Supporting information

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

SUPPLEMENTARY METHODS

Library preparation

100 ng of genomic DNA from each individual was digested with 2.4 U of restriction enzyme SbfI-HF (New-England Biolabs Inc., Beverly MA, USA) at 37ºC for 60 minutes, followed by an inactivation step of 80ºC for 20 minutes. BestRAD Sbf1 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 Sbf1 adapter sequences. Reactions were incubated at 20ºC 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 Sbf1 adaptors. Sbf1-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 20ºC for 15 minutes with Blunt/TA Ligase Master Mix (New-England Biolabs Inc.) and then at 37ºC 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. 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.

SUPPLEMENTARY TABLES & FIGURES

Table S1. List of parameters of the model with default values

* 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 and Wilson 1983).

† Recombination rates observed for parts of the collared-flycatcher (*Ficedula albicollis*) genome (1). 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 and 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.

Parameter	Definition	Value	Source
N_0	Initial number of breeders	400	(2)
ζ	Number of young fledging by clutch	1.9	(2)
S_S	Survival rate at summer	0.96	(3)
S_a	Survival rate at autumn	0.76	(3)
S_w	Survival rate at winter	0.63	(3)
δ	Sex ratio	0.5	(2)
G	Generation time	3	(4)
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	‡

Table S2: Distributional skew of F_{ST} values calculated for each autosomal chromosome.

Only chromosomes with at least 10 x 500kb windows are reported. Comparisons where distributional skew is significantly higher when diverging with gene flow are highlighted in green; whereas comparisons where distributional skew is significantly lower when diverging under isolation (no gene flow) are highlighted in red. Significance determined using a modified randomisation test (see methods).

Chromosome	Early stage		Mid stage	Late stage	
	Gene flow (SI vs. CI)	No gene flow (SI vs. FP)	Gene flow (HI vs. ML)	Gene flow (LF vs. GT)	No gene flow (LH vs. VN)
1	1.964	1.715	1.086	1.371	0.445
1A	1.836	2.339	1.289	0.619	1.446
1B	-	-	-	-	-
2	1.608	1.558	4.483	1.013	0.424
3	1.372	0.785	1.780	0.779	0.443
4	4.886	1.191	0.872	1.088	0.667
4A	1.613	3.074	2.870	0.632	0.110
5	0.866	0.663	0.972	1.655	1.781
6	1.302	0.829	1.517	1.713	0.491
7	0.637	0.941	1.020	1.136	0.762
8	1.288	0.758	0.651	0.259	-0.060
9	2.527	0.441	1.032	2.340	0.529
10	0.872	0.878	0.485	0.624	1.558
11	0.863	1.191	1.397	0.000	-0.407
12	3.384	1.415	0.352	1.774	-0.113
13	1.658	1.220	0.900	0.781	0.186
14	1.251	0.380	1.056	0.673	0.156
15	1.671	1.138	3.334	0.526	-0.001
17	-0.106	-0.427	-0.069	0.672	0.150
18	0.037	1.503	1.970	-0.104	-0.478
19	0.706	-0.665	0.179	0.013	0.227
20	0.291	0.347	0.818	-0.246	1.541
21	0.887	0.988	0.431	0.509	-0.150
22	-	-	-	-	-
23	-	-	-	-	-
24	2.586	1.111	-0.326	1.054	-0.502
25	-	-	-	-	-
26	-	-	-	-	-
27	-	-	-	-	-

28	-0.777	0.623	0.793	-0.238	0.753
LGE22	-	-	-	-	-

Table S3: P-values for pairwise comparisons of F_{ST} distributional skew for simulations, tested using a randomisation test.

Values in grey shaded cells indicate level of skew for a given timeframe.

Gene flow	**Simulated time frame** (Generations)			
	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	-	0.054	0.081
	1000	-	-	-0.352
m = 0.001	30	0.593	0.386	0.1032
	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

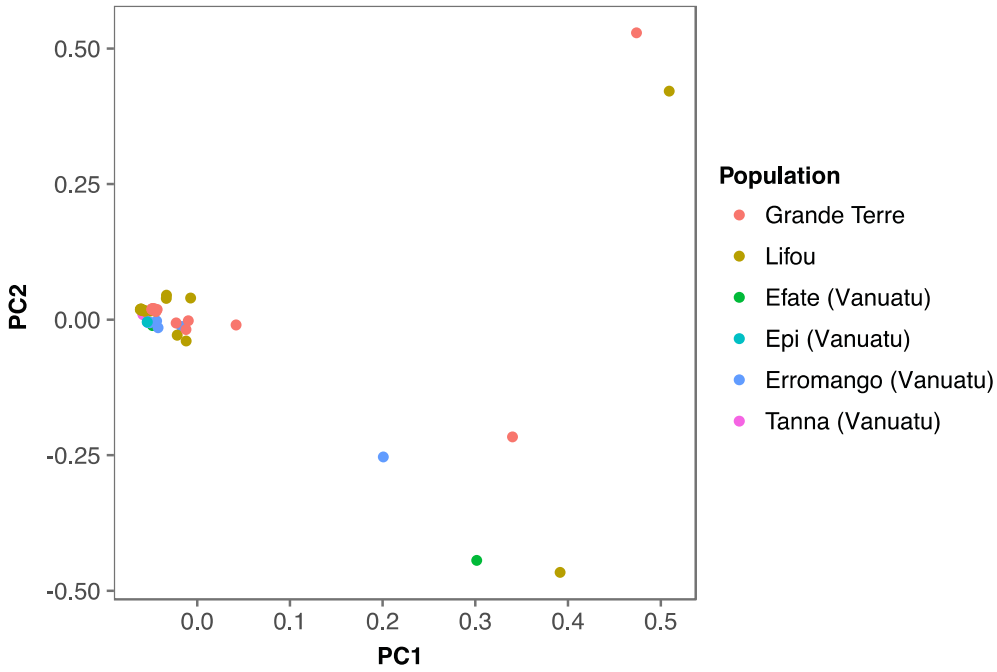

Fig S1: Principle component analysis of genetic variation.

Based on 129,505 SNPs across 56 individuals sampled across 6 Melanesian islands (New Caledonia and Vanuatu). The variance explained by PC1 and PC2 is 16.46% and 10.47%, respectively.

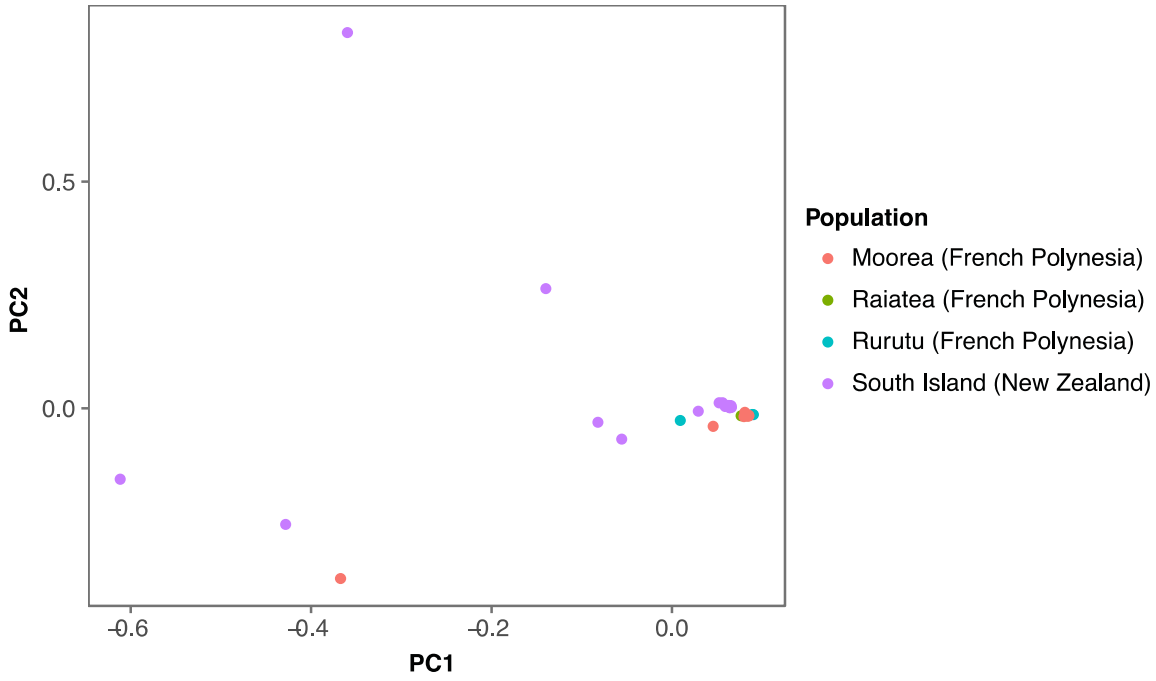

Fig S2: Principle component analysis of genetic variation.

Based on 129,505 SNPs across 37 individuals sampled across three French Polynesian islands and South Island, New Zealand. The variance explained by PC1 and PC2 is 12.99% and 9.01%, respectively.

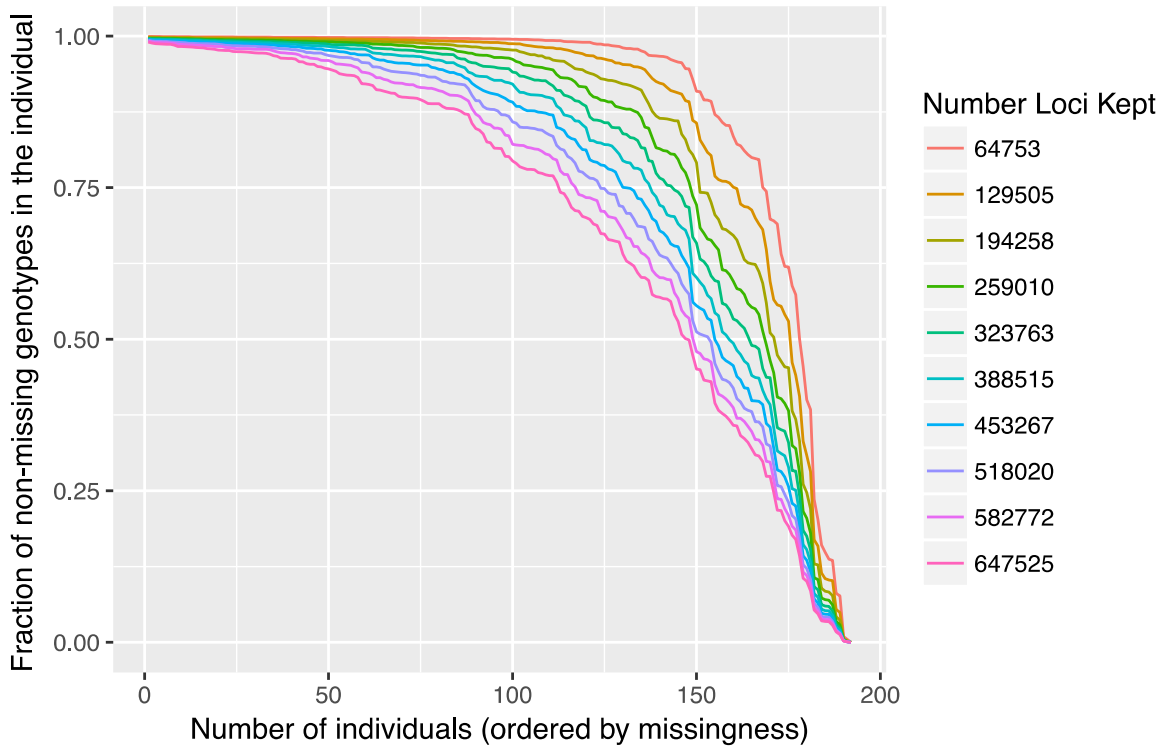

Fig S3: Missingness plot generated using *genoscapeRtools*.

This was used to determine the optimal number of samples and SNPs to retain for downstream analyses.

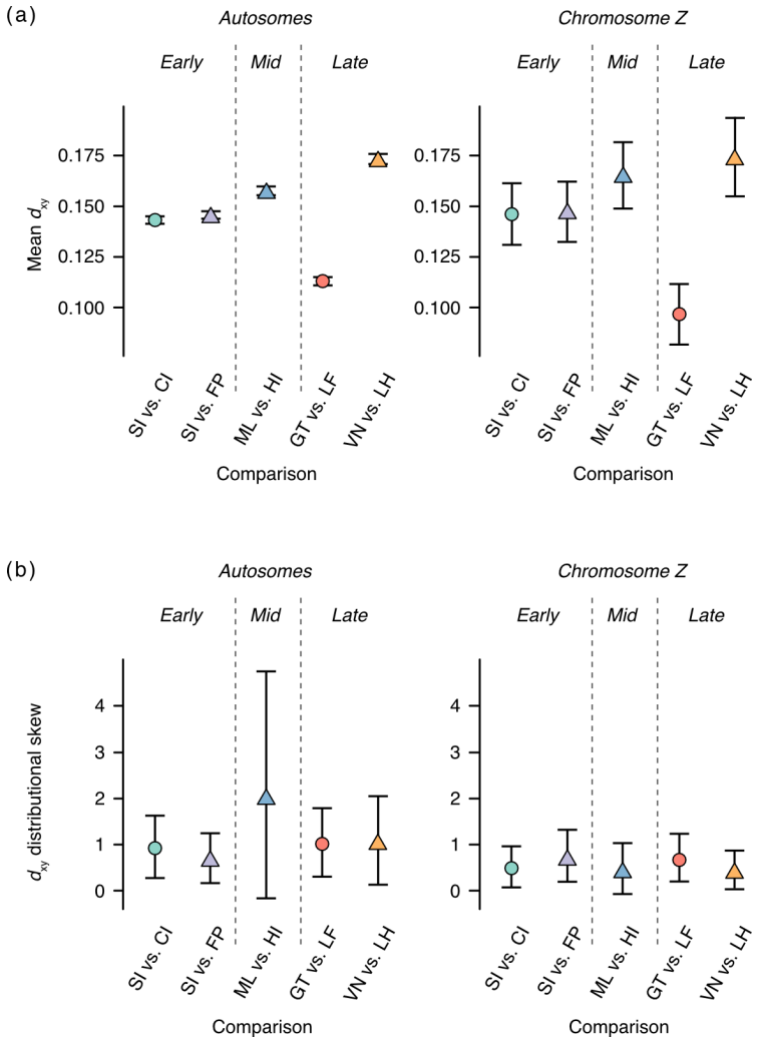

Fig S4: Mean d_{xy} and distributional skew of d_{xy} values.

(A) Mean d_{xy} and **(B)** distributional skew of d_{xy} values calculated in non-overlapping 500kb windows. Calculated for autosomes and chromosome Z separately for each population comparison. 95% confidence intervals obtained via bootstrapping over loci. Circles indicate divergence with gene flow and triangles divergence in isolation.

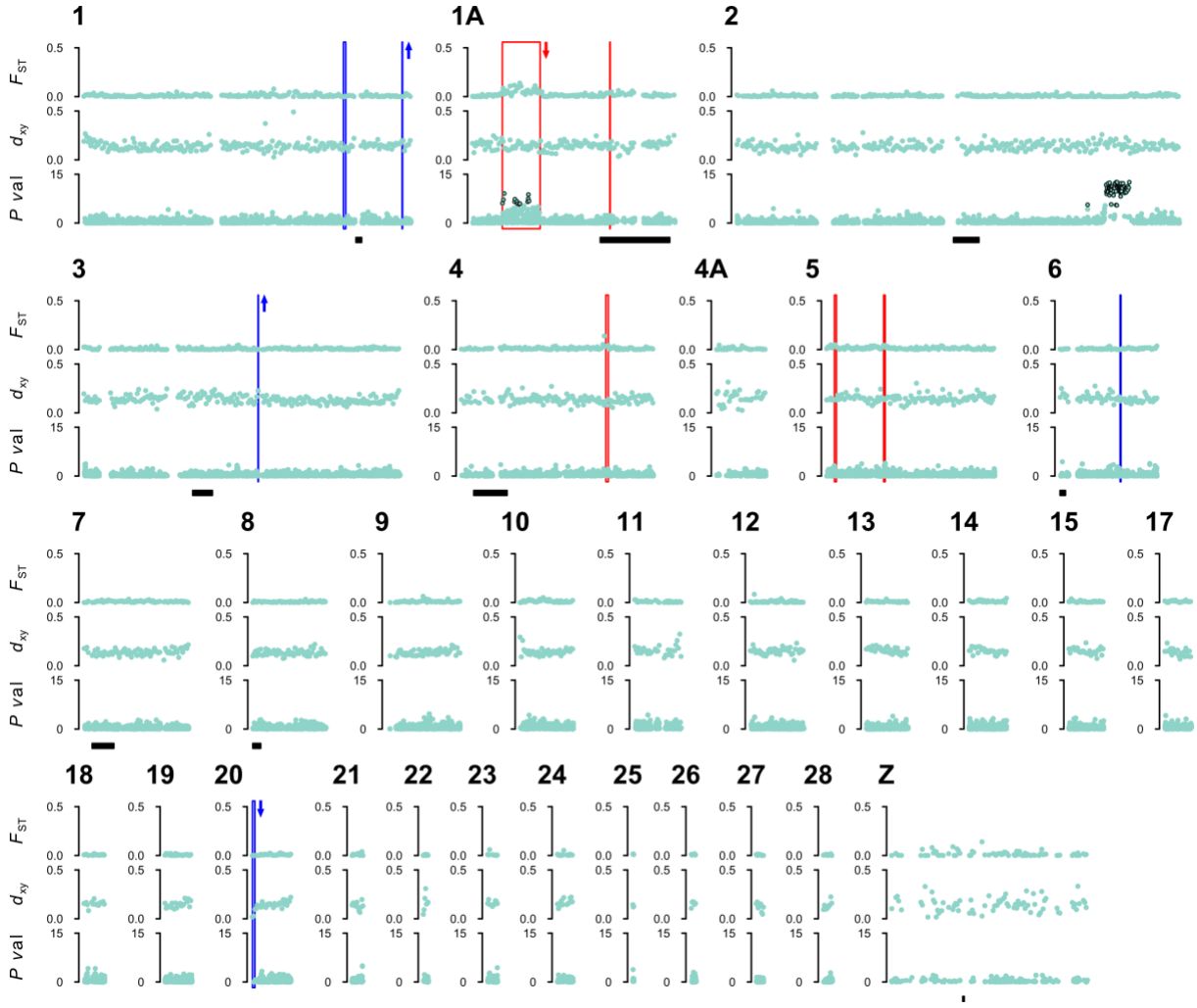

Fig S5: Patterns of nucleotide differentiation and outlier analysis between South Island and Chatham Island silvereyes.

For each chromosome, plots show relative nucleotide differentiation (F_{ST}, top), absolute nucleotide differentiation (d_{xy}, middle) and negative $\log_{10}$ (p-values) estimated using PCAdapt. FST and dxy are calculated in non-overlapping 500kb windows and PCAdapt p-values are calculated on a SNP-by-SNP basis. 50kb windows with 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. Genomic islands and genomic valleys are based on F_{ST} and locations are the same as those shown in Fig 5. Outlier SNPs identified using FDR = 0.001 are outlined. Known centromere locations are indicated in black, as inferred by Warren *et al.* (5) using fluorescence *in situ* hybridization (FISH) probes. Chromosomes are numbered according to the zebra finch nomenclature.

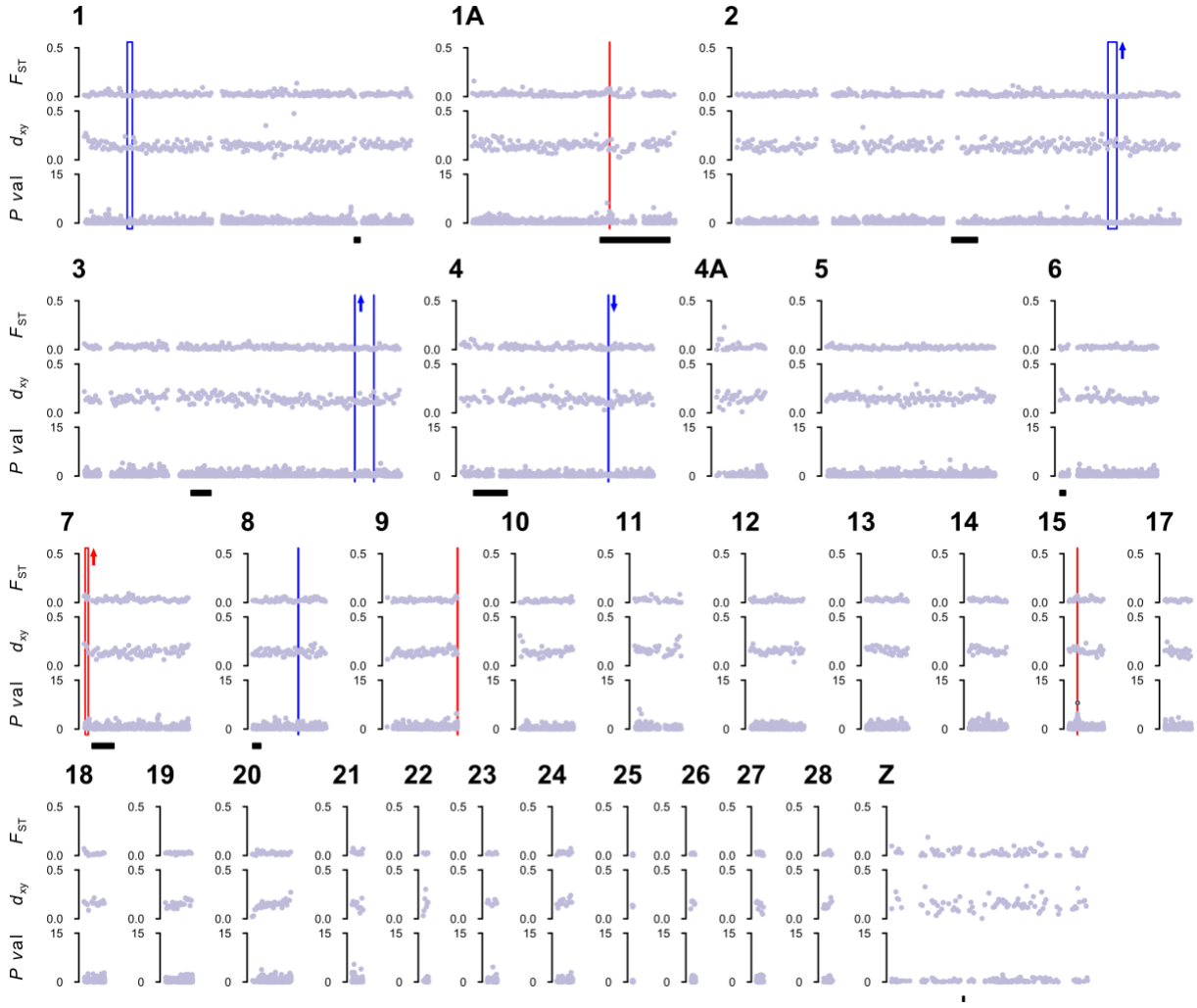

Fig S6: Patterns of nucleotide differentiation and outlier analysis between South Island and French Polynesian silvereyes.

For each chromosome, plots show relative nucleotide differentiation (F_{ST}, top), absolute nucleotide differentiation (d_{xy}, middle) and negative $\log_{10}$ (p-values) estimated using PCAdapt. FST and dxy are calculated in non-overlapping 500kb windows and PCAdapt p-values are calculated on a SNP-by-SNP basis. 50kb windows with 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. Genomic islands and genomic valleys are based on F_{ST} and locations are the same as those shown in Fig 5. Outlier SNPs identified using FDR = 0.001 are outlined. Known centromere locations are indicated in black, as inferred by Warren *et al.* (5) using fluorescence *in situ* hybridization (FISH) probes. Chromosomes are numbered according to the zebra finch nomenclature.

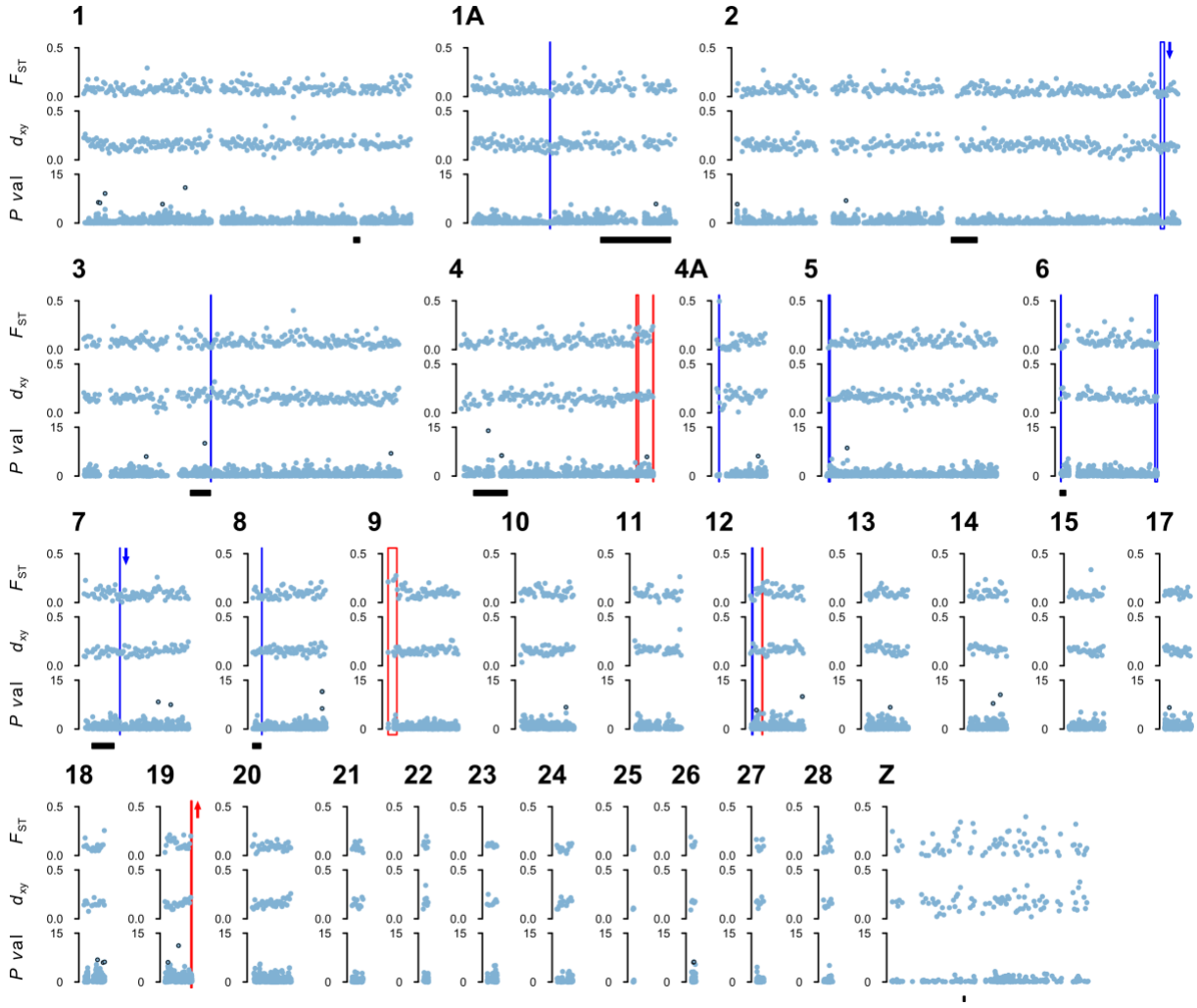

Fig S7: Patterns of nucleotide differentiation and outlier analysis between Mainland and Heron Island silvereyes.

For each chromosome, plots show relative nucleotide differentiation (F_{ST}, top), absolute nucleotide differentiation (d_{xy}, middle) and negative $\log_{10}$ (p-values) estimated using PCAdapt. FST and dxy are calculated in non-overlapping 500kb windows and PCAdapt p-values are calculated on a SNP-by-SNP basis. 50kb windows with 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. Genomic islands and genomic valleys are based on F_{ST} and locations are the same as those shown in Fig 5. Outlier SNPs identified using FDR = 0.001 are outlined. Known centromere locations are indicated in black, as inferred by Warren *et al.* (5) using fluorescence *in situ* hybridization (FISH) probes. Chromosomes are numbered according to the zebra finch nomenclature.

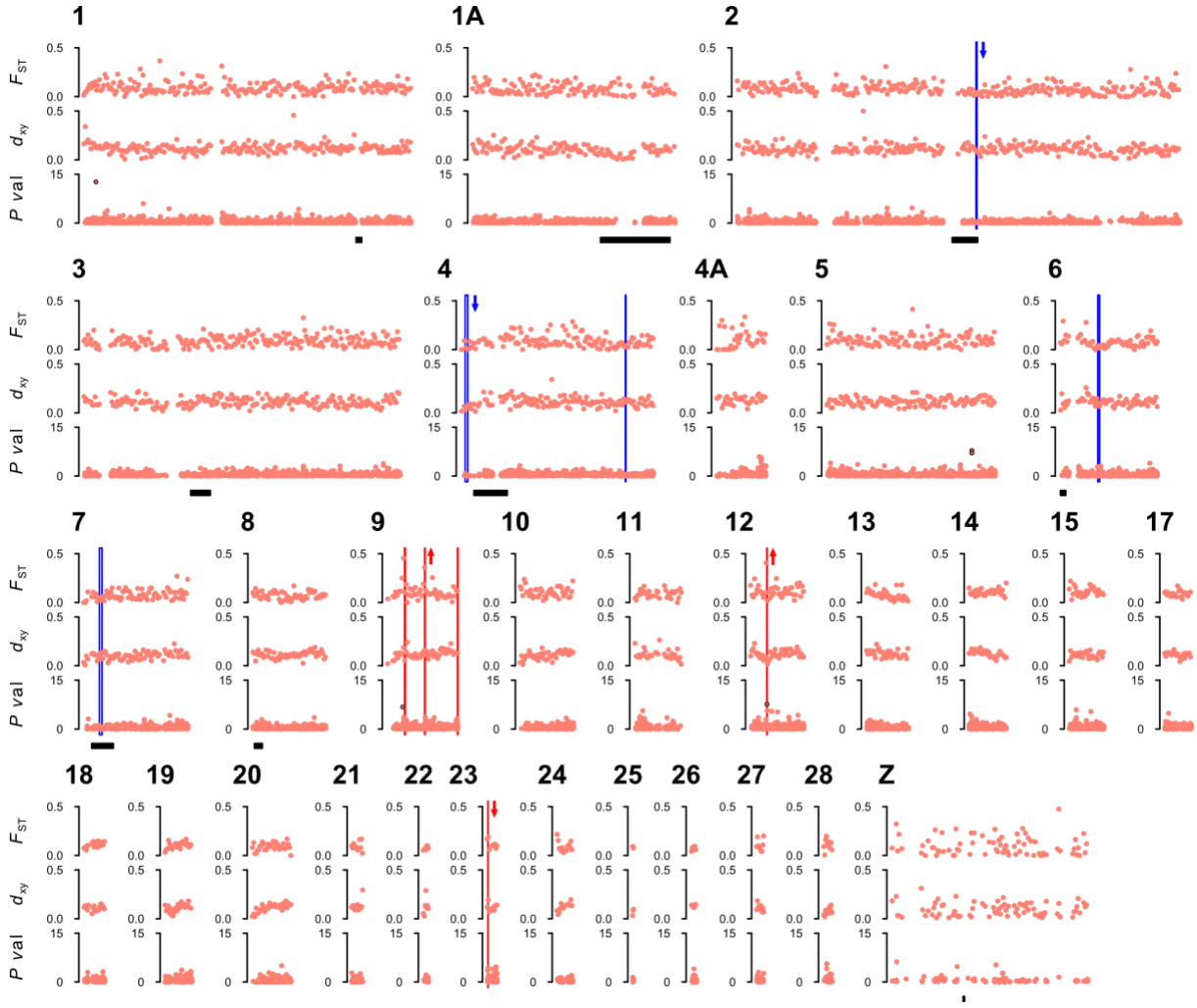

Fig S8: Patterns of nucleotide differentiation and outlier analysis between Grande Terre and Lifou silvereyes.

For each chromosome, plots show relative nucleotide differentiation (F_{ST}, top), absolute nucleotide differentiation (d_{xy}, middle) and negative $\log_{10}$ (p-values) estimated using PCAdapt. FST and dxy are calculated in non-overlapping 500kb windows and PCAdapt p-values are calculated on a SNP-by-SNP basis. 50kb windows with 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. Genomic islands and genomic valleys are based on F_{ST} and locations are the same as those shown in Fig 5. Outlier SNPs identified using FDR = 0.001 are outlined. Known centromere locations are indicated in black, as inferred by Warren *et al.* (5) using fluorescence *in situ* hybridization (FISH) probes. Chromosomes are numbered according to the zebra finch nomenclature.

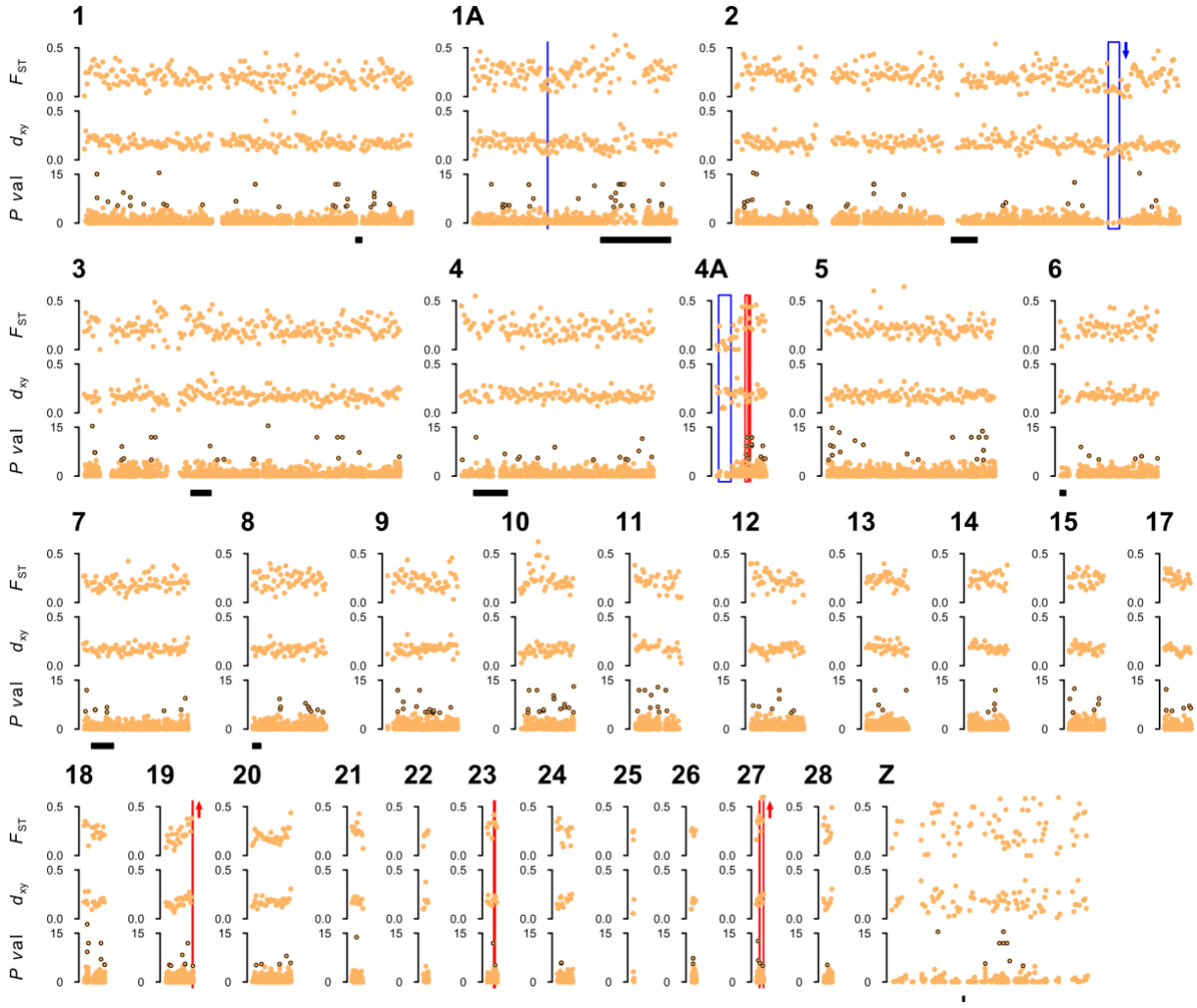

Fig S9: Patterns of nucleotide differentiation and outlier analysis between Vanuatu and Lord Howe Island silvereyes.

For each chromosome, plots show relative nucleotide differentiation (F_{ST}, top), absolute nucleotide differentiation (d_{xy}, middle) and negative $\log_{10}$ (p-values) estimated using PCAdapt. FST and dxy are calculated in non-overlapping 500kb windows and PCAdapt p-values are calculated on a SNP-by-SNP basis. 50kb windows with 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. Genomic islands and genomic valleys are based on F_{ST} and locations are the same as those shown in Fig 5. Outlier SNPs identified using FDR = 0.001 are outlined. Known centromere locations are indicated in black, as inferred by Warren *et al.* (5) using fluorescence *in situ* hybridization (FISH) probes. Chromosomes are numbered according to the zebra finch nomenclature.

REFERENCES

1. Kawakami T, Smeds L, Backström N, Husby A, Qvarnström A, Mugal CF, et al. 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. 2014;23(16):4035–58.

2. Kikkawa J, Wilson JM. Breeding and Dominance Among the Heron Island Silvereyes Zosterops Lateralis Chlorocephala. Emu - Austral Ornithology. 1983;83(3):181–98.

3. Sandvig EM, Coulson T, Kikkawa J, Clegg SM. The influence of climatic variation and density on the survival of an insular passerine Zosterops lateralis. PLOS ONE. 2017;12(4):e0176360.

4. Clegg SM, Frentiu FD, Kikkawa J, Tavecchia G, Owens IPF. 4000 Years of phenotypic change in an island bird: Heterogeneity of selection over three microevolutionary timescales. Evolution. 2008;62(9):2393–410.

5. Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A, et al. The genome of a songbird. Nature. 2010;464(7289):757–62.