept the population within empirically observed bounds for the Heron Island population (100-300 breeding pairs). 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. 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 from an exponential distribution. The crossover points assigned along the chromosomes were compared to the genotype structure of individuals and the position of each SNP. Recombination occurred when the random variable was less than a particular recombination rate. 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 B.1, Appendix B). 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)). Simulations were run in R as described in Quilodrán et al. (2020). R code is available at GitHub: <https://github.com/eriqande/gids>.

2.3 Results

2.3.1 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 ≤ 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 435,456 were biallelic). Of the 168 samples sequenced, 150 were retained after removing individuals where $\geq 30\%$ of data was missing. The number of individuals retained per location ranged from 16 to 21 (Table 2.1).

2.3.2 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} (Fig 2.3A). 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 homogenising nature of gene flow impedes divergence. Compared to autosomes, chromosome Z exhibited higher levels of divergence for all population comparisons (Figure 2.3A). This difference was significant for Mainland vs. Heron Island, and Mainland vs. Lord Howe Island comparisons (Mann-Whitney: all P -values < 0.05).

2.3.3 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 towards large values of F_{ST} (skew = 2.69 and 2.07 respectively) (Figure 2.3B) and characterised by extreme L-shaped distributions with divergence limited to few 50kb windows (Figure 2.4B). Over longer divergence timeframes distributional skew was reduced, with the lowest

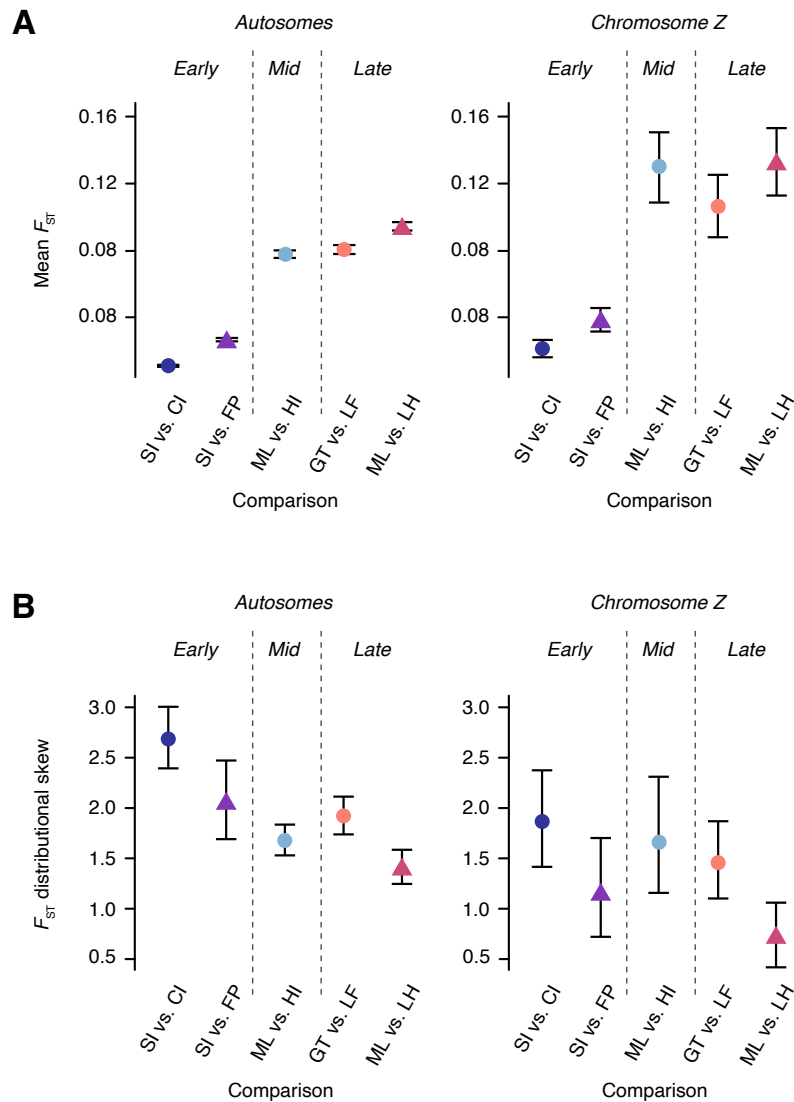


Figure 2.3. (A) Mean F_{ST} and (B) distributional skew of F_{ST} values. Calculated for autosomes and chromosome Z separately for each population comparison. Populations diverging with gene flow are indicated by triangles and populations diverging in isolation by circles. 95% confidence intervals obtained via bootstrapping over loci.

Table 2.2. *P*-values for pairwise comparisons of F_{ST} distributional skew tested using a randomisation 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	-	0.092	0.778	0.344	<0.001
SI vs. FP	0.055	-	0.512	0.509	0.180
ML vs. HI	<0.001	0.010	-	0.666	0.019
GT vs. LF	<0.001	0.472	0.101	-	0.005
ML vs. LH	<0.001	<0.001	0.025	<0.001	-

levels of skew (skew = 1.41) observed for the late stage – no gene flow comparison (Mainland vs. Lord Howe Island)(Figure 2.3B). The accumulation of genome-wide divergence over longer divergence timeframes is clearly visible in Figure 2.4A. When matched for divergence timeframe, the F_{ST} distributions of populations diverging under the influence of gene flow showed higher levels of skew (more localised divergence) than populations diverging in the absence of gene flow. Although not all pairwise differences were significant, recently diverging populations had significantly higher skew for autosomes than late stage diverging populations (Table 2.2). 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 B.2, Appendix B).

Like autosomes, chromosome Z showed a clear trend towards decreasing skew (Figure 2.3B and Figure 2.4B), with divergence visually increasing over longer timeframes (Figure 2.4A). As was the case for autosomes, the F_{ST} distribution for Mainland vs. Lord Howe Island (late stage - no gene flow) was the least skewed, meaning that divergence was more widespread across the Z chromosome. However, this difference in skew was only significant when compared to the South Island vs. Chatham Island comparison (P -value = 0.040, Table 2.2).

2.3.4 Dynamics of genomic islands of divergence and valleys of similarity

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)

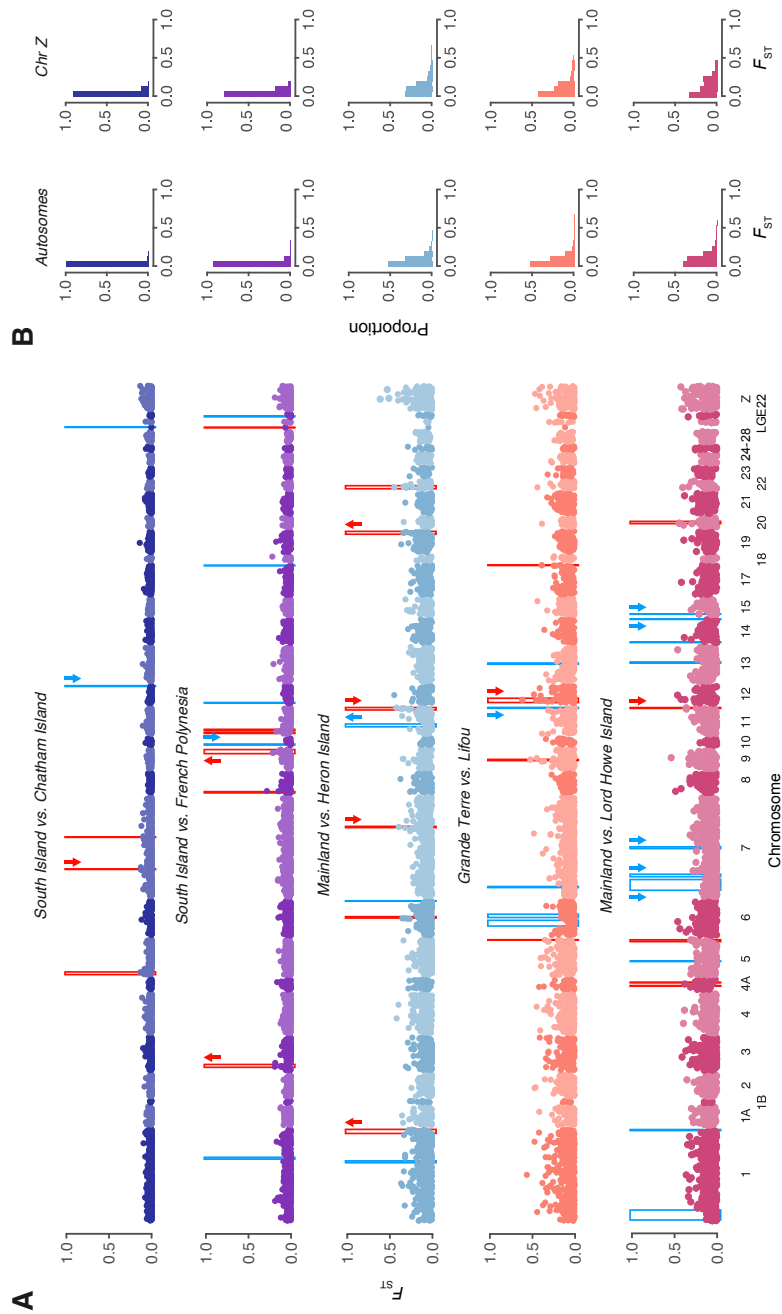


Figure 2.4. (A) Pairwise F_{ST} across the genome for each population comparison calculated in non-overlapping 50kb windows. Regions of elevated differentiation (genomic islands) are highlighted in red and regions with low differentiation (genomic valleys) are highlighted in blue. Arrows beside genomic islands/valleys indicate if d_{xy} within these regions was significantly elevated (upwards pointing arrows) or decreased (downwards 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.

(Figure 2.4). The number of genomic islands identified per population comparison ranged from 3 to 6 (Table 2.3). The number of genomic islands identified did not differ when compared across different gene flow scenarios or across different divergence timeframes. Similarly, genomic island size – although variable (mean size ranged from 233 kb to 567 kb, see Table 2.3 and Figure 2.5) – did not differ significantly across different gene flow contexts (Kruskal-Wallis: $\chi^2_1 = 0.490$, P -value = 0.484) or divergence timescales (Kruskal-Wallis: $\chi^2_2 = 4.258$, P -value = 0.119). We identified two regions of the genome in which genomic islands of divergence occurred at the same location in multiple population comparisons. The size of these ‘shared’ genomic islands is reported in Table 2.4. 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 2.4A).

Genomic valleys were also identified in all diverging population comparisons (Figure 2.4A). Their size, although variable (mean size ranged from 100kb to 665 kb; see Table 2.3 and Figure 2.5), did not differ significantly when compared across divergence timeframes (Kruskal-Wallis: $\chi^2_2 = 1.722$, P -value = 0.423) or gene flow scenarios (Kruskal-Wallis: $\chi^2_1 = 0.153$, P -value = 0.696). We identified a single region of the genome (located on chromosome 7) in which a genomic valley occurred across multiple population comparisons (Table 2.4). 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 identified in small numbers (Figure 2.4A). A single genomic valley with significantly elevated d_{xy} was identified for the Mainland vs. Heron Island comparison.

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

PCAdapt identified between 12 and 90 outlier SNPs in each population comparison (Table 2.5 and Figure 2.6). Outlier SNPs did not correspond to the position of genomic islands or genomic valleys in any comparison.

Table 2.3. 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}
<i>Genomic islands</i>							
Early stage	Gene flow	SI vs. CI	3	233	600	0.031	0.002
Early stage	No-gene flow	SI vs. FP	6	392	950	0.051	0.005
Mid stage	Gene flow	ML vs. HI	6	567	950	0.165	0.006
Late stage	Gene flow	GT vs. LF	4	350	1,000	0.250	0.002
Late stage	No-gene flow	ML vs. LH	5	240	500	0.194	0.005
<i>Genomic valleys</i>							
Early stage	Gene flow	SI vs. CI	2	100	150	0.001	0.004
Early stage	No-gene flow	SI vs. FP	5	200	400	0.008	0.003
Mid stage	Gene flow	ML vs. HI	3	367	650	0.048	0.005
Late stage	Gene flow	GT vs. LF	5	540	1,350	0.016	0.003
Late stage	No-gene flow	ML vs. LH	10	665	2,400	0.066	0.003

Table 2.4. Size of genomic islands and valleys occurring at the same location across multiple population comparisons.

Divergence timeframe	Mode of divergence	Comparison	Genomic island size (kb)		Genomic valley size (kb)
			Chr 5	Chr 11	Chr 7
Early stage	Gene flow	SI vs. CI	-	-	-
Early stage	No-gene flow	SI vs. FP	-	-	-
Mid stage	Gene flow	ML vs. HI	-	550	-
Late stage	Gene flow	GT vs. LF	5	-	200
Late stage	No-gene flow	ML vs. LH	400	150	2,600

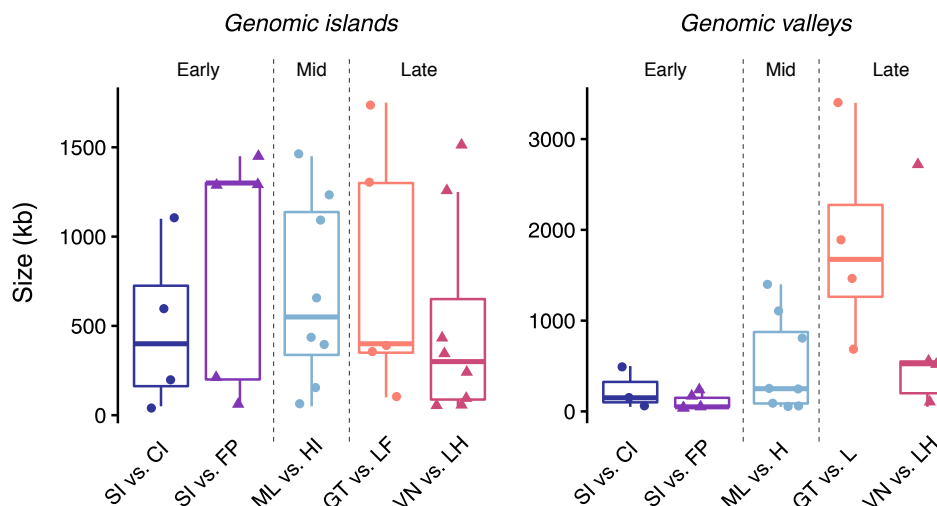


Figure 2.5. Size (kb) of genomic islands and genomic valleys identified for each population comparison. Circles indicate divergence with gene flow and triangles divergence in isolation.

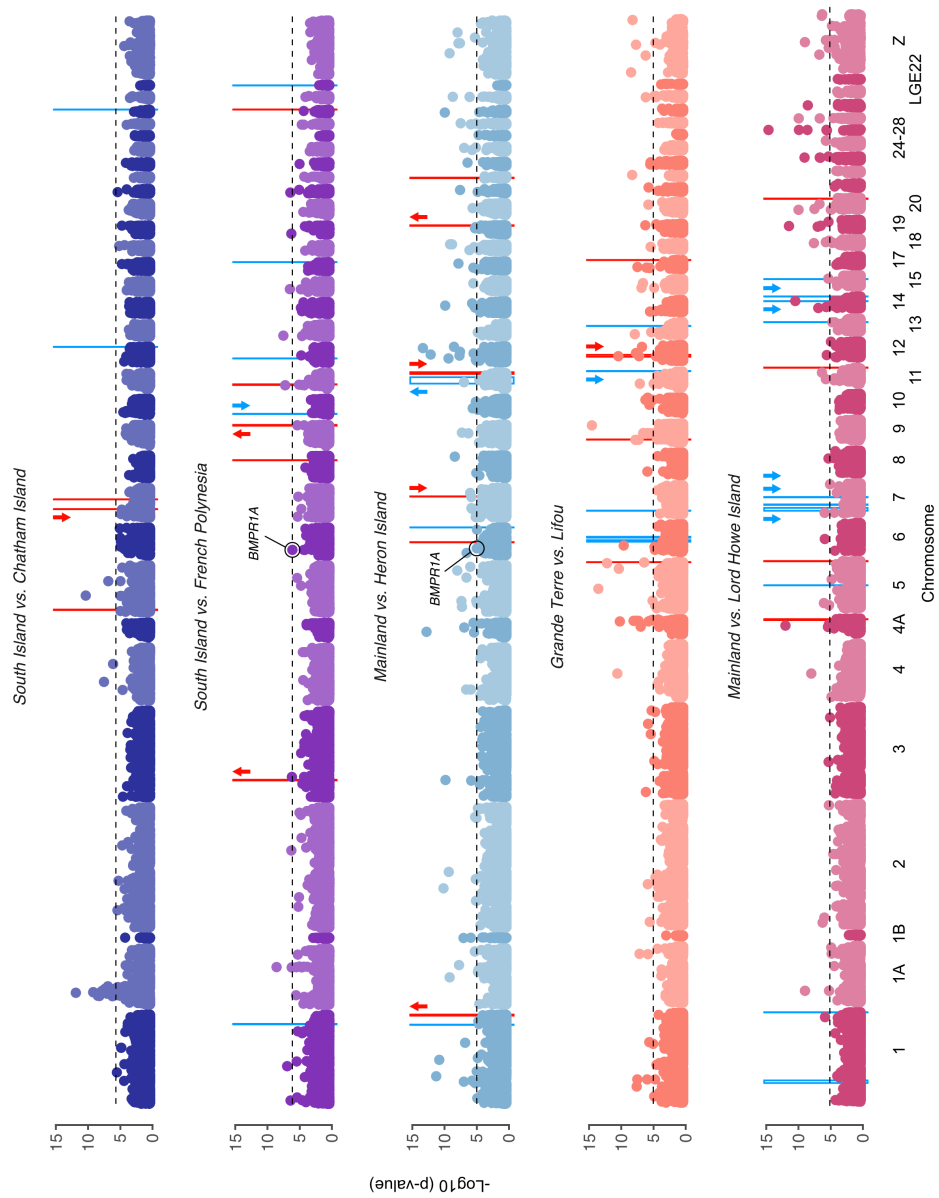


Figure 2.6. Manhattan plot of negative log₁₀ (*P*-values) estimated using *PCAdapt*. Dashed line indicates *P*-value threshold above which SNPs are determined outliers (FDR = 0.01). 50kb windows with elevated differentiation (genomic islands) are highlighted in red and regions with low differentiation (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 Fig 2.4. Arrows beside genomic islands/valleys indicate if d_{xy} within these regions was significantly elevated (upwards pointing arrows) or decreased (downwards pointing arrows) compared to chromosomal background levels. Significance determined using Wilcoxon signed-rank tests. Outlier SNPs located within candidate gene *BMPR1A* are labelled. Chromosomes are numbered according to the zebra finch nomenclature.

Table 2.5. Number of outlier SNPs identified using *PCAdapt* and the number located within genomic islands and 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	24	0	0
Early stage	No-gene flow	SI vs. FP	12	0	0
Mid stage	Gene flow	ML vs. HI	90	0	0
Late stage	Gene flow	GT vs. LF	74	0	0
Late stage	No-gene flow	ML vs. LH	55	0	0

2.3.6 Candidate gene analysis

A review of the literature identified 22 candidate genes associated with body and/or bill size differences in birds, for which we had sequence data or occurred within the 50kb windows used when summarising divergence statistics (Table 2.6). 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 *BMPRI1A* (Figure 2.6 and Table 2.6). This was the only candidate gene to contain outlier SNPs.

2.3.7 Simulations of neutral divergence

Simulations of neutral divergence, using an individual based model, were able to capture the transition from localised to more genome-wide levels of divergence, in which F_{ST} distributions were characterised by low values of F_{ST} during the early stages of divergence, shifting towards higher F_{ST} values over longer divergence timeframes (Figure 2.7). However, differences in distributional skew between simulations at different timescales were not all significant (Table B.3, Appendix B). Simulating divergence under various levels of gene flow ($m = 0$, $m = 0.0001$, $m = 0.001$, and $m = 0.01$) showed that the accumulation of divergence slowed under increasing levels of gene flow (Figure 2.8). Although F_{ST} distributions varied between simulated rates of recombination, we observed no clear pattern (Fig 2.8).

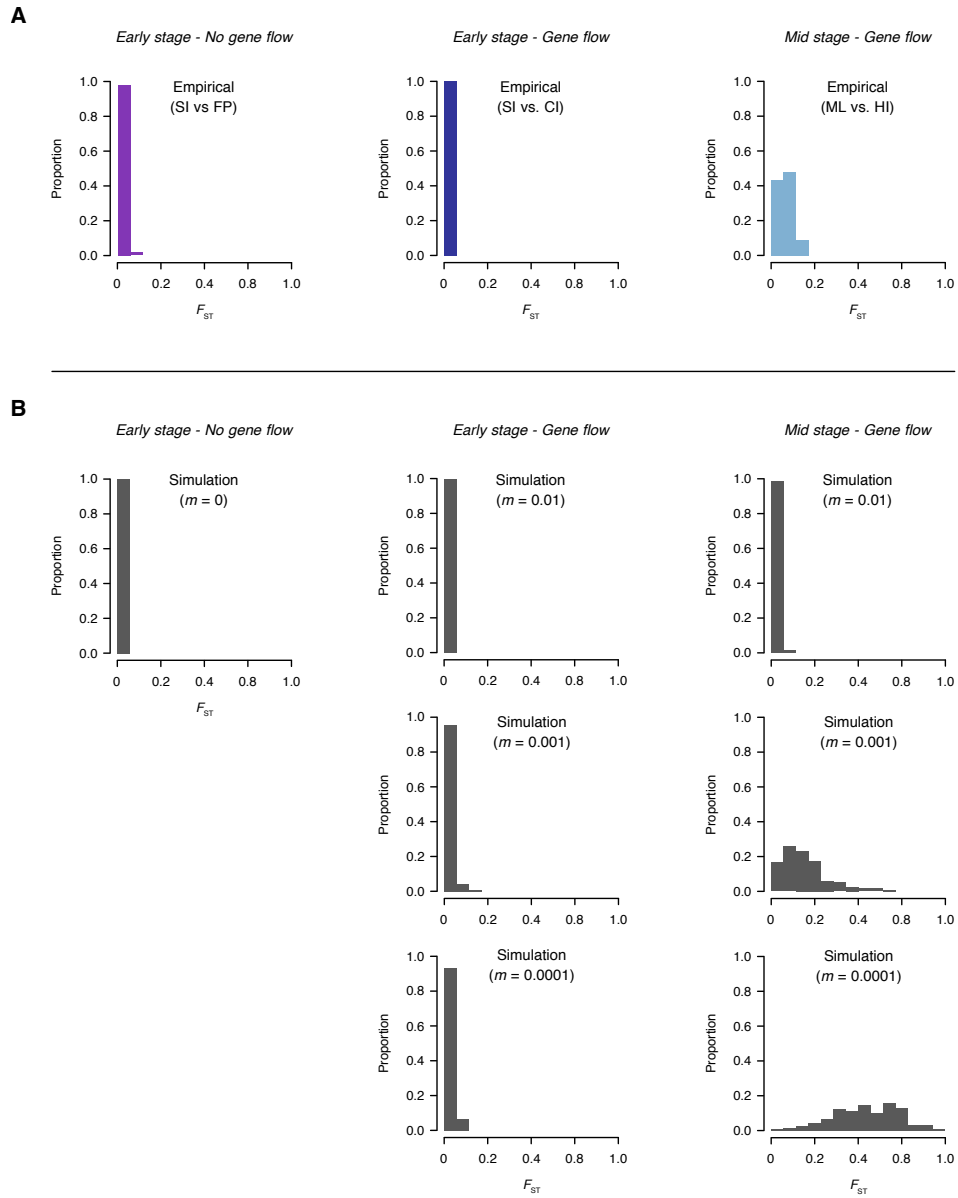


Figure 2.7. Comparison of observed and simulated F_{ST} distributions. (A) Frequency distributions of pairwise differentiation (F_{ST}) for empirical data (chromosome 5 only). (B) Frequency distributions of pairwise differentiation (F_{ST}) for simulated data. The F_{ST} values are calculated in 500kb 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: 30 generations; Early stage - gene flow: 60 generations; Mid stage - gene flow: 1,000 generations).

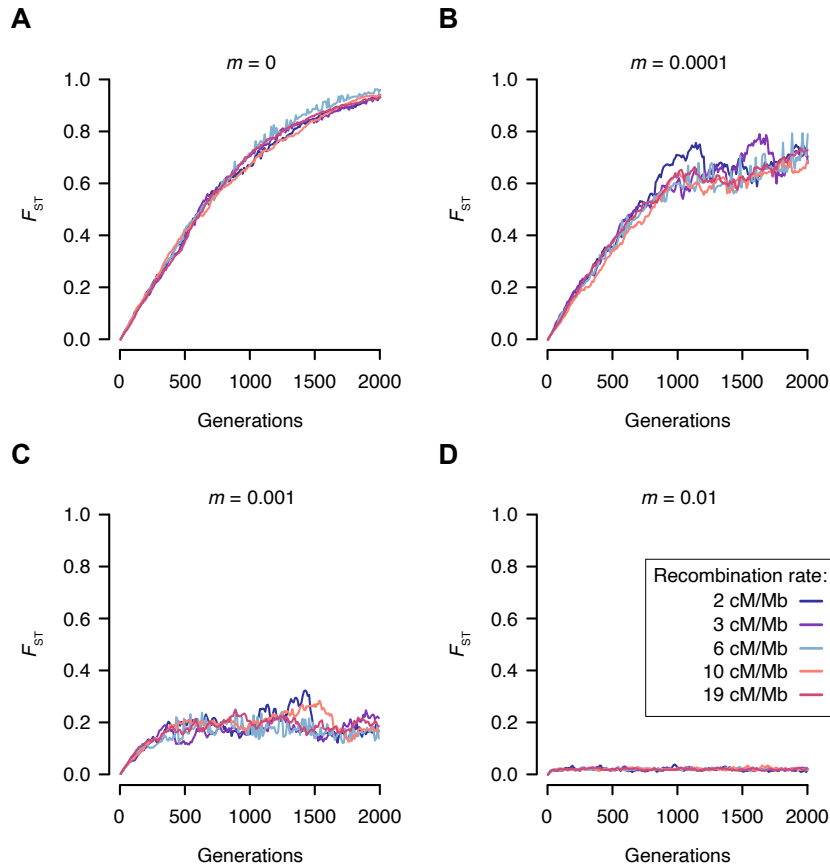


Figure 2.8. Expected F_{ST} values for 2,000 generations of simulated neutral divergence. (A) During divergence without gene flow ($m = 0$) F_{ST} rapidly increases, achieving fixation at 2,000 generations under all recombination rates; (B) During divergence with low levels of migration ($m = 0.0001$) 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 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.

Table 2.6. 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?
<i>ALX1</i>	1A	1	Bill	N	N
<i>ALX4</i>	5	2	Bill	N	N
<i>BMP4</i>	5	3, 4	Bill	N	N
<i>BMPR1A</i>	6	2	Bill	N	Y (SI vs. FP; ML vs. HI)
<i>CACNA1G</i>	18	5	Bill	N	N
<i>COL17A1</i>	6	6	Body	N	N
<i>COL4A5</i>	4A	2	Bill	N	N
<i>COL6A3</i>	7	6	Body	N	N
<i>FGF10</i>	Z	1	Bill	N	N
<i>FOXC1</i>	2	1	Bill	N	N
<i>GSC</i>	5	1	Bill	N	N
<i>HMGA2</i>	1A	7	Bill	N	N
<i>IGF1</i>	1A	8	Bill	N	N
<i>INHBA</i>	2	2	Bill	N	N
<i>ITPR2</i>	1A	6	Body	N	N
<i>ITPR3</i>	26	6	Body	N	N
<i>NELL1</i>	5	2	Bill	N	N
<i>RDH14</i>	3	1	Bill	N	N
<i>SATB2</i>	7	2	Bill	N	N
<i>SIX2</i>	3	2	Bill	N	N
<i>TRPS1</i>	2	2	Bill	N	N
<i>VPS13B</i>	2	1, 2	Bill	N	N

1. Lamichhaney et al. (2015)
2. Bosse et al. (2017)
3. Abzhanov et al. (2004)
4. Campàs et al. (2010)
5. Cheng et al (2017)
6. Cornetti et al. (2015)
7. Lamichhaney et al. (2016)
8. von Holdt et al. (2018)

2.4 Discussion

We take advantage of an exceptional natural system where gene flow scenarios and divergence timescales are known *a priori*, to reveal key insights into how divergence accumulates at the genomic level. Our results provide an empirical characterisation 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 not 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 localised to more genome-wide levels of divergence. The transition from localised divergence (highly negatively 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.

2.4.1 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 towards 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 however, the observed distributions of genomic differences across the genome fit with expectations for how genomic divergence accumulates across the speciation continuum (Nosil & Feder, 2012; Seehausen et al., 2014). However, our simulations demonstrated that the shift from localised 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 localised and ultimately genome-wide divergence (Seehausen et al., 2014). One important difference, however, was that our simulations suggest that the transition from localised 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 considered diverging population consisting of 400 individuals, 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 sizes 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 diverging under gene flow compared to populations diverging in 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 characterisation of this difference in the accumulation of genomic divergence. This moderating effect of gene flow was also apparent when simulating divergence. However, as the number of population comparisons in our study is few, we cannot state that all other populations diverging with gene flow will show this pattern of reduced divergence. Further comparisons will need to be conducted to confirm that this is a general pattern.

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 the 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 & 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 SNPs, 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).

2.4.2 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 assumed to be, genetically isolated (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 (subfamily: *Geospizinae*): (Han et al., 2017); and Yellow-rumped warblers (*Setophaga coronata*): (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. This observation may be consistent with the more recently proposed multi-locus coupling mechanism of divergence (Flaxman et al., 2013; Flaxman et al., 2014; Butlin and Smadja, 2017; Schilling et al., 2018), however more direct evidence is required.

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 two genomic islands were located at overlapping genomic positions across population comparisons, and only one of these occurred across

different divergence timeframes. This genomic island (located on Chromosome 11) was smallest in the late stage comparison. Whole genome analysis would likely provide more shared genomic islands to allow statistical comparisons, but currently we lack support for the idea that expansion of genomic islands of divergence at specific locations over time regularly 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). In fact we did not observe a single genomic island containing outlier SNPs. 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 (i.e. d_{xy} and F_{ST} respectively) can provide further insight into the formation of genomic islands (Nachman & 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. However, as we also observed islands where both types of measures were elevated in populations diverging without gene flow, our results provide support for an alternative explanation for this pattern: that genomic islands with elevated d_{xy} occur due to lineage sorting (Han et al., 2017). 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). Such islands were the most frequently observed across all comparisons, suggesting that genomic islands arise most 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 & 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 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 characterises 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.

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 (Roesti et al., 2014), in no case did the position of outlier SNPs correspond to the position of genomic valleys. As such, at present we have limited evidence to confirm that genomic valleys frequently arise due to effects of parallel 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).

2.4.3 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 (*BMPR1A*) 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 (Mainland, Australia and South Island, New Zealand) (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).

2.5 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-characterised 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.

2.6 Acknowledgements

This work was funded by a grant from the John Fell Oxford University Press Research Fund to TC and SMC, with additional support from an anonymous donation to KCR and a Natural Environment Research Council (NERC) studentship awarded

to ATSP. CSQ acknowledges support from the Early Postdoc Mobility fellowship of the Swiss National Science Foundation (P2GEP3_168973). Sample collection was conducted with the permission of the governments of Australia, French Polynesia, New Caledonia, New Zealand and Vanuatu. 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 acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work. We thank Samuel Flaxman and two anonymous reviewers who provided helpful comments on earlier drafts of this manuscript.

2.7 References

- Abzhanov A, Protas M, Grant BR, Grant PR, Tabin CJ** (2004). Bmp4 and Morphological Variation of Beaks in Darwin's Finches. *Science* **305**: 1462.
- Ali OA, Rourke SMO, Amish SJ, Meek MH, Luikart G, Jeffres C, et al.** (2016). RAD Capture (Rapture): Flexible and efficient sequence-based genotyping. *Genetics* **202**: 389-400.
- Andrew RL, Rieseberg LH** (2013). Divergence is focused on few genomic regions early in speciation: Incipient speciation of sunflower ecotypes. *Evolution* **67**: 2468-2482.
- Andrews S** (2010). FASTQC: A quality control tool for high throughput sequence data. Available at: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Backström N, Forstmeier W, Schielzeth H, Mellenius H, Nam K, Bolund E, et al.** (2010). The recombination landscape of the zebra finch *Taeniopygia guttata* genome. *Genome Research* **20**: 485-95.
- Black RA** (2010). *Phylogenetic and phenotypic divergence of an insular radiation of birds*. PhD Thesis: Imperial College London.
- Bosse M, Spurgin LG, Laine VN, Cole EF, Firth JA, Gienapp P, et al.** (2017). Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science* **358**: 365-368.
- Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, et al.** (2015). Linked selection and recombination rate variation drive the evolution of the

- genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Research* **25**: 1656–1665.
- Butlin R, DeBelle A, Kerth C, Snook RR, Beukeboom LW, Castillo Cajas RF, et al.** (2012). What do we need to know about speciation? *Trends in Ecology and Evolution* **27**: 27–39.
- Butlin RK, Smadja CM** (2017). Coupling, reinforcement, and speciation. *The American Naturalist* **191**: 155–172.
- Campàs O, Mallarino R, Herrel A, Abzhanov A, Brenner MP** (2010). Scaling and shear transformations capture beak shape variation in Darwin's finches. *Proceedings of the National Academy of Sciences* **107**: 3356.
- Carneiro M, Ferrand N, Nachman MW** (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.
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA** (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology* **22**: 3124–3140.
- Charlesworth D** (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics* **2**: 379–384.
- Charlesworth B, Coyne JA, Barton NH** (2017). The relative rates of evolution of sex chromosomes and autosomes. *American Naturalist* **130**: 113–146.
- Chaves JA, Cooper EA, Hendry AP, Podos J, De León LF, Raeymaekers JAM, et al.** (2016). Genomic variation at the tips of the adaptive radiation of Darwin's finches. *Molecular Ecology* **25**: 5282–5295.
- Cheng Y, Gao B, Wang H, Han N, Shao S, Wu S, et al.** (2017). Evolution of beak morphology in the ground tit revealed by comparative transcriptomics. *Frontiers in zoology* **14**: 58.
- Clegg SM, Degnan SM, Kikkawa J, Moritz C, Estoup A, Owens IPF** (2002a). Genetic consequences of sequential founder events by an island-colonising bird. *Proceedings of the National Academy of Sciences* **99**: 8127–8132.
- Clegg SM, Degnan SM, Moritz C, Estoup A, Kikkawa J, Owens IPF, 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.

- Clegg SM, Frentiu FD, Kikkawa J, Tavecchia G, Owens IPF** (2008). 4000 Years of phenotypic change in an island bird: Heterogeneity of selection over three microevolutionary timescales. *Evolution* **62**: 2393–2410.
- Cornetti L, Valente LM, Dunning LT, Quan X, Black R A., Hébert O, et al.** (2015). The genome of the “great speciator” provides insights into bird diversification. *Genome Biology and Evolution* **7**: 2680–2691.
- Coyne JA** (1985). The genetic basis of Haldane’s rule. *Nature* **314**: 736–738.
- Darwin C** (1859). *On the origin of species by means of natural selection*. Murray: London.
- Davey JL, Blaxter MW** (2010). RADseq: Next-generation population genetics. *Briefings in Functional Genomics* **9**: 416–423.
- Degnan SM, Moritz C** (1992). Phylogeography of mitochondrial DNA in two species of white-eyes in Australia. *The Auk* **109**: 800–811.
- Delmore KE, Lugo Ramos JS, Van Doren BM, Lundberg M, Bensch S, Irwin DE, et al.** (2018). Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds. *Evolution Letters* **2**: 76–87.
- Dobzhansky T** (1974). Genetic analysis of hybrid sterility withing the species *Drosophila pseudoobscura*. *Hereditas* **77**: 81–88.
- Ellegren H** (2009). Genomic evidence for a large-Z effect. *Proceedings of the Royal Society B: Biological Sciences* **276**: 361–366.
- Ellegren H** (2010). Evolutionary stasis: The stable chromosomes of birds. *Trends in Ecology and Evolution* **25**: 283–291.
- Estoup A, Clegg SM** (2003). Bayesian inferences on the recent island colonisation history by the bird *Zosterops lateralis lateralis*. *Molecular Ecology* **12**: 657–674.
- Feder JL, Egan SP, Nosil P** (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics* **28**: 342–350.
- Feder JL, Flaxman SM, Egan SP, Comeault AA, Nosil P** (2013). Geographic mode of speciation and genomic divergence. *Annual Review of Ecology, Evolution, and Systematics* **44**: 73–97.
- Feder JL, Nosil P** (2010). The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* **64**: 1729–1747.

- Feulner PGD, Chain FJJ, Panchal M, Huang Y, Eizaguirre C, Kalbe M, et al.** (2015). Genomics of divergence along a continuum of parapatric population differentiation. *PLoS genetics* **11**: e1004966.
- Flaxman SM, Feder JL, Nosil P** (2013). Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. *Evolution* **67**: 2577–2591.
- Flaxman SM, Wacholder AC, Feder JL, Nosil P** (2014). Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Molecular Ecology* **23**: 4074–4088.
- Grabherr MG, Russell P, Meyer M, Mauceli E, Alföldi J, di Palma F, et al.** (2010). Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* **26**: 1145–1151.
- Han F, Lamichhaney S, Rosemary Grant B, Grant PR, Andersson L, Webster MT** (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin’s finches. *Genome Research* **27**: 1004–1015.
- Hofer T, Foll M, Excoffier L** (2012). Evolutionary forces shaping genomic islands of population differentiation in humans. *BMC Genomics* **13**: 107.
- Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA** (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6**: e1000862.
- Hopley D** (1982). *The geomorphology of the Great Barrier Reef: Quaternary development of coral reefs*. John Wiley and Sons: New York.
- Kawakami T, Smeds L, Backström N, Husby A, Qvarnström A, Mugal CF, 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. *Molecular Ecology* **23**: 4035–4058.
- Kearns AM, Restani M, Szabo I, Schröder-Nielsen A, Kim JA, Richardson HM, et al.** (2018). Genomic evidence of speciation reversal in ravens. *Nature Communications* **9**: 906.
- Kikkawa J, Wilson JM** (1983). Breeding and dominance among the Heron Island silvereyes *Zosterops lateralis chlorocephala*. *Emu* **83**: 181–198.

- Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, et al.** (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**: 371–375.
- Lamichhaney S, Han F, Berglund J, Wang C, Almén MS, Webster MT, et al.** (2016). A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science* **352**: 470–474.
- Langmead B, Salzberg SL** (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359.
- Luu K, Bazin E, Blum MGB** (2017). pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources* **17**: 67–77.
- Ma T, Wang K, Hu Q, Xi Z, Wan D, Wang Q, et al.** (2017). Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. *Proceedings of the National Academy of Sciences* **115**: e236–e243.
- Mallet J** (2008). Hybridization, ecological races and the nature of species: Empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**: 2971–2986.
- Mank JE, Nam K, Ellegren H** (2010). Faster-Z evolution is predominantly due to genetic drift. *Molecular Biology and Evolution* **27**: 661–670.
- Marques DA, Lucek K, Meier JI, Mwaiko S, Wagner CE, Excoffier L, et al.** (2016). Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLOS Genetics* **12**: e1005887.
- Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, et al.** (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research* **11**: 1817–1828.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K** (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**: 1297–1303.
- Mees GF** (1969). A systematic review of the Indo-Australian Zosteropidae. *Zoologische Verhandelingen* **102**: 1–390.

- Monnet C, Thibault JC, Varney A** (1993). Stability and changes during the twentieth century in the breeding landbirds of Tahiti (Polynesia). *Bird Conservation International* **3**: 261–280.
- Nachman MW, Payseur BA** (2012). Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**: 409–421.
- Nielsen R** (2005). Molecular signatures of natural selection. *Annual Review of Genetics* **39**: 197–218.
- Nosil P** (2012). *Ecological Speciation*. Oxford University Press: Oxford.
- Nosil P, Feder JL** (2012). Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**: 332–342.
- Nosil P, Feder JL, Flaxman SM, Gompert Z** (2017). Tipping points in the dynamics of speciation. *Nature Ecology & Evolution* **1**: e0001.
- Nosil P, Funk DJ, Ortiz-Barrientos D** (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* **18**: 375–402.
- Oyler-McCance SJ, Cornman RS, Jones KL, Fike JA** (2015). Z chromosome divergence, polymorphism and relative effective population size in a genus of lekking birds. *Heredity* **115**: 452–459.
- Powell THQ, Hood GR, Murphy MO, Heilveil JS, Berlocher SH, Nosil P, et al.** (2013). Genetic divergence along the speciation continuum: The transition from host race to species in rhabdophorini (diptera: Tephritidae). *Evolution* **67**: 2561–2576.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al.** (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics* **81**: 559–575.
- Quilodr  n CS, Ruegg K, Sendell-Price AT, Anderson EC, Coulson T, Clegg SM** (2020). The multiple population genetic and demographic routes to islands of genomic divergence. *Methods in Ecology and Evolution* **11**: 6–21.
- Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlovi   M, et al.** (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary Biology* **30**: 1450–1477.

- Renaut S, Grassa CJ, Yeaman S, Moyers BT, Lai Z, Kane NC, et al.** (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications* **4**: 1827.
- Ricklefs RE, Bermingham E** (2007). The causes of evolutionary radiations in archipelagoes: Passerine birds in the lesser Antilles. *The American Naturalist* **169**: 285–297.
- Riesch R, Muschick M, Lindtke D, Villoutreix R, Comeault AA, Farkas TE, et al.** (2017). Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology & Evolution* **1**: 0082.
- Robertson BC, Degnan SM, Kikkawa J, Moritz CC** (2001). Genetic monogamy in the absence of paternity guards: The Capricorn silvereye, *Zosterops lateralis chlorocephalus*, on Heron Island. *Behavioral Ecology* **12**: 666–673.
- Roesti M, Gavrillets S, Hendry AP, Salzburger W, Berner D** (2014). The genomic signature of parallel adaptation from shared genetic variation. *Molecular Ecology* **23**: 3944–3956.
- Roesti M, Hendry AP, Salzburger W, Berner D** (2012). Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology* **21**: 2852–2862.
- Ruegg K, Anderson EC, Boone J, Pouls J, Smith TB** (2014). A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology* **23**: 4757–4769.
- Sætre GP, Sæther SA** (2010). Ecology and genetics of speciation in *Ficedula* flycatchers. *Molecular Ecology* **19**: 1091–1106.
- Schilling PM, Mullen PS, Kronforst M, Safran JR, Nosil P, Feder LJ, et al.** (2018). Transitions from single- to multi-locus processes during speciation with gene flow. *Genes* **9**: 274.
- Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, et al.** (2014). Genomics and the origin of species. *Nature Reviews Genetics* **15**: 176–192.
- Sendell-Price AT, Ruegg KC, Clegg SM** (2020). Rapid morphological divergence following a human-mediated introduction: The role of drift and directional selection. *Heredity* **124**: 535–549.

- Seutin G, White B, Boag PT** (1991). Preservation of avian blood and tissue samples for DNA analyses. *Canadian Journal of Zoology* **69**: 82–90.
- Smadja C, Galindo J, Butlin R** (2008). Hitching a lift on the road to speciation. *Molecular Ecology* **17**: 4177–4180.
- Stölting KN, Nipper R, Lindtke D, Caseys C, Waeber S, Castiglione S, et al.** (2013). Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Molecular Ecology* **22**: 842–855.
- Storchová R, Reif J, Nachman MW** (2010). Female heterogamety and speciation: Reduced introgression of the z chromosome between two species of nightingales. *Evolution* **64**: 456–471.
- Supple MA, Papa R, Hines HM, McMillan WO, Counterman BA** (2015). Divergence with gene flow across a speciation continuum of *Heliconius* butterflies. *BMC Evolutionary Biology* **15**: 204.
- Thibault J-C, Cibois A** (2017). *Birds of eastern Polynesia: A biogeographic atlas*. Lynx Publishing: Ontario.
- Toews DPL, Brelsford A, Grossen C, Milá B, Irwin DE** (2016). Genomic variation across the Yellow-rumped Warbler species complex. *The Auk* **133**: 698–717.
- Turner TL, Hahn MW, Nuzhdin S V.** (2005). Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology* **3**: 1572–1578.
- Van Doren BM, Campagna L, Helm B, Illera JC, Lovette IJ, Liedvogel M** (2017). Correlated patterns of genetic diversity and differentiation across an avian family. *Molecular Ecology* **26**: 3982–3997.
- Via S** (2012). Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**: 451–460.
- Via S, West J** (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology* **17**: 4334–4345.
- Vijay N, Bossu CM, Poelstra JW, Weissensteiner MH, Suh A, Kryukov AP, et al.** (2016). Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature Communications* **7**: 1–10.

- vonHoldt BM, Kartzinell RY, Huber CD, Le Underwood V, Zhen Y, Ruegg K, et al.** (2018). Growth factor gene IGF1 is associated with bill size in the black-bellied seedcracker *Pyrenestes ostrinus*. *Nature Communications* **9**: 4855–4855.
- Wang J, Street NR, Scofield DG, Ingvarsson PK** (2016). Variation in linked selection and recombination drive genomic divergence during allopatric speciation of European and American aspens. *Molecular Biology and Evolution* **33**: 1754–1767.
- Wolf JBW, Ellegren H** (2016). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics* **18**: 87–100.
- Wright S.** (1931) Evolution in Mendelian populations. *Genetics* **16**: 97–159.
- Wu CI** (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**: 851–865.
- Zhang D, Song G, Gao B, Cheng Y, Qu Y, Wu S, et al.** (2017). Genomic differentiation and patterns of gene flow between two long-tailed tit species (*Aegithalos*). *Molecular Ecology* **26**: 6654–6665.

The idiosyncratic nature of genomic divergence following population founding

Ashley T. Sendell-Price, Kristen C. Ruegg
Bruce C. Robertson & Sonya M. Clegg

3.1 Introduction

The establishment of a new population involves phases of founding and recovery leading to long-term persistence. During each stage, the random sampling effect of drift has the potential to affect the degree of diversity and divergence exhibited by a population (Clegg, 2010). Where populations are founded by few individuals and recovery times are long, a potential consequence suffered by newly founded populations is loss of genetic variation – a ‘founder effect’ (Nei, Maruyama, & Chakraborty, 1975; Wright, 1931). Loss of diversity has important implications on several fronts. For example, it may limit responses to new adaptive landscapes via loss of potentially adaptive alleles (Lande & Barrowdough, 1987), or in a severe form could result in inbreeding depression (Charlesworth & Charlesworth, 1987) and contribute to population extinction (Frankham, 2005). The stochastic effects of population founding may also result in loss or fixation of rare variants, and previously common alleles becoming rare, all of which may be important in directing the trajectory of a colonised population.

While population founding can result in the loss of alleles, especially rare ones, theoretical and empirical work suggests that the potential for substantial reductions in measures of population variation such as heterozygosity, additive genetic variance, and heritability is limited (Coyne & Orr, 2004). The reason being that alleles which make it to the newly established population often begin at intermediate frequencies, and at intermediate frequencies heterozygosity, additive genetic variance, and heritability are maximised (Coyne & Orr, 2004). For example, Nei (1975) argues that even under extreme founding conditions, i.e. the establishment of a new population by a single gravid female, rapid population growth following

founding can maintain levels of heterozygosity. Although empirical evidence for loss of diversity in wild populations does exist (i.e. house finch (*Carpodacus mexicanus*) (Hawley et al., 2006); Coqui frogs (genus: *Eleutherodactylus*) (Peacock et al., 2009); Lupines (genus: *Lupinus*) (Vásquez et al., 2016); and Steller's jay (*Cyanocitta stelleri*) (Burg et al., 2005)), loss of genetic diversity is not ubiquitous with many studies failing to detect clear founder effects (e.g. saddleback (*Philesturnus carunculatus*) (Taylor & Jamieson, 2008); Alpine Ibex (*Capra ibex*) (Biebach & Keller, 2012); white-tailed deer (*Odocoileus virginianus*) (Fuller et al., 2020) and humans (Tabbada et al., 2010).

Several explanations can be proposed for why significant loss of diversity is observed in some studies but not in others. For example, substantial reductions may not be observed in populations that are established from genetically depauperate populations as there is limited diversity to lose (Clegg, 2010; Taylor & Jamieson, 2008). Further, variation in the detection of founder effects may reflect differences in the density and sensitivity of markers used. For example, Short Tandem Repeats (STRs – often referred to as microsatellite markers) provide a poor prediction of genome-wide levels of nucleotide diversity (Roques et al., 2019; Väli et al., 2008). In contrast large numbers of Single Nucleotide Polymorphisms (SNPs) provide a greater power and precision to quantify genomic levels of diversity (Kardos et al., 2016). This is because the measurement of variability of SNPs scattered across a significant fraction of the genome make it possible to reliably predict overall levels of genetic variation (Väli et al., 2008). As many of the studies that fail to detect significant reductions in diversity make use of variation at few microsatellite markers and not SNPs, such lack of founder effects may be attributed to the low sensitivity of the markers used.

Where populations are the product of multiple founding steps, for example during a stepwise range expansion, the potential for loss of diversity is expected to be increased and the magnitude of this loss exaggerated. Contemporary and naturally occurring stepwise colonisations, where populations are the product of multiple founding steps are relatively few. Within the south Pacific, the silvereye bird (*Zosterops lateralis*) has colonised a series of islands from Tasmania during the past 200 years (see Figure 3.1) (Mees, 1969). This historically documented sequential colonisation is a classic of ornithological literature (Lack, 1971; Mayr, 1942). Taking advantage of the well-recorded colonisation history of *Z. lateralis*, using six microsatellite loci, Clegg et al. (2002) demonstrated that although single founder events had limited impact on genetic diversity or differentiation, there were significant losses of allelic diversity and increases in population divergence when populations were separated by multiple founding steps. A similar finding has since

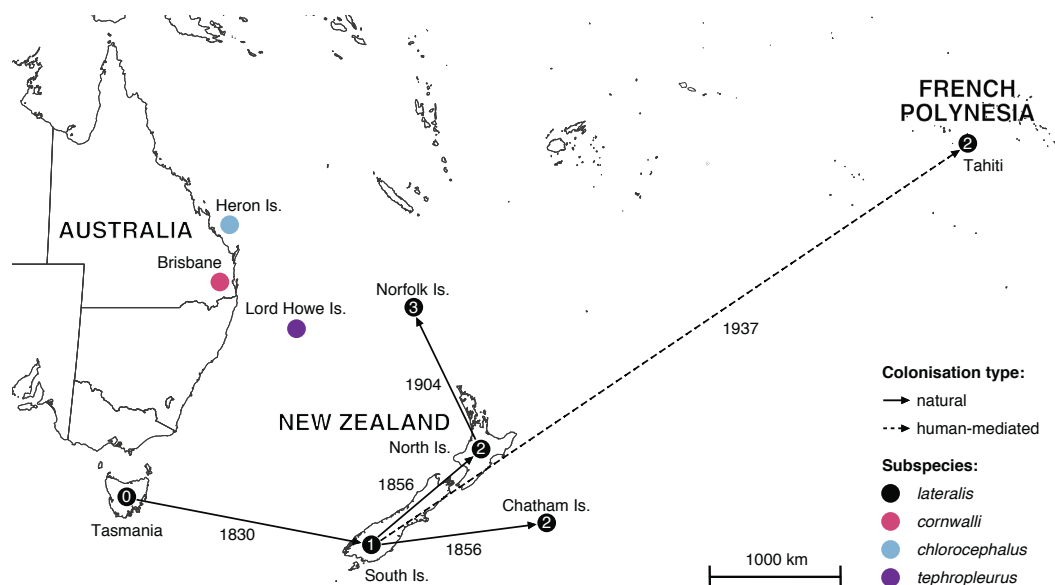


Figure 3.1. The distribution of silvereve forms used in this study. Black circles show sampling locations for the recent coloniser, *Z. l. lateralis*, and coloured circles indicate sampling locations for the ancestral Australian mainland form and evolutionarily old forms (Heron Island and Lord Howe Island). The timing and direction of the recent sequential colonisation by *Z. l. lateralis* is indicated by dates and arrows. Numbers within black circles indicate the number of founding steps separating the Tasmanian population from derived populations.

been reported in studies of song sparrows (*Melospiza melodia*) (Pruett & Winker, 2005) and invasive green anoles (*Anolis carolinensis*) (Michaelides et al., 2018). Both of these studies were also limited to few neutral microsatellite loci.

In the era of genomics, the availability of next generation sequencing technologies now provides the opportunity to re-evaluate the consequences of population founding on genetic diversity across both neutral and non-neutral loci and at a much higher resolution than before. More importantly genomic techniques also provide the opportunity to study the fate of individual loci. For example, the stochastic effects of population founding may result in loss or fixation of rare variants, and previously common alleles becoming rare. Here, using reduced representation sequencing, to generate many thousand SNPs, we revisit the sequential colonisation of the Tasmanian silvereve to reassess such genetic impacts of population founding.

Table 3.1. Sampling information. No. sequenced = number individuals sequenced; no. retained = number individuals retained post-filtering.

Sampling location	Subspecies	Code	Lat / Long	Year sampled	No. (sequenced)	No. (retained)
Mainland	<i>cornwalli</i>	ML	-28.18 / 153.45	2013	20	18
Tasmania	<i>lateralis</i>	TAS	-42.88 / 147.33	1997	39	19
South Island, NZ	<i>lateralis</i>	SZZ	-45.53 / 170.30	1997	23	21
North Island, NZ	<i>lateralis</i>	NNZ	-40.22 / 175.57	1997	17	10
Chatham Island	<i>lateralis</i>	CI	-43.94 / 176.49	1997	22	18
Norfolk Island	<i>lateralis</i>	NI	-29.04 / 167.96	1997	24	12
Tahiti	<i>lateralis</i>	TAH	-17.58 / -149.53	2017	21	9
Heron Island	<i>chlorocephalus</i>	HI	-23.44 / 151.91	2015	20	18
Lord Howe Island	<i>tephropleurus</i>	LH	-31.33 / 159.05	1998	20	20

3.2 Methods

3.2.1 Sample collection

Silvereyes were sampled from nine locations across Australia and islands of the south Pacific (Table 3.1). This included a series of historically colonised populations with known establishment dates, along with a population established from a human-mediated introduction. The populations sampled were: (1) *Z. l. lateralis* from the initial source (Tasmania) of a recent colonisation sequence; (2) *Z. l. lateralis* from the sequentially colonised islands of South Island (New Zealand), Chatham Island, North Island (New Zealand), Norfolk Island, and Tahiti; (3) *Z. l. cornwalli* from the east coast of the Australian mainland used as a representative of a large, outbreeding continental population; and (4) representatives of old island colonisation events, *Z. l. chlorocephalus* (Heron Island), and *Z. l. tephropleurus* (Lord Howe Island). All birds were caught using mist nets or traps and 20–40 μ l of blood collected from the brachial wing vein which was stored in 1 ml of lysis buffer (0.01M Tris-HCl; 0.01M NaCl; 0.01M EDTA; 1% n-lauroylsarcosine, pH 8.0) and frozen at -20°C (Seutin et al., 1991).

3.2.2 Sequencing and bioinformatics

Restriction-site associated DNA paired-end (RAD-PE) sequencing was used to assess the genomic consequences of population founding. This reduced representation sequencing method has the potential to identify many hundreds of thousands of single nucleotide polymorphisms (SNPs) distributed throughout the genome (Davey

& Blaxter, 2010). DNA extraction and library preparation were conducted at the University of California, Los Angeles. Libraries were prepared using restriction enzyme *SbfI*-HF, following the protocol outlined in Appendix A. The resulting libraries were sequenced on two Illumina HiSeq4000 lanes (Illumina, San Diego, CA, USA) at the UC Davis Genome Center using paired-end 150-bp sequence reads. The quality of sequencing reads was checked visually using FASTQC (Andrews, 2010) and reads assigned to individuals using the *process_radtags* script in the STACKS version 1.4 software pipeline (Catchen et al., 2013). Reads containing uncalled bases and/or bases of low quality were also discarded in this step using default quality thresholds (an average Phred score of 10 in sliding windows of 15% of the length of the read), as were reads with possible adapter contamination and/or missing the *SbfI*-HF restriction site. The STACKS *clone_filter* script was used to remove PCR duplicates. The remaining reads were then mapped to the *Zosterops lateralis melanops* genome assembly version 1 (NCBI Assembly GCA_001281735.1 (Cornetti et al., 2015) with BOWTIE2 version 2.2.6. (Langmead & Salzberg, 2012), using end-to-end alignment and default settings (allowing for a maximum of two mismatches in the seed (-n 2)). SNPs were identified using the *HaplotypeCaller* tool from the Genome Analysis Toolkit (GATK) nightly build version 2016-12-05-ga159770 (McKenna et al., 2010) and the outputted Variant Call Format (VCF) file filtered to remove indels and only include biallelic SNPs where the minor allele count was ≥ 2 , minimum genotype quality = 30, minimum depth = 8, and where SNPs were called in at least 50% of individuals. Prior to conducting downstream analyses the VCF file was further filtered using VCFtools to retain a subset of 89,026 SNPs and 145 individuals (those with less than 30% of data missing). The total number of individuals retained per population is reported in Table 3.1.

As the *Z. l. melanops* genome is only assembled to the scaffold level (Cornetti et al., 2015), we mapped *Z. l. melanops* scaffolds 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 *Z. l. melanops* scaffolds to chromosomes and determine their order, location and orientation using 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.

3.2.3 Population structure

We examined patterns of population structure by performing maximum likelihood estimation of individual admixture proportions using the program ADMIXTURE (Alexander et al, 2009), testing K values 1 to 9. For each value of K, we conducted 100 independent runs and summarised runs using CLUMPP v.1.1.2 (Jakobsson and Rosenberg, 2007). As the ADMIXTURE manual recommends avoiding SNPs with high linkage disequilibrium, we used the ‘--indep-pairwise 100 kb 1 0’ command in PLINK (Purcell et al, 2007) to remove one of every pair of SNPs with $r^2 > 0$ within 100 kb sliding windows.

3.2.4 Inferring demographic history

For each recently colonised population, we estimated demographic parameters from the site–frequency spectrum (SFS) using the composite-likelihood and coalescent simulation approach implemented in *fastsimcoal* v. 2.6 (Excoffier et al., 2013; Excoffier & Foll, 2011). As outgroup sequences were unavailable, demographic inference used the distribution of minor allele frequencies. To avoid problems associated with model overparameterisation, we tested two simple demographic models (Figure 3.2) that captured the key features of the process that we were interested in exploring, i.e. changes in population size following colonisation. The first (null model) represents a population of constant size and is defined by a single parameter, the current population size (N_{cur}). The second (bottleneck model) is characterised by five parameters: the ancestral population size (N_{anc}), time of colonisation event and start of population bottleneck (t_{col}), population size during the population bottleneck (N_{bot}), time of population expansion (t_{exp}), and the current population size (N_{cur}). Similar simple demographic models have been used to assess demographic declines in an endangered rattlesnake (*Sistrurus catenatus*) (Sovic et al., 2019) and to assess the impact of urbanisation on white-footed mice (*Peromyscus leucopus*) in New York City (Harris et al., 2016).

Parameter search ranges were selected based on historically recorded colonisation dates and assumed a generation time of 2.5 years (Kikkawa & Degnan, 1998). Using the South Island colonisation step as an example, t_{col} was fixed to 67 generations (based on known colonisation date - 167 years pre-sampling), t_{exp} was fixed between 47 and 65 generations ago (between 2 and 20 generations post colonisation), and N_{bot} was fixed between 2 and 500 individuals (within reasonable bounds for natural colonisation events by this species). For N_{anc} we set initial parameter search

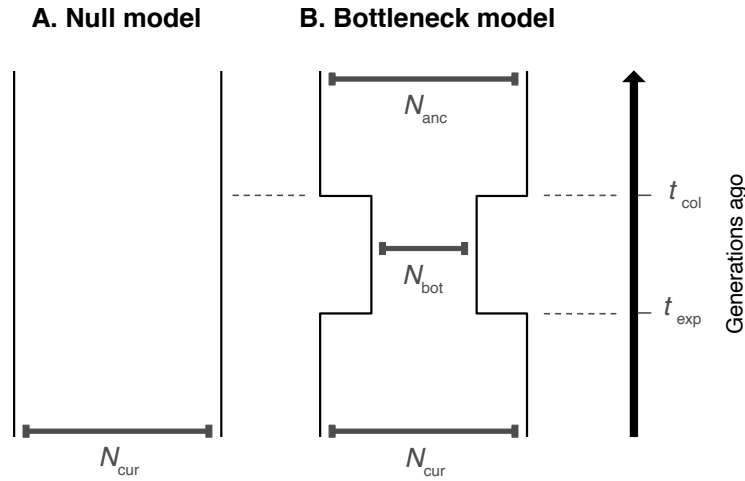


Figure 3.2. Schematic representation of the demographic scenarios modelled with *fastsimcoal2*. (A) Null model with a constant effective population size. (B) Bottleneck model incorporating a reduction in population size at time of colonisation followed by population recovery. t_{col} = time of colonisation event (and start of bottleneck); t_{exp} = time of population expansion (end of bottleneck); N_{anc} = ancestral population size; N_{cur} = current population size; N_{bot} = population size during population bottleneck.

ranges to between 1,000 and 100,000 individuals. Parameter search ranges for N_{cur} were set between 10,000 and 100,000 individuals for geographically large islands (SNZ and NNZ) and 500 and 10,000 individuals for smaller islands (CI, NI and TAH). For each population, we performed 100 independent model runs (40 Expectation/Conditional Maximization (ECM) cycles, 1,000,000 simulations per run) and chose the run with the highest likelihood as the best estimate of demographic parameters. The maximum-likelihood runs under each model were then compared using Akaike Information Criterion (AIC) to select the optimal model for each population.

As mutation rates for silvereyes are unknown, parameter estimates were based on estimated mutation rates for collared flycatchers (*Ficedula albicollis*): 4.6×10^{-9} mutations per nucleotide site per generation (Smeds et al., 2016). As *fastsimcoal2* requires the use of un-linked SNPs and is sensitive to missing data, we further filtered the LD-pruned dataset (see Population Structure section) to include only SNPs present in all individuals. VCFs were then converted to SFS format using the R package *vcf2sfs* (available on GitHub: <https://github.com/shenglin-liu/vcf2sfs>).

3.2.5 Population genomic analyses

Genetic diversity within populations was measured as average expected heterozygosity (H_e) and average nucleotide diversity (π), both of which were estimated using the Populations program of the STACKs software pipeline (Catchen et al, 2013). Confidence intervals across loci were calculated using bootstrap resampling (1,000 iterations). Significance of pairwise differences in heterozygosity and nucleotide diversity between each island population and its immediate source were assessed using Mann-Whitney tests. For the recently colonised populations, Spearman's Rank Correlation was used to test for associations between measures of genetic diversity and the number of colonisation steps proceeding population founding. For example, Tasmania has zero proceeding colonisation steps, whereas Norfolk Island is the product of three island colonisation steps (Tasmania to South Island New Zealand, South Island to North Island New Zealand, North Island to Norfolk Island).

To assess the effects of population founding on loss of rare alleles, for each recently colonised population we used the R package *vcfR* (Knaus and Grünwald, 2017) to calculate the minor allele frequency (MAF) for each SNP. We then identified rare alleles as those SNPs where the minor alleles frequency within the Tasmanian population was ≤ 0.1 and counted the number of rare alleles that were lost (MAF = 0) in subsequently colonised populations.

Genetic divergence between populations was assessed using Wright's fixation index (F_{ST}), a measure of relative divergence. Pairwise F_{ST} values were calculated, on a SNP-by-SNP basis, between each population using VCFtools (Danecek et al, 2011), and 95% confidence intervals estimated using bootstrap resampling (1,000 iterations). As the data is in matrix form, Mantel tests were used to test for associations between the number of colonisation steps separating any two populations and mean F_{ST} values, and p-values estimated using 1,000 randomisations.

3.2.6 Identification of highly diverged genomic regions

To determine how divergence accumulates across the genome as populations undergo multiple founding steps, we also calculated F_{ST} between the Tasmania population and each subsequently colonised population in non-overlapping 500 kb windows, using Python scripts developed by Martin et al (2013) (https://github.com/simonhmartin/genomics_general). We then compared the third moment (skewness) of F_{ST} distributions between population comparisons. Where divergence is limited to few loci, F_{ST} distributions were expected to be highly skewed, whereas

if more loci are involved then distributions are expected to be less skewed. We tested for significant differences in the skew of empirical distributions using a randomisation test, in which we compared absolute differences in skew for observed distributions to absolute differences calculated for 10,000 randomised distributions. Following this, we identified highly diverged genomic regions in each comparison as those 500kb windows with the top 1% of F_{ST} values. To determine if highly diverged regions were maintained in sequentially colonised populations, for each colonisation step we assessed if highly diverged regions identified in the previous colonisation step(s) maintained high levels of F_{ST} .

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). SNPs with high-LD were removed using a window size of 200 SNPs and an r^2 threshold of 0.1. Outlier SNPs were identified using the following settings for all population comparisons: K = 2, MinMAF= 0.2, and FDR = 0.01.

3.3 Results

3.3.1 RAD-PE sequencing and bioinformatics

Overall, RAD-PE sequencing resulted in an average of 3,370,049 paired-end reads per sample ($\pm 365,973$) covering 3,205,474 variable sites. Subsequent quality filtering consisted of removal of indels and only including biallelic SNPs where: the minor allele count was ≥ 2 ; the minimum genotype quality = 30; the minimum depth = 8; and SNPs were called in at least 50% of individuals. This filtering reduced the total number of SNPs to 446,432. Of the 216 samples sequenced, 145 were retained after removing individuals where $\geq 30\%$ of data was missing across a subsample of 89,026 SNPs (full-SNP dataset). The number of individuals retained per location ranged from 9 to 21 (Table 3.1). Reordering of *Zosterops* scaffolds onto *T. guttata* chromosomes, based on output from Satsuma Synteny and removal of unmapped/unoriented scaffolds, resulted in 87,099 SNPs (*T. guttata*-orientated SNP dataset). Filtering of the full-SNP dataset to remove linked SNPs resulted in 5,000

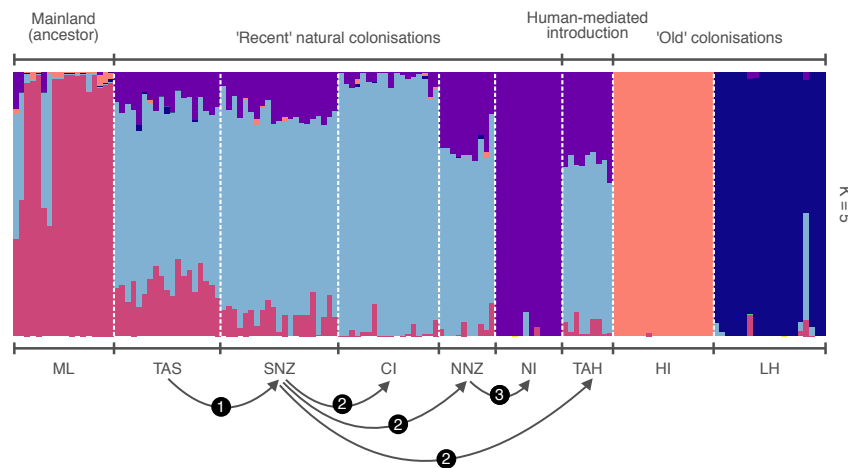


Figure 3.3. Maximum likelihood estimation of individual ancestries calculated with ADMIXTURE (Alexander et al, 2009) for all populations analysed, using K=5 (mean cross-validation error = 0.592). Based on 5,000 LD-filtered SNPs. Arrows indicate the direction of stepwise colonisations.

unlinked SNPs for admixture analysis, and subsequent removal of SNPs with missing data resulted in 1,587 SNPs for demographic modelling.

3.3.2 Population structure

Maximum likelihood estimation of individual ancestries calculated with ADMIXTURE (Alexander et al., 2009), consistently provided the lowest cross-validation error for K = 5 (mean cross validation error across runs = 0.592). At K = 5, individuals were grouped as follows: 1) Mainland (with some exceptions); 2) Tasmania and sequentially colonised populations excluding Norfolk Island; 3) Norfolk Island; 4) Heron Island; and 5) Lord Howe Island (Figure 3.3)

3.3.3 Demographic history

Statistical comparisons of the two demographic models showed strongest support (lowest AIC values) for the model incorporating a population size change as the best-fitting for all populations (Table 3.2). Based on this model, point estimates of the effective founding population sizes for recent colonised silveryeye populations ranged from 17 to 357 effective founders and population recover times (time between colonisation and end of population bottleneck) ranged from 7 to 14 generations

Table 3.2. Best parameter estimates from demographic modelling implemented in *fastsimcoal2*. Model A: Null model with a constant effective population size. Model B: Bottleneck model incorporating a reduction in population size at time of colonisation followed by population recovery. N_{anc} = ancestral population size; N_{cur} = current population size; N_{bot} = population size during population bottleneck; t_{col} = timing of colonisation event; t_{exp} = timing of population expansion (end of bottleneck). All timings are in generations (using a 2.5-year generation time).

Model	Pop.	N_{cur}	N_{bot}	t_{col}	t_{exp}	N_{anc}	Recovery time ($t_{col}-t_{exp}$)	Max estimated likelihood	Max observed likelihood	AIC
A	SNZ	64,403	-	-	-	-	-	-1554.951	-1529.263	7162.814
	CI	7,679	-	-	-	-	-	-1114.051	-1015.758	5132.295
	NNZ	44,393	-	-	-	-	-	-1000.642	-992.031	4610.127
	TAH	8,327	-	-	-	-	-	-873.326	-844.641	4023.985
	NI	2,130	-	-	-	-	-	-1040.830	-1015.758	4795.199
B	SNZ	29,513	139	53	67	84,910	14	-1536.812	-1529.263	7087.405
	CI	957	357	47	56	83,492	9	-1104.225	-1015.758	509.144
	NNZ	84,270	100	45	56	13,002	10	-995.022	-992.031	4592.388
	TAH	3,021	17	25	32	77,642	7	-846.246	-844.641	3907.107
	NI	5,441	48	25	37	24,732	12	-1019.375	-1015.758	4704.395

(Table 3.2). Both the Tahiti and Norfolk Island populations had effective founding populations of less than 50 individuals.

3.3.4 Genetic diversity and differentiation in recently colonised populations

When each recently colonised population was compared to its immediate source, significant reductions in expected heterozygosity were observed in North New Zealand, Chatham Island, Norfolk Island and Tahiti populations (all Mann-Whitney P -values <0.001) but not South New Zealand (Figure 3.4A and Table 3.3). A similar pattern was observed for nucleotide diversity (all Mann-Whitney P -values <0.001), with the exception of the North New Zealand population, which had significantly increased nucleotide diversity compared to its immediate source (South New Zealand) (Mann-Whitney: $V = 977,670,000$, P -value <0.001). As determined using spearman's rank correlation, both measures of genetic diversity were negatively associated with the number of founding steps preceding population establishment (Nucleotide diversity: $r = -0.040$; P -value <0.001 ; Expected heterozygosity: $r = -0.044$; P -value <0.001). All recently colonised populations had significantly higher expected heterozygosity and allelic diversity than evolutionarily old colonisations (all Mann-Whitney P -values <0.001).

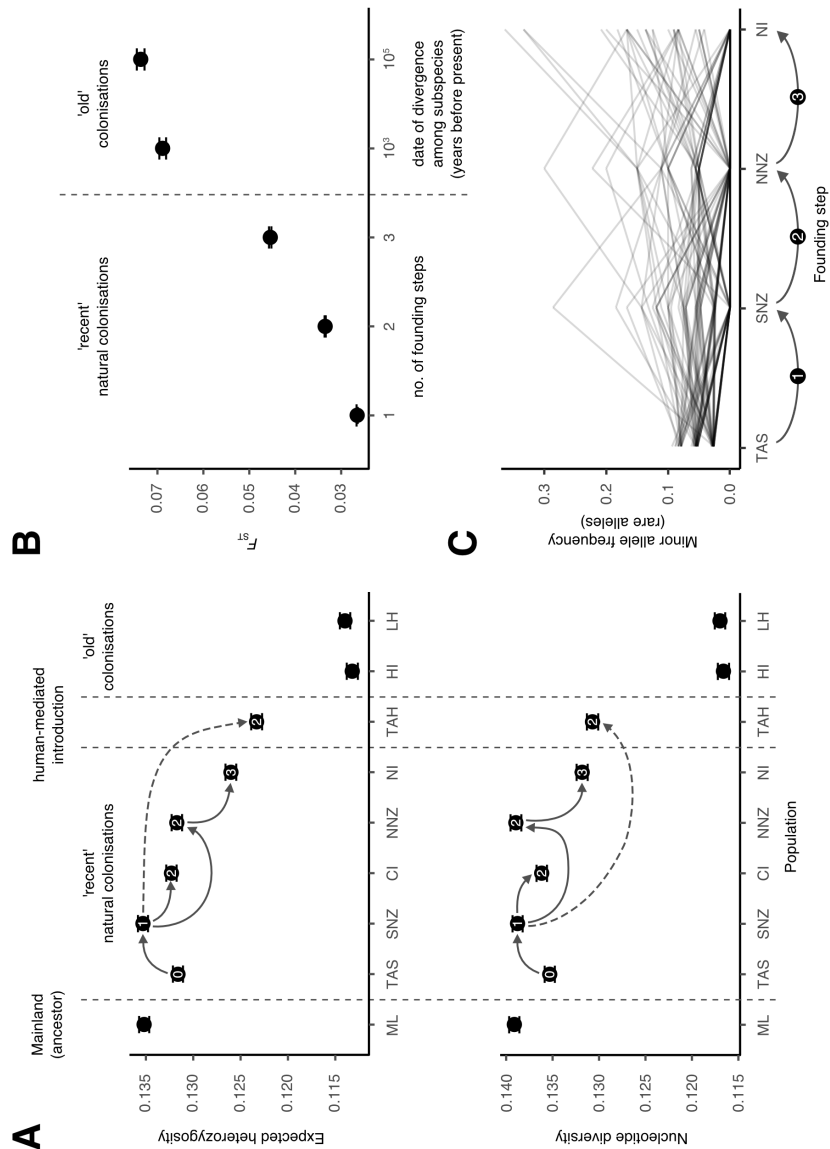


Figure 3.4. (A) Genetic diversity of silvereye forms as measured by expected heterozygosity and nucleotide diversity. Arrows indicate recent colonisation sequence. Error bars indicate 95% bootstrap confidence intervals across SNPs. Abbreviations for populations are the same as in Table 3.1. (B) Mean genetic divergence of silvereye populations as measured by pairwise F_{ST} . Error bars indicate 95% confidence intervals calculated across comparisons. Number of founder events is the number of island colonisations separating any two populations. Age of divergence among subspecies and species were based on molecular and paleobiological evidence (Black, 2010; Degnan & Moritz, 1992; Hopley, 1982). (C) Changes in the Minor Allele Frequency (MAF) of 100 randomly sampled rare alleles for the Tasmania to Norfolk Island colonisation sequence. Rare alleles were identified in the Tasmania population (MAF <0.1) and tracked in subsequent colonisation steps. Arrows indicate the direction of stepwise colonisations.

Table 3.3. Pairwise differences in expected heterozygosity (above diagonal) and nucleotide diversity (below diagonal) between populations. Level of significance, estimated using Mann-Whitney tests, indicated by: * <0.05, ** <0.01, *** < 0.001, NS = not significant. For recent colonisations, populations separated by single founding steps are highlighted in blue, two founding steps in orange, and three founding steps in pink.

	ML	TAS	SNZ	CI	NNZ	NI	TAH	HI	LH
ML	-	0.0036 ***	0.0001 ***	0.0029 ***	0.0035 ***	0.0092 ***	0.0119 ***	0.022 ***	0.0212 ***
TAS	0.0038 ***	-	0.0037 ***	0.0007 NS	0.0002 ***	0.0056 ***	0.0083 ***	0.0184 ***	0.0177 ***
SNZ	0.0004 ***	0.0035 ***	-	0.003 ***	0.0036 ***	0.0093 ***	0.012 ***	0.0221 ***	0.0213 ***
CI	0.003 ***	0.0009 NS	0.0026 ***	-	0.0006 *	0.0063 ***	0.009 ***	0.0191 ***	0.0183 ***
NNZ	0.0002 ***	0.0036 NS	0.0002 ***	0.0028 ***	-	0.0057 ***	0.0084 ***	0.0185 ***	0.0178 ***
NI	0.0073 ***	0.0035 ***	0.007 ***	0.0044 ***	0.0071 ***	-	0.0027 ***	0.0128 ***	0.0121 ***
TAH	0.0084 ***	0.0046 ***	0.0081 ***	0.0055 ***	0.0083 ***	0.0011 *	-	0.0101 ***	0.0093 ***
HI	0.0225 ***	0.0187 ***	0.0222 ***	0.0196 ***	0.0223 ***	0.0152 ***	0.087 ***	-	0.0008 NS
LH	0.0222 ***	0.0183 ***	0.0218 ***	0.0192 ***	0.022 ***	0.0148 ***	0.0137 ***	0.0004 NS	-

As determined using Mantel tests of matrix comparisons, mean pairwise F_{ST} calculated between recently colonised forms was positively associated with the number of colonisation steps separating populations ($r = 0.701$, $P < 0.01$). Mean, pairwise F_{ST} increased as the number of founding steps between populations increased, with levels of divergence between populations separated by three founding steps approaching levels seen between the youngest of the evolutionarily ‘old’ *Zosterops* forms (Heron Island) and the mainland population (1,000 generations of separation) (Figure 3.4B). When treating each pairwise comparison separately, with the exception of the Norfolk Island vs. Tahiti comparison, F_{ST} between recently colonised populations was significantly lower than that observed between the mainland population and evolutionarily ‘old’ *Zosterops* forms (Figure 3.4B; Mann-Whitney P -values <0.01 for all comparisons).

Within the Tasmanian population 29,382 sites were identified as containing rare alleles (sites where the minor allele frequency ≤ 0.1). In subsequent colonisation steps, between 8,857 and 19,039 loci lost rare alleles identified in the Tasmanian population (Table 3.4). For the naturally colonised populations, the number of rare alleles lost increased as the number of founding steps separating them from Tasmania increased. The largest loss of alleles (19,039) observed across all recently

Table 3.4. Mean F_{ST} , F_{ST} distributional skew and the number of rare alleles (identified in the Tasmanian population) that were lost in subsequent colonisation steps (total number of rare alleles identified = 29,382). For comparisons were populations are separated by $\geq$ founding steps. P-values from randomisation tests comparing skewness of F_{ST} distributions to Tasmania vs. South Island step are also reported.

Comparison	No. founding steps separating populations	Mean F_{ST}	F_{ST} distributional skew	P value	No. rare alleles lost
TAS vs. SNZ	1	0.007	3.964	-	8,857
TAS vs. NNZ	2	0.011	3.296	0.2463	14,862
TAS vs. CI	2	0.015	2.626	<0.001	12,567
TAS vs. TAH	2	0.024	2.383	<0.001	19,039
TAS vs. NI	3	0.031	1.817	<0.001	17,877

founded populations occurred for the human-introduced, Tahitian population. For a random subset of 100 rare alleles MAF fluctuated widely between population founding steps (Figure 3.4C).

3.3.5 Genetic diversity and differentiation in evolutionarily old island populations.

evolutionarily old *Zosterops* forms had significantly lower nucleotide diversity (Figure 3.4A; Mann-Whitney P -values <0.01 for all comparisons) and significantly lower expected heterozygosity (Figure 3.4A; Mann-Whitney P -values <0.01 for all comparisons) than their ancestral source population of *Z. l. cornwalli* on the Australian mainland. The lowest levels of expected heterozygosity and nucleotide diversity were observed for the Heron Island (*Z. l. chlorocephalus*) population. Measures of genetic diversity were also lower than that observed for all recently colonised populations (Table 3.3) and levels of divergence were well beyond that observed in recently colonised populations (Figure 3.4B). Of the two evolutionarily old forms, *Z. l. tephroleurus* on Lord Howe Island exhibited the highest levels of divergence from the Mainland population (Figure 3.4B).

3.3.6 Genomic patterns of divergence and outlier loci

The distributions of F_{ST} values calculated in 500kb windows were most skewed for the single-colonisation step comparison (Tasmania vs. South Island; skew = 3.964). Skewness of F_{ST} distributions decreased as the number of colonisation

Table 3.5. Total number of outlier SNPs identified using *PCAdapt* and the number located within windows with the top 1% of F_{ST} values for each population comparison.

Comparison	# outlier SNPs	# within top 1% of F_{ST} windows
TAS vs. SNZ	76	0
TAS vs. CI	52	1
TAS vs. NNZ	7	0
TAS vs. TAH	38	1
TAS vs. NI	43	3

steps separating populations from Tasmania increased, with the lowest levels of distributional skew observed for the three-step comparison (Tasmania vs. Norfolk Island; skew = 1.817) (Table 3.4). With the exception of the Tasmania vs. North Island comparison, differences in distributional skew were significant as determined using randomisation tests (Table 3.4).

The position of 500 kb windows with the top 1% of F_{ST} values varied between population comparisons, and in many instances highly diverged windows showed reduced divergence in subsequent steps (Figure 3.5). Using *PCAdapt*, we identified between 7 and 76 outlier SNPs per population comparison as putatively under directional selection (Table 3.5). In few cases did outlier SNPs correspond to the position of the top 1% of F_{ST} windows (Table 3.5 and Figure 3.6).

3.4 Discussion

A genomic-level comparison of a historically recent, sequential colonisation sequence with more ancient island populations of silvereyes in Australia and the south Pacific has provided novel insights into the impact of founder events on genomic divergence. First, by tracking the fate of individual loci, we find that rare alleles are frequently lost following the establishment of new populations, and that the magnitude of this loss is increased where populations are founded sequentially. Second, by tracking divergence across the genome, we find that population founding has an idiosyncratic effect on the buildup of divergence, with highly diverged genomic regions frequently becoming less diverged and regions of little divergence frequently becoming more diverged, in subsequent founding steps. Furthermore, several findings from the original study of this system using six microsatellites (Clegg et al., 2002) were supported by the larger dataset. These include the general pattern of gradual loss of variation with sequential founder steps, and the lack of rapidly elevated

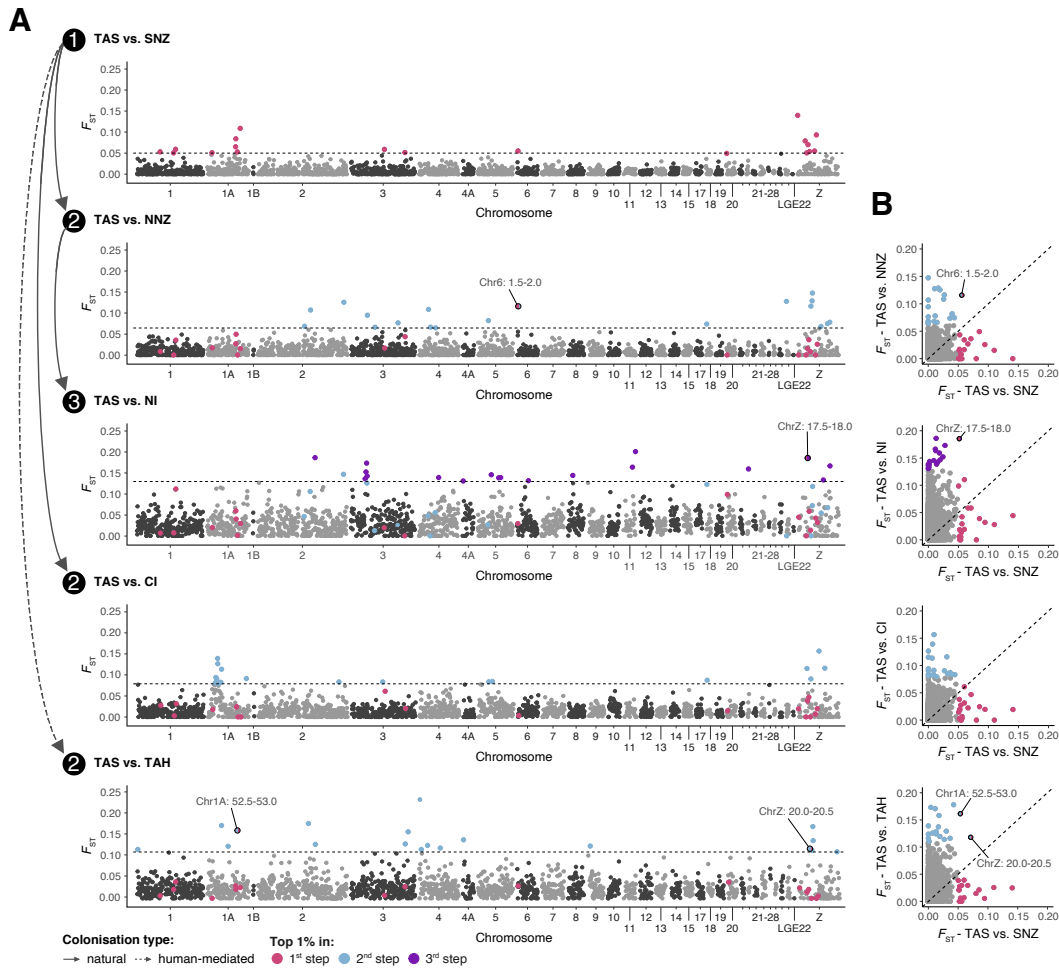


Figure 3.5. (A) Pairwise differentiation (F_{ST}) across the genome for recent stepwise colonisations by the Tasmanian silvereye (*Z. l. lateralis*). F_{ST} values were calculated in 500kb windows. The horizontal dashed line indicates the 99th percentile used to identify highly diverged windows. Points above this line in single step colonisations are highlighted in pink; two step colonisations highlighted in blue; and three step colonisations highlighted in purple. 500 kb windows identified as in the top 1% of F_{ST} windows in multiple comparisons are highlighted and their chromosomal position indicated in megabases. Chromosomes are shown with alternating light and dark shading. Chromosomes are numbered according to the zebra finch nomenclature. Arrows indicate the direction of stepwise colonisations. Numbers within black circles indicate the number of founding steps separating the Tasmanian population from derived populations. (B) Pairwise F_{ST} values calculated between Tasmania and South New Zealand (founding step 1) compared to pairwise F_{ST} values calculated between Tasmania and subsequently colonised populations. F_{ST} values were calculated in 500kb windows. Points in pink indicate top 1% of F_{ST} windows in the step 1 comparison and points in blue or purple indicate top 1% of F_{ST} windows identified in subsequent founding steps. 500kb windows identified as in the top 1% of F_{ST} windows in multiple comparisons are highlighted and their chromosomal position indicated in megabases. Abbreviations for populations are the same as in Table 3.1.

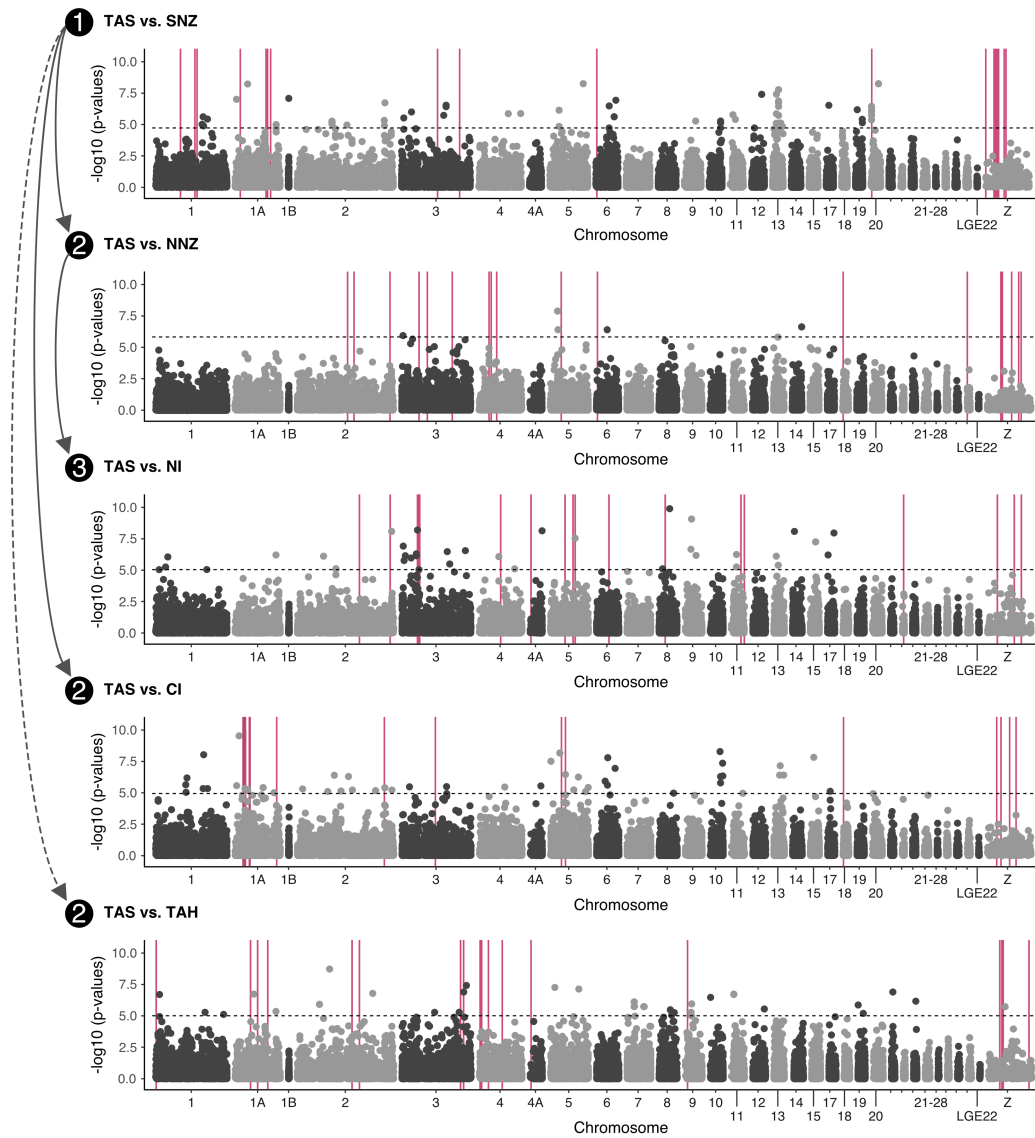


Figure 3.6. Manhattan plot of negative log₁₀ (P -values) estimated using *PCAdapt*. Points above the dashed line indicate outlier SNPs identified using $FDR = 0.01$. Highly diverged regions (500kb windows with the top 1% of F_{ST} values) are highlighted in pink. Chromosomes are shown with alternating light and dark shading. Chromosomes are numbered according to the zebra finch nomenclature. Abbreviations for populations are the same as in Table 3.1.

levels of genetic divergence with single founder steps that would be comparable to evolutionarily old populations. The high number of SNPs produced by this study translated to much more precise estimates of genetic variation, with much narrower confidence intervals compared to those using microsatellites, meaning that statistically significant differences in variation were seen with single founder steps using RAD-seq data as compared to microsatellite data.

3.4.1 The distribution of divergence across the genome

The pattern of increased F_{ST} with the number of founder events separating populations were the same for microsatellite and genomic data – the latter showing average F_{ST} approximately five times higher for populations separated by three founding steps compared to one. But how evenly does this divergence accumulate across the genome? When comparing the genomic landscape of divergence for each colonised population with the original source in Tasmania, the distribution of individual SNPs is more skewed (more localised) for the single step comparison (Tasmania vs. South Island) than for comparisons involving two or three founder steps. This suggests that increasing the number of founding steps separating populations, generates a more genome wide pattern of divergence, rather than average F_{ST} being elevated by a few outlier SNPs.

By tracking the fate of highly diverged loci across comparisons we are also able to determine if highly diverged regions are maintained or diverge further as populations are formed from sequential founding steps. Of the highly diverged regions identified for the single founding step comparison (Tasmania vs. South Island) very few were within the top windows identified in subsequent colonising steps, and in many instances highly diverged windows showed reduced divergence in subsequent steps. This suggests that, although highly diverged genomic regions do rapidly arise, due to the stochastic effects of population founding such differences are frequently lost in subsequent founding steps. Recent simulations of divergence using an individual-based framework suggest that this stochastic nature of genomic divergence is in part driven by the genetic composition and variation of founders (Quilodrán et al., 2019).

A key challenge of studying the genomic landscape of divergence is disentangling the effects of different evolutionary processes (Ravinet et al., 2017; Westram & Ravinet, 2017). During the initial stages, divergence is expected to be localised to regions of the genome where loci are under strong selection, forming peaks of divergence often referred to as ‘genomic islands of divergence’ (Feder et al.,

2013; Seehausen et al., 2014). Therefore, differences in the positions of windows with the top 1% of F_{ST} values between comparisons, may reflect differences in the selective environment. Contrary to this expectation, we found that highly diverged regions of the genome rarely contained outlier SNPs putatively under directional selection. Therefore, keeping in mind the limitations of current annotations of passerine genomes, differences in the position of windows with the top 1% of F_{ST} values between comparisons do not appear to reflect differences in the selective environment between comparisons.

A further noteworthy finding is the lack of continuity in the position of outlier loci across comparisons. Silvereye populations show a repeated pattern of phenotypic change on islands, some of which are encapsulated by the ‘island syndrome’ (Clegg & Owens, 2002; Leisler & Winkler, 2015). These changes include morphology, such as changes in body size; life history, such as having fewer offspring but living longer; and population-level characteristics, such as having high population densities. Given the repeatability of these morphological changes, we had expected to observe some consistency in the position of outlier loci between comparisons. However, in very few cases did outlier loci, putatively under directional selection, occur at the same position across multiple comparisons.

3.4.2 Loss of diversity following population founding

The relatively minor impact of single founder events on silvereye genetic variation, compared to the cumulative effects of sequential founder events previously found with microsatellite variation (Clegg et al., 2002) was also born out using our RAD-Seq dataset. Broadly speaking, measures of genetic diversity (nucleotide diversity, expected heterozygosity and presence of rare alleles) were lowest when populations were separated by multiple founding steps. Evidence that multiple founding steps have such a cumulative effect have accrued from several studies of microsatellite variation in other cases of sequential colonisation (Lambert et al., 2005; Michaelides et al., 2018; Pruett & Winker, 2005; Thulin et al., 2011), though none have been investigated using genome-level data. Such reductions in genetic diversity, in particular loss of rare alleles, could limit the evolutionary potential of populations to adapt to changing environments (Frankham et al., 2002).

The one exception where a single founder event had a larger impact on diversity was the human-mediated introduction of silvereyes to Tahiti, which exhibited a pronounced loss of genetic diversity as measured by both nucleotide diversity and expected heterozygosity. This population was not sampled by Clegg et al. (2002).

Based on the nature of the introduction program carried out by Eastham Guild (Guild, 1938; Thibault & Cibois, 2017; Thibault & Monnet, 1990), it is likely that the introduction of the silvereve consisted of a single release of only a handful of individuals. A small founding population is supported by previous demographic modelling of this population (Sendell-Price et al., 2020). In comparison, demographic modelling of natural silvereve colonisations pointed to relatively large founding flock sizes (100+ individuals) for all populations except for the last island in the colonisation sequence, Norfolk Island, where a smaller founding size was estimated (Clegg et al., 2002; Estoup & Clegg, 2003). A small founding population size could therefore explain the reduced variation noted from a single founder step. Our estimates of effective founding population sizes are consistent with this explanation, indicating a lower number of effective founding for Tahiti, compared to naturally established silvereve populations. However, a small effective founding population size was also estimated for the Norfolk Island population, suggesting that not all natural colonisations by the silvereve are the product of large founding flocks. Like the Tahitian silvereves, the Norfolk Island population also showed substantial loss of diversity, although this may be an artifact of the additional founding steps that this population has undergone. In all, our estimates of founding population sizes were in line with those estimated using microsatellite loci (Clegg et al., 2002).

3.4.3 Genetic divergence and diversity compared to evolutionarily old populations

We tested the potential evolutionary impact of sequential founder events, by comparing the level of divergence and diversity among the recently colonised populations to that seen in evolutionarily old populations, corresponding to colonisations that took place between several thousand and hundreds of thousands of years ago. Consistent with Clegg et al. (2002), evolutionarily old populations exhibited the lowest levels of genetic diversity and highest levels of divergence. However, sequentially founded populations did not achieve levels of differentiation or diversity seen in older colonisation events as was previously observed for the silvereve using microsatellite loci (Clegg et al., 2002). This finding is at odds with existing verbal models that propose that population founding can result in instantaneous levels of differentiation that are comparable to that observed in much older populations (Chakraborty & Nei, 1977; Hedrick, 1999).

3.5 Conclusion

Over the past decade a major challenge in speciation genomics has been to understand how different evolutionary processes shape patterns of divergence at the genomic level. However, most research has focused on the role of selection in shaping the genomic landscape, largely ignoring the role of non-selective processes. Using the sequential colonisation of south Pacific islands by the silvereye bird, we have provided the first assessment of how patterns of divergence are shaped by the stochastic effects of drift. In the silvereye system, highly diverged genomic regions can arise rapidly following population founding; however, such divergence peaks are highly idiosyncratic between founding steps. This system has also allowed us to demonstrate that the stochastic effects of population founding can result in significant loss of genetic diversity. The potential for, and magnitude of, this loss is intensified when populations are the product of multiple founding steps and/or when populations are founded by few individuals.

3.6 Acknowledgements

This work was funded by a Natural Environment Research Council (NERC) studentship awarded to ASP, a Percy Sladen Memorial Fund grant to SMC and Oxford University Press John Fell Fund awarded to SMC. Sample collection was conducted with the permission of the governments of Australia, New Zealand and French Polynesia. The authors would like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work.

3.7 References

- Alexander DH, Novembre J, Lange K** (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**: 1655–1664.
- Andrews S** (2010). FASTQC: A quality control tool for high throughput sequence data. Available at: www.bioinformatics.babraham.ac.uk/projects/fastqc/
- Biebach I, Keller LF** (2012). Genetic variation depends more on admixture than number of founders in reintroduced Alpine ibex populations. *Biological Conservation* **147**: 197–203.

- Burg TM, Gaston AJ, Winker K, Friesen VL** (2005). Rapid divergence and post-glacial colonisation in western North American Steller's jays (*Cyanocitta stelleri*). *Molecular Ecology* **14**: 3745–3755.
- Black RA** (2010). *Phylogenetic and phenotypic divergence of an insular radiation of birds*. PhD Thesis: Imperial College London.
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA** (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology* **22**: 3124–3140.
- Chakraborty R, Nei M** (1977). Bottleneck effects on average heterozygosity and genetic distance with the stepwise mutation model. *Evolution* **31**: 347–356.
- Charlesworth D, Charlesworth B** (1987). Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics* **18**: 237–268.
- Clegg SM** (2010). Evolutionary changes following island colonisation in birds: Empirical insights into the roles of microevolutionary processes. In: Losos JB, Ricklefs RE *The theory of island biogeography revisited*. Princeton University Press, Princeton: 293–325.
- Clegg, Degnan SM, Kikkawa J, Moritz C, Estoup A, Owens IPF** (2002). Genetic consequences of sequential founder events by an island-colonising bird. *PNAS* **99**: 8127–8132.
- Clegg S, Owens P** (2002). The 'island rule' in birds: Medium body size and its ecological explanation. *Proceedings of the Royal Society of London Series B: Biological Sciences* **269**: 1359–1365.
- Cornetti L, Valente LM, Dunning LT, Quan X, Black R A., Hébert O, et al.** (2015). The genome of the "great speciator" provides insights into bird diversification. *Genome Biology and Evolution* **7**: 2680–2691.
- Coyne JA, Orr HA** (2004). *Speciation*. Sinauer Associates: Sunderland, MA.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al.** (2011). The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.
- Davey JW, Blaxter ML** (2010). RADSeq: Next-generation population genetics. *Briefings in Functional Genomics* **9**: 416–423.
- Degnan SM, Moritz C** (1992). Phylogeography of mitochondrial DNA in two species of white-eyes in Australia. *The Auk* **109**: 800–811.

- Estoup A, Clegg SM** (2003). Bayesian inferences on the recent island colonisation history by the bird *Zosterops lateralis lateralis*. *Molecular Ecology* **12**: 657–674.
- Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M** (2013). Robust demographic inference from genomic and SNP data. *PLoS genetics* **9**: e1003905.
- Excoffier L, Foll M** (2011). Fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* **27**: 1332–1334.
- Feder JL, Flaxman SM, Egan SP, Comeault AA, Nosil P** (2013). Geographic mode of speciation and genomic divergence. *Annual Review of Ecology, Evolution, and Systematics* **44**: 73–97.
- Frankham R** (2005). Genetics and extinction. *Biological Conservation* **126**: 131–140.
- Frankham R, Briscoe DA, Ballou JD** (2002). *Introduction to conservation genetics*. Cambridge university press: Cambridge.
- Fuller J, Ferchaud AL, Laporte M, Le Luyer J, Davis TB, Côté SD, Bernatchez L** (2020). Absence of founder effect and evidence for adaptive divergence in a recently introduced insular population of white-tailed deer (*Odocoileus virginianus*). *Molecular Ecology* **29**: 86–104.
- Grabherr MG, Russell P, Meyer M, Mauceli E, Alföldi J, di Palma F, et al.** (2010). Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* **26**: 1145–1151.
- Guild E** (1938). Tahitian aviculture: Acclimation of foreign birds. *Avicultural Magazine* **3**: 8–11.
- Harris SE, Xue AT, Alvarado-Serrano D, Boehm JT, Joseph T, Hickerson MJ, Munshi-South J** (2016). Urbanization shapes the demographic history of a native rodent (the white-footed mouse, *Peromyscus leucopus*) in New York City. *Biology Letters* **12**: 2015098.
- Hawley DM, Hanley D, Dhondt AA, Lovette IJ** (2006). Molecular evidence for a founder effect in invasive house finch (*Carpodacus mexicanus*) populations experiencing an emergent disease epidemic. *Molecular Ecology* **15**: 263–275.
- Hedrick PW** (1999). Perspective: Highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**: 313–318.

- Hopley D** (1982). *The geomorphology of the Great Barrier Reef: Quaternary development of coral reefs*. John Wiley and Sons: New York.
- Jakobsson M, Rosenberg NA** (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**: 1801–1806.
- Kardos M, Taylor HR, Ellegren H, Luikart G, Allendorf FW** (2016). Genomics advances the study of inbreeding depression in the wild. *Evolutionary Applications* **9**: 1205–1218.
- Kikkawa J, Degnan S** (1998). Effective population size of an island passerine. In: *Proceedings of the 22nd International Ornithological Congress*. Durban: South Africa
- Knaus BJ, Grünwald NJ** (2017). vcfr: A package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources* **17**: 44–53.
- Lack DL** (1971). *Ecological isolation in birds*. Wiley-Blackwell: New Jersey.
- Lambert DM, King T, Shepherd LD, Livingston A, Anderson S, Craig JL** (2005). Serial population bottlenecks and genetic variation: Translocated populations of the New Zealand saddleback (*Philesturnus carunculatus rufusater*). *Conservation Genetics* **6**: 1–14.
- Lande R, Barrowdough G** (1987). Effective population size, genetic variation, and their use in population. In: Soulé ME (Ed) *Viable Populations for Conservation*, Cambridge University Press: Cambridge, pp 87–124.
- Langmead B, Salzberg SL** (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359.
- Leisler B, Winkler H** (2015). Evolution of island warblers: Beyond bills and masses. *Journal of avian biology* **46**: 236–244.
- Luu K, Bazin E, Blum MGB** (2017). pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources* **17**: 67–77.
- Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, et al.** (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research* **23**: 1817–1828.

- Mayr E** (1942). *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press: Cambridge.
- Mayr E** (1954). Change of genetic environment and evolution. In: Huxley J, Hardy AC, Ford EB (eds) *Evolution as a process*, Allen and Unwin: London, 157–180.
- Mayr E** (1966). *Animal species and evolution*. Belknap Press of Harvard University Press: Cambridge.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K** (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**: 1297–1303.
- Mees GF** (1969). A systematic review of the Indo-Australian Zosteropidae. *Zoologische Verhandelingen* **102**: 1–390.
- Michaelides SN, Goodman RM, Crombie RI, Kolbe JJ** (2018). Independent introductions and sequential founder events shape genetic differentiation and diversity of the invasive green anole (*Anolis carolinensis*) on Pacific Islands. *Diversity and Distributions* **24**: 666–679.
- Nei M, Maruyama T, Chakraborty R** (1975). The bottleneck effect and genetic variability in populations. *Evolution* **29**: 1–10.
- Peacock MM, Beard KH, O'Neill EM, Kirchoff VS, Peters MB** (2009). Strong founder effects and low genetic diversity in introduced populations of Coqui frogs. *Molecular Ecology* **18**: 3603–3615.
- Pruett CL, Winker K** (2005). Northwestern song sparrow populations show genetic effects of sequential colonisation. *Molecular Ecology* **14**: 1421–1434.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al.** (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics* **81**: 559–575.
- Quilodrán CS, Ruegg K, Sendell-Price AT, Anderson EC, Coulson T, Clegg SM** (2020). The multiple population genetic and demographic routes to islands of genomic divergence. *Methods in Ecology and Evolution* **11**: 6–21.
- Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlović M, et al.** (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology* **30**: 1450–1477.

- Roques S, Chancerel E, Boury C, Pierre M, Acolas ML** (2019). From microsatellites to single nucleotide polymorphisms for the genetic monitoring of a critically endangered sturgeon. *Ecology and Evolution*, **9**: 7017–7029.
- Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, et al.** (2014). Genomics and the origin of species. *Nature Reviews Genetics* **3**: 176–92.
- Sendell-Price AT, Ruegg KC, Clegg SM** (2020). Rapid morphological divergence following a human-mediated introduction: The role of drift and directional selection. *Heredity* **124**: 535–549.
- Seutin G, White B, Boag PT** (1991). Preservation of avian blood and tissue samples for DNA analyses. *Canadian Journal of Zoology* **69**: 82–90.
- Smeds L, Qvarnström A, Ellegren H** (2016). Direct estimate of the rate of germline mutation in a bird. *Genome Research* **26**: 1211–1218.
- Sovic M, Fries A, Martin SA, Gibbs HL** (2019). Genetic signatures of small effective population sizes and demographic declines in an endangered rattlesnake, *Sistrurus catenatus*. *Evolutionary Applications* **12**: 664–678.
- Tabbada KA, Trejaut J, Loo JH, Chen YM, Lin M, Mirazón-Lahr M, et al.** (2010). Philippine mitochondrial DNA diversity: A populated viaduct between Taiwan and Indonesia? *Molecular Biology and Evolution* **27**: 21–31.
- Taylor SS, Jamieson IG** (2008). No evidence for loss of genetic variation following sequential translocations in extant populations of a genetically depauperate species. *Molecular Ecology* **17**: 545–556.
- Thibault, Cibois A** (2017). *Birds of Eastern Polynesia. A biogeographic Atlas*. Lynx Edicions: Barcelona.
- Thibault, Monnet C** (1990). Le Zosterops à poitrine grise *Zosterops lateralis* (Latham) en Polynésie: Un oiseau introduit devenu colonisateur actif. *Oiseau et la Revue Française d'Ornithologie* **60**: 233–240.
- Thulin C-G, Englund L, Ericsson G, Spong G** (2011). The impact of founder events and introductions on genetic variation in the muskox *Ovibos moschatus* in Sweden. *Acta Theriologica* **56**: 305–314.
- Väli Ü, Einarsson A, Waits L, Ellegren H** (2008). To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology* **17**: 3808–3817.

- Van Doren BM, Campagna L, Helm B, Illera JC, Lovette IJ, Liedvogel M** (2017). Correlated patterns of genetic diversity and differentiation across an avian family. *Molecular Ecology* **26**: 3982-3997.
- Vásquez DLA, Balslev H, Hansen MM, Sklenář P, Romoleroux K** (2016). Low genetic variation and high differentiation across sky island populations of *Lupinus alopecuroides* (Fabaceae) in the northern Andes. *Alpine Botany* **126**: 135–142.
- Westram AM, Ravinet M** (2017). Land ahoy? Navigating the genomic landscape of speciation while avoiding shipwreck. *Journal of Evolutionary Biology* **30**: 1522–1525.
- Wright S** (1931). Evolution in Mendelian populations. *Genetics* **16**: 97-159.

Rapid morphological divergence following a human-mediated introduction: The role of drift and directional selection

Ashley T. Sendell-Price, Kristen C. Ruegg
& Sonya M. Clegg



Published as:

Sendell-Price AT, Ruegg KC, Clegg SM (2020). Rapid morphological divergence following a human-mediated introduction: The role of drift and directional selection. *Heredity* 124: 535-549.



Rapid morphological divergence following a human-mediated introduction: the role of drift and directional selection

Ashley T. Sendell-Price¹ · Kristen C. Ruegg^{1,2,3} · Sonya M. Clegg^{1,4}

Received: 4 September 2019 / Revised: 31 January 2020 / Accepted: 3 February 2020 / Published online: 20 February 2020
© The Author(s) 2020. This article is published with open access

Abstract

Theory predicts that when populations are established by few individuals, random founder effects can facilitate rapid phenotypic divergence even in the absence of selective processes. However, empirical evidence from historically documented colonisations suggest that, in most cases, drift alone is not sufficient to explain the rate of morphological divergence. Here, using the human-mediated introduction of the silvereye (*Zosterops lateralis*) to French Polynesia, which represents a potentially extreme example of population founding, we reassess the potential for morphological shifts to arise via drift alone. Despite only 80 years of separation from their New Zealand ancestors, French Polynesian silvereyes displayed significant changes in body and bill size and shape, most of which could be accounted for by drift, without the need to invoke selection. However, signatures of selection at genes previously identified as candidates for bill size and body shape differences in a range of bird species, also suggests a role for selective processes in driving morphological shifts within this population. Twenty-four SNPs in our RAD-Seq dataset were also found to be strongly associated with phenotypic variation. Hence, even under population founding extremes, when it is difficult to reject drift as the sole mechanism based on rate tests of phenotypic shifts, the additional role of divergent natural selection in novel environments can be revealed at the level of the genome.

Introduction

The speed at which divergent evolution proceeds is highly variable. However, the underlying processes driving differences in evolutionary rates are not well understood (Provine 1989; Barton 1998; Emerson et al. 2001; Price

2008; Emerson and Gillespie 2008). Instances of rapid phenotypic divergence following the establishment of new populations are often attributed to the strong selective pressures provided by novel biotic and abiotic environments (Millien 2006; Mathys and Lockwood 2011; Jensen et al. 2017). However, rapid divergence of phenotypic traits does not necessarily need to involve selection. Theory predicts that when new populations are established, the random sampling effect of drift during the founding event has the potential to facilitate rapid phenotypic divergence (Mayr 1942; Carson and Templeton 1984). Such rapid phenotypic changes could reflect the phenotypic attributes of founders themselves (Grant and Grant 1995; Berry 1998; Kliber and Eckert 2005; Baker et al. 2006) or arise via more complex means in which population founding results in the reorganisation of quantitative genetic variation, catalysing divergence (Mayr 1954; Carson and Templeton 1984). Despite this, studies that attribute phenotypic divergence in naturally established populations to drift and founder effects are rarer (e.g. Armstrong et al. 2018), than the numerous publications that invoke divergent natural selection (e.g. Clegg et al. 2002b; Balakrishnan and Edwards 2009; Illera et al. 2007; Rasner et al. 2004; Yeh 2004).

Supplementary information The online version of this article (<https://doi.org/10.1038/s41437-020-0298-8>) contains supplementary material, which is available to authorized users.

✉ Ashley T. Sendell-Price
ashley.sendell-price@zoo.ox.ac.uk

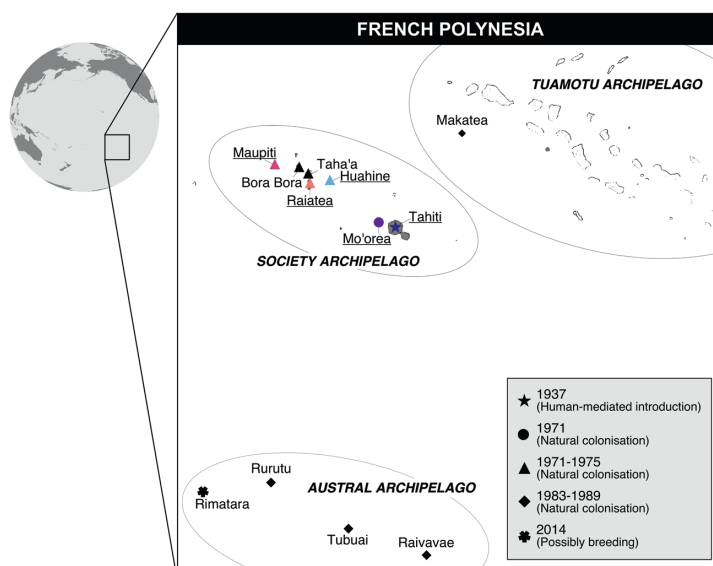
¹ Department of Zoology, Edward Grey Institute of Field Ornithology, University of Oxford, Oxford OX1 3PS, UK

² Department of Biology, Colorado State University, Fort Collins, CO, USA

³ Center for Tropical Research, Institute of the Environment and Sustainability, University of California, Los Angeles, Los Angeles, CA, USA

⁴ Environmental Futures Research Institute, Griffith University, Queensland 4111, Australia

Fig. 1 Colonisation history of the silvereeye (*Zosterops lateralis*) across islands of the Austral, Society and Tuamotu archipelagos of French Polynesia following the human-mediated introduction to Tahiti in 1937. Based on distribution data in Thibault and Monnet (1990) and Thibault and Cibois (2017). Sampled islands are underlined and coloured.



As the effects of drift are expected to be most pronounced when effective population sizes are small, recovery times are long and long-term effective population sizes are limited (Wright 1931; Nei et al. 1975; Clegg 2010), the opportunity for founder-induced phenotypic divergence following natural colonisation events may be limited. For example, in the silvereeye (*Zosterops lateralis*) the absence of founder-induced morphological divergence following the natural colonisation of islands across the South Pacific may be explained by: establishment by founding flocks in excess of hundreds of individuals (Clegg et al. 2002a; Estoup and Clegg 2003), rapid population recovery within two to three generations (Clegg 2010) and occasional immigration from the source population (Clegg and Phillimore 2010). In contrast to natural colonisation events, human-mediated introductions may be more likely to produce rapid phenotypic shifts via founder-induced divergence, as the number of founding individuals are often smaller, recovery times often longer, long-term effective population sizes are often more limited than in natural colonisations, and where introductions are geographically remote from source populations, gene flow does not occur (Blackburn et al. 2009). In this study we examine the potential for the more extreme founding conditions associated with human-mediated introductions to produce rapid morphological divergence.

The islands of French Polynesia contain a rich assemblage of introduced bird species following documented releases of at least fifty-nine species during the late 19th and early 20th centuries (Lockwood et al. 1999), with thirteen establishing breeding populations (Thibault and Cibois 2017). The silvereeye was introduced from the South Island

of New Zealand to the island of Tahiti in 1937 by aviculturist Eastham Guild (Guild 1938, 1940). Following introduction, the silvereeye persisted in low numbers until the late 1950s after which it rapidly increased in population size and expanded into all habitat types on the island (Thibault and Monnet 1990; Monnet et al. 1993; Thibault and Cibois 2017). Following this, Tahitian silvereeyes naturally dispersed to ten other islands across French Polynesia, including islands within the Austral, Society and Tuamotu archipelagos (Thibault and Monnet 1990; Thibault and Cibois 2017) (Fig. 1). Although the exact number of silvereeyes introduced to Tahiti is unknown, Guild's writings in the Avicultural Magazine (Guild 1940), suggest the founding population likely consisted of only a handful of individuals (the individuals released were a gift from Mr G. Rowland Hutchinson, the President of the Avicultural Society of New Zealand). A small founding population size combined with a long population recovery and geographic isolation from source and neighbouring populations (the nearest silvereeye population *Z. l. flaviceps* in Fiji is located over 3000 km away), provide the potential for rapid phenotypic divergence, such as changes in body/bill size and shape.

Here, we combine genome-wide tests for selection with morphological analyses to test the hypothesis that rapid phenotypic divergence can result from drift alone when populations establish under more extreme conditions than those associated with natural colonisation (very few founding individuals and slower population recovery). We address the following specific questions: (1) to what extent did the French Polynesian population experience a strong and sustained population bottleneck following introduction?

Table 1 Sampling information.

Island	Code	Latitude/Longitude	Morphology ^a	Sequenced ^b	Retained ^c
South Island	SNZ	−45.53/170.30	91	23	21
Tahiti	TAH	−17.58/−149.53	52	21	10
Mo'orea	MOO	−17.49/−149.83	75	21	8
Huahine	HUA	−16.82/−150.98	12	12	8
Raiatea	RAI	−16.77/−151.43	30	21	14
Maupiti	MAU	−16.45/−152.27	21	21	11

^aNumber of samples for which morphological data were collected.^bNumber of individuals included in RAD library.^cNumber of individuals retained in genomic analyses post-filtering.

(2) Can the rate of phenotypic divergence between French Polynesian and New Zealand silvereyes be explained by drift alone? and (3) What is the genomic evidence for selective processes? Commensurate with the introduction of a handful of individuals, we show that compared with natural colonisations by the silvereye, the French Polynesian population has undergone a more substantial population bottleneck followed by slower population recovery. In most instances we could not rule out drift as the sole cause of observed morphological shifts. However, the identification of outlier loci at genes previously associated with morphological divergence in birds, also suggests a strong role for directional selection.

Materials and methods

Sample collection

We collected blood samples and morphological data from 190 silvereyes across the islands of Tahiti, Mo'orea, Huahine, Raiatea and Maupiti (Society Archipelago) in April and May 2017. Birds were caught using mist nets and 20–40 µl of blood collected from the brachial wing vein, stored in 1 ml of lysis buffer (0.01M Tris-HCl; 0.01M NaCl; 0.01M EDTA; 1% n-lauroylsarcosine, pH 8.0) and frozen at −20 °C (Seutin et al. 1991). Following this, six morphological traits were measured. Wing length (mm) was measured as maximum flattened cord of the longest primary feather using a butted metal rule. Dial callipers (accuracy ±0.1 mm) were used to measure metatarsal length (tarsus length), mandible length to proximal edge of the nasal opening (Bill length) and mandible width and depth at the distal edge of the nasal opening (Bill width and Bill depth) (mm). Tail length (mm) of central tail feathers was measured from base to tip using either dividers or a tail rule. Wing and tail measurements were excluded from subsequent analyses if the individual was moulting. Bill length measures were excluded if the mandible was broken or overgrown. In addition, blood samples and morphological

data previously collected from South Island (New Zealand) were used in the study. New Zealand samples were collected for Clegg et al. (2002a) and were collected in the same manner as French Polynesian samples. Sampling locations and the number of samples in analyses are shown in Table 1. French Polynesian birds were measured by A.T. S.P. and S.M.C., whereas New Zealand birds were measured by S.M.C. only. A comparison of morphological measurements conducted by A.T.S.P. and S.M.C. of the same six individuals showed no significant effect of data collector on any of the morphological traits measured (all two-sample *t*-test *P* values > 0.05) and therefore suggesting no strong effect attributable to different observers.

RAD-PE sequencing and SNP calling

For 96 French Polynesian samples, genomic DNA was extracted using Qiagen DNEasy blood and tissue extraction kits (Qiagen, Venlo, Netherlands) and restriction site-associated DNA paired-end (RAD-PE) sequencing libraries constructed at the University of California, Los Angeles following the protocol outlined in Supplementary Material. The resulting libraries were sequenced on a single Illumina HiSeq4000 lane (Illumina, San Diego, CA, USA) at the UC Davis Genome Center using paired-end 150-bp sequence reads. Following sequencing, the quality of sequencing reads was checked visually using FASTQC (Andrews 2010). Sequence reads were assigned to individuals using the *process_radtags* script in the STACKS version 1.4 software pipeline (Catchen et al. 2013). 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 windows of 15% of the length of the read). Sequences with possible adaptor contamination and/or missing the *SbfI* restriction site were also discarded. Reads were filtered to remove 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 GCA_001281735.1) (Corretti 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)). Individual sample bam files were merged with existing RAD-PE sequencing reads for 23 New Zealand samples (unpublished raw data) to form a single bam file using Picard Tools version 2.7 (<http://broadinstitute.github.io/picard/>). Single Nucleotide Polymorphisms (SNPs) were identified using the *Haplotype-Caller* tool from the Genome Analysis Toolkit (GATK) nightly build version 2016-12-05-ga159770 (McKenna et al. 2010) and the resulting output filtered using VCFtools (Danecek et al. 2011) to remove indels and only include biallelic SNPs where the minor allele frequency > 0.01, minimum genotype quality = 30, minimum depth = 8, and where SNPs were called in at least 50% of individuals. Although more PCR duplicates would have been removed had duplicate filtering been applied following mapping of reads, given that in other widely used protocols such as double-digest RAD-sequencing (Peterson et al. 2012) duplicates cannot be removed at all, and also given that a recent study suggests PCR duplicate removal has little effect on genotyping (Euclide et al. 2020), remaining PCR duplicates are not likely to provide a substantial source of bias in our dataset.

To determine the optimum number of SNPs and individuals to retain for downstream analyses, data missingness was visualised (Fig. S1) using *genoscapeRtools* (<https://github.com/eriqande/genoscapeRtools>; <https://doi.org/10.5281/zenodo.848279>). The VCF file was then further filtered using VCFtools to include the optimum number of SNPs and only those individuals where less than 30% of data were missing.

As the *Z. l. melanops* genome is only assembled to the scaffold level (Cornetti et al. 2015), we mapped *Z. l. melanops* scaffolds 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). Further custom scripts (see 'Data availability') were used to reorder the GATK outputted VCF file accordingly.

Population structure

To investigate genetic structure among samples we conducted a Principle Component Analysis (PCA) using the full SNP dataset. We also examined patterns of population structure by performing maximum likelihood estimation of individual admixture proportions using the program *ADMIXTURE* (Alexander et al. 2009), testing *K* values 1–6. For each value of *K*, we conducted 20 independent runs and summarised runs using *CLUMPP* v.1.1.2 (Jakobsson and Rosenberg 2007). As the *ADMIXTURE* manual recommends avoiding SNPs with high linkage disequilibrium, we

used the '-indep-pairwise 100 kb 1 0' command in *PLINK* (Purcell et al. 2007) to remove one of every pair of SNPs with $r^2 > 0$ within 100 kb sliding windows.

Demographic history inference

In order to infer effective population size changes in French Polynesia, we estimated demographic parameters from the joint Site Frequency Spectrum (SFS) using the composite-likelihood and coalescent simulation approach implemented in *fastsimcoal* v. 2.6 (Excoffier and Foll 2011; Excoffier et al. 2013). As outgroup sequences were unavailable, demographic inference used the distribution of minor allele frequencies (folded-SFS). Based on findings from population structure analysis (see Fig. 2) we tested two different demographic models: a two-population demographic model which treated French Polynesia as a single population (Fig. 3a) and a three-population demographic model which incorporated a within French Polynesia population split in which Maupiti was treated as a separate population (Fig. 3b). Parameter search ranges were selected based on historical records. Assuming a generation time of 2.5 years (Kikkawa and Degnan 1998), and a known introduction time of 80 years ago, t_{intro} was fixed to 32 generations. Population size following introduction (N_{intro}) was fixed between 2 and 100 individuals (within reasonable bounds for a single event, human-mediated introduction). Population recovery following introduction (t_{exp}) was fixed between 19 and 26 generations ago (in line with timing of documented population expansion within French Polynesia). For the three-population model the establishment of the Maupiti population (t_{col}) was fixed between 11 and 14 generations ago (in line with the documented colonisation timeframe) and founding population size of the Maupiti population (N_{col}) fixed between 2 and 500 individuals. The three-population demographic model also incorporated a population size change within the Maupiti population (t_{change}) which was fixed between 2 and 11 generations ago. Contemporary population sizes were estimated for each population with initial search ranges for SNZ (N_{SNZ}), all French Polynesia ($N_{\text{All FP}}$) and FP cluster 1 (N_{FP1}) set to between 10,000 and 10,000,000 individuals. Given the very small geographic size of the island of Maupiti, search ranges for FP cluster 2 (Maupiti only) were bounded between 5 and 1,000 individuals. For each demographic model, we performed 100 independent runs (100 expectation/conditional maximisation cycles, 1,000,000 simulations per run), and chose the run with the highest likelihood as the best parameter estimates. As mutation rates for silvereyes are unknown, parameter estimates were based on estimated mutation rates for collared flycatchers (*Ficedula albicollis*): 4.6×10^{-9} mutations per nucleotide

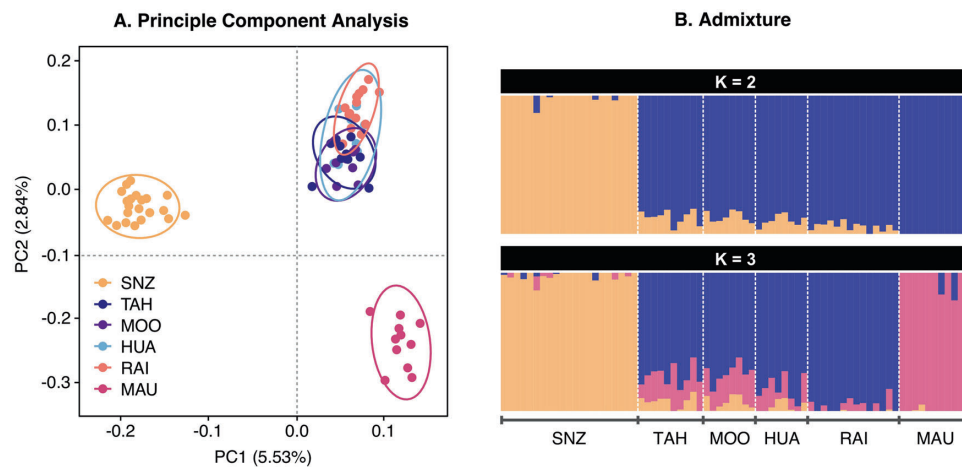


Fig. 2 Population structure and demographic history of New Zealand (South Island) and French Polynesian silvereyes. **a** Principle component analysis of genetic variation based on 5,414 LD-filtered SNPs. The variance explained by PC1 and PC2 is 5.53% and 2.84%, respectively. **b** Maximum likelihood estimation of individual

ancestries calculated with *ADMIXTURE* (Alexander et al. 2009) for all populations analysed, testing K values from 2 to 3 and based on 5,414 LD-filtered SNPs. Mean cross-validation error: $K_2 = 0.613$; $K_3 = 0.617$. SNZ South Island, New Zealand; TAH Tahiti; MOO Mo'orea; HUA Huahine; RAI Raiatea; MAU Maupiti.

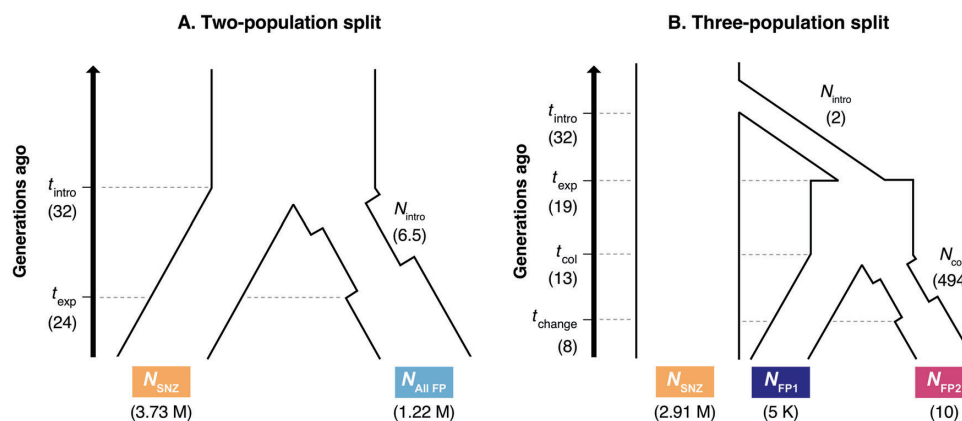


Fig. 3 Demographic models tested with *fastsimcoal2*. **a** Two-population split where FP is treated as a single population. **b** Three-population split where FP is treated as two separate populations (FP1 = Tahiti, Mo'orea, Huahine and Raiatea; FP2 = Maupiti only). t_{intro} = time of introduction to Tahiti; t_{exp} = time of population expansion following introduction; t_{col} = time of colonisation of Maupiti; t_{change} = time of switch to long-term population size of Maupiti

population; N_{SNZ} = current SNZ population size; $N_{Ali FP}$ = current population size of French Polynesian population when treated as a single population; N_{FP1} = current population size of cluster 1 (Tahiti, Mo'orea, Raiatea and Huahine); N_{FP2} = current population size of cluster 2 (Maupiti); N_{intro} = population size following introduction; N_{col} = population size following colonisation of Maupiti. Values in parentheses indicate parameter estimates from best model runs.

site per generation (Smeds et al. 2016). As *fastsimcoal2* requires the use of unlinked SNPs and is sensitive to missing data, we further filtered the LD-pruned dataset (see 'Population structure' section) to include only SNPs present in all individuals. VCFs were then converted to SFS format using the python script *EasySFS.py* (<https://github.com/isaacovercast/easySFS>) using flag-a to include all sites.

Morphological divergence

To summarise morphological divergence between the French Polynesian silvereye population(s) and representatives of its New Zealand ancestor, we conducted a PCA of morphological data collected from 189 silvereyes sampled across the Society Archipelago (French Polynesia) and 91 from Dunedin (South Island, New Zealand). PCA was conducted on

log-transformed data for wing length, tail length, bill length, depth and width. Differences in PC scores between South Island and French Polynesian silvereyes were tested for significance using analysis of variance (ANOVA).

Morphological shifts between New Zealand and French Polynesia were calculated for each morphological trait measured. Morphological data were log-transformed, and shifts expressed as standard deviation units, using pooled standard deviation estimates. To determine if random processes alone could account for observed shifts in morphological traits, we calculated Lande's (1976) statistic N_e^* , the effective population size required to explain morphological shifts by drift alone. N_e^* was calculated for each morphological trait as follows:

$$N_e^* = \frac{(1.96)^2 h^2 t}{(\bar{z})^2}$$

where h^2 = narrow-sense heritability of trait in question, t = number of generations, z = mean morphological shift and σ = phenotypic standard deviation of the colonised population. Calculations of heritability (h^2) were based on parent offspring, full-sibling comparisons and cross-fostering experiments in the Heron Island silvereye (*Z. l. chlorocephalus*) population. Heritability (h^2) estimates for morphological traits, ranged from 0.2 and 0.6 (Clegg et al. 2002b). Based on the mean breeding age of the Heron Island silvereye, estimated from a 15-year demographic study (Kikkawa, unpublished data, as cited in Clegg et al. (2002b)), generation time was estimated at between 2 and 3 years. Using these ranges in heritability and generation time we calculated a lower estimate (2-year generation time and $h^2 = 0.2$) and upper estimate (3-year generation time and $h^2 = 0.6$) of N_e^* for each morphological trait. N_e^* estimates were compared with estimates of N_e for French Polynesia calculated as harmonic means referred to as $N_{e(h)}$. The harmonic mean is always lower than the arithmetic mean and is especially sensitive to low values of N_e . Use of the harmonic mean is therefore relevant when two related demographic scenarios are suspected, an in situ population bottleneck or a founder event (Wright 1938). Where $N_{e(h)} > N_e^*$, drift cannot be the sole mechanism of divergence for a morphological trait and therefore selection must be invoked (Lande 1976). Both N_e^* and $N_{e(h)}$ were calculated when treating French Polynesia as a single population (All FP), and treating French Polynesia as two separate populations (FP1 and FP2).

Identifying signatures of selection and candidate gene analysis

To detect signatures of selection at the genomic level, 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 the choice of K principal components, based on inspection of a scree plot, where 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 comparing each French Polynesian island individually to the New Zealand population: $K = 2$, $\text{MinMAF} = 0.1$ and $\text{FDR} = 0.01$. As *PCAdapt* does not included the option to assign individuals to populations, outlier detection was conducted using individual French Polynesian islands to ensure that identified outliers reflected differences between New Zealand and French Polynesia rather than within French Polynesia differences.

Known, novel or predicted genes of the *T. guttata* genome occurring within 10,000 bp of outlier SNPs were identified using the Ensembl *BioMart* database (<https://www.ensembl.org/biomart>). Genes that had been previously associated with morphological variation in birds were determined to be candidate genes underlying the observed divergence in body size and beak shape in silvereyes. Genes known to be involved in craniofacial variation/disease in non-avian species were also considered candidates.

Genetic associations with morphological traits

To associated variation at SNPs with PC1 scores, we used a Bayesian sparse linear-mixed model (BSLMM) as implemented in the software package *GEMMA* (Zhou and Stephens 2012). As BSLMM combines linear-mixed models with sparse regression models, this method is well suited to situations where the underlying genetic architecture of the trait is unknown (Zhou et al. 2013). This genome-wide association approach controls for population structure by incorporating a relatedness matrix as a covariate in the mixed model. The model was run for 20 million iterations with a burn-in of 5 million iterations. This was repeated ten times, and the resulting hyperparameters averaged across runs. As a conservative approach to identify SNPs that were significantly associated with PC1 scores, we filtered for candidate SNPs using a strict posterior inclusion probability (PIP) ≥ 0.05 . This strict cut-off is five times higher than the widely used level of 0.01. *T. guttata* genes occurring within 10,000 bp of associated SNPs were identified using *BioMart*, and those previously associated with morphological variation in birds, or where a strong case could be made based on associations in other taxa, were considered candidate genes.

Results

RAD-PE sequencing and bioinformatics

Overall, RAD-PE sequencing resulted in an average of 227,092 paired-end reads per sample ($\pm 16,784$) covering 2,128,552 variable sites. Subsequent quality filtering (removal of indels and only including biallelic SNPs where: the minor allele count was ≥ 2 ; minimum genotype quality = 30; minimum depth = 8 and SNPs were called in at least 50% of individuals) reduced the total number of SNPs to 215,543. Of the 119 samples sequenced, 72 were retained after removing individuals where $\geq 30\%$ of data were missing across a subsample of 64,663 SNPs. The number of individuals retained per location ranged from 8 to 21 (Table 1). Reordering of *Zosterops* scaffolds onto *T. guttata* chromosomes, based on output from *Satsuma Synteny* and removal of unmapped/unoriented scaffolds, resulted in 63,849 SNPs (full SNP dataset). The LD-filtered dataset contained 5,414 unlinked SNPs for admixture analysis, and subsequent removal of SNPs with missing data resulted in 1587 SNPs for demographic modelling.

Population structure

A PCA showed clear separation of individuals into three distinct clusters when plotted against PC1 and PC2 (Fig. 2a). The three clusters consisted of: (1) New Zealand only; (2) Tahiti, Mo'orea, Huahine and Raiatea combined (subsequently referred to as FP cluster 1) and (3) Maupiti only (subsequently referred to as FP cluster 2). Maximum likelihood estimation of individual ancestries calculated with ADMIXTURE (Alexander et al. 2009), consistently provided the lowest cross-validation error for $K = 2$ (mean cross-validation error across runs = 0.613) with New Zealand and French Polynesia forming distinct groups. At $K = 3$ (mean cross-validation error across runs = 0.617) admixture results were consistent with those of the PCA (Fig. 2b).

Demographic history

Based on demographic modelling with *fastsimcoal2*, we estimated the effective population size immediately following introduction to be between 2 and 6.5 individuals, and population recovery estimated to take between 8 and 13 generations depending on the Model used. For model A (two-population split), contemporary effective population sizes for New Zealand and French Polynesia were estimated as 3.73 million and 1.22 million individuals, respectively (Fig. 3a). For model B (three-population split) effective population sizes for New Zealand, FP cluster 1 (Tahiti, Mo'orea, Huahine and Raiatea) and FP cluster 2 (Maupiti) were estimated as 2.91 million, ~5100 and 10 individuals,

respectively (Fig. 3b). Based on parameter estimates, the Maupiti population (which forms its own distinct population cluster) diverged from other French Polynesia islands ~13 generations ago. The effective population size of the Maupiti population immediately following colonisation was estimated as ~494 individuals.

Patterns and rates of morphological divergence

Compared with their New Zealand ancestors, French Polynesian silvereyes had significantly longer tail lengths, longer and deeper, but narrower bills, and shorter wings (all Wilcoxon P values < 0.05 ; see Fig. 4a). Differences in tarsus length were non-significant ($W = 7723.5$, $P = 0.189$). These differences were also maintained when treating French Polynesia as two populations, with the exception wing length for FP2 (Maupiti only) which was not significantly different from New Zealand ($W = 1192.5$, $P = 0.074$; see Fig. 4a).

The first two principle components (PCs) accounted for 29.38% and 21.75% of variance, respectively. PC1 broadly summarised body/bill size (all factor loadings had the same sign, although they were not entirely even), whereas PC2 summarised body/bill shape (primarily contrasting bill size and body size traits) (see Table S1 for factor loading). Differentiation between New Zealand and French Polynesian silvereyes along PC1 and PC2 is visible in the PCA summary plot (Fig. 4b). An ANOVA indicated significant variation between population units (PC1 scores: $F_{(6,441)} = 28.59$, $P < 0.001$; PC2 scores: $F_{(6,441)} = 10.66$, $P < 0.001$). PC1 and PC2 scores differed significantly between New Zealand and French Polynesian populations, with French Polynesian silvereyes showing a trend towards larger PC1 scores (increased body/bill size) and lower PC2 scores (larger longer/deeper bills, and shorter wings) (post hoc Tukey tests, $P < 0.001$ in both cases) (Fig. 4c).

Rate tests indicated that, in most instances, drift alone could not be dismissed as the sole cause of observed morphological shifts. For example, the effective population size required to explain the observed shift in wing length when treating French Polynesia as a single population ($N_e^* = 101\text{--}456$) was well above the harmonic mean of effective population sizes at each generation ($N_{e(h)} = 10.39$). However, when using the least conservative (i.e. smallest) estimates of N_e^* and treating French Polynesia as a single population, drift could be rejected as the sole mechanism driving differences in bill depth and PC1 scores. The minimum effective population sizes required to explain shifts in bill length ($N_e^* = 7$) and PC1 scores ($N_e^* = 9$) were both lower than the harmonic mean of effective population sizes ($N_{e(h)} = 10.39$). When treating French Polynesia as two populations, in no case was $N_{e(h)} > N_e^*$ (Table 2).

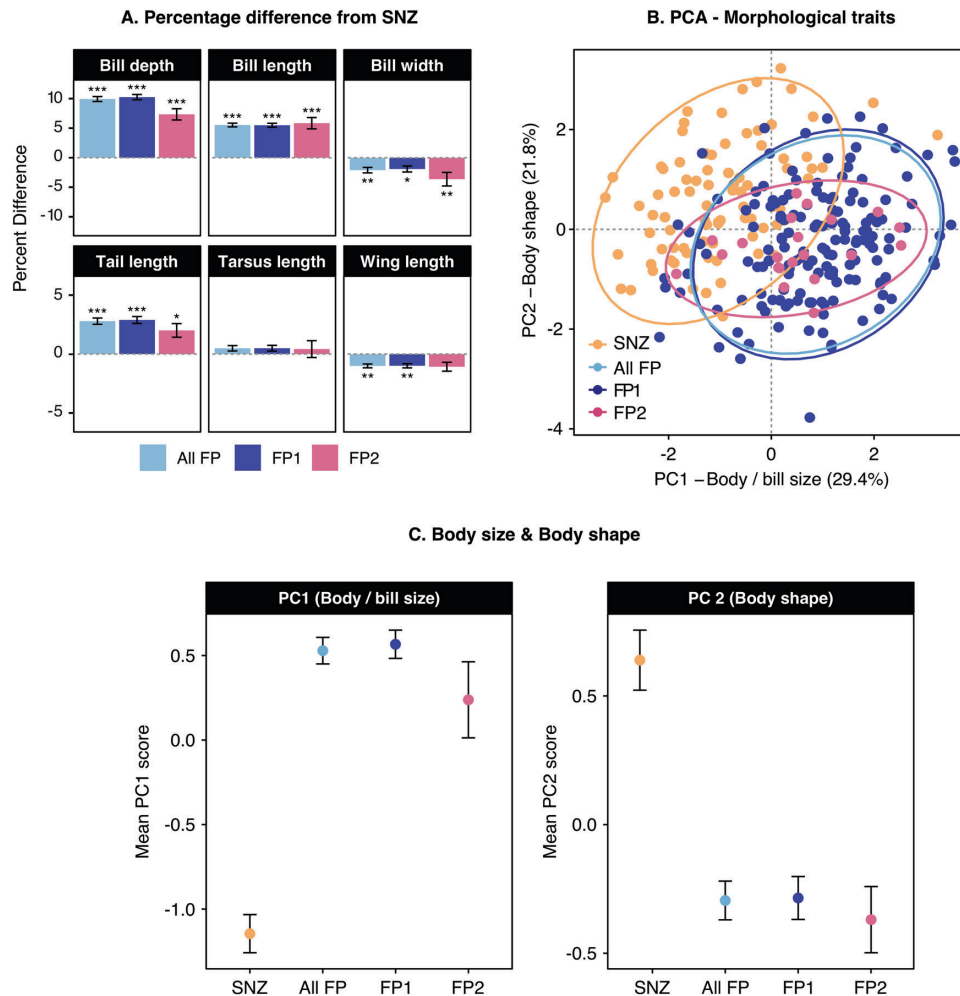


Fig. 4 Morphological divergence of French Polynesian silvereyes. a The percent difference between New Zealand and French Polynesian morphological trait means. A negative percent difference in this case refers to a trait being smaller in the French Polynesian population compared with the New Zealand population, whereas a positive percent difference refers to a trait being larger in the French Polynesian population. Bars around means represent 95% confidence intervals. Asterisks indicate significant difference as determined using Wilcoxon rank sum tests: $P < 0.001$ (***); $P < 0.01$ (**); $P < 0.05$ (*). **b** Principal components analysis of morphological variation. Based on six morphological traits (wing length, tail length, tarsus length, bill length

to posterior nasal opening and bill width/depth at anterior nasal opening) across 189 individuals of *Z. lateralis* from five islands in the Society Archipelago (French Polynesia) and 91 individuals from the South Island (New Zealand) source population. The variance explained by PC1 and PC2 is 29.4% and 21.8%, respectively (see Table S1). **c** PC1 scores (mean $\pm$ SE) and PC2 scores (mean $\pm$ SE); both calculated for New Zealand, French Polynesia combined and individual French Polynesian clusters, respectively. PC1 broadly summarises body size (all factor loadings had the same sign). PC2 broadly summarises variation in body shape (primarily contrasting bill size and body size traits).

Signatures of selection and candidate genes

Using *PCadapt* we identified between 15 and 509 outlier SNPs, putatively under selection (Fig. 5a). One hundred and sixty-two known, novel or predicted genes occurred within 10 kb of outlier SNPs (Table S2), of which eleven (*E2F4*, *FREM2*, *NFIA*, *OSR2*, *PBX3*, *PTDSS1*, *RALGPS1*, *TMC6*, *VPS13B*, *VPS50* and *ZMYND11*) have been previously

associated with bill/body size differences in birds or craniofacial morphogenesis in non-avian species (Table S3). The BSLMM identified 24 SNPs with a strong association with PC1 scores ($PIP \geq 0.5$) (Fig. 5b). These highly associated SNPs were located within 10,000 bp of 15 genes, of which six (*CDK14*, *OBSL1*, *IGF1R*, *INPP4B*, *RUNX3* and *ZMYND11*) were previously associated with morphological variation or craniofacial defects in vertebrates, including

birds (Table S4). Genes not previously associated with morphological variation included: *ATXN7*, *FNIP2*, *INHA*, *LDLRAP1*, *NOS1*, *SGK3*, *THEMIS* and *TOM1L2*. A single SNP (Chr 2: 10,960,228 bp) was identified as both a *PCAdapt* outlier and strongly associated with PC1 scores in the BSLMM. This SNP was located within the *ZMYND11* (Zinc Finger MYND-Type Containing 11).

Discussion

The introduction of the silvereye to French Polynesia provides a contemporary case in which to assess the potential for founder-induced drift to facilitate morphological divergence under more extreme founding conditions than those observed in natural colonisations by this species. Consistent with the silvereye's capacity for rapid differentiation (Frentiu et al. 2007; Clegg et al. 2002b; Clegg et al. 2008), we observed morphological shifts of significant magnitude, and rate tests suggested these shifts could be explained solely by drift in most cases. This result demonstrates the role of stochastic effects of population founding extremes in facilitating rapid morphological shifts at the very early stage of divergence. However, a role for divergent natural selection of ecologically relevant traits was also implicated from the genomic analysis, highlighting the joint role of the two microevolutionary processes, at the early stage of divergence.

Demographic history

While the exact date of introduction of silvereyes to French Polynesia was historically documented (Guild 1938; Thibault and Monnet 1990; Monnet et al. 1993; Thibault and Cibois 2017), Guild did not record the number of individuals he released in 1937, and as such population size immediately following founding is unknown. Combining historical records with demographic inference using the observed SFS, we were able to infer that the French Polynesian population was founded by few individuals and following introduction the population remained bottlenecked for several generations, and recovery relatively slow. This finding confirms that, compared with natural island colonisations by the silvereye which are thought to most often be established by flocks in excess of 100 individuals (Clegg et al. 2002a; Estoup and Clegg 2003) and population recovery thought to occur within 2–3 generations (Clegg 2010), the silvereye's introduction to French Polynesia offers a more extreme case in which to assess the potential for founder-induced drift to facilitate morphological divergence. Such a small founding population size (<10 effective founders) and slow recovery (8–13 generations post introduction) is also in keeping with historical

records—sightings of silvereyes were limited to around the release site until the late 1950s but becoming widespread across Tahiti by 1971 (Thibault and Cibois 2017). Although a more complete demographic model would provide further certainty regarding the demographic history of French Polynesian silvereyes, current sampling prevents this as only five of the eleven islands have been sampled. Nevertheless simple demographic models can be informative and have been used to effectively reconstruct recent known demographic histories (e.g. McCoy et al. (2014)).

The role of drift

We observed rapid shifts in multiple morphological traits, with French Polynesian silvereyes overall showing significant increases in body and bill size and changes in body shape (longer and deeper bills, longer tails and shorter wings). Unlike previous studies of natural island colonisations by birds, in most instances we were unable to dismiss drift as the sole mechanism through which morphological shifts may have arisen using rate tests. Whereas natural island colonisations by birds may have reduced potential for founder-induced morphological divergence due to the establishment by large flocks (Vincek et al. 1997; Clegg et al. 2002a; Estoup and Clegg 2003; Illera et al. 2007), rapid population recovery following founding (Clegg 2010) and continued immigration from the source population (Clegg and Phillimore 2010), our results suggest that founder-induced morphological divergence may occur under more extreme founding conditions. This finding is at odds with other studies of avian introductions (Lima et al. 2012; Valentin et al. 2018; Gleditsch and Sperry 2019), all of which were unable to explain observed morphological shifts by drift alone despite small founding population sizes.

Table 2 Estimates of N_e^* , the effective population size required to explain morphological shifts by drift alone.

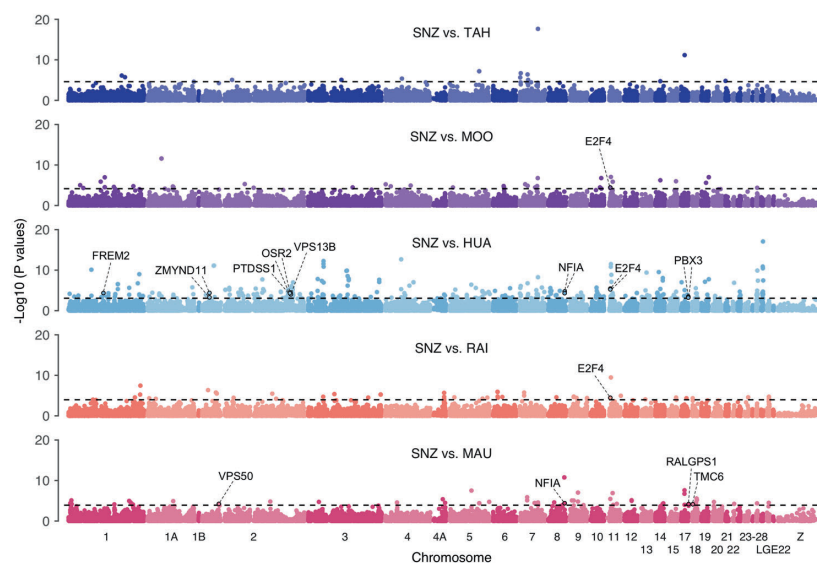
Trait	All FP	FP1	FP2
Wing	101–456	108–489	54–242
Tail	34.4–155	34–153	36–161
Tarsus	835–3,766	804–3,625	1,215–5,480
Bill length (P)	13–57	13–58	11–52
Bill depth (A)	<u>7–31</u>	7–30	7–32
Bill width (A)	192–865	241–1,086	43–195
PC1	<u>9–38</u>	8–37	11–51
PC2	25–113	28–126	7–32

Lower estimates are calculated with a generation time of three years and heritability estimates of 0.2, and upper estimates calculated with a generation time of 2 years and heritability estimates of 0.6. Underlined values show cases where $N_{e(h)} > N_e^*$, and drift can be rejected as the sole mechanism of differentiations.

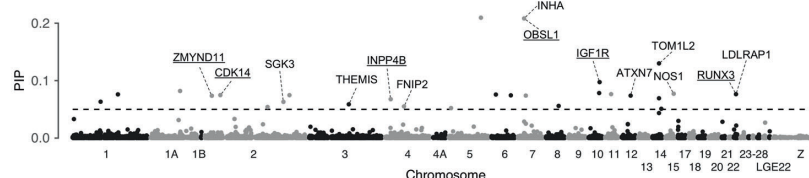
Fig. 5 Signatures of selection in the French Polynesian silvereye population. a

Manhattan plot of negative $\log_{10}(P \text{ values})$ estimated using *PCAdapt*. Points above the dashed line indicate outlier SNPs identified using $FDR = 0.01$. Genes containing outlier SNPs or within 10,000 bp of outlier SNPs and thought to be associated with morphological divergence are labelled. The distance of outlier SNPs from candidate genes is reported in Table S3. **b** Manhattan plot of BSLMM analysis of PC1 (body/bill size) scores. The dashed line indicates the threshold of posterior inclusion probability ($PIP = 0.5$). Genes within 10,000 bp of SNPs with $PIP \geq 0.5$ are labelled. Genes previously associated with morphological variation in birds, or where a strong case for inclusion as candidates can be made, are underlined. Chromosomes are numbered according to the *T. guttata* nomenclature. SNZ South Island, New Zealand; TAH Tahiti; MOO Mo'orea; HUA Huahine; RAI Raiatea; MAU Maupiti.

A. PCAdapt outlier analysis



B. Genome-wide association study



Non-random founding population

While a combination of founder-induced drift and strong selective pressures associated with island-living provides the most likely explanation for morphological shifts in French Polynesian silvereyes, there is the possibility that the founding individuals were a non-random phenotypic subset of the source population. An alternative explanation for increased body size is size biased dispersal, in which only the largest of individuals survive long-range colonisation across water barriers (Clegg et al. 2002a). Parallel 'size selective' mechanisms could operate in human-mediated introductions; for example, Guild may have exerted size bias when releasing birds to Tahiti, or only the largest birds survived to be released. Unfortunately, such bias cannot be tested for.

The role of natural selection

The pattern of morphological divergence we observed is broadly consistent with the island syndrome in which, amongst other repeated patterns of change, birds on islands

exhibit larger body size and have more robust bills than their mainland conspecifics (Clegg and Owens 2002; Leisler and Winkler 2015). While we can only speculate on the proximate causes of selection in French Polynesia, the island syndrome is thought to extend from exposure to a common suite of biotic conditions on islands; a combination of reduced predation (Losos and Ricklefs 2009), and a shift in the balance of inter- versus intra-specific competition (Blondel 2000), features that fundamentally change selective pressures in predictable ways (Grant 1998). Given the paucity of endemic land birds on oceanic islands, the absence of competitors in French Polynesia may favour the exploitation of a greater breadth of resources, which may be facilitated by a larger body size (Grant 1968; Lack 1971). The role of interspecific competition in shaping morphology is well-evidenced (Schluter et al. 1985; Kirschel et al. 2009) and can occur remarkably rapidly, even within timescales as short as a couple of decades (Grant and Grant 2006). Alternatively, larger body size may confer an advantage in high-density island populations where aggression is expected to be increased and territoriality expected to be year-round (Kikkawa 1980a, b; Robinson-Wolrath and

Owens 2003). Finally, as small size usually makes it easier to escape or hide from predators, larger body size may also result from relaxed risk of predation on islands (Lomolino 1985). Tests of these hypotheses in the Capricorn silveryeye (*Z. l. chlorocephalus*) on Heron Island suggest that, in this species, increased intraspecific competition is the prevailing mechanism driving increased body size of insular forms (Kikkawa 1980b; Scott et al. 2003; Robinson-Wolrath and Owens 2003).

Despite French Polynesian silvereyes having increased PC1 scores (larger overall size) compared with their New Zealand ancestor, their wing lengths were shorter and tarsus lengths unchanged. Smaller wing size may suggest reduced reliance on flight and/or increase use of densely vegetated habitat (Savile 1957; Winkler and Leisler 1985; Leisler and Winkler 2015). However, as decreased wing length was not accompanied by an increase in tarsus length—as would be expected to aid terrestrial movement (Zeffer and Norberg 2003; Zeffer et al. 2003), this result is difficult to interpret. Interestingly, decreased wing length but no change in tarsus length has recently been reported for non-native frugivores on the Hawaiian island of O'ahu (Gleditsch and Sperry 2019).

As the observed morphological shifts have taken place within a maximum of 80 years (~32 generations), our study supports the hypothesis that the bulk of differences can develop rapidly, potentially within the first couple of hundred generations after colonisation as proposed for the Capricorn silveryeye (Clegg et al. 2008). It would therefore be valuable to capture the divergence process at an even earlier stage e.g. during population establishment, as has been done for Eurasian blackbirds on the island of Heligoland (Engler et al. 2019). This may be possible in this system as the silvereye continues to expand its range in French Polynesia, for example a population has established on Rimatara (Austral archipelago) as recently as 2014 (Thibault and Cibois 2017).

Candidate genes

Using genome-wide sequencing data we are able to identify signatures of selective processes operating within French Polynesia. Outlier SNP detection allowed us to identify 12 candidate genes potentially underlying morphological differences, eight of which were mapped to directly by SNPs putatively under directional selection and four which occurred within 10 kb of outliers. These included: *VPS50*—previously associated with bill length in Berthelot's pipit (*Anthus berthelotii*) (Armstrong et al. 2018); *VPS13B*—identified as under directional selection between species of Darwin's finches (Lawson and Petren 2017); *NFIA*—associated with bill length in the house sparrow (*Passer domesticus*) (Lundregan et al. 2018) and craniofacial

abnormality in humans (Rao et al. 2014); *PTDSSI*—which is under directional selection in birds of paradise (Prost et al. 2018); *OSR2*—which has been experimentally demonstrated to play a role in beak development in birds (Brugmann et al. 2010) and craniofacial defects in mice (Stanier and Moore 2004); and *E2F4*, *FREM2*, *PBX3*, *RALGPS1*, *TMC6* and *ZMYND11*, which are all associated with craniofacial variation/disorders in a range of non-avian species including house mouse (*Mus musculus*), baboons (genus: *Papio*), European glass eels (*Anguilla anguilla*) and humans (Humbert et al. 2000; Cobben et al. 2014; Amaya 2015; Joganic 2016; De Meyer et al. 2017). The low overlap of genes identified across French Polynesian islands could reflect different selective pressures operating between islands, or alternatively may reflect differences in our power to detect outliers between sub-populations.

By performing a genome-wide association analysis we aimed to link phenotypic variation across French Polynesian and New Zealand silvereyes to the genomic variation in our SNP dataset. We identified several SNPs with strong associations with PC1 (body size) scores. Although not all strongly associated SNPs occurred within gene coding regions, a finding which is perhaps not surprising given the density of our marker set, several did occur within or close to gene coding regions including genes that have previously been associated with morphological variation in birds, non-avian vertebrates or implicated in craniofacial disease in humans. This included: *OBSL1*—a cytoskeletal adaptor protein previously associated with body size in humans (Hanson et al. 2009) and cetaceans (Sun et al. 2019); and *IGFIR*—a transmembrane receptor of insulin-like growth factor 1 (*IGF1*). The IGF1 pathway has a well-known role in body size scaling in a range of taxa, including chickens (Beccavin et al. 2001; Beckman et al. 2003), cattle (Grossi et al. 2015), brown house snakes (Sparkman et al. 2010) and dogs (Sutter et al. 2007); and has been previously associated with bill size in the black-bellied seedcracker (onHoldt et al. 2018). *RUNX3*—which, in the zebrafish, modulates bone morphogenetic protein (BMP) signalling during craniofacial development (Dalcq et al. 2012). BMPs such as BMP4 have been strongly associated with morphological variation of beaks in Darwin's finches (Abzhanov et al. 2004). Finally, *CDK14*, *ZMYND11* and *INPP4B*, which have been associated with craniofacial morphology in finches, humans and house mouse, respectively (Cobben et al. 2014; Amaya 2015; Pallares et al. 2015; Lawson and Petren 2017). *ZMYND11* was the only candidate gene to be both associated with PC1 scores and also contain outlier SNPs.

In Darwin's finches bill size is thought to be modulated by the transcriptional regulator *HMG2*: 'high mobility group AT-hook 2' (Lamichhaney et al. 2015) and beak shape strongly associated with *ALX1*: 'ALX homeobox 1' haplotypes (Lamichhaney et al. 2016). However, as SNPs

within our RAD-Seq dataset did not overlap with the position of these genes, we are unable to speculate on the role of *HMGA2* and *ALX1* in modulating bill divergence in silvereyes. Whole genome analyses would likely provide the opportunity to identify further candidate genes and allow us to address the role of *HMGA2* and *ALX1* in silvereye morphological divergence. Although not addressed here, the French Polynesian silvereye's ability to undergo rapid morphological divergence may be further modulated by epigenetic changes. For example, in the absence of genotypic changes, epigenetic variation has been shown to facilitate rapid change in the bill size and shape of Darwin's finches in response to sudden environmental changes such as drought or human disturbance (Skinner et al. 2014; Lamichhaney et al. 2015; McNew et al. 2017). Future work should therefore consider the role of epigenetics in facilitating rapid differentiation.

Conclusion

Given that knowledge of contemporary natural colonisation events is limited, evolutionary insights into the early stages of divergence have been predominantly retrospective or laboratory based. However, contemporary introductions of birds to islands provide ideal systems in which to study the genomic and phenotypic changes that take place during the early stages of divergence. Taking advantage of the well-documented introduction of the silvereye to French Polynesia, we set out to assess the role of founder-induced drift in driving phenotypic divergence under more extreme founding conditions than provided by natural colonisation events in this species. While we were unable to dismiss drift as the sole cause of significant shifts in morphological traits, given that island-living is known to exert strong selective pressure on avian morphology and French Polynesian silvereyes show body size/shape changes consistent with the 'island syndrome', observed phenotypic shifts within the French Polynesian population likely result from a combination of drift and selective processes. Studies of colonisation in action would provide the opportunity to tease apart the roles of drift and selection in driving morphological shifts at different stages of divergence. Although, such events are extremely rare, the potential establishment of a silvereye population on the French Polynesian island Rimatara (Austral archipelago) may provide such an opportunity. Genetic and phenotypic changes associated with population founding could also be further investigated in other species introduced to French Polynesia, such as common waxbill (*Estrilda astrild*), red-browed firetail (*Neochmia temporalis*) and chestnut-breasted mannikin (*Lonchura castaneothorax*), allowing a contrast between a range of species.

Data availability

Resequencing data from this study have been submitted to the European Nucleotide Archive (ENA; <https://www.ebi.ac.uk/ena>) under accession numbers PRJEB36361 and PRJEB25440. VCF files and custom scripts required to replicate analyses are available on dryad (<https://doi.org/10.5061/dryad.7d7wm37rm>).

Acknowledgements This work was funded by a Natural Environment Research Council (NERC) studentship and Genetics Society Heredity Fieldwork Grant awarded to ATSP and, a John Fell Fund grant awarded to SMC and Tim Coulson, and a Percy Sladen Memorial Fund grant to SMC. Sample collection was conducted with the permission of the governments of French Polynesia and New Zealand. We thank Neil Davies and Marie Fourdrigniez for logistical assistance and Jeffrey Tamati, Cael Sendell-Price and Olivia Grant for assistance during sample collection. The authors would like to acknowledge the use of the University of Oxford's Advanced Research Computing (ARC) facility and the UC Davis Genome Center in carrying out this work. We thank Barbara Mable and three anonymous reviewers who provided helpful comments on earlier drafts of this manuscript.

Author contributions ATSP and SMC jointly conceived the project and SMC supervised the project. Fieldwork was conducted by ATSP and SMC. Bioinformatics and analyses were carried out by ATSP with assistance from SMC and KCR. The manuscript was written by ATSP and SMC, with input from KCR.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

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

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

References

- Abzhanov A, Protas M, Grant BR, Grant PR, Tabin CJ (2004) Bmp4 and morphological variation of beaks in Darwin's finches. *Science* 305:1462
- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19:1655–1664
- Amaya L (2015) Genetic architecture of craniofacial shape in the house mouse: a genetic and morphological perspective. PhD Thesis, University of Kiel

- Andrews S (2010) FASTQC: a quality control tool for high throughput sequence data. www.bioinformatics.babraham.ac.uk/projects/fastqc/
- Armstrong C, Richardson DS, Hipperson H, Horsburgh GJ, Küpper C, Percival-Alwyn L et al. (2018) Genomic associations with bill length and disease reveal drift and selection across island bird populations. *Evol Lett* 2:22–36
- Baker MC, Baker MSA, Tilchman LM (2006) Differing effects of isolation on evolution of bird songs: examples from an island-mainland comparison of three species. *Biol J Linn Soc* 89:331–342
- Balakrishnan CN, Edwards SV (2009) Nucleotide variation, linkage disequilibrium and founder-facilitated speciation in wild populations of the zebra finch (*Taeniopygia guttata*). *Genetics* 181:645
- Barton NH (1998) Natural selection and random genetic drift as causes of evolution on islands. In: Grant PR (ed) *Evolution on islands*. Oxford University Press, Oxford, p 102–123
- Beccavin C, Chevalier B, Cogburn LA, Simon J, Duclos MJ (2001) Insulin-like growth factors and body growth in chickens divergently selected for high or low growth rate. *J Endocrinol* 168:297–306
- Beckman BR, Larsen DA, Dickhoff WW (2003) Life history plasticity in chinook salmon: relation of size and growth rate to autumnal smolting. *Aquaculture* 222:149–165
- Berry R (1998) Evolution of small mammals. In: Grant PR (ed) *Evolution on islands*. Oxford University Press, Oxford, p 35–50
- Blackburn TM, Lockwood JL, Cassey P (2009) *Avian invasions: the ecology and evolution of exotic birds*. Oxford University Press, Oxford
- Blondel J (2000) Evolution and ecology of birds on islands: trends and prospects. *Vie et Milieu* 50:205–220
- Brugmann SA, Powder KE, Young NM, Goodnough LH, Hahn SM, James AW et al. (2010) Comparative gene expression analysis of avian embryonic facial structures reveals new candidates for human craniofacial disorders. *Hum Mol Genet* 19:920–930
- Carson HL, Templeton AR (1984) Genetic revolutions in relation to speciation phenomena: the founding of new populations. *Annu Rev Ecol Syst* 15:97–132
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Mol Ecol* 22:3124–3140
- Clegg SM (2010) Evolutionary changes following island colonisation in birds: empirical insights into the roles of microevolutionary processes. In: Losos JB, Ricklefs RE (eds) *The theory of island biogeography revisited*. Princeton University Press, New Jersey, p 293–325
- Clegg DSM, Kikkawa J, Moritz C, Estoup A, Owens IPF (2002a) Genetic consequences of sequential founder events by an island-colonizing bird. *PNAS* 99:8127–8132
- Clegg SM, Degnan SM, Moritz C, Estoup A, Kikkawa J, Owens IPF 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
- Clegg SM, Frentiu FD, Kikkawa J, Tavecchia G, Owens IPF (2008) 4000 Years of phenotypic change in an island bird: heterogeneity of selection over three microevolutionary timescales. *Evolution* 62:2393–2410
- Clegg S, Owens P (2002) The ‘island rule’ in birds: medium body size and its ecological explanation. *Proc R Soc B Biol Sci* 269:1359–1365
- Clegg SM, Phillimore AB (2010) The influence of gene flow and drift on genetic and phenotypic divergence in two species of *Zosterops* in Vanuatu. *Philos Trans R Soc B Biol Sci* 365:1077–1092
- Cobben J, Weiss M, Van Dijk F, De Reuver R, de Kruiff C, Pondaag W et al. (2014) A de novo mutation in ZMYND11, a candidate gene for 10p15.3 deletion syndrome, is associated with syndromic intellectual disability. *Eur J Med Genet* 57:636–638
- Cornetti L, Valente LM, Dunning LT, Quan X, Black Ra, Hébert O et al. (2015) The genome of the “Great Speciator” provides insights into bird diversification. *Genome Biol Evol* 7:2680–2691
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158
- Dalcq J, Pasque V, Ghaye A, Larbuisson A, Motte P, Martial JA et al. (2012) RUNX3, EGR1 and SOX9B form a regulatory cascade required to modulate BMP-signaling during cranial cartilage development in zebrafish. *PLoS ONE* 7:e50140–e50140
- De Meyer J, Maes GE, Dirks RP, Adriaens D (2017) Differential gene expression in narrow- and broad-headed European glass eels (*Anguilla anguilla*) points to a transcriptomic link of head shape dimorphism with growth rate and chemotaxis. *Mol Ecol* 26:3943–3953
- Emerson BC, Gillespie RG (2008) Phylogenetic analysis of community assembly and structure over space and time. *Trends Ecol Evol* 23:619–630
- Emerson BC, Paradis E, Thébaud C (2001) Revealing the demographic histories of species using DNA sequences. *Trends Ecol Evol* 16:707–716
- Engler JO, Sacher T, Coppack T, Bairlein F (2019) Assortative mating frames establishment in a young island bird population. *R Soc Open Sci* 6:190050
- Estoup A, Clegg SM (2003) Bayesian inferences on the recent island colonisation history by the bird *Zosterops lateralis lateralis*. *Mol Ecol* 12:657–674
- Euclide PT, McKinney GJ, Bootsma M, Tarsa C, Meek MH, Larson WA (2020) Attack of the PCR clones: Rates of clonality have little effect on RAD-seq genotype calls. *Mol Ecol Resour* 20:66–78
- Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLoS Genet* 9:e1003905
- Excoffier L, Foll M (2011) Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 27:1332–1334
- Frentiu FD, Clegg SM, Blows MW, Owens IPF (2007) Large body size in an island-dwelling bird: a microevolutionary analysis. *J Evol Biol* 20:639–649
- Gleditsch JM, Sperry JH (2019) Rapid morphological change of nonnative frugivores on the Hawaiian island of O’ahu. *Evolution* 73:1456–1465
- Grabherr MG, Russell P, Meyer M, Mauceli E, Alföldi J, di Palma F et al. (2010) Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* 26:1145–1151
- Grant PR (1968) Bill size, body size, and the ecological adaptations of bird species to competitive situations on islands. *Syst Biol* 17:319–333
- Grant PR (1998) *Evolution on islands*. Oxford University Press, New York, NY
- Grant PR, Grant BR (1995) The founding of a new population of Darwin’s finches. *Evolution* 49:229–240
- Grant PR, Grant BR (2006) Evolution of character displacement in Darwin’s finches. *Science* 313:224–226
- Grossi D, do A, Buzanskas ME, Grupioni NV, de Paz CCP, Regitano LC, de A, de Alencar MM et al. (2015) Effect of IGF1, GH, and PIT1 markers on the genetic parameters of growth and reproduction traits in Canchim cattle. *Mol Biol Rep* 42:245–251
- Guild E (1938) Tahitian aviculture: acclimation of foreign birds. *Avic Mag* 3:8–11
- Guild E (1940) Western bluebirds in Tahiti. *Avic Mag* 5:284–285
- Hanson D, Murray PG, Sud A, Temtamy SA, Aglan M, Superti-Furga A et al. (2009) The primordial growth disorder 3-M syndrome connects ubiquitination to the cytoskeletal adaptor OBSL1. *Am J Hum Genet* 84:801–806

- Humbert PO, Rogers C, Ganiatsas S, Landsberg RL, Trimarchi JM, Dandapani S et al. (2000) E2F4 is essential for normal erythrocyte maturation and neonatal viability. *Mol Cell* 6:281–291
- Illera JC, Emerson BC, Richardson DS (2007) Population history of Berthelot's pipit: colonisation, gene flow and morphological divergence in Macaronesia. *Mol Ecol* 16:4599–4612
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806
- Jensen AJ, Hansen LP, Johnsen BO, Karlsson S (2017) Rapid evolution of genetic and phenotypic divergence in Atlantic salmon following the colonisation of two new branches of a watercourse. *Genet Sel Evol* 49:22
- Joganic JL (2016) A quantitative genetic analysis of craniofacial variation in baboons. PhD Thesis, Washington University in St. Louis, Missouri
- Kikkawa J (1980a) Weight changes in relation to social hierarchy in captive flocks of silvereyes, *Zosterops lateralis*. *Behaviour* 74:92–100
- Kikkawa J (1980b) Winter survival in relation to dominance classes among silvereyes *Zosterops lateralis chlorocephala* of Heron Island, great barrier reef. *Ibis* 122:437–446
- Kikkawa J, Degnan S (1998) Effective population size of an island passerine. In: Adams NJ, Slotow RH (eds) Proceedings of the 22nd international ornithological congress, Birdlife South Africa, Johannesburg, p 213
- Kirschel AN, Blumstein DT, Smith TB (2009) Character displacement of song and morphology in African tinkerbirds. *PNAS* 106:8256–8261
- Kliber A, Eckert CG (2005) Interaction between founder effect and selection during biological invasion in an aquatic plant. *Evolution* 59:1900–1913
- Lack DL (1971) Ecological isolation in birds. Harvard University Press, Massachusetts
- Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A et al. (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* 518:371–375
- Lamichhaney S, Han F, Berglund J, Wang C, Almén MS, Webster MT et al. (2016) A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science* 352:470–474
- Lande R (1976) Natural selection and random genetic drift in phenotypic evolution. *Evolution* 30:314–314
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359
- Lawson LP, Petren K (2017) The adaptive genomic landscape of beak morphology in Darwin's finches. *Mol Ecol* 19:4978–4989
- Leisler B, Winkler H (2015) Evolution of island warblers: beyond bills and masses. *J Avian Biol* 46:236–244
- Lima MR, Macedo RHF, Martins TLF, Schrey AW, Martin LB, Bensch S (2012) Genetic and morphometric divergence of an invasive bird: the introduced house sparrow (*Passer domesticus*) in Brazil. *PLoS ONE* 7:e53332
- Lockwood JL, Moulton MP, Balet KL (1999) Introduced avifaunas as natural experiments in community assembly. In: Weiher E, Keddy P (eds) Ecological assembly: Advances, perspectives, retreats. Cambridge University Press, Cambridge, p 108–125
- Lomolino MV (1985) Body size of mammals on islands: the island rule reexamined. *Am Nat* 125:310–316
- Losos JB, Ricklefs RE (2009) The theory of island biogeography revisited. Princeton University Press, New Jersey
- Lundregan SL, Hagen IJ, Gohli J, Niskanen AK, Kempainen P, Ringsby TH et al. (2018) Inferences of genetic architecture of bill morphology in house sparrow using a high-density SNP array point to a polygenic basis. *Mol Ecol* 27:3498–3514
- Luu K, Bazin E, Blum MGB (2017) pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Mol Ecol Resour* 17:67–77
- Mathys BA, Lockwood JL (2011) Contemporary morphological diversification of passerine birds introduced to the Hawaiian archipelago. *Proc R Soc B Biol Sci* 278:2392–2400
- Mayr E (1942) Systematics and the origin of species, from the viewpoint of a zoologist. Harvard University Press, Massachusetts
- Mayr E (1954) Change of genetic environment and evolution. In: Huxley J, Hardy AC, Ford EB (eds) Evolution as a process. Allen and Unwin, London, p 157–180
- McCoy RC, Garud NR, Kelley JL, Boggs CL, Petrov DA (2014) Genomic inference accurately predicts the timing and severity of a recent bottleneck in a nonmodel insect population. *Mol Ecol* 23:136–150
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303
- McNew SM, Beck D, Sadler-Riggelman I, Knutie SA, Koop JA, Clayton DH et al. (2017) Epigenetic variation between urban and rural populations of Darwin's finches. *BMC Evol Biol* 17:183
- Millien V (2006) Morphological evolution is accelerated among island mammals. *PLoS Biol* 4:1863–1868
- Monnet C, Thibault JC, Varney A (1993) Stability and changes during the twentieth century in the breeding landbirds of Tahiti (Polynesia). *Bird Conserv Int* 3:261–280
- Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution* 29:1–10
- Pallares LF, Carbonetto P, Gopalakrishnan S, Parker CC, Ackert-Bicknell CL, Palmer AA et al. (2015) Mapping of craniofacial traits in outbred mice identifies major developmental genes involved in shape determination. *PLoS Genet* 11:e1005607–e1005607
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* 7:e37135
- Price T (2008) Speciation in birds. Roberts and Company, Colorado
- Prost S, Armstrong EE, Nylander J, Thomas GW, Suh A, Petersen B et al. (2018) Genome analyses show strong selection on coloration, morphological and behavioral phenotypes in birds-of-paradise. <https://doi.org/10.1101/287086>
- Provine W (1989) Founder effects and genetic revolutions in microevolution and speciation: an historical perspective. In: Giddings LV, Kaneshiro KY, Anderson WW (eds) Genetics, speciation and the founder principle. Oxford University Press, New York, NY, p 43–76
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575
- Rao A, O'Donnell S, Bain N, Meldrum C, Shorter D, Goel H (2014) An intragenic deletion of the NFIA gene in a patient with a hypoplastic corpus callosum, craniofacial abnormalities and urinary tract defects. *Eur J Med Genet* 57:65–70
- Rasner CA, Yeh P, Eggert LS, Hunt KE, Woodruff DS, Price TD (2004) Genetic and morphological evolution following a founder event in the dark-eyed junco, *Junco hyemalis thurberi*. *Mol Ecol* 13:671–681
- Robinson-Wolrath SI, Owens IPF (2003) Large size in an island-dwelling bird: intraspecific competition and the dominance hypothesis. *J Evol Biol* 16:1106–1114
- Savile OBO (1957) Adaptive evolution in the avian wing. *Evolution* 11:212–224
- Schluter D, Price TD, Grant PR (1985) Ecological character displacement in Darwin's finches. *Science* 227:1056–1059

- Scott SN, Clegg SM, Blomberg SP, Kikkawa J, Owens IPF (2003) Morphological shifts in island-dwelling birds: the roles of generalist foraging and niche expansion. *Evolution* 57:2147–2156
- Seutin G, White B, Boag PT (1991) Preservation of avian blood and tissue samples for DNA analyses. *Can J Zool* 69:82–90
- Skinner MK, Gurerrero-Bosagna C, Haque MM, Nilsson EE, Koop JAH, Knutie SA et al. (2014) Epigenetics and the evolution of Darwin's finches. *Genome Biol Evol* 6:1972–1989
- Smeds L, Qvarnström A, Ellegren H (2016) Direct estimate of the rate of germline mutation in a bird. *Genome Res* 26:1211–1218
- Sparkman AM, Byars D, Ford NB, Bronikowski AM (2010) The role of insulin-like growth factor-1 (IGF-1) in growth and reproduction in female brown house snakes (*Lamprophis fuliginosus*). *Gen Comp Endocrinol* 168:408–414
- Stanier P, Moore GE (2004) Genetics of cleft lip and palate: syndromic genes contribute to the incidence of non-syndromic clefts. *Hum Mol Genet* 13:R73–R81
- Sun Y, Liu Y, Sun X, Lin Y, Yin D, Xu S et al. (2019) Insights into body size variation in cetaceans from the evolution of body-size-related genes. *BMC Evolut Biol* 19:157
- Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, Zhu L et al. (2007) A single IGF1 allele is a major determinant of small size in dogs. *Science* 316:112
- Thibault JC, Cibois A (2017) Birds of eastern Polynesia. A biogeographic Atlas. Lynx Publishing, Ontario
- Thibault JC, Monnet C (1990) Le Zosterops à poitrine grise *Zosterops lateralis* (Latham) en Polynésie: un oiseau introduit devenu colonisateur actif L'Oiseau et la Rev Française d'Ornithologie 60:233–240
- Valentin RE, Lockwood JL, Mathys BA, Fonseca DM (2018) Influence of invasion history on rapid morphological divergence across island populations of an exotic bird. *Ecol Evol* 8:5291–5302
- Van Doren BM, Campagna L, Helm B, Illera JC, Lovette IJ, Liedvogel M (2017) Correlated patterns of genetic diversity and differentiation across an avian family. *Mol Ecol* 26:3982–3997
- Vincek V, O'Huigin C, Satta Y, Takahata N, Boag PT, Grant PR et al. (1997) How large was the founding population of Darwin's finches? *Proc R Soc Lond Ser B Biol Sci* 264: 111–118
- vonHoldt BM, Kartzinell RY, Huber CD, Le Underwood V, Zhen Y, Ruegg K et al. (2018) Growth factor gene IGF1 is associated with bill size in the black-bellied seedcracker *Pyrenestes ostrinus*. *Nat Commun* 9:4855–4855
- Winkler H, Leisler B (1985) Morphological aspects of habitat selection in birds. In: Cody ML (ed) *Habitat selection in birds*. Academic Press Inc, Orlando, p 415–434
- Wright S (1931) Evolution in Mendelian populations. *Genetics* 16:97
- Wright S (1938) Size of population and breeding structure in relation to evolution. *Science* 87:430–431
- Yeh PJ (2004) Rapid evolution of a sexually selected trait following population establishment in a novel habitat. *Evolution* 58:166–174
- Zeffer A, Johansson LC, Marmebro A (2003) Functional correlation between habitat use and leg morphology in birds (Aves). *Biol J Linn Soc* 79:461–484
- Zeffer A, Norberg UM (2003) Leg morphology and locomotion in birds: requirements for force and speed during ankle flexion. *J Exp Biol* 206:1085–1097
- Zhou X, Carbonetto P, Stephens M (2013) Polygenic Modelling with Bayesian sparse linear mixed models. *PLOS Genet* 9:e1003264
- Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 44:821

General Discussion

Since the publication of ‘On the Origin of Species’ a major challenge in evolutionary biology has been to determine the relative importance of different evolutionary mechanisms, such as natural selection, genetic drift, gene flow and mutation in the speciation process (van den Berg & Weissing, 2015). While the goal of identifying unambiguous signatures of particular evolutionary processes in patterns of genomic divergence remains elusive, using the well-characterised silveryeye system this thesis has furthered understanding of how such processes shape patterns of divergence at the level of the genome and how divergence unfolds following the establishment of new populations. As the preceding chapters each contained their own discussion, in this chapter I briefly summarise the main results, highlight some general points that emerge, and discuss future directions.

5.1 Summary of results

A key goal of the genomic era has been to infer the evolutionary dynamics at play during the process of speciation by analysing the genomic landscape of divergence (Ravinet et al., 2017). However, empirical assessments of genomic landscapes under varying evolutionary scenarios are few, limiting the ability to achieve this goal. In Chapter 2, I combined RAD-sequencing and individual-based simulations to evaluate the genomic landscape in the silveryeye. Using comparisons matched for divergence timeframe and gene flow context, this chapter documented how genomic patterns accumulate as populations progress along the speciation continuum.

In contrast to previous predictions, results from Chapter 2 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 no were SNPs putatively under selection tightly associated with the position of genomic islands of divergence. Furthermore, by using individual-based simulations, Chapter 2 demonstrates that the transition from localised to more genome-wide levels of divergence can occur in the absence of selection. Taken together, these results

challenge the ubiquity of existing verbal models that explain the accumulation of genomic differences across the speciation continuum (namely the divergence hitchhiking model of speciation) and instead support the idea that divergence both within and outside of genomic islands is important during the speciation process. This observation is perhaps more consistent with the recently proposed multi-locus coupling mechanism of divergence (Flaxman et al., 2013, 2014; Butlin and Smadja, 2017; Schilling et al., 2018). Such alternative models of divergence should therefore be explored in future empirical work.

Over the past decade a major challenge in speciation genomics has been to understand how different evolutionary processes shape patterns of divergence at the genomic level (Ravinet et al., 2017). However, most research has focused on the role of selection in shaping the genomic landscape, largely ignoring the role of non-selective processes. Using the recent sequential colonisation of south Pacific islands by the Tasmanian silvereye subspecies (*Z. l. lateralis*), Chapter 3 provides the first assessment of how patterns of genomic divergence are shaped by the stochastic effects of drift following population founding. By tracking divergence across the genome, I demonstrate that population founding has an idiosyncratic effect on the buildup of divergence, with highly diverged genomic regions frequently becoming less diverged and regions of little divergence frequently becoming more diverged, in subsequent founding steps.

Further, by tracking the fate of individual loci, this chapter demonstrates that rare alleles are frequently lost following the establishment of new populations, and that the magnitude of this loss is increased where populations are founded sequentially. Such loss of rare alleles, is of importance as it may limit the evolutionary potential of populations to adapt to changing environments (Frankham et al., 2002). Several findings from the original study of this system using six microsatellites (Clegg et al., 2002a) were also supported by this chapter. These include the general pattern of gradual loss of variation with sequential founder steps, and the lack of rapidly elevated levels of genetic divergence with single founder steps that would be comparable to evolutionarily old populations. The latter finding is at odds with verbal models that propose that population founding can result in instantaneous levels of differentiation that are comparable to that observed in much older populations (Chackraborty & Neu, 1977; Hedrick, 1999).

Advancements in genomic techniques also provide the opportunity to assess the role of different evolutionary processes in driving phenotypic shifts following the establishment of new populations. Studies of insular species often highlight the role of natural selection in driving rapid phenotypic divergence following colonisation

(Millien, 2006; Mathys & Lockwood, 2011; Jensen et al., 2017). However, theory predicts that when populations are established by few individuals, random founder effects can also facilitate rapid phenotypic divergence even in the absence of selective processes (Mayr, 1942; Carson & Templeton, 1984). Despite this, empirical evidence from historically documented colonisations suggest that the conditions required for drift alone to drive morphological divergence are rarely met (Clegg et al., 2002b; Yeh, 2004; Rasner et al., 2004; Balakrishnan & Edwards, 2009). In Chapter 4, using the human-mediated introduction of the silvereye (*Zosterops lateralis*) to French Polynesia as a model, I first inferred demographic history using genomic-level data to show that, compared to natural colonisation by this species, French Polynesian silvereyes offer a more extreme case of population founding in which to assess the potential for morphological shifts to arise via drift alone. Demographic modelling suggested that the French Polynesian population was established by very few individuals and following introduction the population remained bottlenecked for several generations, and recovery relatively slowly.

Despite only 80 years of separation (~40 generations) from their New Zealand ancestors, French Polynesian silvereyes displayed significant changes in body/bill size and shape, which in most instances rate tests could account for by drift alone. However, as rate tests do not definitively rule out selection, the identification of outlier loci at genes previously identified as candidates for bill size and body shape differences in a range of bird species, also suggests a role for selective processes in driving morphological shifts within this population. Twenty-four SNPs in our RAD-Seq dataset were also found to be strongly associated with phenotypic variation.

Taken together the above results suggest that under population founding extremes the potential for morphological shifts to be driven by founder-induced drift is more likely, however, the action of natural selection in novel environments is also important in driving phenotypic changes. Studies of colonisation in action would provide the opportunity to further tease apart the roles of drift and selection in driving morphological shifts. Genetic and phenotypic changes associated with population founding could also be further investigated in other species introduced to French Polynesia, such as common waxbill (*Estrilda astrild*), red-browed firetail (*Neochmia temporalis*), and chestnut-breasted mannikin (*Lonchura castaneothorax*), allowing a contrast between a range of species.

By making use of the temporal and geographic variation that exists within the silvereye sub-species complex this thesis has provided several key insights into how divergence accumulates at the levels of the genome, how patterns of divergence can be shaped by different evolutionary processes, and how divergence unfolds following

the establishment of new populations. However, such evolutionary insights represent the tip of the iceberg of what can be learnt using this well-characterised system.

5.2 Future directions

The study of island-colonising birds has played a pivotal role in the development of evolutionary theory ever since Charles Darwin's observations of finches in the Galápagos (Darwin, 1878) and Alfred Wallace's work on the birds of the Malay Archipelago (Wallace, 1869) informed their co-discovery of natural selection (Darwin & Wallace, 1858). However, as contemporary island colonisations are rare, evolutionary insights into the very early stages of divergence have been predominantly retrospective – past processes are inferred from the vantage point of the present. This approach has been referred to as the “spyglass” as processes are studied long after they have taken place (Via, 2009). However, as past processes cannot be seen clearly through the veil of time, the spyglass approach is unlikely to reveal much about the initial changes that take place following the establishment of new populations, leaving many important evolutionary questions un-answered. Such questions include: Do changes early on in the divergence process constrain the evolutionary trajectory of a population? and Why do some species succeed in establishing new populations, but others fail?

In this thesis, using the contemporary natural colonisations of south Pacific islands by the Tasmanian silvereye and the more recent human-mediated introduction of the silvereye to French Polynesia, I have been able to study incipient divergence from a much more recent and clearer vantage point than before. However, these populations have still been diverging for some time - at least 40 generations in the case of French Polynesia - and therefore do not quite capture those very initial changes post-colonisation. Future work should therefore consider using the more recent, subsequent natural colonisation of individual French Polynesian islands that followed this species' introduction to Tahiti to observe these changes from even closer vantage points. This includes the rare opportunity to study colonisation in action, following the very recent natural colonisation of Rimatara, a remote island within French Polynesia's Austral archipelago within the past four years (~2 generations since colonisation).

Introductions of birds to oceanic islands not only provide the opportunity to study incipient divergence in the wild, but also the opportunity to study the factors that influence population establishment and long-term population persistence following

colonisation. During the late 19th and early 20th centuries at least 68 other species were introduced to the islands of French Polynesia (Lever, 2005; Long, 1981; Low, 2002; Thibault & Cibois, 2017). Many of these introductions were the work of prolific aviculturist Eastham Guild, whom in the 1930-40s, single handedly released over 7,000 individuals of 59 non-native species to Tahiti (Guild, 1938; Guild, 1940). Unsurprisingly many of these introductions did not result in long-term persistence, and today only 13 introduced species have established breeding populations (Thibault & Cibois, 2017). Whether or not the introduction of a novel species results in long-term population establishment is difficult to predict. Many failed introductions to French Polynesia likely result from a mismatch between the species' ecological requirements and those ecological niches provided by French Polynesia's unique biotic and abiotic environment. A striking example of such a mismatch is the introduction of 800 individuals of the Gouldian Finch, an arid-zone seed-specialist, which failed to establish post-introduction due to the lack of grass species present on Tahiti at the time (Low, 2002). But other species appear to have failed in spite of suitable habitat, owing instead to the stochastic nature of the colonisation process and the particular vulnerabilities of small populations. Such vulnerabilities include increased risk of inbreeding depression and the loss of adaptive diversity (Charlesworth & Charlesworth, 1987; Lande & Barrowclough, 1987). Light could be shed on the factors that increase colonisation success and long-term persistence using the success and demise of French Polynesia's avian introductions as case studies. This could be achieved by using historical records to determine the role of founding population size, founding effort (single introductions vs. multiple introduction attempts) and habitat suitability in establishment success. For successfully introduced species, this could be further bolstered using genetic data to infer the demographic changes that have taken place since introduction.

A further avenue for future research, is to determine the genetic underpinnings of a repeated evolutionary pattern seen not only in the silvereye but in a wide range of island fauna - the so-called 'island syndrome' (Adler & Levins, 1994). Following the colonisation of islands common changes include initially small species evolving larger body size (and vice versa) (Lomolino, 2005), reduced dispersal capacity, increased tameness (Beauchamp, 2004), and a slowed pace of life (longer-lived with slower reproduction) (MacArthur et al., 1972). It is thought that the island syndrome ultimately extends from exposure to a common suite of biotic conditions on islands; a combination of reduced predation (Losos & Ricklefs, 2009), and a shift in the balance of within- versus between-species competition (Blondel, 2000), features that fundamentally change the selective pressures in predictable ways. How are these changes manifested at the genetic level? Are island syndrome characteristics

achieved via selection at the same or different genes? As the silvereye displays varying degrees of island syndrome characteristics, including larger body size, increased tameness and slower life history, depending on the time since colonisation (Clegg & Owens, 2002), this system could be used to elucidate the genetic mechanisms that underpin this widespread evolutionary phenomenon.

5.3 References

- Adler GH, Levins R** (1994). The Island Syndrome in Rodent Populations. *Quarterly Review of Biology* **69**: 473-490.
- Balakrishnan CN, Edwards SV** (2009). Nucleotide variation, linkage disequilibrium and founder-facilitated speciation in wild populations of the zebra Finch (*Taeniopygia guttata*). *Genetics* **181**: 645-660.
- Beauchamp G** (2004) Reduced flocking by birds on islands with relaxed predation. *Proceedings of the Royal Society B: Biological Sciences* **271**: 1039-1042.
- Blondel J** (2000). Evolution and ecology of birds on islands: Trends and prospects. *Vie et Milieu* **50**: 205-220.
- Carson HL, Templeton AR** (1984). Genetic revolutions in relation to speciation phenomena: The founding of new populations. *Annual Review of Ecology, Evolution, and Systematics* **15**: 97-132.
- Chakraborty R, Nei M** (1977). Bottleneck effects on average heterozygosity and genetic distance with the stepwise mutation model. *Evolution* **31**: 347-356.
- Charlesworth D, Charlesworth B** (1987) Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology, Evolution, and Systematics* **18**: 237-68.
- Clegg SM, Degnan SM, Kikkawa J, Moritz C, Estoup A, Owens IPF** (2002a). Genetic consequences of sequential founder events by an island-colonising bird. *PNAS* **99**: 8127-8132.
- Clegg SM, Degnan SM, Moritz C, Estoup A, Kikkawa J, Owens IPF, 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.

- Clegg SM, Owens, PF.** (2002) The ‘island rule’ in birds: medium body size and its ecological explanation. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **269**: 1359-1365.
- Covas R** (2012). Evolution of reproductive life histories in island birds worldwide. *Proceedings of the Royal Society B: Biological Sciences* **279**: 1531-1537.
- Darwin C** (1878) *Journal of researches into the natural history and geology of the countries visited during the voyage of HMS beagle round the World, under the command of Capt. Fitz Roy, R.N.* Appleton and Company: New York.
- Darwin C, Wallace A** (1858) On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Zoological Journal of the Linnean Society* **3**: 45–62.
- Guild E** (1938) Tahitian aviculture: acclimatization of foreign birds. *Avicultural Magazine* **3**: 8–11.
- Guild E** (1940) Western bluebirds in Tahiti. *Avicultural Magazine* **5**: 284–5.
- Hedrick PW** (1999). Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution* **53**: 313–318.
- Jensen AJ, Hansen LP, Johnsen BO, Karlsson S** (2017). Rapid evolution of genetic and phenotypic divergence in Atlantic salmon following the colonisation of two new branches of a watercourse. *Genetics Selection Evolution* **49**: 22.
- Lande R, Barrowclough G** (1987) Effective population size, genetic variation, and their use in population. In: Soulé ME (Ed) *Viable populations for conservation*. Cambridge University Press: Cambridge, 87-124.
- Lever C** (2005) *Naturalised birds of the World*. A&C Black Publishers Ltd: London.
- Lomolino MV** (2005). Body size evolution in insular vertebrates: Generality of the island rule. *Journal of Biogeography* **32**: 1683-1699.
- Long JL** (1981) *Introduced birds of the world: The worldwide history, distribution and influence of birds introduced to new environments*. David & Charles Publishers: Exeter.
- Losos JB, Ricklefs RE** (2009). *The theory of island biogeography revisited*. Princeton University Press: New Jersey
- Low T** (2002) *Feral future: The untold story of Australia’s exotic invaders*. University of Chicago Press: Chicago.

- MacArthur RH, Diamond JM, Karr JR** (1972). Density compensation in island faunas. *Ecology* **53**: 330-342.
- Mathys BA, Lockwood JL** (2011). Contemporary morphological diversification of passerine birds introduced to the Hawaiian archipelago. *Proceedings of the Royal Society B: Biological Sciences* **278**: 2392–2400.
- Mayr E** (1942). *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press, Cambridge.
- Millien V** (2006). Morphological evolution is accelerated among island mammals. *PLoS Biology* **4**: 1863–1868.
- Rasner CA, Yeh P, Eggert LS, Hunt KE, Woodruff DS, Price TD** (2004). Genetic and morphological evolution following a founder event in the dark-eyed junco, *Junco hyemalis thurberi*. *Molecular Ecology* **13**: 671–681.
- Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlović M, et al.** (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology* **30**: 1450–1477.
- Thibault JC, Cibois A** (2017) *Birds of Eastern Polynesia. A biogeographic Atlas*. Lynx Edicions: Barcelona.
- Via S** (2009) Natural selection in action during speciation. *PNAS* **106**: 9939–46.
- Wallace A** (1869) *The Malay Archipelago: The land of the orang-utan and the bird of paradise; a narrative of travel, with studies of man and nature*. Macmillan and Co: London.
- Yeh PJ** (2004). Rapid evolution of a sexually selected trait following population establishment in a novel habitat. *Evolution* **58**: 166–174.

Library preparation

Library preparation for Restriction-site Associated DNA (RAD) sequencing was conducted as follows: 100 ng of genomic DNA from each individual was digested with 2.4 U of restriction enzyme *Sbf*I-HF (New-England Biolabs Inc., Beverly MA, USA) at 37°C for 60 minutes, followed by an inactivation step of 80°C for 20 minutes. BestRAD *Sbf*I adapters were then ligated to the overhanging ends of the products of the restriction reactions by adding 50 nM to each sample with 320 U T4 DNA Ligase (New-England Biolabs Inc.). Barcoding of samples was achieved with a set of index nucleotides within the BestRAD *Sbf*I adapter sequences. Reactions were incubated at 20°C for 16 hours and then heat-inactivated by holding at 65°C for 20 minutes. The reactions were then pooled, cleaned with AMPure XP beads, and the products randomly sheared to a mean size of 500 bp by sonication, using a BioRuptor NGS (Diagenode). Following sonication, Dynabeads M-280 streptavidin magnetic beads (Life Technologies) were used to bind the biotinylated ends of the BestRAD *Sbf*I adaptors. *Sbf*I-HF was used once more to release adapter bound fragments and AMPure XP beads were used to clean the samples. After the bead clean up, ends were treated with Blunt End-Repair Mix found in the NEBNext Ultra DNA Library Prep Kit (New-England Biolabs Inc.) to remove overhangs. NEBNext Adaptors for Illumina sequencing were then ligated to blunt ends following a two-step incubation; first incubated at 20°C for 15 minutes with Blunt/TA Ligase Master Mix (New-England Biolabs Inc.) and then at 37°C for 15 minutes with USER Enzyme (New-England Biolabs Inc.). Following size selection using AMPure XP beads to isolate fragments within the size range 300–700 bp, the library was PCR enriched using 10 uM of P1 index primer and 10 uM of Universal PCR Primer (NEB). First a test PCR was run using 5 ul of DNA with 15 PCR cycles. Depending on the brightness of the band produced, the final PCR cycle number was adjusted from 9-12 cycles. PCR products were cleaned using AMPure XP beads.

All library preparation was conducted at the University of California, Los Angeles. Libraries were sequenced on Illumina HiSeq4000 lanes (Illumina, San Diego, CA, USA) at the UC Davis Genome Center, using paired-end 150-bp sequence reads.

Supplementary Material: Chapter 2

Table B.1. List of parameters of the model with default values.

Parameter	Definition	Value	Source
N_0	Initial number of breeders	400	Kikkawa et al. (1983)
ζ	Number of young fledging by clutch	1.9	Kikkawa et al. (1983)
S_s	Survival rate at summer	0.96	Sandvig et al. (2017)
S_a	Survival rate at autumn	0.76	Sandvig et al. (2017)
S_w	Survival rate at winter	0.63	Sanvig et al. (2017)
δ	Sex ratio	0.5	Kikkawa et al. (1983)
G	Generation time	3	Clegg et al. (2008)
n_L	Number of SNPs	2533	Chromosome 17
σ_{dem}	Stochastic demographic variant	0.0013	*
θ	Recombination rate	2, 3, 6, 10, 19	†
m	Migration rate	0.0001, 0.001, 0.01	‡

* This value gives an average of 200 breeding pairs per generation, in line with the size of the breeding population observed on Heron Island (Kikkawa & Wilson 1983).

† Recombination rates observed for parts of the collared-flycatcher (*Ficedula albicollis*) genome (Kawakami et al., 2014). The value of 2 cM/Mb is the average for chromosomes >100Mb, 3 cM/Mb is the genome-wide average; 6 cM/Mb the average for chromosome 17, 10 cM/Mb the average for chromosomes <10Mb, and 19 cM/Mb the maximum average observed in a collared-flycatcher chromosome.

‡ A migration rate of 0.01 was empirically estimated among Heron Island and neighbouring islands (Brook & Kikkawa, 1998). This was taken as the upper estimate of migration rate, but as the population comparisons considered here include more isolated islands we also ran simulations assuming migration rates one and two magnitudes lower.

Table B.2. Distributional skew of F_{ST} values calculated for each autosomal chromosome. Only chromosomes with at least 10 x 50kb windows are reported. Comparisons where distributional skew is significantly higher when diverging with gene flow are highlighted in pink. Significance determined using a modified randomisation test (see Chapter 2 methods).

Chromosome	Early Stage		Mid stage	Late stage	
	Gene flow (SI vs. CI)	No gene flow (SI vs. FP)		Gene flow (GT vs. LF)	No gene flow (ML vs. LH)
1	2.19	1.93	1.52	2.02	1.37
1A	1.54	1.38	1.79	1.88	0.87
1B	-	-	-	-	-
2	3.13	1.11	1.00	2.31	0.97
3	2.32	2.30	1.56	1.61	1.35
4	2.66	1.72	1.58	1.94	1.92
4A	1.24	1.68	1.14	1.73	0.85
5	3.04	1.37	1.75	1.36	1.13
6	3.17	1.60	1.51	1.60	1.46
7	2.99	1.56	1.55	2.07	1.30
8	1.91	3.28	1.70	0.77	1.63
9	1.25	2.41	1.83	1.61	1.54
10	1.28	0.78	0.23	0.82	0.58
11	2.23	1.80	2.22	1.76	1.48
12	1.72	0.93	2.23	1.66	1.57
13	2.82	2.32	1.13	1.34	1.55
14	1.64	1.20	1.01	1.32	1.26
15	1.32	1.28	1.76	1.93	1.27
17	1.79	1.73	1.32	1.60	1.33
18	1.99	3.26	0.49	0.82	0.17
19	4.37	1.91	1.16	0.52	1.13
20	1.45	1.82	2.05	2.53	1.34
21	2.06	0.89	1.76	1.05	1.06
22	1.72	0.81	1.71	2.21	1.61
23	1.30	0.56	1.57	1.89	-0.18
24	1.56	0.92	0.38	0.99	0.31
25	1.54	1.28	1.68	0.16	0.50
26	2.33	0.25	0.64	0.79	0.47
27	-	-	-	-	-
LGE22	2.21	0.90	1.08	-0.17	0.65

Table B.3. P-values for pairwise comparisons of F_{ST} distributional skew for simulations, tested using a randomisation test. Values in pink shaded cells indicate level of skew for a given timeframe.

Gene flow	Simulation timeframe			
	Generations	30	60	1000
$m = 0$	30	1.094	0.025	0.001
	60	-	0.473	0.021
	1000	-	-	-0.064
$m = 0.0001$	30	1.170	0.001	0.001
	60	-	0.001	0.081
	1000	-	-	-0.352
$m = 0.001$	30	0.593	0.386	0.103
	60	-	1.983	0.079
	1000	-	-	1.235
$m = 0.01$	30	0.936	0.627	0.511
	60	-	1.730	0.875
	1000	-	-	1.830

References

- Clegg SM, Frentiu FD, Kikkawa J, Tavecchia G, Owens IPF** (2008) 4000 Years of phenotypic change in an island bird: Heterogeneity of selection over three microevolutionary timescales. *Evolution* **62**: 2393–2410.
- Kawakami T, Smeds L, Backström N, Husby A, Qvarnström A, Mugal CF, 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. *Molecular Ecology* **23**: 4035–4058.
- Kikkawa J, Wilson JM** (1983) Breeding and dominance among the Heron Island silvereyes *Zosterops Lateralis Chlorocephalus*. *Emu* **83**: 181–198.
- Sandvig EM, Coulson T, Kikkawa J, Clegg SM** (2017) The influence of climatic variation and density on the survival of an insular passerine *Zosterops lateralis*. *PLOS ONE* **12**: e0176360.
- Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A, et al.** (2010) The genome of a songbird. *Nature* **464**: 757–762.

Supplementary Material: Chapter 4

Table C.1. Principal components analysis of morphological data. Coefficients of correlation between morphological variables and the derived principal components extracted by principal components analysis.

Morphological trait	Factor loadings	
	PC1	PC2
Wing length	0.265	0.719
Tail length	0.487	0.429
Tarsus length	0.297	-
Culmen length (P)	0.525	-0.419
Culmen depth (A)	0.571	-0.284
Culmen width (A)	-	0.182
Variance explained	29.38%	21.75%

Table C.2. Known, novel or predicted genes occurring within 10kb of outlier SNPs identified, using *PCAdapt*, between New Zealand and French Polynesian islands. Population codes as follows: NZ = South Island, New Zealand; FP = all French Polynesian islands; TAH = Tahiti, MOO = Mo'orea; HUA = Huahine; RAI = Raiatea; MAU = Maupiti.

Ensembl Gene stable ID	Chrom	Gene start (bp)	Gene end (bp)	Gene name	Comp (SNZ vs.)
ENSTGUG00000006887	1	9154708	10148444	<i>DMD</i>	HUA
ENSTGUG00000010796	1	35582832	35585680		HUA
ENSTGUG00000010849	1	37962335	37970867	<i>KDELC1</i>	RAI
ENSTGUG00000011664	1	49704389	49861236		MOO
ENSTGUG00000011735	1	50626080	50796176	<i>FRY</i>	HUA
ENSTGUG00000011967	1	53819513	53939566	<i>FREM2</i>	HUA
ENSTGUG00000012117	1	55948726	55958266	<i>SPRYD7</i>	MOO
ENSTGUG00000012422	1	59579943	59591090	<i>MTRF1</i>	HUA
ENSTGUG00000012549	1	69793105	69992841		MOO
ENSTGUG00000013001	1	86711122	86758565	<i>NARS2</i>	TAH
ENSTGUG00000013496	1	101653625	102054107	<i>EPHA6</i>	RAI
ENSTGUG00000000764	2	10897250	11003676	<i>ZMYND11</i>	HUA
ENSTGUG00000001094	2	17223251	17392565	<i>MKX</i>	HUA
ENSTGUG00000001311	2	21974685	22109468	<i>CUBN</i>	RAI
ENSTGUG00000001501	2	25251229	25335000	<i>VPS50</i>	MAU
ENSTGUG00000004189	2	62221688	62308845	<i>NEK11</i>	HUA
ENSTGUG00000004894	2	64731646	64741764	<i>TGM4</i>	MOO
ENSTGUG00000007906	2	85435608	85489383	<i>VWC2</i>	HUA

ENSTGUG00000008631	2	93799185	93925718		HUA
ENSTGUG00000010074	2	106598017	106623803		RAI
ENSTGUG00000011815	2	130703700	130717968	<i>DECRI</i>	HUA
ENSTGUG00000011869	2	132302035	132307232	<i>PDP1</i>	HUA
ENSTGUG00000011966	2	133397397	133422104	<i>PTDSS1</i>	HUA
ENSTGUG00000011987	2	133895664	133963069	<i>LAPTM4B</i>	HUA
ENSTGUG00000011992	2	134004617	134048324	<i>MATN2</i>	HUA
ENSTGUG00000012027	2	134428506	134431204	<i>OSR2</i>	HUA
ENSTGUG00000012030	2	134494587	134931563	<i>VPS13B</i>	HUA
ENSTGUG00000012148	2	136194280	136200691	<i>CTHRC1</i>	HUA
ENSTGUG00000003612	3	16425219	16471280	<i>PLEKHH2</i>	MAU
ENSTGUG00000010902	3	50723474	50824543	<i>ARFGEF3</i>	TAH
ENSTGUG00000011615	3	59131319	59181152	<i>MOXD1</i>	HUA
ENSTGUG00000011773	3	61861362	61921523		HUA
ENSTGUG00000012218	3	71122911	71252155	<i>BVES</i>	HUA
ENSTGUG00000013229	3	108349385	108357841	<i>PLA2G7</i>	HUA
ENSTGUG00000002514	4	10348610	10684047		HUA
ENSTGUG00000004549	4	25456400	25591830		HUA
ENSTGUG00000005716	5	6271419	6285186	<i>CAT</i>	MOO
ENSTGUG00000005980	5	6538590	6567985	<i>KDM2A</i>	HUA
ENSTGUG00000006012	5	6571621	6577897	<i>GRK2</i>	HUA
ENSTGUG00000006094	5	6582669	6585026	<i>ANKRD13D</i>	HUA
ENSTGUG00000008025	5	8537039	8729742		HUA
ENSTGUG00000009146	5	12881930	13003509	<i>NAP1L4</i>	HUA
ENSTGUG00000012373	5	41137345	41175645	<i>TSHR</i>	HUA
ENSTGUG00000012820	5	51054854	51135675	<i>RCOR1</i>	HUA
ENSTGUG00000012959	5	55239178	55306635	<i>PPP2R5E</i>	HUA
ENSTGUG00000013194	5	60520695	60523997	<i>VCCKMT</i>	HUA
ENSTGUG00000004487	6	3175115	3314084		HUA
ENSTGUG00000004592	6	3250009	3269651	<i>PALD1</i>	HUA
ENSTGUG00000007660	6	16443466	16474660	<i>SEC31B</i>	MOO
ENSTGUG00000007917	6	16864802	16865718		MOO
ENSTGUG00000003583	7	2408013	2681518		HUA
ENSTGUG00000004748	7	7834568	8039714	<i>CNTNAP5</i>	RAI
ENSTGUG00000007152	7	13113900	13170539		HUA, TAH
ENSTGUG00000007262	7	13252086	13272166	<i>CSRNP3</i>	MOO
ENSTGUG00000007583	7	13965237	14081484	<i>CERS6</i>	HUA
ENSTGUG00000010653	7	23648780	23744147		MAU
ENSTGUG00000011260	7	28607527	28615834	<i>LSS</i>	HUA, MOO
ENSTGUG00000011322	7	28963908	29034843		HUA, MAU, TAH
ENSTGUG00000011444	7	29360824	29444139	<i>CPS1</i>	MAU
ENSTGUG00000003826	8	2916739	2967087		HUA
ENSTGUG00000018373	8	3334159	3339163	<i>ZBTB37</i>	HUA
ENSTGUG00000005495	8	8944758	9180702	<i>DPYD</i>	MAU
ENSTGUG00000006332	8	12485119	12513156	<i>SELENOF</i>	RAI
ENSTGUG00000006382	8	12638237	12741483		HUA
ENSTGUG00000007625	8	17464143	17689061	<i>RNF220</i>	HUA
ENSTGUG00000009063	8	22081785	22093628	<i>ZYG11A</i>	MAU
ENSTGUG00000009700	8	24544351	24566662		MAU
ENSTGUG00000009745	8	24945852	25164582	<i>NFIA</i>	HUA, MAU
ENSTGUG00000005665	9	3776452	3976275	<i>HS6ST1</i>	MAU
ENSTGUG00000007084	9	6573476	6609786	<i>PPP2R3A</i>	RAI
ENSTGUG00000008270	9	10768020	10786905		RAI
ENSTGUG00000009820	9	17638634	17660277		HUA
ENSTGUG00000010675	9	20069493	20079705	<i>GNB4</i>	HUA

ENSTGUG00000006654	10	8523659	8598997	MYO5A	HUA
ENSTGUG00000008476	10	14621964	14731878	SLCO3A1	MOO
ENSTGUG00000009006	10	18257909	18275693		HUA
ENSTGUG00000017715	10	20485775	20485848	MIR184	HUA
ENSTGUG00000010006	10	20492147	20493497	ZNF710	HUA
ENSTGUG00000010009	10	20501862	20505986	SEMA4B	HUA
ENSTGUG00000005417	11	4815708	4825887	E2F4	HUA, MOO, RAI
ENSTGUG00000005428	11	4827918	4834507	ELMO3	HUA, MOO, RAI
ENSTGUG00000005766	11	5056834	5058661	TPPP3	HUA
ENSTGUG00000008994	11	14711166	14741310		MAU
ENSTGUG00000008404	12	11785828	11797268	CCDC174	MAU
ENSTGUG00000010102	12	19614606	19766886		HUA
ENSTGUG00000000209	13	1338072	1344091	IL12B	HUA
ENSTGUG00000000776	13	6519325	6562614	FAT2	RAI
ENSTGUG00000000816	13	6633705	6783751		HUA
ENSTGUG00000001246	13	9525620	9569022	H2AFY	HUA
ENSTGUG00000004540	14	3169643	3207271	KIAA0556	HUA
ENSTGUG00000008135	14	12554331	12555206		MAU
ENSTGUG00000009821	15	11316478	11424936		MOO
ENSTGUG00000009830	15	11366499	11374215		MOO
ENSTGUG00000009905	15	11507233	11538464		HUA
ENSTGUG00000003398	17	5138627	5267364		MAU
ENSTGUG00000003550	17	5370218	5372236		MAU
ENSTGUG00000003552	17	5374125	5374580	DPM2	MAU
ENSTGUG00000003554	17	5376833	5488169	ST6GALNAC4	HUA, MAU, TAH
ENSTGUG00000003556	17	5380700	5381723		MAU
ENSTGUG00000003559	17	5383752	5387428	AK1	MAU
ENSTGUG00000003704	17	5499738	5503313	TRUB2	HUA, TAH
ENSTGUG00000007364	17	10940614	11046375	PBX3	HUA
ENSTGUG00000007428	17	11515416	11577459	RALGPS1	MAU
ENSTGUG00000003385	18	1691137	1696099	TMC6	MAU
ENSTGUG00000003410	18	1699283	1707780	HGS	MAU
ENSTGUG00000003446	18	1708980	1712466	MRPL12	MAU
ENSTGUG00000006238	18	5544437	5559002		MAU
ENSTGUG00000006545	18	5571232	5573803		MAU
ENSTGUG00000006551	18	5585671	5600256		MAU
ENSTGUG00000007675	18	6697273	6703781	CYBC1	HUA
ENSTGUG00000007809	18	7731908	7741193	JMJD6	HUA
ENSTGUG00000007819	18	7751042	7828425	ST6GALNAC2	HUA
ENSTGUG00000007923	18	7960889	7981633	UBE2O	MAU
ENSTGUG00000008958	19	11474946	11478895		HUA, MOO
ENSTGUG00000008962	19	11481381	11485166		HUA, MOO
ENSTGUG00000008966	19	11496529	11534638	PITPNM3	HUA, MOO
ENSTGUG00000002039	21	109799	116290	ZBTB17	TAH
ENSTGUG00000004728	22	2773733	2777396		HUA
ENSTGUG00000004741	22	2784936	2792494	ASH2L	HUA
ENSTGUG00000001038	23	2978316	2979139		HUA
ENSTGUG00000001039	23	2987758	2989220		HUA
ENSTGUG00000001130	23	3280382	3312215	RUNX3	HUA
ENSTGUG00000001160	23	3534172	3567527	MYOM3	HUA
ENSTGUG00000001176	23	3568981	3570931		HUA
ENSTGUG00000001265	23	3743972	3744625	HEYL	HUA
ENSTGUG00000001266	23	3755982	3768257	PABPC4	HUA
ENSTGUG00000001438	23	4294754	4297891	LUZP1	HUA
ENSTGUG00000001329	26	2526723	2529254		MAU

ENSTGUG00000001330	26	2537936	2542499	<i>TNNT2</i>	MAU
ENSTGUG00000001345	26	2697635	2701249	<i>TMEM9</i>	HUA
ENSTGUG00000001363	26	2749706	2780727	<i>PPFIA4</i>	MOO
ENSTGUG00000001381	26	2801182	2823090	<i>ADORA1</i>	HUA
ENSTGUG00000001511	26	3571959	3682353	<i>PLXNA2</i>	HUA
ENSTGUG00000002482	27	2319216	2319366		HUA
ENSTGUG00000002515	27	2388654	2627252		HUA
ENSTGUG00000002656	27	2509069	2631604	<i>VPS25</i>	HUA
ENSTGUG00000002727	27	2596517	2601876	<i>TUBG1</i>	HUA
ENSTGUG00000002735	27	2609963	2615943	<i>CNTNAP1</i>	HUA
ENSTGUG00000002970	27	2944147	2950604		HUA
ENSTGUG00000000551	28	2758875	2784928		MAU
ENSTGUG00000000664	28	3047464	3054358	<i>WDR18</i>	RAI
ENSTGUG00000000667	28	3061236	3072627		RAI
ENSTGUG00000002908	1A	12508602	12668585	<i>LHFPL3</i>	RAI
ENSTGUG00000007680	1A	39135461	39219032	<i>OTOGL</i>	MAU
ENSTGUG000000012520	1A	70536304	70554685	<i>ADA2</i>	TAH
ENSTGUG00000005324	4A	16132112	16150679	<i>CLCN5</i>	HUA, RAI

Samples sequenced

Table D.1. Samples sequenced as part of this thesis. Only samples that passed quality filtering are shown.

Sample ID	Location	Species / subspecies	Lat	Long	Year collected	Ch2	Ch3	Ch4
CDH16	Brisbane	Z. l. cornwalli	-28.18	153.45	2013	X		X
CDH17	Brisbane	Z. l. cornwalli	-28.18	153.45	2013	X		X
CDH18	Brisbane	Z. l. cornwalli	-28.18	153.45	2013	X		X
CDH19	Brisbane	Z. l. cornwalli	-28.18	153.45	2013	X		X
CDH20	Brisbane	Z. l. cornwalli	-28.18	153.45	2013	X		X
CDH22	Brisbane	Z. l. cornwalli	-28.18	153.45	2013	X		X
CDH23	Brisbane	Z. l. cornwalli	-28.18	153.45	2013	X		X
CDH24	Brisbane	Z. l. cornwalli	-28.18	153.45	2013	X		X
CDH25	Brisbane	Z. l. cornwalli	-28.18	153.45	2013	X		X
CDH26	Brisbane	Z. l. cornwalli	-28.18	153.45	2013	X		X
CDH43	Brisbane	Z. l. cornwalli	-28.18	153.45	2013	X		X
CDH44	Brisbane	Z. l. cornwalli	-28.18	153.45	2013	X		X
CDH46	Brisbane	Z. l. cornwalli	-28.18	153.45	2013	X		X
CDH47	Brisbane	Z. l. cornwalli	-28.18	153.45	2013	X		X
CDH48	Brisbane	Z. l. cornwalli	-28.18	153.45	2013	X		X
CDH49	Brisbane	Z. l. cornwalli	-28.18	153.45	2013	X		X
CDH50	Brisbane	Z. l. cornwalli	-28.18	153.45	2013	X		X
CDH54	Brisbane	Z. l. cornwalli	-28.18	153.45	2013	X		X
CI10	Chatham Island	Z. l. lateralis	-43.94	-176.49	1997	X		X
CI11	Chatham Island	Z. l. lateralis	-43.94	-176.49	1997	X		X
CI21	Chatham Island	Z. l. lateralis	-43.94	-176.49	1997	X		X
CI23	Chatham Island	Z. l. lateralis	-43.94	-176.49	1997	X		X
CI24	Chatham Island	Z. l. lateralis	-43.94	-176.49	1997			X
CI25	Chatham Island	Z. l. lateralis	-43.94	-176.49	1997	X		X
CI26	Chatham Island	Z. l. lateralis	-43.94	-176.49	1997	X		X
CI27	Chatham Island	Z. l. lateralis	-43.94	-176.49	1997	X		X
CI28	Chatham Island	Z. l. lateralis	-43.94	-176.49	1997	X		X
CI29	Chatham Island	Z. l. lateralis	-43.94	-176.49	1997	X		X
CI30	Chatham Island	Z. l. lateralis	-43.94	-176.49	1997	X		X
CI32	Chatham Island	Z. l. lateralis	-43.94	-176.49	1997	X		X
CI33	Chatham Island	Z. l. lateralis	-43.94	-176.49	1997	X		X
CI34	Chatham Island	Z. l. lateralis	-43.94	-176.49	1997	X		X
CI35	Chatham Island	Z. l. lateralis	-43.94	-176.49	1997	X		X
CI36	Chatham Island	Z. l. lateralis	-43.94	-176.49	1997	X		X
CI37	Chatham Island	Z. l. lateralis	-43.94	-176.49	1997	X		X
CI39	Chatham Island	Z. l. lateralis	-43.94	-176.49	1997	X		X
GT120	Grande Terre	Z. l. griseonotus	-21.79	166.02	2014	X		
GT132	Grande Terre	Z. l. griseonotus	-21.79	166.02	2014	X		
GT136	Grande Terre	Z. l. griseonotus	-21.79	166.02	2014	X		
GT150	Grande Terre	Z. l. griseonotus	-21.79	166.02	2014	X		

GT156	Grande Terre	Z. l. griseonotus	-21.79	166.02	2014	X	
GT203	Grande Terre	Z. l. griseonotus	-21.79	166.02	2014	X	
GT224	Grande Terre	Z. l. griseonotus	-21.79	166.02	2014	X	
GT239	Grande Terre	Z. l. griseonotus	-21.79	166.02	2014	X	
N107	Grande Terre	Z. l. griseonotus	-20.70	165.00	2006	X	
N108	Grande Terre	Z. l. griseonotus	-20.70	165.00	2006	X	
N109	Grande Terre	Z. l. griseonotus	-20.70	165.00	2006	X	
N110	Grande Terre	Z. l. griseonotus	-20.70	165.00	2006	X	
N206	Grande Terre	Z. l. griseonotus	-20.70	165.00	2006	X	
N208	Grande Terre	Z. l. griseonotus	-20.70	165.00	2006	X	
N211	Grande Terre	Z. l. griseonotus	-20.70	165.00	2006	X	
N212	Grande Terre	Z. l. griseonotus	-20.70	165.00	2006	X	
N213	Grande Terre	Z. l. griseonotus	-20.70	165.00	2006	X	
12-050	Heron Island	Z. l. chlorocephalus	23.44	151.91	2015	X	X
12-051	Heron Island	Z. l. chlorocephalus	23.44	151.91	2015	X	X
12-052	Heron Island	Z. l. chlorocephalus	23.44	151.91	2015	X	X
12-053	Heron Island	Z. l. chlorocephalus	23.44	151.91	2015	X	X
12-055	Heron Island	Z. l. chlorocephalus	23.44	151.91	2015	X	X
12-056	Heron Island	Z. l. chlorocephalus	23.44	151.91	2015	X	X
12-057	Heron Island	Z. l. chlorocephalus	23.44	151.91	2015	X	X
12-059	Heron Island	Z. l. chlorocephalus	23.44	151.91	2015	X	X
12-061	Heron Island	Z. l. chlorocephalus	23.44	151.91	2015	X	X
12-063	Heron Island	Z. l. chlorocephalus	23.44	151.91	2015	X	X
12-064	Heron Island	Z. l. chlorocephalus	23.44	151.91	2015	X	X
12-065	Heron Island	Z. l. chlorocephalus	23.44	151.91	2015	X	X
12-066	Heron Island	Z. l. chlorocephalus	23.44	151.91	2015	X	X
12-068	Heron Island	Z. l. chlorocephalus	23.44	151.91	2015	X	X
12-049	Heron Island	Z. l. chlorocephalus	23.44	151.91	2015		X
12-058	Heron Island	Z. l. chlorocephalus	23.44	151.91	2015		X
12-060	Heron Island	Z. l. chlorocephalus	23.44	151.91	2015		X
12-062	Heron Island	Z. l. chlorocephalus	23.44	151.91	2015		X
LH101	Heron Island	Z. l. chlorocephalus	23.44	151.91	2015		X
LH119	Heron Island	Z. l. chlorocephalus	23.44	151.91	2015		X
HUA13	Huahine	Z. l. lateralis	16.82	-150.98	2017		X
HUA18	Huahine	Z. l. lateralis	16.82	-150.98	2017		X
HUA19	Huahine	Z. l. lateralis	16.82	-150.98	2017		X
HUA2	Huahine	Z. l. lateralis	16.82	-150.98	2017		X
HUA20	Huahine	Z. l. lateralis	16.82	-150.98	2017		X
HUA23	Huahine	Z. l. lateralis	16.82	-150.98	2017		X
HUA4	Huahine	Z. l. lateralis	16.82	-150.98	2017		X
HUA7	Huahine	Z. l. lateralis	16.82	-150.98	2017		X
HUA8	Huahine	Z. l. lateralis	16.82	-150.98	2017		X
L519	Lifou	Z. l. melanops	-20.79	167.12	2006	X	
L536	Lifou	Z. l. melanops	-20.79	167.12	2006	X	
L537	Lifou	Z. l. melanops	-20.79	167.12	2006	X	
L539	Lifou	Z. l. melanops	-20.79	167.12	2006	X	
L549	Lifou	Z. l. melanops	-20.79	167.12	2006	X	
L550	Lifou	Z. l. melanops	-20.79	167.12	2006	X	
L591	Lifou	Z. l. melanops	-20.79	167.12	2006	X	
L593	Lifou	Z. l. melanops	-20.79	167.12	2006	X	
L594	Lifou	Z. l. melanops	-20.79	167.12	2006	X	
L595	Lifou	Z. l. melanops	-20.79	167.12	2006	X	
L596	Lifou	Z. l. melanops	-20.79	167.12	2006	X	
L604	Lifou	Z. l. melanops	-20.79	167.12	2006	X	
L613	Lifou	Z. l. melanops	-20.79	167.12	2006	X	

L618	Lifou	Z. l. melanops	-20.79	167.12	2006	X	
L619	Lifou	Z. l. melanops	-20.79	167.12	2006	X	
L645	Lifou	Z. l. melanops	-20.79	167.12	2006	X	
L656	Lifou	Z. l. melanops	-20.79	167.12	2006	X	
LH102	Lord Howe Island	Z. l. tephropleurus	-31.60	159.10	1998	X	X
LH103	Lord Howe Island	Z. l. tephropleurus	-31.60	159.10	1998	X	X
LH104	Lord Howe Island	Z. l. tephropleurus	-31.60	159.10	1998	X	X
LH105	Lord Howe Island	Z. l. tephropleurus	-31.60	159.10	1998	X	X
LH106	Lord Howe Island	Z. l. tephropleurus	-31.60	159.10	1998	X	X
LH107	Lord Howe Island	Z. l. tephropleurus	-31.60	159.10	1998	X	X
LH108	Lord Howe Island	Z. l. tephropleurus	-31.60	159.10	1998	X	X
LH113	Lord Howe Island	Z. l. tephropleurus	-31.60	159.10	1998	X	X
LH114	Lord Howe Island	Z. l. tephropleurus	-31.60	159.10	1998	X	X
LH115	Lord Howe Island	Z. l. tephropleurus	-31.60	159.10	1998	X	X
LH116	Lord Howe Island	Z. l. tephropleurus	-31.60	159.10	1998	X	X
LH117	Lord Howe Island	Z. l. tephropleurus	-31.60	159.10	1998	X	X
LH118	Lord Howe Island	Z. l. tephropleurus	-31.60	159.10	1998	X	X
LH128	Lord Howe Island	Z. l. tephropleurus	-31.60	159.10	1998	X	X
LH129	Lord Howe Island	Z. l. tephropleurus	-31.60	159.10	1998	X	X
LH130	Lord Howe Island	Z. l. tephropleurus	-31.60	159.10	1998	X	X
LH131	Lord Howe Island	Z. l. tephropleurus	-31.60	159.10	1998	X	X
LH132	Lord Howe Island	Z. l. tephropleurus	-31.60	159.10	1998	X	X
MAU11	Maupiti	Z. l. lateralis	-16.45	-152.27	2017		X
MAU12	Maupiti	Z. l. lateralis	-16.45	-152.27	2017		X
MAU14	Maupiti	Z. l. lateralis	-16.45	-152.27	2017		X
MAU15	Maupiti	Z. l. lateralis	-16.45	-152.27	2017		X
MAU16	Maupiti	Z. l. lateralis	-16.45	-152.27	2017		X
MAU17	Maupiti	Z. l. lateralis	-16.45	-152.27	2017		X
MAU19	Maupiti	Z. l. lateralis	-16.45	-152.27	2017		X
MAU21	Maupiti	Z. l. lateralis	-16.45	-152.27	2017		X
MAU22	Maupiti	Z. l. lateralis	-16.45	-152.27	2017		X
MAU23	Maupiti	Z. l. lateralis	-16.45	-152.27	2017		X
MAU25	Maupiti	Z. l. lateralis	-16.45	-152.27	2017		X
MAU26	Maupiti	Z. l. lateralis	-16.45	-152.27	2017		X
MAU27	Maupiti	Z. l. lateralis	-16.45	-152.27	2017		X
MAU28	Maupiti	Z. l. lateralis	-16.45	-152.27	2017		X
MAU29	Maupiti	Z. l. lateralis	-16.45	-152.27	2017		X
MAU30	Maupiti	Z. l. lateralis	-16.45	-152.27	2017		X
MAU31	Maupiti	Z. l. lateralis	-16.45	-152.27	2017		X
MAU32	Maupiti	Z. l. lateralis	-16.45	-152.27	2017		X
MAU33	Maupiti	Z. l. lateralis	-16.45	-152.27	2017		X
SE109	Mo'orea	Z. l. lateralis	-17.51	-149.91	2005	X	
SE110	Mo'orea	Z. l. lateralis	-17.51	-149.91	2005	X	
SE113	Mo'orea	Z. l. lateralis	-17.51	-149.91	2005	X	
SE115	Mo'orea	Z. l. lateralis	-17.51	-149.91	2005	X	
SE118	Mo'orea	Z. l. lateralis	-17.51	-149.91	2005	X	
SE119	Mo'orea	Z. l. lateralis	-17.51	-149.91	2005	X	
SE121	Mo'orea	Z. l. lateralis	-17.51	-149.91	2005	X	
SE122	Mo'orea	Z. l. lateralis	-17.51	-149.91	2005	X	
SE124	Mo'orea	Z. l. lateralis	-17.51	-149.91	2005	X	
SE127	Mo'orea	Z. l. lateralis	-17.51	-149.91	2005	X	
MOO4	Mo'orea	Z. l. lateralis	-17.49	-149.83	2017		X
MOO42	Mo'orea	Z. l. lateralis	-17.49	-149.83	2017		X
MOO45	Mo'orea	Z. l. lateralis	-17.49	-149.83	2017		X
MOO46	Mo'orea	Z. l. lateralis	-17.49	-149.83	2017		X

MOO48	Mo'orea	Z. l. lateralis	-17.49	-149.83	2017	X		
MOO52	Mo'orea	Z. l. lateralis	-17.49	-149.83	2017	X		
MOO55	Mo'orea	Z. l. lateralis	-17.49	-149.83	2017	X		
MOO56	Mo'orea	Z. l. lateralis	-17.49	-149.83	2017	X		
MOO59	Mo'orea	Z. l. lateralis	-17.49	-149.83	2017	X		
MOO63	Mo'orea	Z. l. lateralis	-17.49	-149.83	2017	X		
MOO65	Mo'orea	Z. l. lateralis	-17.49	-149.83	2017	X		
NOR102	Norfolk Island	Z. l. lateralis	-29.04	167.96	1998		X	
NOR103	Norfolk Island	Z. l. lateralis	-29.04	167.96	1998		X	
NOR104	Norfolk Island	Z. l. lateralis	-29.04	167.96	1998		X	
NOR110	Norfolk Island	Z. l. lateralis	-29.04	167.96	1998		X	
NOR111	Norfolk Island	Z. l. lateralis	-29.04	167.96	1998		X	
NOR112	Norfolk Island	Z. l. lateralis	-29.04	167.96	1998		X	
NOR113	Norfolk Island	Z. l. lateralis	-29.04	167.96	1998		X	
NOR114	Norfolk Island	Z. l. lateralis	-29.04	167.96	1998		X	
NOR117	Norfolk Island	Z. l. lateralis	-29.04	167.96	1998		X	
NOR121	Norfolk Island	Z. l. lateralis	-29.04	167.96	1998		X	
NOR126	Norfolk Island	Z. l. lateralis	-29.04	167.96	1998		X	
NOR130	Norfolk Island	Z. l. lateralis	-29.04	167.96	1998		X	
NOR124	Norfolk Island	Zosterops tenuirostris	-29.04	167.96	1998		X	
NOR132	Norfolk Island	Zosterops tenuirostris	-29.04	167.96	1998		X	
NOR134	Norfolk Island	Zosterops tenuirostris	-29.04	167.96	1998		X	
NOR140	Norfolk Island	Zosterops tenuirostris	-29.04	167.96	1998		X	
NOR141	Norfolk Island	Zosterops tenuirostris	-29.04	167.96	1998		X	
PN1	North Island	Z. l. lateralis	-40.22	175.57	1997		X	
PN10	North Island	Z. l. lateralis	-40.22	175.57	1997		X	
PN11	North Island	Z. l. lateralis	-40.22	175.57	1997		X	
PN13	North Island	Z. l. lateralis	-40.22	175.57	1997		X	
PN14	North Island	Z. l. lateralis	-40.22	175.57	1997		X	
PN15	North Island	Z. l. lateralis	-40.22	175.57	1997		X	
PN2	North Island	Z. l. lateralis	-40.22	175.57	1997		X	
PN3	North Island	Z. l. lateralis	-40.22	175.57	1997		X	
PN4	North Island	Z. l. lateralis	-40.22	175.57	1997		X	
PN7	North Island	Z. l. lateralis	-40.22	175.57	1997			X
PO7-064	Raiatea	Z. l. lateralis	-16.82	-151.46	2010	X		
PO7-066	Raiatea	Z. l. lateralis	-16.82	-151.46	2010	X		
RAI17	Raiatea	Z. l. lateralis	-16.77	-151.42	2017		X	
RAI18	Raiatea	Z. l. lateralis	-16.77	-151.42	2017		X	
RAI2	Raiatea	Z. l. lateralis	-16.77	-151.42	2017		X	
RAI20	Raiatea	Z. l. lateralis	-16.77	-151.42	2017		X	
RAI21	Raiatea	Z. l. lateralis	-16.77	-151.42	2017		X	
RAI24	Raiatea	Z. l. lateralis	-16.77	-151.42	2017		X	
RAI25	Raiatea	Z. l. lateralis	-16.77	-151.42	2017		X	
RAI26	Raiatea	Z. l. lateralis	-16.77	-151.42	2017		X	
RAI31	Raiatea	Z. l. lateralis	-16.77	-151.42	2017		X	
RAI32	Raiatea	Z. l. lateralis	-16.77	-151.42	2017		X	
RAI34	Raiatea	Z. l. lateralis	-16.77	-151.42	2017		X	
RAI6	Raiatea	Z. l. lateralis	-16.77	-151.42	2017		X	
RAI7	Raiatea	Z. l. lateralis	-16.77	-151.42	2017		X	
RAI9	Raiatea	Z. l. lateralis	-16.77	-151.42	2017		X	
PO2-092	Rurutu	Z. l. lateralis	-22.48	-151.34	2012	X		
PO2-095	Rurutu	Z. l. lateralis	-22.48	-151.34	2012	X		
PO2-098	Rurutu	Z. l. lateralis	-22.48	-151.34	2012	X		
D10	South Island	Z. l. lateralis	-45.83	170.47	1997	X	X	X
D12	South Island	Z. l. lateralis	-45.83	170.47	1997	X	X	X

D13	South Island	Z. l. lateralis	-45.83	170.47	1997	X	X	X
D14	South Island	Z. l. lateralis	-45.83	170.47	1997	X	X	X
D15	South Island	Z. l. lateralis	-45.83	170.47	1997	X	X	X
D2	South Island	Z. l. lateralis	-45.83	170.47	1997	X	X	X
D20	South Island	Z. l. lateralis	-45.83	170.47	1997	X	X	X
D22	South Island	Z. l. lateralis	-45.83	170.47	1997	X	X	X
D23	South Island	Z. l. lateralis	-45.85	170.48	1997	X	X	X
D24	South Island	Z. l. lateralis	-45.85	170.48	1997	X	X	X
D25	South Island	Z. l. lateralis	-45.85	170.48	1997	X	X	X
D26	South Island	Z. l. lateralis	-45.85	170.48	1997	X	X	X
D27	South Island	Z. l. lateralis	-45.85	170.48	1997	X	X	X
D28	South Island	Z. l. lateralis	-45.85	170.48	1997	X	X	X
D29	South Island	Z. l. lateralis	-45.85	170.48	1997	X	X	X
D30	South Island	Z. l. lateralis	-45.85	170.48	1997	X	X	X
D5	South Island	Z. l. lateralis	-45.83	170.47	1997	X	X	X
D9	South Island	Z. l. lateralis	-45.83	170.47	1997	X	X	X
D16	South Island	Z. l. lateralis	-45.83	170.47	1997		X	
D18	South Island	Z. l. lateralis	-45.83	170.47	1997		X	X
D21	South Island	Z. l. lateralis	-45.83	170.47	1997		X	X
D3	South Island	Z. l. lateralis	-45.83	170.47	1997		X	
D8	South Island	Z. l. lateralis	-45.83	170.47	1997		X	X
TAH14	Tahiti	Z. l. lateralis	-17.57	-149.53	2017		X	X
TAH16	Tahiti	Z. l. lateralis	-17.57	-149.53	2017		X	X
TAH17	Tahiti	Z. l. lateralis	-17.57	-149.53	2017		X	X
TAH18	Tahiti	Z. l. lateralis	-17.57	-149.53	2017		X	
TAH38	Tahiti	Z. l. lateralis	-17.57	-149.53	2017		X	
TAH43	Tahiti	Z. l. lateralis	-17.57	-149.53	2017		X	X
TAH45	Tahiti	Z. l. lateralis	-17.57	-149.53	2017		X	X
TAH46	Tahiti	Z. l. lateralis	-17.57	-149.53	2017		X	X
TAH48	Tahiti	Z. l. lateralis	-17.57	-149.53	2017		X	X
TAH49	Tahiti	Z. l. lateralis	-17.57	-149.53	2017		X	X
TAH53	Tahiti	Z. l. lateralis	-17.57	-149.53	2017		X	
TAH54	Tahiti	Z. l. lateralis	-17.57	-149.53	2017		X	
TAH58	Tahiti	Z. l. lateralis	-17.57	-149.53	2017		X	X
TAS19	Tasmania	Z. l. lateralis	-42.88	147.33	1997			X
TAS2	Tasmania	Z. l. lateralis	-42.88	147.33	1997			X
TAS26	Tasmania	Z. l. lateralis	-42.88	147.33	1997			X
TAS3	Tasmania	Z. l. lateralis	-42.88	147.33	1997			X
TAS42	Tasmania	Z. l. lateralis	-42.88	147.33	1997			X
TAS45	Tasmania	Z. l. lateralis	-42.88	147.33	1997			X
TAS52	Tasmania	Z. l. lateralis	-42.88	147.33	1997			X
TAS53	Tasmania	Z. l. lateralis	-42.88	147.33	1997			X
TAS55	Tasmania	Z. l. lateralis	-42.88	147.33	1997			X
TAS64	Tasmania	Z. l. lateralis	-42.88	147.33	1997			X
TAS68	Tasmania	Z. l. lateralis	-42.88	147.33	1997			X
TAS71	Tasmania	Z. l. lateralis	-42.88	147.33	1997			X
TAS72	Tasmania	Z. l. lateralis	-42.88	147.33	1997			X
TAS74	Tasmania	Z. l. lateralis	-42.88	147.33	1997			X
TAS76	Tasmania	Z. l. lateralis	-42.88	147.33	1997			X
TAS78	Tasmania	Z. l. lateralis	-42.88	147.33	1997			X
TAS79	Tasmania	Z. l. lateralis	-42.88	147.33	1997			X
TAS8	Tasmania	Z. l. lateralis	-42.88	147.33	1997			X
TAS81	Tasmania	Z. l. lateralis	-42.88	147.33	1997			X

The multiple population genetic and demographic routes to islands of genomic divergence

Claudio S. Quilodrán, Kristen C. Ruegg,
Ashley T. Sendell-Price, Eric C. Anderson,
Tim Coulson & Sonya M. Clegg

Published as:

Quilodrán CS, Ruegg KC, Sendell-Price AT, Anderson EC, Coulson T, Clegg SM (2020). The multiple population genetic and demographic routes to islands of genomic divergence. *Methods in Ecology and Evolution* 11: 6-21.



RESEARCH ARTICLE

The multiple population genetic and demographic routes to islands of genomic divergence

Claudio S. Quilodrán¹ | Kristen Ruegg^{1,2,3} | Ashley T. Sendell-Price¹ |
Eric C. Anderson⁴ | Tim Coulson¹ | Sonya M. Clegg¹

¹Department of Zoology, University of Oxford, Oxford, UK

²Center for Tropical Research, Institute of the Environment and Sustainability, University of California, Los Angeles, Los Angeles, CA, USA

³Department of Biology, Colorado State University, Fort Collins, CO, USA

⁴Fisheries Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, Santa Cruz, CA, USA

Correspondence

Claudio S. Quilodrán

Email: claudio.quilodran@zoo.ox.ac.uk

Funding information

Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: P2GEP3_168973 and P400PB_183930

Handling Editor: Susan Johnston

Abstract

1. The way that organisms diverge into reproductively isolated species is a major question in biology. The recent accumulation of genomic data provides promising opportunities to understand the genomic landscape of divergence, which describes the distribution of differences across genomes. Genomic areas of unusually high differentiation have been called genomic islands of divergence. Their formation has been attributed to a variety of mechanisms, but a prominent hypothesis is that they result from divergent selection over a small portion of the genome, with surrounding areas homogenized by gene flow. Such islands have often been interpreted as being associated with divergence with gene flow. However, other mechanisms related to genomic structure and population history can also contribute to the formation of genomic islands of divergence.
2. We currently lack a quantitative framework to examine the dynamics of genomic landscapes under the complex and nuanced conditions that are found in natural systems. Here, we develop an individual-based simulation to explore the dynamics of diverging genomes under various scenarios of gene flow, selection and genotype-phenotype maps.
3. Our modelling results are consistent with empirical observations demonstrating the formation of genomic islands under genetic isolation. Importantly, we have quantified the range of conditions that produce genomic islands. We demonstrate that the initial level of genetic diversity, drift, time since divergence, linkage disequilibrium, strength of selection and gene flow are all important factors that can influence the formation of genomic islands. Because the accumulation of genomic differentiation over time tends to erode the signal of genomic islands, genomic islands are more likely to be observed in recently divergent taxa, although not all recently diverged taxa will necessarily exhibit islands of genomic divergence. Gene flow primarily slows the swamping of islands of divergence with time.
4. By using this framework, further studies may explore the relative influence of particular suites of events that contribute to the emergence of genomic islands under sympatric, parapatric and allopatric conditions. This approach represents

Tim Coulson and Sonya M. Clegg are joint senior authors.

a novel tool to explore quantitative expectations of the speciation process, and should prove useful in elucidating past and projecting future genomic evolution of any taxa.

KEYWORDS

divergent selection, evolution, gene flow, genomic divergence, genomic landscape, glads r package, individual-based model, island of genomic divergence

1 | INTRODUCTION

A major aim of evolutionary biology is to understand mechanisms associated with the divergence of organisms between populations and the emergence of new species. This motivated Charles Darwin and Alfred Wallace 160 years ago when they advanced the Theory of Natural Selection (Darwin & Wallace, 1858). Since then, the increasing accumulation of genetic, genomic and computational tools has allowed a better understanding of the genetic basis of the speciation process, resulting in the rise of a new era of evolutionary research (Chanderbali, Berger, Howarth, Soltis, & Soltis, 2016; Hughes, 2009; Roux et al., 2016). Patterns of divergence at the level of the genome have been characterized for an increasing number of taxa, but the extent to which observed patterns are informative about evolutionary processes is actively debated (e.g. Burri et al., 2015; Ellegren et al., 2012; Renaut et al., 2013; Ruegg, Anderson, Boone, Pouls, & Smith, 2014).

The genomic landscape of divergence describes the distribution of differences across the genomes of diverging organisms. The genome of a diverging taxon does not change uniformly, with some regions changing at higher rates than others (Ravinet et al., 2017; Seehausen et al., 2014). If a single process uniquely generates a particular divergence pattern, then identification of that pattern can confidently be interpreted as representing a particular evolutionary history. In contrast, if multiple processes can generate the same patterns of genomic divergence, then identification of the pattern will not point to a specific process, although the suite of candidate processes may be narrowed. In these cases, additional information beyond patterns of genomic divergence, such as the ecological and evolutionary context of a given divergence, will be required to understand patterns of evolutionary divergence.

Genomic islands of divergence – highly differentiated regions of the genome that are surrounded by regions of low differentiation – are a particularly intriguing pattern of genomic divergence (Harr, 2006; Nosil, Funk, & Ortiz-Barrientos, 2009; Turner, Hahn, & Nuzhdin, 2005). Genomic islands are often a focus of studies using genomic scans to identify target genes of natural selection (Lotterhos & Whitlock, 2014). Initially, genomic island formation was attributed to the action of divergent natural selection on particular loci, which contribute to reproductive isolation via pre- or post-zygotic barriers to a greater extent than elsewhere within the genome (Wu, 2001; Wu & Ting, 2004). This process is hypothesized to create elevated regions of genomic differentiation,

containing the selected loci and other physically linked loci, surrounded by regions homogenized by gene flow (Nosil et al., 2009). Several authors consequently interpreted the presence of genomic islands of divergence as a signal of divergence with gene flow (e.g. Feder, Flaxman, Egan, Comeault, & Nosil, 2013), concluding that the speciation process in sympatric or parapatric conditions may be more common than previously thought (e.g. Fraïsse, Roux, Welch, & Bierne, 2014; Nosil, 2008; Soria-Carrasco et al., 2014). However, empirical studies have also proposed that genomic islands of divergence can arise in the absence of gene flow due to a variety of causes, such as the structure of the diverging genomes (e.g. variation in the recombination rate) and the action of genetic drift, background selection and adaptation to local environmental conditions (Campagna, Gronau, Silveira, Siepel, & Lovette, 2015; Cruickshank & Hahn, 2014; Noor & Bennett, 2009). Genomic regions of reduced recombination, such as areas close to centromeres or inside chromosome rearrangements that experience background or divergent selection have received particular attention (Noor & Bennett, 2009). Selection acting in regions with low recombination rates is expected to reduce the effective population size of these genomic regions to a greater extent than elsewhere in the genome, which may result in the creation of genomic islands (Feder & Nosil, 2009; Turner & Hahn, 2010). Note that those regions of reduced recombination may still harbour important genes related to the divergence process (Samuk et al., 2017). Furthermore, regions of low divergence may occur because of incomplete lineage sorting rather than homogenization by gene flow, with peaks of genetic divergence being an artefact of a loss of nucleotide diversity within populations after divergent selection (Cruickshank & Hahn, 2014).

There have been some previous attempts to model the structure of genomic landscapes that can be applied to the formation of genomic islands of divergence. However, these models either simulate single bi-allelic selected loci (Charlesworth, Nordborg, & Charlesworth, 1997; Sedghifar, Brandvain, & Ralph, 2016) or consider a small number of simulated loci (Feder, Gejji, Yeaman, & Nosil, 2012; Feder & Nosil, 2009, 2010). They also provide a static view of the divergence process by summarizing selection as a single parameter (sometimes referred to as s). While informative, such models represent specific stages when populations have already achieved a given level of differentiation. Flaxman, Feder, and Nosil (2013) used an individual-based model to project this dynamic forward in time, but their model was constrained to a uniform distribution of loci

with constant recombination rates. Recently developed tools allow the simulation of large quantities of genomic information in an efficient coalescent framework (Haller, Galloway, Kelleher, Messer, & Ralph, 2019; Haller & Messer, 2019; Kelleher, Etheridge, & McVean, 2016; Lohse, 2017). While simulations based on coalescence are particularly useful when exploring the dynamics of neutral evolution, they are less informative for small populations that experience selection and where reproduction is not panmictic (Currat, Gerbault, Di, Nunes, & Sanchez-Mazas, 2015). Nonetheless, a combination of recently developed tools have helped to perform simulations beyond this original limitation of the coalescence theory (see Haller et al., 2019; Haller & Messer, 2019). A flexible quantitative framework is desirable to help investigate the dynamics of genomic landscapes, thus further increasing the utility of existing genomic datasets (Feder et al., 2013; Seehausen et al., 2014). We develop a quantitative individual-based modelling approach to simulate the dynamic of a genomic landscape of divergence (R package 'glads': genomic landscape of divergence simulations). Our approach does not include a backward coalescent algorithm to reconstruct genetic diversity of a subsample of the population. Instead, all individuals in a population are simulated forward in time, allowing the exploration of complex demographic and genetic consequences of selection. The model simulates any number of loci and allelic polymorphisms and can be motivated or parameterized using data. There is a range of individual-based frameworks that allow the simulation of all individuals in a population (e.g. Currat et al., 2015; Peng & Kimmel, 2005). A major difference between our approach and these tools is the inclusion of a fitness function making our approach compatible with many structured ecological and evolutionary models (Hohenlohe, Bassham, Currey, & Cresko, 2012; Nosil et al., 2009). This allows us to explore the population and genomic dynamics associated with the process of divergence between populations before an equilibrium is reached, in a way that is often not possible when selection is fixed as a constant parameter. Basically, the fitness function allows to link selection (and hence evolution) to ecology (e.g. population dynamics) in a way that approaches based on parameters describing selection do not. The fitness function is, therefore, extending genetic variation right through to population dynamics (and potentially community dynamics) in way that has not been done before (see Coulson, Potter, & Felmy, 2019). Our approach is highly flexible and can be constructed for any genotype-phenotype map and any configuration of recombination rates between neighbouring loci, can be constructed for deterministic and stochastic environments, and incorporates any desired system of mating. The simulation method presented here provides a flexible framework to examine the dynamics of diverging genomic landscapes under various scenarios of gene flow and selection on single genes or networks of multiple interacting genes. Our approach represents a novel tool to evaluate quantitative expectations in genomic landscapes. It is useful to elucidate the influence of a range of demographic and evolutionary scenarios, including divergence with or without gene flow, the divergence timeframe and the structure of target genomes.

2 | MATERIALS AND METHODS

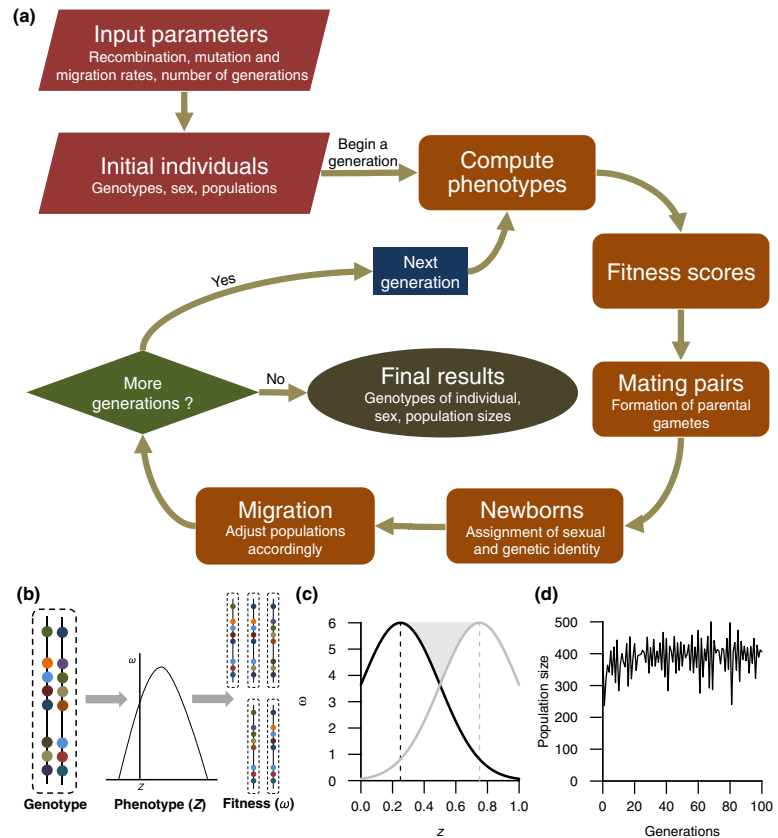
2.1 | General description of the model

The purpose of our model is to provide insight into how a range of genetic and demographic processes can generate genomic signatures and patterns of genomic divergence between populations. Our primary motivation was to explore factors associated with the emergence of genomic islands of divergence, but our approach can be applied to many questions about the structure of genomes, genomic landscapes and the evolution of divergent organisms.

The model is individual based and consists of two populations that may or may not be linked by gene flow. Our model is composed of three hierarchical levels: genotypes, phenotypic traits and demographic rates. The dynamics of the populations, the distributions of genotypes at each locus and the phenotypic traits are all emergent properties of the model. The model tracks the multivariate distribution of multi-locus genotypes and phenotypes. We simulate individuals that are characterized by sex and genetic identity (Figure 1a). The genotype and the environment determine the phenotypic trait values of an individual via a genotype-phenotype map. The phenotypic trait values influence an individual's expected demography (i.e. survival, mate choice and reproductive success). For example, assuming a per generation time step, the potential number of offspring produced by each individual depends on its phenotype $w = f(z)$, which in turn depends on the individual genotype and on the environment $z = g(G, E)$ (Figure 1b). G is a numeric value determined by an individual's genotype, representing the genetic value of the genotype. In the case of an additive genetic map, the genetic value of a genotype will be a breeding value. E represents the effect of the environment on phenotypic expression, and this allows us to capture the effects of plasticity on phenotypic expression. The environmental effect is important when simulating real-life eco-evolutionary dynamics because it almost always interacts with the genotype to determine the expression of a phenotypic trait (Bradshaw, 1965; Kokko et al., 2017). The realized demography and, consequently, summary statistics, such as the sex ratio and the number of offspring, are obtained by sampling from a distribution whose expected value is the expected demography. Once mating pairs are formed, the genotype of the young is determined by merging haploid gametes produced by each parent. Genetic variation in the offspring is determined by recombination and mutation.

We describe how the model is implemented in the next section. Our starting point is the distribution of individuals classified by genotype, sex and population (Figure 1a). First, we generate the phenotypic trait of each individual given its genotype (and potentially the environment). Second, we calculate individual fitness given an individual's phenotype, the population it is in and potentially the environment. Mating pairs are formed (randomly, assortatively or disassortatively) based on these individual fitness scores. Parental gametes are then produced given recombination and mutation rates, before segregating within mating pairs to generate offspring genotypes. The offspring can disperse to the neighbouring population

FIGURE 1 General description of the simulation framework. (a) Main steps of the general modelling approach. The red polygons represent the starting conditions. The orange squares are the different computing steps on each generation. The green polygon is a condition variable stating either the running of a next generation (blue square) or the end of simulations (olive green circle). (b) Relationship between genotypes, phenotypes, and fitness. The genetic variation is represented in different colours. The space between points represents unequal cM distances. (c) Two different fitness functions with different phenotypic optima. (d) Example of population size across time



with a given probability. The loop is then repeated for the next offspring generation.

2.2 | Individual-based framework

We assume organisms are diploid and composed of males and females. Each individual i is characterized by a two-dimensional array that represents a pair of homologous chromosomes. Multiple pairs of arrays may also be constructed to allow the characterization of any number of chromosomes. Similarly, variation in the number of dimensions of the arrays may be introduced to extend this framework to haploid or polyploid organisms. Each element of the array is an integer defining the copy of a given allele at a given locus. Individuals are also classified into populations. We assume random mating within a population, although this assumption can easily be relaxed (Ellner, Childs, & Rees, 2016; Schindler et al., 2015). Populations i and j are linked by migration rates (m_{ij} and m_{ji}) describing movement from population i to j and vice versa. We assume that individuals that migrate and reproduce successfully pass their genes into the other population hence incorporating gene flow into the model. The genotype-phenotype and phenotype-demography map can differ between populations if required.

The model proceeds in discrete time steps representing generations. It is a forward simulation that includes reproduction and migration at each time step. Density dependence regulates the population growth rate, influencing the probability of successful reproduction (Figure 1a).

The fitness of individuals is associated with a phenotype (z). We only focus on an additive genetic genotype-phenotype map here, but maps including epistasis, pleiotropy and dominance are possible. In the additive case, the sum of values of alleles at each locus gives a breeding value (b_v) for each individual at that locus. The sum of breeding values across loci gives a breeding value for the phenotype. Therefore,

$$z = \sum_{v=1}^{n_a} b_v + \epsilon_{\text{env}}(0, \sigma_{\text{env}}), \quad (1)$$

where n_a is the number of additive loci. In our simulations, the environmental contribution (ϵ_{env}) is assumed to be stochastic and normally distributed, with mean 0 and standard variation σ_{env} . ϵ_{env} may also be dependent on population density or any other environmental driver (Coulson et al., 2017).

The fitness function (ω) defines the phenotype-fitness map and consequently the type of selection influencing the divergence

between populations. Once a population has colonized a novel area, new phenotype–environment interactions appear on the phenotype–demography map, shifting the distribution of phenotypes that are expected to have higher fitness (i.e. phenotypic optima). The difference in phenotypic optima between the populations drives the strength of ‘divergent selection’ (grey area, Figure 1c). Populations exposed to equal phenotypic optima are considered to be under ‘concordant selection’. The fitness function we use has the form:

$$\omega = b_0 e^{-\frac{1}{2} \left(\frac{4z - b_1 n_a}{b_2 n_a} \right)^2} - b_3 N + \epsilon_{\text{dem}}(0, \sigma_{\text{dem}}) \quad (2)$$

The first part on the right-hand side of Equation 2 is based on a Gaussian-distribution determining the relation between the phenotypic trait value (z) and fitness (ω). The parameters b_0 , b_1 and b_2 define the maximum number of offspring produced, the phenotypic optima, and the variance of the Gaussian curve, respectively. The second part of Equation 2 determines the intensity of density dependence (b_3) on the fitness of individuals that are members of a population of size N . The final part of the equation introduces a stochastic demographic variant with mean 0 and standard variation σ_{dem} . The last two parts of the equation thus determine the increasing or decreasing variation in fitness due to fluctuations in population size and demographic stochasticity. Any other form of fitness function could be introduced to account for specific relationships between phenotypes (e.g. weight, height, bill size, colour pattern) and the expected number of offspring produced.

The number of breeding events is regulated by the number of females present in the population. Males are randomly selected according to the number of breeding females. The genotypes of both parents participate in the genomic structure of their offspring by transmitting a haploid copy of genetic material. The offspring differs from the parents by carrying half of the genome of each parent and by specific rules defining the recombination rate (ρ) between homologous chromosomes. We do not explore the effect of new mutations here, because we are primarily interested in the emergence of genomic islands at relative early stages of evolutionary diversification. However, mutation can easily be incorporated by generating a novel polymorphism at a random locus at a given rate per generation (see example in Appendix S1).

The genetic variation in the new generation is determined by the recombination rate during the segregation of haploid gametes of each parent. Segregation starts with a randomly selected copy of a chromosome (i.e. one of the two dimensions of the individual array defining its genotypes). The recombination rate may either be a fixed value between neighbouring loci or may be specified as a per-base-pair rate with crossover points distributed on the physical chromosome via a Poisson process. In the first case, when a recombination map is available, a vector of $n_t - 1$ elements has to be supplied with the recombination rate (ρ) between each pair of neighbouring loci. Each time a gamete is segregated in the simulation, a recombination occurs (1) or not (0) at a position between loci by comparing the

recombination rate (ρ) between the loci to a uniformly distributed random variable (U) on the unit interval. A separate U is simulated for each pair of neighbouring loci, and $U < \rho$ means a recombination occurs. There is no recombination between homologous chromosomes when $\rho = 0$, meaning both loci are completely linked (e.g. within an inversion or situated close to centromeres), while with a value of $\rho = 0.5$, the recombination rate is completely random (i.e. both loci are very distant on the same chromosome or are located on different chromosomes). A value of $\rho < 0.5$ means the loci are physically linked. In the second case, the physical positions of the markers on the chromosome are specified, and then crossovers are simulated as a Poisson process along the physical distance of the chromosome. The simulator allows recombination to be a homogeneous Poisson process, with rate specified by the expected per-base-pair recombination rate (e.g. 1e-08, corresponding to 1 cM per megabase. This latter method may be preferred when trying to fit a large dataset of genomic information with an unknown recombination rate between neighbouring loci (e.g. single-nucleotide polymorphisms). Because we are primarily interested in the effects of various levels of linkage disequilibrium in the formation of islands of genomic divergence, we present results using the first approach, but an example with the second method is also shown in supporting information (Appendix S1).

The offspring represent individuals with the potential to reproduce in the next generation. We assume an equal sex ratio at birth and assign the sexes to offspring by sampling with replacement, with an equal probability of assignment to each sex. A weighted probability could be supplied when unequal sex ratios are considered in the simulations.

The final step is the migration of offspring to neighbouring populations. The probability of migration of each individual is obtained from a uniform distribution, so each individual has the same expected probability of migration. The final number of individuals of population i dispersing to population j is defined by the migration rate m_{ij} . Individuals of i having migration probability smaller than m_{ij} move to population j . A value of $m_{ij} = 0$ means no migration and thus no gene flow between populations, while a value of 0.5 means random migration (and hence random reproduction) between them.

The final number of individuals in population i at time $t + 1$ can be estimated as the sum of fitness values of all females present in the population at time t (N_t^f) and the number of migrants from population j (males and females, $N_t^f m_{ji}$):

$$N_{t+1}^i = \sum_{l=1}^{N_t^f} \omega_{l,t} + N_t^f m_{ji} \quad (3)$$

The model is implemented in a R package called ‘glads’ (R Development Core Team, 2017), with some functions written in C++ and integrated to R by using the RCPP package (Eddelbuettel et al., 2011). The package is available on GitHub (<https://github.com/eriqua/ndeglads>), and is easily modifiable for further applications simulating data based on allelic information, SNPs or DNA sequences. An example is available in Appendix S1 and a user manual is available in

Appendix S2. Below we describe a number of simulations with different parameterizations to explore how the signatures of genomic divergence are generated by various processes.

2.3 | Initialization

We start by simulating how two populations of diploid individuals with equal intra-genomic variation at the beginning of the simulations diverge. The migration rate between the two populations was varied across different simulations to explore divergence without gene flow (i.e. $m_{ij} = 0$) and divergence with gene flow (i.e. $m_{ij} \neq 0$). The demographic and genetic parameter values were chosen to describe two fitness functions that can either have identical or contrasting phenotypic optima, but with a similar number of individuals in each population during the simulation (Figure 1c,d, Table 1).

The mean population sizes of the two populations were always around 400 (Figure 1d). This is also the initial number of individuals at the beginning of the simulations. The genomic structure of individuals was characterized by genotypes across 300 loci ($n_L = 300$), that were either strongly linked ($\rho = 0.0001$) or completely unlinked ($\rho = 0.5$). This range of linkage allows us to explore the dynamic of genomic landscapes across more contiguous or distantly related loci. Because previous simulations on the formation of genomic islands of divergence were restricted to bi-allelic loci (e.g. Feder, Gejji, et al., 2012; Flaxman et al., 2013), we ran simulations with a higher number of alleles to allow for greater allelic variation (Table 1). The genomic identities of individuals were randomly assigned at the beginning of

each simulation by setting the seed of the random number generator in R. This allows us to explore the same founder population under different evolutionary scenarios.

Fifty of the 300 simulated loci were chosen to have a non-zero and additive variation in allelic contributions to the phenotypic trait value. This fraction of loci is potentially subject to selection. By operating on the phenotype, selection changes the distribution of genotypes at each locus that contributes to the phenotype in the simulation. The remaining 250 loci not influencing phenotypes are neutral and were used to examine the effect of drift and linkage on the appearance of genomic islands. This allowed us to account for both adaptive and neutral evolution simultaneously. The phenotypes were always computed from 50 additive loci ($n_a = 50$), 10 of which were always linked. These 50 additive loci contributed to the phenotypic trait values of individuals, with the additive value of each allele ranging between 0 and 1. The sum of additive values was then used to compute the phenotype, and then the fitness score, for each individual. However, further studies may expand this procedure to include any required genotype-phenotype map. In summary, we have four classes of genes: (a) unlinked genes contributing to the phenotype; (b) linked genes where both loci contribute to the phenotype; (c) unlinked genes that do not contribute to the phenotype; and (d) linked genes that do not contribute to the phenotype. The first two categories of genes are under selection, and the last two are not.

The number of mating pairs depends on the number of breeding females. Female reproductive success was determined first, before male mates were assigned to father each offspring. In this simulation, we assumed random mating, although other mating patterns are possible (e.g. Schindler et al., 2015). Offspring sex was assigned randomly, with probability .5 (Table 1).

TABLE 1 List of parameters of the model with default values

Symbol	Definition	Value ^a
N_i	Number of individuals in population i	Initial size: $N_1 = N_2 = 400$
n_L	Number of loci	300
n_a	Number of additive loci	50
A_p	Number of alleles at locus p	20
B_v	Breeding values of additive loci	[0, 1]
m_{ij}	Migration rate of population i to population j	[0, 0.5]
ρ_{pq}	Recombination rate between loci p and q	[0, 0.5] ^b
b_0	Maximum generated offspring	$P_1 = P_2 = 6$
b_1	Phenotypic optima	$P_1 = 0.25$; $P_2 = 0.75$
b_2	Variance of the fitness curve	$P_1 = P_2 = 0.5$
b_3	Density-dependent demographic effect	$P_1 = 0.01$; $P_2 = 0.005$
σ_{env}	Stochastic environmental variant	0.01
σ_{dem}	Stochastic demographic variant	1

^a P_1 and P_2 refer to the value for population 1 and 2, respectively.

^bIt may also represent a single average value for the whole chromosome (see methods).

2.4 | Genomic divergence

We measured pairwise F_{ST} at each locus to estimate genetic differentiation between populations. F_{ST} is a widely used measure of differentiation across divergent genomes in studies of genomic islands of divergence (e.g. Ellegren et al., 2012; Kusakabe et al., 2017). We computed F_{ST} at each simulated locus using the R package 'pegas' (Paradis, 2010). Genetic differentiation averaged across multiple loci was calculated using the approach of Nei (1973), as implemented in the R package 'mmmod' (Winter, 2012).

We identified significant genomic islands of divergence using the approach of Marques et al. (2016), modified from the original method of Hofer, Foll, and Excoffier (2012). Marques et al. (2016) used Hidden Markov Models (HMM) to identify three hidden states: low, intermediate and high genomic differentiation. The regions of intermediate states are considered 'genomic background' that arise randomly under a hierarchical island model. The regions of extremely high differentiation are considered to be genomic islands of divergence. The Baum-Welch algorithm is applied from 1,000 different randomly chosen initial values for the parameters, and the final parameter estimates are taken to be those from the run producing the highest likelihood. Subsequently, using those estimates, the Viterbi

algorithm is used to assign the most likely sequence of states across the genome (Hofer et al., 2012). The significance of the clustering was assessed by randomly permuting islands 10,000 times across the whole simulated genomic area.

2.5 | Simulations

We conducted simulation experiments using a wide suite of parameter values (see Table S1). These were designed to examine how various scenarios of linkage between loci, drift, selection and time since divergence influence the formation of genomic islands of divergence in both the presence and the absence of gene flow. Parameter values are presented in Table 1 and the supplementary information provides more details for the choice of each parameter set (Table S1). The first simulation characterizes the effect of the founder population on the resulting genetic divergence (F_{ST}) at an early stage of independent evolution (100 generations, $m_{ij} = 0$). We then explored in more detail, the effect of physical linkage, gene flow and time since initial divergence. Drift is included in all simulations through the group of genes that are not involved with the phenotype trait value, and through the random selection of gametes at birth. We assigned a name to each group of simulations and will briefly describe their structure.

2.5.1 | Random sampling of founders and concordant selection

These simulations were designed to examine how random sampling of the founder population influenced genomic divergence. Both populations were exposed to neutral evolution and concordant selection, with identical phenotypic optima (equal to population 1, Table 1). We ran 50 simulations with different random initial founder genotypes. Founder genotypes were determined by sampling a uniform distribution with replacement.

2.5.2 | Random sampling of founders and divergent selection

We considered the same 50 founder populations as before but added divergent selection. The selective pressures generating evolutionary divergence between populations were generated by their respective fitness functions. The amount of difference between phenotypic optima measures the strength of 'divergent selection' (Figure 1c, Table 1).

2.5.3 | Levels of heterozygosity in the founder population

Our third set of simulations was designed to explore how variable levels of heterozygosity among founder populations influenced the

variance of genomic divergence at the end of the simulation. The level of heterozygosity in the founder population was varied by sampling alleles at a locus with variable frequencies of replacement (see Table S1 for more information). The variable frequency of replacement represented the weighted probability of a random sampling with replacement among the 20 polymorphisms available for each locus. This ranged from 1 (an equal probability of allelic sampling and more heterozygous) to 100 (an unequal probability of allelic sampling and more homozygous). As this value becomes higher, it increases the probability for individuals to carry the same allele on both copies of their genes. Because linked loci are hypothesized to be more likely to be involved in the formation of genomic islands and we are analysing this factor separately, we excluded these loci in the final estimation of variability in genetic differentiation (an analysis without the exclusion of linked loci is included in supporting information).

2.5.4 | Physical genomic linkage

Having characterized how initial conditions might influence results, we next examined the effect of physical linkage on the formation of islands of genomic divergence. We ran 100 simulations with equal founder populations, but changed the recombination rate between linked loci, ranging from nearly complete linkage ($\rho = 0.0001$) to no linked loci ($\rho = 0.5$).

2.5.5 | Strong selection at a single, unlinked locus

We next explored the effect of strong selection on unlinked genes of large phenotypic effect. Fifty additive loci contributed to phenotypic expression, but one locus contributed 10 times more than the others. This means that rather than having multiple linked loci affecting the trait there is, in particular, one locus of very large effect that is unlinked to the other loci that influence phenotype. We considered the same founder population as in 4 (genomic linkage).

2.5.6 | Time since divergence with and without gene flow

To explore how time since divergence influenced the formation of genomic islands, we ran simulations with the same 50 founder populations of our previous analysis 'random sampling of founders and divergent selection', but for different lengths of time: 100, 500, 1,000 and 2,000 generations. We repeated these simulations in the presence ($m = 0.01$) and the absence ($m = 0$) of gene flow.

2.5.7 | Physical linkage and gene flow

Finally, we explored how linkage and gene flow combined to influence the formation of genomic islands. We simulated various

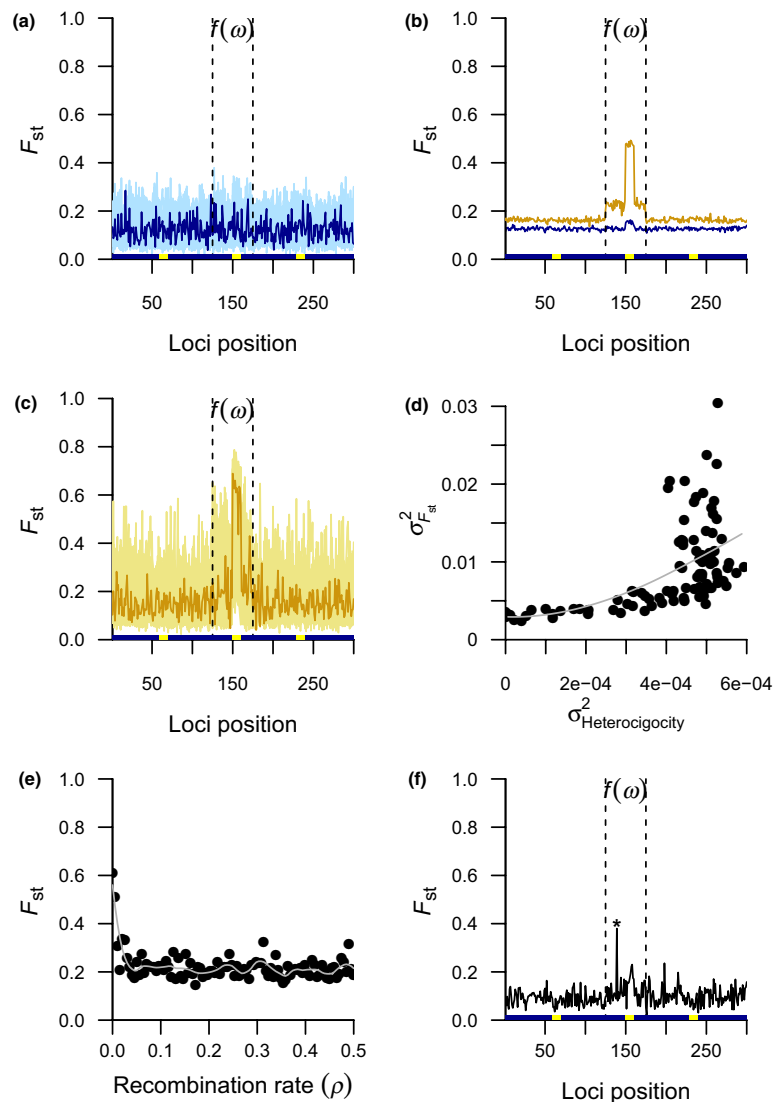


FIGURE 2 Simulations of genetic differentiation between two simulated populations without gene flow. (a) Random sampling of founders and concordant selection. The sky-blue lines represent the resulting genetic differentiation (F_{ST}) on 50 comparisons with different random sampling of founders. The dark blue line illustrates the resulting values for a single founder population randomly chosen. The yellow squares on the horizontal axis represent physically linked loci ($\rho = 0.0001$) and the blue rectangles independent ones ($\rho = 0.5$). The dotted vertical lines delimit loci participating in the computation of phenotypes. (b) Mean divergence by loci on the 50 founder populations. The blue and the orange lines represent divergent selection and concordant selection, respectively. (c) Random sampling of founders and divergent selection. The yellow lines and the orange line represent equal founder populations as in Figure 2a, but adding divergent selection to the analysis. (d) Levels of heterozygosity in the founder population. Influence of the heterozygosity variance at the beginning of the simulations on the variance of F_{ST} at the end of the simulations. (e) Genomic linkage. Effect of the strength of linkage on the formation of a genomic island. F_{ST} values are averaged over the 10 linked loci influencing the computation of phenotypes and using the same starting conditions as the dark yellow line in Figure 2c,f. Strong selection at a single, unlinked locus. A single independent locus with a stronger additive effect on the computation of phenotypes (*). All data are presented after 100 generations of independent evolution

rates of migration and recombination, using the same 50 founder populations as in 4 (physical genomic linkage). We recorded average F_{ST} at 10 linked loci affecting the expression of phenotypes (positions 150 to 159) and 10 independent loci not related to

phenotype expression (positions 90 to 99). This allowed us to determine the magnitude of differentiation between regions of linked divergent selection and the genomic background of neutral evolution.

3 | RESULTS

3.1 | Random sampling of founders and concordant selection

Our first simulations explored the effect of initial conditions on divergence and the formation of genomic islands under equal selective pressures (i.e. concordant selection). The sky-blue lines in Figure 2a show the resulting F_{ST} values for all 50 random initial populations. When considering the average F_{ST} values by loci across the 50 pairwise comparisons, linked genes that contribute to the phenotype have a slightly higher F_{ST} than unlinked genes (dark blue line, Figure 2b). However, independent of the type of loci (i.e. under selection or neutral), all positions have almost the same probability of becoming an area of higher or lower genomic divergence.

Different genotypes coding for identical phenotypes influence the dynamics of genetic differentiation with time (Figure 2a). F_{ST} values across the whole genome ranged between 0 and about 0.3. Interactions between the genotype-phenotype map and the phenotype-demographic map influence the development of genetic differentiation between populations. The dark blue line in Figure 2a represents an example of a single founder population with a typical, heterogeneous genomic landscape that has formed over 100 generations. The Hidden Markov Model identified genomic islands in various, but not all of the 50 founder populations (Figure S2). These areas of higher genomic divergence appear seemingly randomly across the whole genome. Among all simulations that generated genomic islands (84% of simulations), 14% did not generate islands on selected linked loci affecting phenotypes. The variance in F_{ST} we observed within and across loci reveal that the genotype-phenotype map of the founder populations influences the patterns of genomic divergence.

In our simulations, populations differentiate after 100 generations, even when both populations have equal phenotypic optima under concordant selection (Figure 2a). This is because there are many ways to generate the same additive phenotypic trait value. The time since initial divergence increases the likelihood of generating these different outcomes, therefore, with enough generations of isolated reproduction, populations can still be highly differentiated even when they are exposed to the same fitness peak.

3.2 | Random sampling of founders and divergent selection

Genomic islands of divergence are, on average, more likely to be observed for genes that contribute to a phenotypic trait that experiences divergent selection across the two populations (orange line, Figure 2b). The range of change in F_{ST} values was also higher under divergent selection than under concordant selection (Figure 2c). The values of F_{ST} across loci ranged between 0 and about 0.8, and this seemed to affect the average F_{ST} across non-selected loci (compare orange and blue line, Figure 2b). The same single founder population illustrated in Figure 2a,c (blue and orange lines) provides an example

of where genomic islands usually form at some linked loci experiencing divergent selection. The Hidden Markov Model identified genomic islands in various, but not all of the 50 founder populations (Figure S3). A single high island of genomic divergence usually emerges at selected loci, but some islands appeared in neutral loci and other populations did not generate any pattern of genomic islands (Figure S3). Among all simulations that generated genomic islands (84% of simulations), only 2% did not generate islands on selected linked loci affecting phenotypes. In general, due to the large variation between simulations, divergent selection did not necessarily generate islands of genomic divergence at loci under selection (yellow lines, Figure 2c and Figure S3).

3.3 | Levels of heterozygosity in the founder population

As the variance of heterozygosity in the founder population increases, so too does the variance in F_{ST} across the genome after 100 generations of independent evolution (Figure 2d). This variance reflects an increase in F_{ST} of loci not under selection. F_{ST} at these loci can be as large as for genes under direct selection. A similar pattern is observed when physically linked selected loci are included in the analysis (Figure S4). This result reveals that the appearance of a pattern of genomic islands at early stages of differentiation can be caused by genetic variation at specific loci in the founder populations.

3.4 | Physical genomic linkage

We ran simulations with the same founder population and parameter values used to generate the dark yellow line in Figure 2c that resulted in an island of genomic divergence, except now we varied the recombination rate (ρ) among linked selected genes. The average F_{ST} of those linked genes was much higher with nearly complete levels of linkage ($\rho < 0.02$), but tended to the average value of neutral genes when the recombination rate was higher, even when they were still physically linked (compare Figure 2e,b). These results show that strong linkage may facilitate the appearance of genomic islands when those genes are affected by divergent selection, even in the absence of gene flow. Extreme physical linkage, therefore, tends to increase the F_{ST} value of genes under selection. The combined effect of divergent selection and linkage is consequently important for the development of genomic islands of larger sizes.

3.5 | Strong selection at a single, unlinked locus

The previous simulations revealed that divergent selection on linked selected loci could sometime result in islands of genomic divergence. We, therefore, next considered a founder population in which an island of genomic divergence formed (Figure 2c), yet altered the genotype-phenotype map such that one independent locus ($\rho = 0.5$) contributed disproportionately to the phenotypic value. This locus

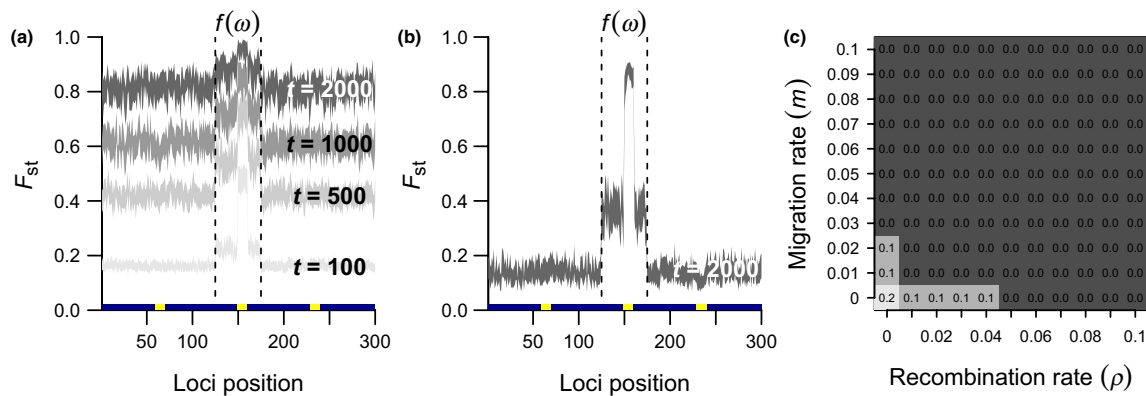


FIGURE 3 Time since divergence and gene flow. (a) Divergence without gene flow. The yellow squares on the horizontal axis represent physically linked loci ($\rho = 0.0001$) and the blue rectangles unlinked ones ($\rho = 0.5$). The dotted vertical lines delimit the loci participating in the computation of phenotypes. The coloured areas represent a confidence interval at 95% of F_{ST} values, estimated over 50 simulations with randomly assigned genetic identity of individuals at the beginning of the divergence ($t =$ generations). (b) Divergence with gene flow. This is similar to the previous figure, but allows for gene flow between populations ($m = 0.01$). (c) Linkage and gene flow. Combined effect of migration rate and recombination rate on the magnitude of a genomic island. The numbers inside the squares represent the difference between mean F_{ST} estimated at the 10 linked loci influencing the computation of phenotypes (i.e. genomic island, positions 150 to 159) and 10 loci not related to fitness and independent (i.e. genomic background, positions 90 to 99). These numbers represent the average difference over the same starting conditions used to estimate the confidence interval of Figure 3a. The data in this last figure are presented after 100 generations of divergent evolution

resulted in a level of F_{ST} of more than twice that observed elsewhere in the genome, including on linked selected genes (Figure 2f). This result reveals that patterns of genomic divergence are not necessarily determined by strongly linked genes of similar effect but can also emerge when one gene of large effect is linked to other markers.

3.6 | Time since divergence with and without gene flow

We extended our 50 previous simulations of 'random sampling of founders and divergent selection' by running them for longer (100 to 2,000 generations). Without gene flow, the trend of higher genomic differentiation in selected loci, particularly in genes that are linked, is more evident at early stages of divergence (100 generations, Figure 3a). The length of time that independent evolution has to act influences genome-wide divergence, masking signals of genomic islands that arise from single or linked loci. The pattern of heterogeneous genomic differentiation is, therefore, less evident and tends to disappear as the numbers of generations since divergence without gene flow increases (2,000 generations, Figure 3a).

3.7 | Physical linkage and gene flow

All previous simulations were performed in the absence of gene flow. Gene flow increases the number of generations over which genomic islands of divergence are apparent. The genomic islands of higher F_{ST} are still present after 2,000 generations, when performing the same simulations as in Figure 3a, but allowing a level of migration between populations ($m = 0.01$, Figure 3b). However, as the level

of gene flow increases, the prevalence of islands of genomic divergence decreases (see the zero values in Figure 3c).

We performed the simulations of divergent selection using the same 50 founder populations (Figure 2a,c), while varying both the migration rate between populations and the recombination rate among linked loci. The numbers inside the grey squares in Figure 3c indicate the magnitude of difference between independent neutral loci (i.e. genomic background) and linked selected loci (i.e. genomic islands). The largest differences were present under conditions with extreme physical linkage ($\rho < 0.04$) and a low migration rate ($m < 0.02$), and ranged from 0.1 to 0.2. Those differences are negligible under concordant selection (Figure S2). Overall, these results show that gene flow may influence the persistence of genomic islands but is not the only factor determining their emergence.

4 | DISCUSSION

4.1 | Genomic islands of divergence

The application of our quantitative framework to model the generation of genomic islands of divergence has revealed that while there are several routes that can result in genomic islands, the conditions required to generate islands are relatively narrow, and importantly, there is no single set of circumstances that guarantee their emergence. For instance, the formation of large genomic islands requires a combination of divergent selection and strong linkage, regardless of the gene flow scenario. In contrast, smaller genomic islands can form via drift in the early stages of divergence in particular. However, in both cases, genomic islands can also fail

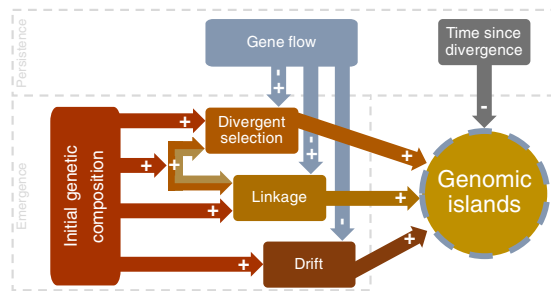


FIGURE 4 Factors influencing the patterns of genomic islands of divergence. Genomic islands may emerge under the influence of linkage, divergent selection, an interaction between these two factors, or drift depending upon the initial genetic composition of the starting populations (positive effects). Gene flow and time since divergence have an effect on the persistence of islands once formed. Gene flow has an indirect effect by interacting with factors influencing the emergence of this pattern. At early stages of divergence, gene flow can lengthen the time that genomic islands are visible (positive effect), but too high a level of gene flow can erase genomic island patterns (negative effect). The time since divergence has a negative effect on genomic islands, which are more visible under earlier rather than later stages of genomic differentiation

to form even when these conditions are met, because outcomes are highly dependent on the initial genetic composition of the diverging populations. Our simulations suggest that genomic islands are most obvious during the early stages of divergence, and tend to disappear with the accumulation of genome-wide divergence over time. If present, gene flow can slow this loss up to a point; however, including gene flow is not necessary to explain genomic island formation. The importance of evolutionary processes that were modelled (divergent selection, drift, gene flow), along with influencing factors of initial genetic composition, degree of genetic linkage, time since divergence and their interactions are summarized in Figure 4. In addition to these factors, previous simulation studies have also highlighted the challenge of identifying genomic islands due to sampling design (Lotterhos & Whitlock, 2014) or the statistic used to identify divergence (Matthey-Doret & Whitlock, 2018). Our modelling approach provides a nuanced understanding of how genomic islands arise, yet it is not possible to confidently interpret a particular process from genomic data on its own (Feder et al., 2013; Jiggins & Martin, 2017; Nosil, 2008; Nosil, Feder, Flaxman, & Gompert, 2017; Seehausen et al., 2014; Turner et al., 2005). This is particularly true in the case of divergence in the presence of gene flow and divergent selection, which is not a necessary condition for the emergence of genomic islands, as concluded by initial studies in the field (Nosil, 2008). Recent simulation studies have also highlighted the importance of multiple processes, in addition to gene flow and selection, such as drift and linkage, in generating genomic islands (Yeaman, Aeschbacher, & Bürger, 2016). While this complexity is already known, our contribution is to show the relationship among those factors, highlighting that its presence does not guarantee the emergence of this pattern, because the time since divergence and

the genetic composition at the beginning of the divergence process are also important.

4.2 | Initial genetic composition and drift influence the generation of genomic islands

Our model revealed two ways that random effects can influence the formation of genomic islands of divergence. First, a previously unappreciated but critical factor influencing their generation was the genetic composition of the initial populations. This was evident from comparison of simulations with identical parameter values but different starting populations, that is, different genetic compositions. In some simulations, genomic islands were generated and in others they failed to form. Furthermore, the starting values influenced island appearance even in regions of the genome that were not influenced by selection, linkage or gene flow – all of which are thought to be important in genomic island formation as discussed below (Feder et al., 2013; Flaxman et al., 2013). Second, drift alone could generate a pattern of numerous islands of small size particularly in the early stages of divergence. Lineage sorting is a slow process when explained only by drift (Cruickshank & Hahn, 2014), in which case ancestral polymorphisms remain shared between recently divergent taxa, reducing genetic diversity and potentially inducing a pattern of genomic islands (Ma et al., 2018). Recent studies exploring the distribution of genomic islands have also advanced the idea that islands arise from neutral processes without a major contribution from divergent selection (Campagna et al., 2015; Wang, Wan, Lim, & Yue, 2016) and the results of our model identify the scenarios where this is particularly likely to be the case. Some studies document a small number of very prominent islands (e.g. Turner et al., 2005; Wang et al., 2016); however, finding multiple islands of low relief is also common (e.g. Ellegren et al., 2012; Feulner et al., 2015; Rugg et al., 2014; Soria-Carrasco et al., 2014). Furthermore, comparisons often involve recently diverged populations (e.g. Nadeau et al., 2012; Rugg et al., 2014; Via, 2012), with some divergence time-scales as short as 100 generations (e.g. Marques et al., 2016). Our modelling suggests that these patterns and types of comparisons could be explained without recourse to explanations that invoke selection, but initial diversity of the founder populations and drift.

Another scenario that may be particularly prone to stochastic effects is where one of the diverging populations experiences a geographic expansion. Klopstein, Currat, and Excoffier (2006) suggest that the effect of drift is stronger in expanding populations because of 'allelic surfing', where alleles that happen to be at the expansion front may incidentally increase in frequency (see Excoffier, Quilodrán, & Currat, 2014; Hofer et al., 2012). This, in turn, impacts genetic composition and variation (i.e. Arenas, François, Currat, Ray, & Excoffier, 2012; Branco & Arenas, 2018; Branco et al., 2018), and if occurring very early during the divergence process, the combination of early differences in genetic composition and drift could generate highly stochastic patterns of islands of divergence.

We have shown that populations diverge through time even under the equal selective pressures of concordant selection. Indeed, highly polygenic traits may also express divergence based on which alleles of the genes under selection in the founder population end up increasing in frequency. Selection will tend to create shorter coalescence times around those selected loci, meaning a lower effective population size and hence greater drift (Nordborg, 1997). However, it should be noted that the specification of the additive genotype–phenotype map we use means there are multiple genotypes that will produce the same phenotypic value. This explains why populations with identical selection regimes can diverge, with some developing islands of genomic divergence, and others not. The nature of our genotype–phenotype map in the model could also underpin the influence of initial genetic composition on our results. Future work will explore whether the same conclusions hold with genotype–phenotype maps that do not assume small additive contributions to the phenotype from genotypes at multiple loci. However, the genotype–phenotype map we use is widely assumed in quantitative genetics, and given that many traits are highly polygenic, is an appropriate initial map to assume in simulations.

4.3 | Linkage and divergent selection generate islands of divergence independent of gene flow

Extreme linkage in combination with divergent selection was necessary, although not sufficient, for the development of the most prominent genomic islands, regardless of whether gene flow occurred or not. These findings are consistent with observations of prominent genomic islands between populations presumed to be under strong divergent selection, and not connected by gene flow (Burri et al., 2015; Zhang et al., 2017). The occurrence of candidate genes, hypothesized to be under natural selection, associated with genomic islands of divergence also supports the role of selection (Kusakabe et al., 2017; Sousa & Hey, 2013). However, empirical results also provide examples where SNPs under selection are not associated with islands of divergence (e.g. Han et al., 2017; Riesch et al., 2017; Ruegg et al., 2014).

The importance of linkage in the appearance of genomic islands has been highlighted in both theoretical and empirical studies (Feder & Nosil, 2010; Flaxman, Wacholder, Feder, & Nosil, 2014; Renaut et al., 2013), with extreme linkage, such as that found near centromeres or within genomic inversions, often associated with the most prominent genomic islands of divergence (Ellegren et al., 2012; Feder & Nosil, 2009; Kawakami et al., 2014). Selection acting in these zones of low rates of recombination (i.e. linked selection) reduces the effective population size of these genomic regions to a greater degree than in the rest of the genomes, generating genomic islands (Feder & Nosil, 2009; Turner & Hahn, 2010).

We did not explore the effect of linkage on deleterious variants (i.e. background selection) in our simulations. However, previous studies have shown that selection on both adaptive and deleterious mutations has a similar effect of reducing within population diversity

(Comeron, 2014; Nordborg, Charlesworth, & Charlesworth, 1996; Slatkin & Wiehe, 1998), and influencing the formation of genomic islands (Cruickshank & Hahn, 2014).

4.4 | The effect of gene flow on the generation and persistence of genomic islands

The idea that genomic islands of divergence were generated primarily by antagonistic effects of divergent selection and gene flow was a favoured explanation until recently (Feder, Egan, & Nosil, 2012; Nosil, 2008; Turner et al., 2005). According to this mechanism, genomic islands form around selected loci involved with the divergence process, and genes physically linked to them, while adjacent neutral or weakly selected regions are homogenized by gene flow (Flaxman et al., 2013; Kawakami et al., 2014; Turner & Hahn, 2010). Our modelling provides further support that the presence of gene flow is not an essential condition; however, an additional insight is that when gene flow does occur, it can lengthen the time that genomic islands are visible. Verbal models of changing genomic landscapes over time predicted that genomic islands would disappear with the accumulation of genome-wide divergence over time (Nosil, 2012; Nosil & Feder, 2012; Wu & Ting, 2004).

Empirical support that this is indeed the case is provided from studies where genomic islands are more frequently documented in recently diverged versus distantly related taxa (e.g. Marques et al., 2016; Nadeau et al., 2012; Via, 2012). Our results reveal that this dynamic can be moderated by gene flow, where a limited amount of gene flow serves to slow down the swamping of genomic islands over time, whereas large amounts of gene flow tend to erase the pattern of islands of divergence altogether.

4.5 | Advantage and limitations of our framework

Our framework is designed to identify how various processes can generate patterns of divergence between populations. Discriminating whether divergence has occurred in the presence or absence of gene flow can be challenging (Feder et al., 2013), and our simulation tool provides a useful way to discriminate between the two scenarios. There is one other package available in R to simulate the genetics and demography of populations, but it is restricted to one locus (Andrello & Manel, 2015). However, in other programming languages, there is a range of individual-based tools simulating the evolution of large genetic datasets forward in time (e.g. Currat et al., 2015; Haller & Messer, 2019; Peng & Kimmel, 2005), but none of these allow simulation via a fitness function as presented here, which is important to explore the range of ecological and evolutionary dynamics associated with the divergence between populations. We are confident our approach works, as it gave results consistent with SimuPOP, a tool widely used in ecology and evolution (Figure S6) (Peng & Amos, 2008; Peng & Kimmel, 2005).

The main limitation of our simulation approach lies in the amount of genetic and ecological information required to parameterize it for a field system. Empirical information is needed to identify fitness functions, specify the genotype–phenotype map or estimate rates of migration. Model organisms with short generation time that have been extensively studied in the past represent a source of data for potential application (e.g. Mackay, 2014). Adapting parameter values from sister species for which information is available may be of use for non-model organisms. This limitation is expected to become less important in the future as the rapid accumulation of freely available ecological and genomic datasets grow (Ellegren, 2014; Jones, Clutton-Brock, Coulson, & Godfray, 2008). However, in the absence of sufficient information to parameterize a fitness function, this framework is still useful to elucidate neutral evolution, which can be simulated in the framework by replacing the fitness function with a random distribution (e.g. Poisson) in order to generate the next generation of offspring (see example in Appendix S1). While mutation is not explored here, as we were mostly interested in divergence at relatively early stages of evolution, its incorporation would not likely change any of the patterns observed in this study (see example in Appendix S1).

The elapsed time of simulations may be a potential limitation for studies simulating a huge number of individuals and genetic information. The time depends on the number of individuals, the number of simulated generations and the amount of genetic information in the analysis. For instance, simulation of a single population of 100 individuals, with 100 SNPs, over 100 generations take approximately 0.2 s, while the same analysis takes 1 s for 1,000 SNPs and approximately 10 s when simulating 1,000 SNPs over 1,000 generations (see Figure S7). However, the computation time may be largely reduced by parallelizing the number of required simulations by the number of available cores in the computer. For instance, a single simulation of divergence between two populations, composed of 100 individuals with 1,000 SNPs, as the one presented in Figure S6 (supporting information), takes around 2.58 s, while eight simulations using 8 cores parallelized with the library 'parallel' of R takes 5.9 s (0.7 s by simulation). All analyses of computation time were performed by using the graphical version of R, in a MAC OS X 10.13 64 bits with 3.1GHZ CPU.

5 | CONCLUSIONS

We have developed a quantitative framework to explore the dynamics of genomic landscapes and identify how various processes can generate patterns of divergence between populations. Our work builds on previous insights (Akerman & Bürger, 2014; Charlesworth et al., 1997; Feder & Nosil, 2009; Flaxman et al., 2013; Sedghifar et al., 2016). We have been able to demonstrate that the formation of genomic islands of divergence is not a deterministic phenomenon, but that they can arise via a number of routes. Those routes are not mutually exclusive and further studies may explore how the interaction of multiple processes at once may influence the observation of a pattern of genomic islands. We urge extreme caution in inferring

a particular ecological or evolutionary process when a particular genomic pattern is observed. Narrowing down the potential cause of a particular signature will likely require ancillary information beyond the genome sequence or modelling exercises that examine the processes that have the potential to generate such a pattern. The methods described here provide a modelling framework which helps to depict such signatures of past evolution, as well as potential routes of future evolution for any divergent taxa.

ACKNOWLEDGEMENTS

This study was financed by a fellowship from the Swiss National Science Foundation (n°P2GEP3_168973 and P400PB_183930 to CSQ). We thank Laurent Excoffier and David Marques for their support on the Hidden Markov Model method. We also thank Jose Manuel Nunes for his suggestions regarding the R package code. Finally, we are also grateful to two anonymous reviewers for their comments and suggestions on an earlier version of this manuscript.

AUTHORS' CONTRIBUTIONS

S.M.C., T.C., E.C.A., K.R., A.T.S.-P. and C.S.Q. conceived the original idea and designed the simulations. C.S.Q., E.C.A. and T.C. coded the R package. C.S.Q. performed the simulations. C.S.Q., T.C. and S.M.C. wrote the first draft of the manuscript and all authors provided comments and edits on manuscript versions.

DATA AVAILABILITY STATEMENT

All the data necessary to repeat the simulations are described in Table 1 and can be reproduced by using the R package 'glads'. Appendix S1 and Appendix S2 provide examples and descriptions of the R package, and are deposited in Dryad Digital Repository <https://doi.org/10.5061/dryad.12jm63xt0>. The R package code is available on GitHub and deposited in Zenodo: <https://doi.org/10.5281/zenodo.3516224> (Anderson, Quilodran, & Coulson, 2019).

ORCID

Claudio S. Quilodrán  <https://orcid.org/0000-0001-7197-9154>
Kristen Ruegg  <https://orcid.org/0000-0001-5579-941X>

REFERENCES

- Akerman, A., & Bürger, R. (2014). The consequences of gene flow for local adaptation and differentiation: A two-locus two-deme model. *Journal of Mathematical Biology*, 68, 1135–1198. <https://doi.org/10.1007/s00285-013-0660-z>
- Anderson, E. C., Quilodran, C. S., & Coulson, T. (2019). eriqande/glads: Companion to MIEE article, Quilodran et al. (Version v1.0). *Zenodo*. <https://doi.org/10.5281/zenodo.3516224>
- Andrello, M., & Manel, S. (2015). MetaPopGen: An R package to simulate population genetics in large size metapopulations. *Molecular Ecology Resources*, 15, 1153–1162.
- Arenas, M., François, O., Currat, M., Ray, N., & Excoffier, L. (2012). Influence of admixture and paleolithic range contractions on current European diversity gradients. *Molecular Biology and Evolution*, 30, 57–61. <https://doi.org/10.1093/molbev/mss203>
- Bradshaw, A. D. (1965). Evolutionary significance of phenotypic plasticity in plants. *Advances in Genetics*, 13, 115–155.

- Branco, C., & Arenas, M. (2018). Selecting among alternative scenarios of human evolution by simulated genetic gradients. *Genes*, 9, 506. <https://doi.org/10.3390/genes9100506>
- Branco, C., Velasco, M., Benguigui, M., Currat, M., Ray, N., & Arenas, M. (2018). Consequences of diverse evolutionary processes on american genetic gradients of modern humans. *Heredity*, 121, 548. <https://doi.org/10.1038/s41437-018-0122-x>
- Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., ... Garamszegi, L. Z. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula flycatchers*. *Genome Research*, 25, 1656–1665.
- Campagna, L., Gronau, I., Silveira, L. F., Siepel, A., & Lovette, I. J. (2015). Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation. *Molecular Ecology*, 24, 4238–4251. <https://doi.org/10.1111/mec.13314>
- Chanderbali, A. S., Berger, B. A., Howarth, D. G., Soltis, P. S., & Soltis, D. E. (2016). Evolving ideas on the origin and evolution of flowers: New perspectives in the genomic era. *Genetics*, 202, 1255–1265. <https://doi.org/10.1534/genetics.115.182964>
- Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*, 70, 155–174.
- Comeron, J. M. (2014). Background selection as baseline for nucleotide variation across the *Drosophila* genome. *PLoS Genetics*, 10, e1004434. <https://doi.org/10.1371/journal.pgen.1004434>
- Coulson, T., Kendall, B. E., Barthold, J., Plard, F., Schindler, S., Ozgul, A., & Gaillard, J.-M. (2017). Modeling adaptive and nonadaptive responses of populations to environmental change. *The American Naturalist*, 190, 313–336. <https://doi.org/10.1086/692542>
- Coulson, T., Potter, T., & Felmy, A. (2019). Fitness functions, genetic and non-genetic inheritance, and why ecological dynamics and evolution are inevitably linked. *bioRxiv*, 762658.
- Cruikshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, 23, 3133–3157. <https://doi.org/10.1111/mec.12796>
- Currat, M., Gerbault, P., Di, D., Nunes, J. M., & Sanchez-Mazas, A. (2015). Forward-in-time, spatially explicit modeling software to simulate genetic lineages under selection. *Evolutionary Bioinformatics*, 11, 27–30.
- Darwin, C., & Wallace, A. (1858). On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Zoological Journal of the Linnean Society*, 3, 45–62. <https://doi.org/10.1111/j.1096-3642.1858.tb02500.x>
- Edelbuettel, D., François, R., Allaire, J., Ushey, K., Kou, Q., Russel, N., ... Bates, D. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40, 1–18.
- Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, 29, 51–63. <https://doi.org/10.1016/j.tree.2013.09.008>
- Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., ... Wolf, J. B. W. (2012). The genomic landscape of species divergence in *Ficedula flycatchers*. *Nature*, 491, 756–760. <https://doi.org/10.1038/nature11584>
- Ellner, S. P., Childs, D. Z., & Rees, M. (2016). *Data-driven modelling of structured populations: A practical guide to the integral projection model*. Basel, Switzerland: Springer International Publishing.
- Excoffier, L., Quilodrán, C. S., & Currat, M. (2014). Models of hybridization during range expansions and their application to recent human evolution. In A. Derevianko & M. Shunkov (Eds.), *Cultural developments in the eurasian paleolithic and the origin of anatomically modern humans* (pp. 122–137). Department of the Institute of Archaeology and Ethnography SB RAS: Novosibirsk, Russia.
- Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, 28, 342–350. <https://doi.org/10.1016/j.tig.2012.03.009>
- Feder, J. L., Flaxman, S. M., Egan, S. P., Comeault, A. A., & Nosil, P. (2013). Geographic mode of speciation and genomic divergence. *Annual Review of Ecology, Evolution, and Systematics*, 44, 73–97. <https://doi.org/10.1146/annurev-ecolsys-110512-135825>
- Feder, J. L., Gejji, R., Yeaman, S., & Nosil, P. (2012). Establishment of new mutations under divergence and genome hitchhiking. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 367, 461–474.
- Feder, J. L., & Nosil, P. (2009). Chromosomal inversions and species differences: When are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions? *Evolution*, 63, 3061–3075. <https://doi.org/10.1111/j.1558-5646.2009.00786.x>
- Feder, J. L., & Nosil, P. (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., Chain, F. J. J., Panchal, M., Huang, Y., Ezaguirre, C., Kalbe, M., ... Milinski, M. (2015). Genomics of divergence along a continuum of parapatric population differentiation. *PLoS Genetics*, 11, e1004966. <https://doi.org/10.1371/journal.pgen.1004966>
- Flaxman, S. M., Feder, J. L., & Nosil, P. (2013). Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. *Evolution*, 67, 2577–2591. <https://doi.org/10.1111/evo.12055>
- Flaxman, S. M., Wacholder, A. C., Feder, J. L., & Nosil, P. (2014). Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Molecular Ecology*, 23, 4074–4088. <https://doi.org/10.1111/mec.12750>
- Fraïsse, C., Roux, C., Welch, J. J., & Bierne, N. (2014). Gene-flow in a mosaic hybrid zone: Is local introgression adaptive? *Genetics*, 197, 939–951. <https://doi.org/10.1534/genetics.114.161380>
- Haller, B. C., Galloway, J., Kelleher, J., Messer, P. W., & Ralph, P. L. (2019). Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. *Molecular Ecology Resources*, 19, 552–566. <https://doi.org/10.1111/1755-0998.12968>
- Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. *Molecular Biology and Evolution*, 36, 632–637. <https://doi.org/10.1093/molbev/msy228>
- Han, F., Lamichhaney, S., Grant, B. R., Grant, P. R., Andersson, L., & Webster, M. T. (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Research*. <https://doi.org/10.1101/gr.215222.116>
- Harr, B. (2006). Genomic islands of differentiation between house mouse subspecies. *Genome Research*, 16, 730–737. <https://doi.org/10.1101/gr.5045006>
- Hofer, T., Foll, M., & Excoffier, L. (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., Bassham, S., Currey, M., & Cresko, W. A. (2012). Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 367, 395–408.
- Hughes, A. L. (2009). Evolution in the post-genome era. *Perspectives in Biology and Medicine*, 52, 332–337. <https://doi.org/10.1353/pbm.0.0090>
- Jiggins, C., & Martin, S. (2017). Glittering gold and the quest for Isla de Muerta. *Journal of Evolutionary Biology*, 30, 1509–1511. <https://doi.org/10.1111/jeb.13110>
- Jones, O. R., Clutton-Brock, T., Coulson, T., & Godfray, H. C. J. (2008). A web resource for the UK's long-term individual-based time-series (LITS) data. *Journal of Animal Ecology*, 77, 612–615. <https://doi.org/10.1111/j.1365-2656.2008.01382.x>
- Kawakami, T., Backström, N., Burri, R., Husby, A., Olason, P., Rice, A. M., ... Ellegren, H. (2014). Estimation of linkage disequilibrium and

- interspecific gene flow in *Ficedula flycatchers* by a newly developed 50k single-nucleotide polymorphism array. *Molecular Ecology Resources*, 14, 1248–1260.
- Kelleher, J., Etheridge, A. M., & McVean, G. (2016). Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Computational Biology*, 12, e1004842. <https://doi.org/10.1371/journal.pcbi.1004842>
- Klopfstein, S., Currat, M., & Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, 23, 482–490. <https://doi.org/10.1093/molbev/msj057>
- Kokko, H., Chaturvedi, A., Croll, D., Fischer, M. C., Guillaume, F., Karrenberg, S., ... Stapley, J. (2017). Can evolution supply what ecology demands? *Trends in Ecology & Evolution*, 32(3), 187–197. <https://doi.org/10.1016/j.tree.2016.12.005>
- Kusakabe, M., Ishikawa, A., Ravinet, M., Yoshida, K., Makino, T., Toyoda, A., ... Kitano, J. (2017). Genetic basis for variation in salinity tolerance between stickleback ecotypes. *Molecular Ecology*, 26, 304–319.
- Lohse, K. (2017). Come on feel the noise—from metaphors to null models. *Journal of Evolutionary Biology*, 30, 1506–1508.
- Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, 23, 2178–2192.
- Ma, T., Wang, K., Hu, Q., Xi, Z., Wan, D., Wang, Q., ... Liu, J. (2018). Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. *Proceedings of the National Academy of Sciences*, 115, E236–E243. <https://doi.org/10.1073/pnas.1713288114>
- Mackay, T. F. (2014). Epistasis and quantitative traits: Using model organisms to study gene–gene interactions. *Nature Reviews Genetics*, 15, 22. <https://doi.org/10.1038/nrg3627>
- Marques, D. A., Lucek, K., Meier, J. I., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2016). Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLoS Genetics*, 12, e1005887. <https://doi.org/10.1371/journal.pgen.1005887>
- Matthey-Doret, R., & Whitlock, M. C. (2018). Background selection and the statistics of population differentiation: consequences for detecting local adaptation. *bioRxiv*, 326256.
- Nadeau, N. J., Whibley, A., Jones, R. T., Davey, J. W., Dasmahapatra, K. K., Baxter, S. W., ... Mallet, J. (2012). Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 343–353.
- Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, 70, 3321–3323. <https://doi.org/10.1073/pnas.70.12.3321>
- Noor, M. A., & Bennett, S. M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, 103, 439. <https://doi.org/10.1038/hdy.2009.151>
- Nordborg, M. (1997). Structured coalescent processes on different time scales. *Genetics*, 146, 1501–1514.
- Nordborg, M., Charlesworth, B., & Charlesworth, D. (1996). The effect of recombination on background selection. *Genetics Research*, 67, 159–174. <https://doi.org/10.1017/S0016672300033619>
- Nosil, P. (2008). Speciation with gene flow could be common. *Molecular Ecology*, 17, 2103–2106. <https://doi.org/10.1111/j.1365-294X.2008.03715.x>
- Nosil, P. (2012). *Ecological speciation: Oxford series in ecology and evolution*. Oxford: Oxford University Press.
- Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1587), 332–342. <https://doi.org/10.1098/rstb.2011.0263>
- Nosil, P., Feder, J. L., Flaxman, S. M., & Gompert, Z. (2017). Tipping points in the dynamics of speciation. *Nature Ecology & Evolution*, 1, 0001. <https://doi.org/10.1038/s41559-016-0001>
- Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, 18, 375–402. <https://doi.org/10.1111/j.1365-294X.2008.03946.x>
- Paradis, E. (2010). pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, 26, 419–420. <https://doi.org/10.1093/bioinformatics/btp696>
- Peng, B., & Amos, C. I. (2008). Forward-time simulations of non-random mating populations using simuPOP. *Bioinformatics*, 24, 1408–1409. <https://doi.org/10.1093/bioinformatics/btn179>
- Peng, B., & Kimmel, M. (2005). simuPOP: A forward-time population genetics simulation environment. *Bioinformatics*, 21, 3686–3687. <https://doi.org/10.1093/bioinformatics/bti584>
- R Development Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ravinet, M., Faria, R., Butlin, R., Galindo, J., Bierne, N., Rafajlović, M., ... Westram, A. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, 30, 1450–1477. <https://doi.org/10.1111/jeb.13047>
- Renaut, S., Grassa, C., Yeaman, S., Moyers, B., Lai, Z., Kane, N., ... Rieseberg, L. (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications*, 4, 1827. <https://doi.org/10.1038/ncomms2833>
- Riesch, R., Muschick, M., Lindtke, D., Villoutreix, R., Comeault, A. A., Farkas, T. E., ... Nosil, P. (2017). Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology & Evolution*, 1, 0082. <https://doi.org/10.1038/s41559-017-0082>
- Roux, C., Fraisse, C., Romiguier, J., Anciaux, Y., Galtier, N., & Bierne, N. (2016). Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biology*, 14, e2000234. <https://doi.org/10.1371/journal.pbio.2000234>
- Ruegg, K., Anderson, E. C., Boone, J., Pouls, J., & Smith, T. B. (2014). A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology*, 23, 4757–4769. <https://doi.org/10.1111/mec.12842>
- Samuk, K., Owens, G. L., Delmore, K. E., Miller, S. E., Rennison, D. J., & Schluter, D. (2017). Gene flow and selection interact to promote adaptive divergence in regions of low recombination. *Molecular Ecology*, 26, 4378–4390. <https://doi.org/10.1111/mec.14226>
- Schindler, S., Gaillard, J. M., Grüning, A., Neuhaus, P., Traill, L. W., Tuljapourkar, S., & Coulson, T. (2015). Sex-specific demography and generalization of the Trivers–Willard theory. *Nature*, 526, 249. <https://doi.org/10.1038/nature14968>
- Sedghifar, A., Brandvain, Y., & Ralph, P. (2016). Beyond clines: Lineages and haplotype blocks in hybrid zones. *Molecular Ecology*, 25, 2559–2576.
- Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., ... Brännström, Å. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, 15, 176.
- Slatkin, M., & Wiehe, T. (1998). Genetic hitch-hiking in a subdivided population. *Genetics Research*, 71, 155–160. <https://doi.org/10.1017/S001667239800319X>
- Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., ... Nosil, P. (2014). Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, 344, 738–742. <https://doi.org/10.1126/science.1252136>
- Sousa, V., & Hey, J. (2013). Understanding the origin of species with genome-scale data: Modelling gene flow. *Nature Reviews Genetics*, 14, 404–414. <https://doi.org/10.1038/nrg3446>
- Turner, T. L., & Hahn, M. W. (2010). Genomic islands of speciation or genomic islands and speciation? *Molecular Ecology*, 19, 848–850.
- Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, 3, e285. <https://doi.org/10.1371/journal.pbio.0030285>

- Via, S. (2012). Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 367, 451–460. <https://doi.org/10.1098/rstb.2011.0260>
- Wang, L., Wan, Z. Y., Lim, H. S., & Yue, G. H. (2016). Genetic variability, local selection and demographic history: Genomic evidence of evolving towards allopatric speciation in Asian seabass. *Molecular Ecology*, 25, 3605–3621. <https://doi.org/10.1111/mec.13714>
- Winter, D. J. (2012). MMod: An R library for the calculation of population differentiation statistics. *Molecular Ecology Resources*, 12, 1158–1160.
- Wu, C. I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, 14, 851–865. <https://doi.org/10.1046/j.1420-9101.2001.00335.x>
- Wu, C.-I., & Ting, C.-T. (2004). Genes and speciation. *Nature Reviews Genetics*, 5, 114. <https://doi.org/10.1038/nrg1269>
- Yeaman, S., Aeschbacher, S., & Bürger, R. (2016). The evolution of genomic islands by increased establishment probability of linked alleles. *Molecular Ecology*, 25, 2542–2558.
- Zhang, D., Song, G., Gao, B., Cheng, Y., Qu, Y., Wu, S., ... Lei, F. (2017). Genomic differentiation and patterns of gene flow between two long-tailed tit species (*Aegithalos*). *Molecular Ecology*, 26(23), 6654–6665.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Quilodrán CS, Ruegg K, Sendell-Price AT, Anderson EC, Coulson T, Clegg SM. The multiple population genetic and demographic routes to islands of genomic divergence. *Methods Ecol Evol.* 2020;11:6–21. <https://doi.org/10.1111/2041-210X.13324>

