Standing genetic variation and *de novo* mutations underlie parallel evolution of island bird phenotypes

Andrea Estandía[1,✉], Ashley T. Sendell-Price[1,2], Bruce C. Robertson[3], and Sonya M. Clegg[1]

[1]Edward Grey Institute of Field Ornithology, Department of Biology, University of Oxford, United Kingdom
[2]Bioinformatics Research Technology Platform, University of Warwick, Coventry, CV4 7AL, United Kingdom
[3]Department of Zoology, University of Otago, Dunedin, New Zealand

Parallel evolution occurs when the same trait evolves in closely related lineages in response to similar ecological contexts and provides some of the best examples of determinism in evolutionary biology. However, from a genetic standpoint, this process can be driven by either new mutations that appear independently in each diverging population or by selection on existing genetic variation common to both lineages. Small bodied birds, for example, tend to increase in size after they colonise a new island, following what is known as the 'island rule'. Such is the case of the Silvereye, a prolific natural coloniser of southwest Pacific islands. Island forms of this bird species increase in body size after they establish, with the pattern and pace of change consistent with directional natural selection and evident even in the most recent colonisations within the last 200 years. The system provides an exceptional opportunity to explore the genomic basis of repeated body size evolution. We sequenced 377 whole genomes from 31 different Silvereye populations, which revealed that both mechanisms are at play: in some lineages, new mutations are highly associated with body and bill size, but there are also highly associated polymorphisms present across all populations. Our research sheds light on the genomic basis of repeated body size evolution and emphasises that multiple molecular mechanisms can underlie similar evolutionary trajectories even within a single taxon.

parallel evolution | adaptation | island rule | GWAS | inversion | structural rearrangements | mutations | body size | standing genetic variation | birds

Correspondence: ***andrea.estandia@biology.ox.ac.uk***

Introduction

Parallel evolution refers to the independent emergence of similar traits in different lineages that share a common ancestor[1], and illustrates deterministic evolutionary outcomes that are widespread across the Tree of Life[2]. Natural selection is typically responsible for parallel evolution, as genetic drift is unlikely to produce such similar changes[3–6]. Selection can act on standing genetic variation to produce the same phenotype[7–9], on different *de novo* mutations in each population[10], or a combination of the two[11]. However, a faster rate of adaptive change is more often facilitated from standing variation than from mutations as the latter are mostly neutral or mildly deleterious[12,13]. For example, the rapid repeated evolution of stickleback low-plated phenotypes is driven by reassembling ancestral variation[14,15]. Additionally, standing genetic variation can be supplemented with genetic variants that contribute to parallel evolution shared via admixture among recently diverged lineages[16,17]. The same processes can affect single polymorphisms or full blocks in the genome: chromosomal rearrangements, like inversions, can play a role in parallel evolution by keeping sets of adaptive alleles together in strong linkage disequilibrium[18,19]. Understanding the relative importance of alternative genetic mechanisms behind parallel evolution has received considerable interest over the last few decades[20,21]. Genomic methods now allow us to investigate examples spanning diverse systems, providing empirical data that is necessary to draw generalities.

Islands provide environments where comparable evolutionary pressures can result in the rapid emergence of similar phenotypes[22]. The 'island syndrome' refers to the phenomenon where sets of morphological, physiological, and behavioural traits evolve in a predictable way when organisms colonise islands[22]. The syndrome ultimately results from the exposure to a common suite of biotic conditions on islands; a combination of reduced predation and parasitism, and a shift in the balance of intra- and interspecific competition, features that fundamentally change the selective pressures in predictable and repeatable ways[22–25]. Arguably the best characterised component of this syndrome is the 'island rule' that describes evolution towards medium body sizes in insular settings; small colonising species evolve larger body sizes and vice versa[23,26–28]. There is evidence for the island rule across vertebrate taxa, including birds[26]. Bill size also follows this rule whereby short-billed species show an increase in bill size[23]. Body and bill size are complex multivariate quantitative traits predicted to be highly polygenic[29–31], but some studies in natural populations show that a few genes of major effect can also impact these traits[32–34].

Members of the silvereye subspecies complex (*Zosterops lateralis*, Aves) have repeatedly colonised islands across the South Pacific from the Australian mainland and Tasmania, with up to 17 morphological subspecies currently recognised[35]. Insular silvereyes are larger than mainland ones, and the degree of this shift increases with time since colonisation[36,37], a pattern that has been shown to be a consequence of directional selection rather than drift[23,37]. The island populations represent at least four independent colonisations from the Australian mainland: to Southern Melanesia (∼150,000 years ago), Lord Howe Island (∼100,000 years ago), southern Great Barrier Reef Islands, including Heron Island (∼4000 years ago) and a recent historically recorded sequential colonisation sequence from Tasmania to New Zealand and outlying islands (<200 years ago)[38]. The populations arising from these independent colonisations are genetically isolated from the original mainland source

and other colonisations, though gene flow occurs among Southern Melanesian islands and among recently colonised populations[39]. The combination of independent and known colonisation sequences, along with broadly repeated shifts in body and bill size in similar ecological contexts provides an excellent system to explore the genetic mechanisms of parallel evolution.

Here, we ask whether *de novo* mutations or standing genetic variation lead to the same phenotype: increased body and bill size in small-bodied species after island colonisation. We find that both are associated with body and bill size variation in Southern Melanesia, whereas only standing genetic variation appears important for repeated patterns in all other populations. This demonstrates empirically that two genetic mechanisms hypothesised to underlie repeated evolution can operate simultaneously in a single system.

Results

The genetic basis of repeated body and bill size evolution.

To describe and quantify the bill and body size shifts in the populations studied here, we summarised univariate traits as principal component scores, resulting in a PC1 for body, and a PC1 for bill that represented size. In both cases 75% of the variation was explained by PC1 and roughly equal trait loadings of univariate traits (Fig. 1; Table S1). We statistically confirmed increases in body and bill size (Fig. S1).

To examine the genetic basis of repeated evolution, we generated low-to-intermediate coverage (5X) whole-genome sequencing data of 377 silvereyes from 24 islands and sampled across seven Australian States and Territories representative of most of the silvereye range (Table S2). After alignment of the sequenced individuals, we analysed genetic variation in a probabilistic framework accounting for low coverage (ANGSD[40]) and obtained 14.3 million SNPs with minor allelic frequencies higher than 5% for differentiation analyses. Population structure was described by a genomic PCA based on whole-genome data that revealed three major groups (Fig. S2). PC1 explained 20.18% of the total variance and separated two clusters: one containing individuals from Southern Melanesia and the other containing the remaining individuals. PC2 explained 6.8% of the variance and separated the two main archipelagos within Southern Melanesia: New Caledonia and Vanuatu. This population structure aligns with previous analysis using fewer individuals[39].

After controlling for population structure, the GWAS identified 31 loci linked to body size variation, and 15 of them (48%) were classified as *de novo* mutations exclusive to islands in Southern Melanesia (Fig. 2A). Thirteen were part of an association peak on Chromosome 2, which also contained one annotated gene, *GRHPR*, important in metabolic processes[41]. Univariate traits, including tarsus and tail were also associated with this peak (Fig. S3). For the top outlier, those individuals with the genotype TT showed a five times increase in median body size relative to those with the CC genotype (Table S3; Fig. 2A,C). Other size-associated loci that formed part of the standing genetic variation showed

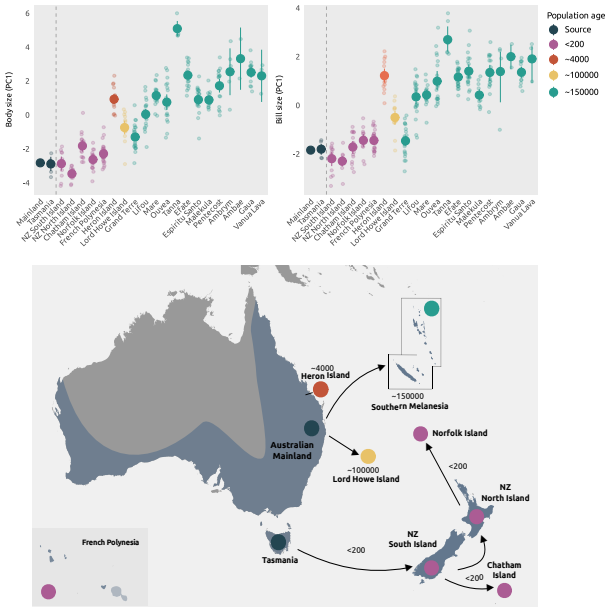

Fig. 1. Morphological divergence in body and bill size (PC1) among silvereye forms. The source populations are placed left to the vertical dashed line. Colonisations are arranged chronologically: the most recent ones on the left and the most ancient on the right. The map below shows the Silvereye distribution highlighted in blue and the time of colonisation of each island or archipelago. French Polynesia is shown as inset.

a similar effect in direction and magnitude on body size (Fig. S3). Among these, were genes involved in calmodulin binding, cytoskeletal motor activity, circadian regulation and other metabolic processes. Additionally, we identified genes associated with particular univariate traits involved in metabolism, bone formation, and cartilage formation (Table S4).

Sixty-three loci were associated with bill size, with 21 of them being *de novo* mutations, present only in Southern Melanesia. These *de novo* mutations were not part of a specific association peak. However, most bill size-associated variants were present across all populations. A number of adjacent variants on a region on Chromosome 18 were found in several of the large- bill-size populations: Ouvea and Tanna, two islands in Southern Melanesia, as well as Heron Island and Lord Howe Island. Three annotated genes, *PLA2G4A*, *CDH23*, and *DNAH9*, were located in this region (Fig. 2B,D; Fig. S6; Table S4).

We quantified the proportion of phenotypic variance explained by all genome-wide SNPs with GCTA[42]. These narrow-sense heritability estimates were high but showed wide confidence intervals (CI) (Table S5). Tarsus length is a highly heritable trait based on pedigree data, and is estimated to have a heritability of 0.7 for Heron Island silvereyes[37], a higher value than our SNP-based estimate (h^2=0.46 (95% CI =0.22-0.7)).

Structural variation: a neo-sex chromosome and multiple putative inversions.

To examine the structural variation along the genome, we identified clusters of PCA windows (local PCAs) based on

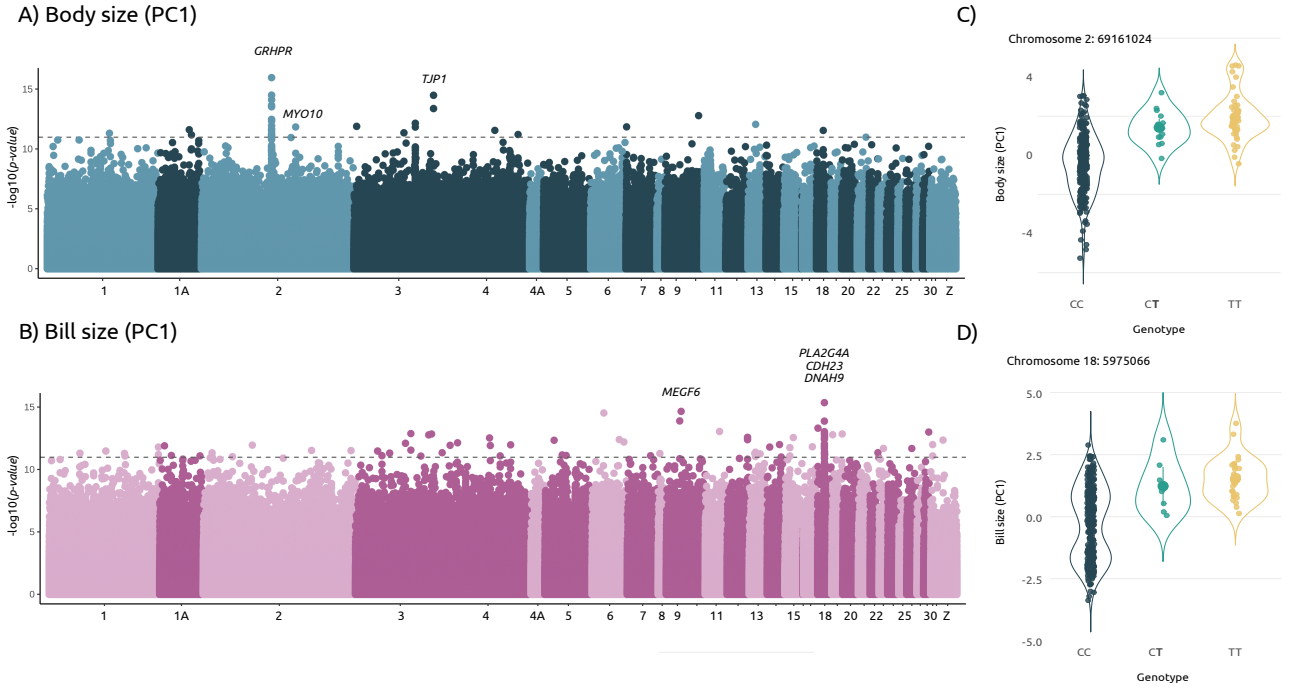

Fig. 2. A, B) Manhattan plots for body size (PC1) and bill size (PC1). Chromosomes are shown with alternating light and dark shading and numbered according to the zebra finch nomenclature. C) Correlation between allele frequency (top ranked SNP) and body size, and D) bill size.

Multidimensional Scaling (MDS) (Fig. 3-4). This method uses common patterns to identify structural variation such as non-recombining haplotype blocks[43,44]. We found a large outlier region on Chromosome 4A (Fig. 3A-3B). MDS1 separated two distinct groups that aligned with sex (Fig. 3C), where sex was independently confirmed by PCR-based sexing of individuals.

To specifically search for putative inversions, we filtered those sections of the genome with at least five adjacent outlier windows, pooled the windows and repeated MDS analysis for each separate region. Several chromosomes showed large outlier regions that formed three discrete equidistant clusters along MDS1 (Fig. 4B). This type of pattern is highly suggestive of a potential inversion with each alternative homokaryotype separated by the intermediate heterokaryotype[20]. Consistent with the occurrence of inversions, LD was high inside these regions contrasting to the LD along the rest of the chromosome (Fig. S7), indicating limited recombination in heterokaryotypes. For the potential inversion on Chromosome 27, the AA karyotype was present at high frequencies in all intermediate and old colonisations (Southern Melanesia, Lord Howe Island and Heron Island), and at very low frequency on the mainland. The potential inversion on Chromosome 28 showed a different pattern with the AA karyotype exclusively present in Southern Melanesia. On Chromosome 29, BB was only recovered from individuals in Southern Melanesia, but some individuals from New Caledonia displayed the AA karyotype (Fig. 4). Body size was not associated with karyotype for Chromosomes 27 and 29 after controlling for population structure. Body size was associated with the AA karyotype in Southern Melanesia, but we could

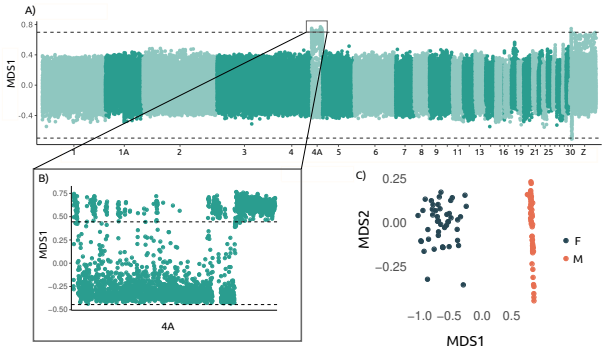

Fig. 3. A) Local PCA results (MDS1) plotted across the genome B) where chromosome 4A shows outlier windows each representing 100 SNPs. C) MDS1 and MDS2 separate by sex in this region.

not control for population structure as karyotype and neutral variation structure were entirely coincident (Fig. S8). Using our custom annotation, we found genes involved in various metabolic processes on Chromosomes 27 and 28, and lipogenesis, adipocyte differentiation and muscle tissue development on Chromosome 29 (Table S6).

Discussion

Selection on standing genetic variation is the main route to repeated evolution of larger bill and body sizes in island-colonising silvereye birds. Size-associated variants were shared across colonisation histories spanning hundreds to over a hundred thousand years. However, some *de novo* mutations present only in the Southern Melanesian silvereye

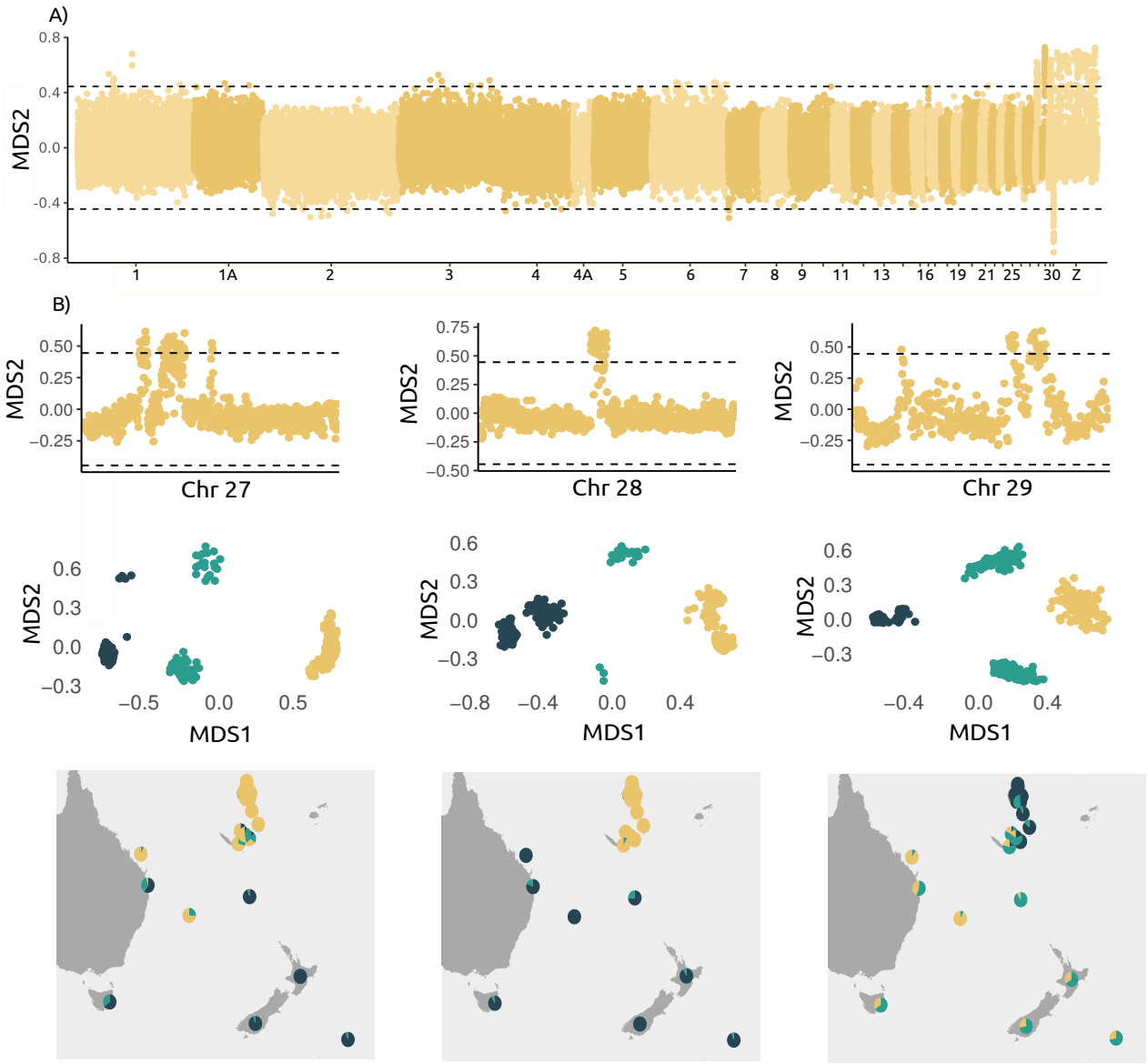

Fig. 4. A) Local PCA results (MDS2) are displayed across the entire genome. B) Outlier regions in chromosomes 27, 28, and 29. Each dot corresponds to 100 SNPs. MDS1 identifies three distinct clusters within the outlier regions, suggesting potential inversions. Chartpies provide information about the relative frequencies of each cluster within different populations.

populations were also linked to increased bill and body sizes. This empirical system demonstrates the joint role of standing genetic variation and *de novo* mutations in producing a repeated island phenotype.

We identified several candidate genes common across all populations, associated overall body size, bill size, or particular univariate traits. This pattern suggests that variation present in ancestral source populations was represented in multiple, independent colonising flocks, and maintained across sequential island colonisations. Given that population founding can result in loss of genetic variation and the fate of rare alleles can be highly stochastic especially when populations remain small[45,46], such an outcome might seem unlikely. However, silvereyes colonise in reasonably large flocks, and they recover rapidly from small population sizes, minimising loss of variation during founding and population establishment[36,45]. Historically colonised silvereye populations and Heron Island silvereyes have experienced rapid trait evolution driven by natural selection[37,47,48], and such rapid evolution is more likely to stem from standing genetic variation than *de novo* mutations[49].

However, multiple large effect *de novo* mutations were associated with body size in the oldest colonisation (Southern Melanesia). These mutations were not fixed across all populations: some of them had fully fixed BB genotypes associated with very large body sizes, while others lacked the B allele entirely, and many displayed all genotypes. Variation could be maintained across the archipelagos if differ-

ent degrees of size are favoured in different populations connected by gene flow[39,50] as seen across Southern Melanesian silvereye populations[39]. While larger size of island silvereyes appears generally to be favoured, other species and indeed other white-eye species that co-occur, could impose local constraints on size. For example, on Lifou, an island in New Caledonia, three *Zosterops* species occur in sympatry: the relatively small-sized endemic *Z. minutus*, the large endemic *Z. inornatus*, with the silvereye being intermediate in body size between the two[51]. An alternative but untested possibility is that variants arise from introgression with sympatric *Zosterops* species. Other mechanisms for maintaining variation, such as negative frequency dependent selection, are more difficult to reconcile in systems like islands where the usual drivers, e.g. predation risk, are limited[52].

Even though linking annotated gene function to phenotypes in non-model organisms is highly speculative, any associations can provide the basis for future investigation. The *de novo* association peak on Chromosome 2 associated with body size has a hit for *GRHPR*. This gene encodes the enzyme glyoxylate reductase/hydroxypyruvate reductase, and is involved in key metabolic processes, including serine metabolism for gluconeogenesis and oxalate metabolism[41]. No other previous studies have linked body size variation with this gene, however, avian comparative genomic studies suggest that enzymes involved in glyoxylate metabolic routes might be responsible for diet-related adaptations in birds[53]. The function of other genes associated with body size is variable, from bone and muscle development to metabolic processing, and even though their role in body size is feasible, exact mechanisms remain to be found.

The gene most strongly linked to variation in bill size is *PLA2G4A*, located on chromosome 18. This gene encodes an enzyme known as phospholipase A2 involved in lipid metabolism[54]. Suppression of *PLA2* induces the disturbance of calcium and phosphorus metabolism and has been shown to cause beak deformities in birds[55,56]. We also found that *CDH23* and *DNAH9*, also on Chromosome 18, are both expressed by cochlear cells and are associated with sound perception[57]. In addition, *MEGF6*, in Chromosome 9, is involved in cell adhesion but was found to be associated with song learning in Zebra finches[58]. Changes in bill size can affect how birds vocalise and perceive sounds[59] and silvereye songs vary significantly across different islands due to cultural drift and adaptations to acoustic environments[60]. However, it is difficult to draw any conclusions from this association; the interplay of song, bill morphology and selection for large size should be investigated in more detail. Interestingly, we did not we did not find significant associations with *ALX1* or *HMGA2*, two large-effect genes that have been repeatedly demonstrated to be implicated in bill morphology variation, including in Darwin's finches[32,33]. However, different genetic architectures are not entirely surprising; even other species of finches show different pathways underlying bill size variation[34].

We identified signatures of structural variants, one of which could potentially contribute to body size variation on chromosome 28, but others where such a relationship was clearly absent. Rearrangements on chromosomes 27, 28 and 29 each showed a pattern consistent with an inversion: three different clusters separated by MDS1 and high LD within the putative inversion compared with the rest of each chromosome[20,44]. Our conservative size estimates of the putative inversions observed here (approximately 200kb for those on Chromosomes 28 and 29 and 650 kb for Chromosome 27) places them in the range of inversion sizes known to be most frequently associated with adaptive phenotypes (100kb to several Mb)[18]. Reduced recombination in the heterokaryotypes preserves linkage among alleles despite mating between alternative homokaryotypes, maintaining the adaptive benefit of the haplotype blocks[18,20]. For example, marine snails (*Littorina saxatilis*) and seaweed flies (*Coelopa frigida*) show inverted regions linked to adaptive phenotypes[61,62]; the latter has an inversion under balancing selection that works as a supergene for body size. The karyotype distribution of the putative inversion identified on Chromosome 28 was associated with body size, but was conflated with geographic population structure. The function of genes inside this region varied widely, some of them are involved in immunity, others in cholesterol metabolism and some are even directly linked to fitness proxies in birds[63], like egg quality[64]. However, it is unclear whether they contributed to the evolution of large body sizes in Southern Melanesia. To further study the role of structural variants we could generate long-read sequencing and recombination maps that help to confirm that they are indeed inversions, and to determine breakpoint locations[44,65].

The three putative inversions displayed karyotype distribution patterns that suggested varying evolutionary histories. The inversion on Chromosome 28 could have emerged after colonisation from the Australian mainland and could have been subsequently fixed in the majority of Southern Melanesian populations. Aside from sampling considerations, alternative scenarios are more complex, requiring different combinations of karyotype loss and maintenance that seem less likely. For putative inversions on both Chromosomes 27 and 29, the maintenance of the heterokaryotype and alternative homokaryotype in many populations suggest they could represent ancient inversions under balancing selection. Such polymorphism can be maintained when the inversion provides a fitness advantage against different selective pressures: if there are epistatic interactions among alleles at different loci[66], if there is associative overdominance[67] or if a locus that is under balancing selection is located within the inversion[18]. Similar to alternative processes that could explain SNP variation, karyotypes could appear to be balanced if there is introgression resulting from gene flow among different sympatric *Zosterops* species occurring across Southern Melanesia, on Norfolk Island and prior to 20th century extinction, also on Lord Howe Island, but not Heron Island or New Zealand.

A sizeable structural variant detected via local PCA and visualised by MDS was on chromosome 4A, coinciding with a known neo-sex chromosome in the Zosteropidae[68], formed via fusion between an autosome and a sex chromosome ap-

proximately 24 Ma in the Sylvioidea superfamily to which the Zosteropidae belong[69,70]. Neo-sex chromosomes can play a role in speciation as they can contain loci for traits that contribute to reproductive isolation (e.g. male courtship display behavioural traits[71]). However, the role of this neo-sex chromosome in speciation within the Sylvioidea, including in the Zosteropidae, is unclear[70].

The study of non-model systems in the wild has inherent practical and bioinformatic limitations. Like any empirical study, we had to balance sample size and genomic coverage. GWAS usually relies on a high number of whole genomes with attached individual data. Capturing and taking samples from the birds used in this study covering a vast geographical distribution was a tremendous effort that spanned decades, and sequencing hundreds of genomes remains relatively expensive. Additionally, body and bill sizes are complex multivariate phenotypes that correlate with several other traits that we did not investigate. Lastly, identifying the causal genes is a major challenge and often needs extra steps following GWAS to confirm the causality of the candidate loci, for example by doing knock-outs, which is not possible to do in wild birds.

Despite these caveats and limitations, work on non-model systems is essential to gain a broad perspective on the different genetic routes that lead to similar phenotypes. The application of genomic methods to the Silvereye system, which encompasses a wide range of island colonisation times and repeated evolution of body and bill size, provides a rare example of the role of standing genetic variation on both short and longer timescales of colonisation, with *de novo* mutations contributing only the oldest island populations.

Methods

Sample collection.

We sampled 377 silvereyes (*Z. lateralis*) from 31 islands or sites on the Australian mainland between 1997 and 2017 (Table S2). Birds were caught in the wild using mist nets or traps and we collected 20–40 µl of blood from the brachial wing vein and stored it in 1 ml of Queen's lysis buffer[77]. We took a suite of morphological measurements, including the wing length using a metal ruler to determine the maximum flattened chord of the longest primary feather, and the length of the central tail feathers from the base to the tip using dividers. We used dial calipers (accuracy ±0.1mm) to measure metatarsal length, head length from the rear of the skull to the bill's tip. Additionally, we took various bill measurements using dial calipers, including mandible length, width, and depth at the posterior nostril opening, and mandible length, width, and depth at the anterior nostril opening. Finally, we measured the body weight. Sex cannot be determined in silvereyes with morphological measurements. We selected a random subset of individuals and sexed them using molecular methods described in the section below. All sampling was non-destructive, and we released the birds at point of capture. Complete sampling information is provided in Table S2.

DNA extraction, molecular sexing, library preparation, and sequencing.

We extracted DNA from blood samples using a standard Phenol-chloroform protocol[77]. We added 100 uL of blood stored in Queen's buffer[77] to a microcentrifuge tube containing 250 uL of DIGSOL extraction buffer (0.02 M EDTA, 0.05 M Tris-HCl (pH 8.0), 0.4 M NaCl, 0.5% sodium dodecyl sulphate (SDS)) and 10uL of Proteinase K (20mg/mL). Samples were incubated at 55C overnight. Following incubation, we added 250 uL of phenol:chloroform:isoamyl alcohol (25:24:1) and gently mixed the samples for 10 minutes. We then centrifuged the samples at 10,000 rpm for 10 minutes. We transferred the aqueous layer to a new microcentrifuge tube and repeated the previous step by adding the phenol mixture, mixing and centrifuging it. We recovered a new aqueous phase, transferred it to a new tube, added 250uL of chloroform:isolamyl alcohol (24:1), mixed and centrifuged the samples as in the previous steps. We precipitated the DNA by adding 2 volumes of ice-cold 100% ethanol, 1 volume of 2.5M ammonium acetate and 2 uL of glycogen, leaving it overnight at -20C. The next morning we centrifuged the tubes for 10 min at 15,000 rpm at 4C and removed the supernatant. We rinsed the precipitate with 500 uL of ice-cold 70% ethanol and centrifuged the tubes as in the previous step. We removed the supernantant and the left the precipitated DNA to dry at room temperature. Once dried, the DNA was resuspended in 50 uL of TE (TrisEDTA) buffer (0.01 M Tris-HCL (pH 8.0), 0.0001 M EDTA). We quantified the DNA concentration with a Qubit 2.0. DNA extracts were sent to Novogene UK for library preparation and whole genome sequencing at 5X on the Illumina Novaseq 6000 platform (Illumina, San Diego) using paired-end 150bp sequencing reads. We molecularly sexed each sample by amplifying the chromohelicase-DNA-binding (*CHD1*) gene using a standard polymerase chain reaction (PCR) protocol and following the method of Griffiths et al. (1998)[89].

Construction of a pseudochromosome assembly.

We used the Chromosembler tool available in Satsuma2[78] to order and orient scaffolds of the *Z. l. melanops* genome assembly (GCA_001281735.1)[79] into 33 pseudochromosomes according to synteny with the Vertebrate Genome Project's Zebra finch (*Taeniopygia guttata*) genome assembly (GCA_003957565.1)[80].

To reconstruct a transcriptome that we could use to generate hints for gene prediction, we used available RNA-seq data for *Zosterops borbonicus*[68]. We filtered out low-quality sequence reads (phred quality score < 20) and adaptor sequences using TrimGalore[81] and mapped them onto our pseudo-chromosome assembly using HISAT2[82]. We used Trinity[83] to assemble the transcriptome.

Assembly annotation.

We annotated the resulting pseudo-chromosome assembly using a multifaceted approach of repeat identification and *ab initio* gene prediction using AUGUSTUS[84]. We incorporated hints containing extrinsic evidence about the location and structure of three types of genomic features: repeti-

tive regions, short-read transcriptome sequences, and single-copy orthologs. To obtain our hints, we first identified and masked repetitive and low-complexity DNA sequences using RepeatMasker[85] with the chicken (*Gallus gallus*) database of repeats. Second, we used BLAT[86] to incorporate RNA-seq information as hints. Third, we used single-copy genes in the Aves lineage group odb10 to detect conserved orthologous genes with BUSCO[87]. We then used BLASTx[88] in the AUGUSTUS output to obtain functional information.

Quality control and genotype calling.

We trimmed reads to remove adapter content and base calls of low quality using fastp[90] with default settings and 10 bp trimmed from the start of each read. We aligned cleaned reads of each individual to our pseudo-chromosome assembly using the Burrows-Wheeler Aligner (BWA)-mem algorithm[91]. Genotype calling can result in bias when using low-coverage sequencing[92]. Given the moderately low sequencing depth of our dataset (5X) downstream analyses were based on genotype probabilities rather than directly-called genotypes. We calculated genotype probabilities in ANGSD[40] using the samtools model and keeping SNPs with a minimum allele frequency (MAF) of 0.05. All analyses described below were run considering only biallelic SNPs with a mapping quality score 20. We filtered the bam files by excluding SNPs that showed significant deviation from Hardy-Weinberg equilibrium, recovering a total of 14,330,920 SNPs.

The genetic basis of parallel evolution.

To test whether new mutations or ancestral standing genetic variation led to similar phenotypes, we did a Genome-wide Association Study (GWAS) under the idea that if the associated variants are common to all populations then they are likely due to standing genetic variation but if they are present in a specific population or specific lineage, they may represent *de novo* mutations that have accumulated since the lineage split. This framework assumes no potential gene flow among colonisations, which is supported by genome-wide data[39].

Statistical tests of repeated increase in body and bill size.

Body size is most often represented by multivariate measures, such as Principal Component Analysis (PCA)[93,94], that account for the covariation among characters and summarise most of the variation in the first two axes, the first representing body size and the second body shape[94]. We conducted a PCA that included all morphological measurements that represent body size: wing length, tail length, tarsus length, and head length; and separately bill size: bill length to posterior nostril opening, bill width, and depth at anterior nostril opening. We tested whether colonisations resulted in changes in body size and bill size relative to the mainland using *brms*[95]. Model 1 included PC1 of body size as the response variable and Model 2 PC1 included bill size. Colonisation time was set as a categorical population-level effect with a weak prior in the form of a normal distribution centered in 0 with a standard deviation of 10. This prior gives equal support to the idea that body size can remain unchanged, increase or decrease after colonisation. We controlled for population by setting it as a group-level effect. We used a Gaussian family and the default link function. We ran four chains each with 5000 iterations which included a warm-up of 1000.

Linkage disequilibrium pruning.

Linkage disequilibrium (LD) refers to the tendency of alleles at linked loci to be inherited together[96]. Accounting for LD in GWAS is important and it can be done by applying pruning methods whereby a subset of markers that are in approximate linkage equilibrium (i.e. not statistically associated and more likely to be inherited independently) are selected. To obtain a set of unlinked SNPs we used ngsLD[97]. We first tested where LD decays within a maximum distance of 100 kb. We ran ngsLD assuming that SNPs could be in LD if they were within distances of 5 kb as r^2 was consistently below 0.1, obtaining 9,508,972 unlinked SNPs that we used for GWAS.

Genome-wide Association Study.

We used the dataset of unlinked SNPs to do GWAS for each of the following univariate variables: wing length, tail length, tarsus length, head length, bill length to posterior nostril opening, bill width, and depth at anterior nostril opening. We also performed a GWAS on the first two PCs of body size (body size PC1 and body size PC2) and bill size (bill size PC1 and bill size PC2). We used ANGSD latent genotype model where genotype is introduced as a latent variable and then the likelihood is maximised using weighted least squares regression98. Population structure is one of the most common confounders in GWAS[99] and needs to be controlled for. To do this, we generated a covariance matrix with PCAngsd[100] and decomposed it into principal components by using a scaling transformation. We applied a Bonferroni correction to classify variants as significant using a P-value of $1x10^{-4}$. For those significant variants, we visually explored the effect of the genotype on the trait and the distribution of the genotypes across populations. We excluded samples with genotype likelihoods lower or equal than 0.33, which indicates that all genotypes have the same probability of being the true genotype. If the variant had an effect on the trait but was present across all lineages, we defined it as a variant arising due to standing genetic variation, while if it was only present on an independent lineage we interpreted it as a putative *de novo* mutation. We used our custom annotation to find the regions where the outliers were located. To confirm that our annotation was robust, we pulled 1000 bp upstream and downstream the outlier from the reference genome and performed a BLAST search against all organisms. To estimate the proportion of phenotypic variance explained by all genome-wide SNPs while correcting for LD bias we used GCTA[42]. As GCTA does not take genotype probabilities as input, we obtained a BED file with PLINK[101]. We created four quartiles of LD scores and used each individual SNP's LD score for binning as recommended by Evans et al. 2018. We also controlled for population structure using the principal components arising from the PCAngsd analysis. We computed the h^2-SNP 95% CI as +- 1.96 * SE.

Detection of structural variants.

To detect structural variants, we conducted a localised PCA of genetic variation based on the methodology outlined in Li and Ralph (2019)[43]. We used PCAngsd and the R package *lostruct* to identify non-recombining haploblocks across the genome. Structural variants like inversions are characterised by high levels of linkage disequilibrium because recombination is often to be negligible in heterokaryotes[102]. For this reason, we used the unpruned SNP dataset containing 14 million SNPs. First, we split each chromosome into non-overlapping windows containing 100 SNPs each, ran PCAngsd and obtained a covariance matrix for each window. We then descomposed each matrix by using scaling transformation and generated a pairwise distance matrix based on PC1 and PC2 for each window. Finally, we mapped similarity using MDS of up to 20 axes. We defined outlier regions as windows with a coefficient of correlation above four standard deviations from the mean. Adjacent outliers with less than 10 windows between them were pooled, and clusters with less than 5 windows were not considered following Mérot et al. (2021)[103]. A typical signature of a polymorphic inversion is three groups of individuals appearing on a PCA: the two homokaryotypes for the alternative arrangements and, as an intermediate group, the heterokaryotypes[20]. All clusters of outlier windows were examined by MDS as single blocks. We then used K-means clustering to identify putative groups of haplotypes. We explored LD patterns in heterokaryotype versus homokarytype individuals for the chromosomes where we found outlier regions. We calculated pairwise r^2 values with ngsLD based on genotype likelihoods obtained by ANGSD and created a heatmap where we could visually explore LD levels. We built a model in *brms* to test whether karyotype identity was associated with body size after controlling for population structure. Karyotype was set as a categorical population-level effect and body size as the explanatory variable. Population was set as group-level effect. We used a Gaussian family and the default link function. We ran four chains each with 5000 iterations which included a warm-up of 1000.

ACKNOWLEDGEMENTS

The samples used in this study were collected over two decades and we thank the many people who helped in numerous ways to facilitate the work. We are grateful to the chiefs and landholders of Vanuatu and New Caledonia for granting access to field sites, and field and logistic assistance from many people including the following: N. Clark, D. Treby, J. LeBreton, F. Cugny , W. Waheoneme, O. Hebert and A. Rouquie (New Caledonia); F. Robertson and C. Sendell-Price (Heron Island); S. Geiger, O. Boissier, R. Hills, S. Totterman, O. Drew, K. Ser, W. Ser, J. Saksak (Vanuatu), I.P.F. Owens, N. Clark (Australian mainland); J. Kikkawa (Norfolk Island); I.P.F. Owens (Lord Howe Island); P. Park, A. Fletcher, P. Gray (Tasmania); P. Schweigman, D. Onely (New Zealand); M. Bell (Chatham Island); N. Davies, O. Grant, C. Sendell-Price (Moorea, French Polynesia). For additional samples we thank A. Phillimore, R. Black, M. Massaro, D. Potvin and N. Clark. The work was conducted under permits from the Direction de l'Environnement Province Sud and Direction Du Développement Economique (New Caledonia and Loyalty Islands with thanks to G. Kakue); Vanuatu Environment Unit letters of permission and permits provided by E. Bani and we further thank D. Kalfatak and T. Tiwok for their assistance (Vanuatu); Délégations régionales à la recherche et à la technologie (French Polynesia); Lord Howe Island Board; Environment Australia (Norfolk Island); Queensland Department of Environment and Resource Management; Parks and Wildlife Service (Tasmania). New Zealand Department of Conservation Te Papa Atawhai; Australian Bird and Bat Banding Scheme project and individual permits to SMC. Ethics clearances were provided by University of Queensland ethics committee (ZOO/165/94/ARC, ZOO/520/96/ARC/PHD, ZOO/520/97/ARC/PHD) and Griffith University ethics committee (ENV/01/12/AEC, ENV/07/16/AEC, ENV/06/20/AEC, ENV/24/13/AEC) to SMC. We thank the funders of this work: Marsden Fund New Zealand to BCR and SMC to support fieldwork on mainland Australia and Heron Island portions of the molecular work; National Geographic Society Committee for Research and Exploration Grant (9383-13) to SMC to support fieldwork in New Caledonia; Natural Environment Research Council (NERC) postdoctoral fellowship to SMC to support fieldwork in Vanuatu and New Caledonia; Percy Sladen Memorial fund to SMC to support fieldwork in French Polynesia; the Department of Zoology, Oxford start-up to SMC to support sequencing. ASP was supported by a NERC studentship (NE/L002612/1); AE was supported by a NERC studentship (NE/S007474/1) and a St John's College Graduate Scholarship.

Author contributions

SMC and AE formulated the idea and SMC supervised the project. SMC, BCR, and ATSP conducted fieldwork. AE, ATSP and BCR conducted laboratory work. AE conducted the genetic and statistical analyses with input from ATSP. AE drafted the manuscript with SMC and all authors provided comments.

Data and code availability

Code to reproduce all analyses is available at https://github.com/andreaestandia/1.0_island_rule_GWAS

References

- 1.Cerca, J. (2022). A simple conceptual framework and nomenclature for studying repeated, parallel and convergent evolution. EcoEvoRxiv. 10.32942/osf.io/g8u39.
- 2.Blount, Z.D., Lenski, R.E., and Losos, J.B. (2018). Contingency and determinism in evolution: Replaying life's tape. Science 362, eaam5979. 10.1126/science.aam5979.
- 3.Stayton, C.T. (2008). Is convergence surprising? An examination of the frequency of convergence in simulated datasets. Journal of Theoretical Biology 252, 1–14. 10.1016/j.jtbi.2008.01.008.
- 4.Arendt, J., and Reznick, D. (2008). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? Trends in Ecology Evolution 23, 26–32. 10.1016/j.tree.2007.09.011.
- 5.Losos, J.B. (2011). Convergence, Adaptation, and Constraint. Evolution 65, 1827–1840. 10.1111/j.1558-5646.2011.01289.x.
- 6.McGhee, G.R. (2011). Convergent Evolution: Limited Forms Most Beautiful (MIT Press).
- 7.Lai, Y.-T., Yeung, C.K.L., Omland, K.E., Pang, E.-L., Hao, Y., Liao, B.-Y., Cao, H.-F., Zhang, B.-W., Yeh, C.-F., Hung, C.-M., et al. (2019). Standing genetic variation as the predominant source for adaptation of a songbird. Proc. Natl. Acad. Sci. U.S.A. 116, 2152–2157. 10.1073/pnas.1813597116.
- 8.Therkildsen, N.O., Wilder, A.P., Conover, D.O., Munch, S.B., Baumann, H., and Palumbi, S.R. (2019).

Contrasting genomic shifts underlie parallel phenotypic evolution in response to fishing. Science 365, 487–490. 10.1126/science.aaw7271.

- 9.Konečná, V., Bray, S., Vlček, J., Bohutínská, M., Požárová, D., Choudhury, R.R., Bollmann-Giolai, A., Flis, P., Salt, D.E., Parisod, C., et al. (2021). Parallel adaptation in autopolyploid Arabidopsis arenosa is dominated by repeated recruitment of shared alleles. Nat Commun 12, 4979. 10.1038/s41467-021-25256-5.
- 10.Fulgione, A., Neto, C., Elfarargi, A.F., Tergemina, E., Ansari, S., Göktay, M., Dinis, H., Döring, N., Flood, P.J., Rodriguez-Pacheco, S., et al. (2022). Parallel reduction in flowering time from de novo mutations enable evolutionary rescue in colonizing lineages. Nat Commun 13, 1461. 10.1038/s41467-022-28800-z.
- 11.Ament-Velásquez, S.L. The Dynamics of Adaptation to Stress from Standing Genetic Variation and de novo Mutations.
- 12.Hedrick, P.W. (2013). Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Molecular Ecology 22, 4606–4618. 10.1111/mec.12415.
- 13.Barrett, R.D.H., and Schluter, D. (2008). Adaptation from standing genetic variation. Trends in Ecology Evolution 23, 38–44. 10.1016/j.tree.2007.09.008.
- 14.Colosimo, P.F., Hosemann, K.E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R.M., Schluter, D., and Kingsley, D.M. (2005). Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles. Science 307, 1928–1933. 10.1126/science.1107239.
- 15.Archambeault, S.L., Bärtschi, L.R., Merminod, A.D., and Peichel, C.L. (2020). Adaptation via pleiotropy and linkage: Association mapping reveals a complex genetic architecture within the stickleback Eda locus. Evolution Letters 4, 282–301. 10.1002/evl3.175.
- 16.Schluter, D., and Nagel, L.M. (1995). Parallel Speciation by Natural Selection. The American Naturalist. 10.1086/285799.
- 17.Montejo-Kovacevich, G., Meier, J.I., Bacquet, C.N., Warren, I.A., Chan, Y.F., Kucka, M., Salazar, C., Rueda-M, N., Montgomery, S.H., McMillan, W.O., et al. (2022). Repeated genetic adaptation to altitude in two tropical butterflies. Nat Commun 13, 4676. 10.1038/s41467-022-32316-x.
- 18.Faria, R., Johannesson, K., Butlin, R.K., and Westram, A.M. (2019). Evolving Inversions. Trends in Ecology Evolution 34, 239–248. 10.1016/j.tree.2018.12.005.
- 19.Harringmeyer, O.S., and Hoekstra, H.E. (2022). Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. Nat Ecol Evol 6, 1965–1979. 10.1038/s41559-022-01890-0.
- 20.Westram, A.M., Faria, R., Johannesson, K., Butlin, R., and Barton, N. (2022). Inversions and parallel evolution. Philosophical Transactions of the Royal Society B: Biological Sciences 377, 20210203. 10.1098/rstb.2021.0203.
- 21.MacPherson, A., and Nuismer, S.L. (2017). The probability of parallel genetic evolution from standing genetic variation. Journal of Evolutionary Biology 30, 326–337. 10.1111/jeb.13006.
- 22.Baeckens, S., and Van Damme, R. (2020). The island syndrome. Current Biology 30, R338–R339. 10.1016/j.cub.2020.03.029.
- 23.Clegg, S.M., and Owens, P.F. (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. 10.1098/rspb.2002.2024.
- 24.Lomolino, M.V. (2005). Body size evolution in insular vertebrates: generality of the island rule. J Biogeography 32, 1683–1699. 10.1111/j.1365-2699.2005.01314.x.
- 25.Meiri, S., Dayan, T., and Simberloff, D. (2004). Body Size of Insular Carnivores: Little Support for the Island Rule. The American Naturalist 163, 469–479. 10.1086/382229.
- 26.Benítez-López, A., Santini, L., Gallego-Zamorano, J., Milá, B., Walkden, P., Huijbregts, M.A.J., and Tobias, J.A. (2021). The island rule explains consistent patterns of body size evolution in terrestrial vertebrates. Nat Ecol Evol 5, 768–786. 10.1038/s41559-021-01426-y.
- 27.Foster, J.B. (1964). Evolution of Mammals on Islands. Nature 202, 234–235. 10.1038/202234a0.
- 28.Van Valen, L. (1973). Body Size and Numbers of Plants and Animals. Evolution 27, 27–35. 10.2307/2407116.
- 29.Yang, J., Weedon, M.N., Purcell, S., Lettre, G., Estrada, K., Willer, C.J., Smith, A.V., Ingelsson, E., O'Connell, J.R., Mangino, M., et al. (2011). Genomic inflation factors under polygenic inheritance. Eur J Hum Genet 19, 807–812. 10.1038/ejhg.2011.39.
- 30.Duntsch, L., Tomotani, B.M., de Villemereuil, P., Brekke, P., Lee, K.D., Ewen, J.G., and Santure, A.W. (2020). Polygenic basis for adaptive morphological variation in a threatened Aotearoa | New Zealand bird, the hihi (Notiomystis cincta). Proc. R. Soc. B. 287, 20200948. 10.1098/rspb.2020.0948.

- 31.Bosse, M., Spurgin, L.G., Laine, V.N., Cole, E.F., Firth, J.A., Gienapp, P., Gosler, A.G., McMahon, K., Poissant, J., Verhagen, I., et al. (2017). Recent natural selection causes adaptive evolution of an avian polygenic trait. Science 358, 365–368. 10.1126/science.aal3298.
- 32.Lamichhaney, S., Berglund, J., Almén, M.S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C.-J., Wang, C., Zamani, N., et al. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature 518, 371–375. 10.1038/nature14181.
- 33.Lamichhaney, S., Han, F., Berglund, J., Wang, C., Almén, M.S., Webster, M.T., Grant, B.R., Grant, P.R., and Andersson, L. (2016). A beak size locus in Darwin's finches facilitated character displacement during a drought. Science 352, 470–474. 10.1126/science.aad8786.
- 34.vonHoldt, B.M., Kartzinel, R.Y., Huber, C.D., Le Underwood, V., Zhen, Y., Ruegg, K., Lohmueller, K.E., and Smith, T.B. (2018). Growth factor gene IGF1 is associated with bill size in the black-bellied seedcracker Pyrenestes ostrinus. Nat Commun 9, 4855. 10.1038/s41467-018-07374-9.
- 35.Clements, J.F., Schulenberg, T.S., Iliff, M.J., Fredericks, T.A., Gerbracht, J.A., Lepage, D., Billerman, S.M., Sullivan, B.L., and Wood, C.L. (2021). The eBird/Clements checklist of Birds of the World: v2021.
- 36.Clegg, S.M., Degnan, S.M., Kikkawa, J., Moritz, C., Estoup, A., and Owens, I.P.F. (2002). Genetic consequences of sequential founder events by an island-colonizing bird. Proceedings of the National Academy of Sciences 99, 8127–8132.
- 37.Clegg, S.M., Frentiu, F.D., Kikkawa, J., Tavecchia, G., and Owens, I.P.F. (2008). 4000 Years of Phenotypic Change in an Island Bird: Heterogeneity of Selection Over Three Microevolutionary Timescales. Evolution 62, 2393–2410. 10.1111/j.1558-5646.2008.00437.x.
- 38.Black, R.A. (2010). Phylogenetic and phenotypic divergence of an insular radiation of birds.
- 39.Estandía, A., Sendell-Price, A.T., Oatley, G., Robertson, F., Potvin, D., Massaro, M., Robertson, B.C., and Clegg, S.M. (2023). Candidate gene length polymorphisms are linked to dispersive behaviour: searching for a mechanism behind the "paradox of the great speciators." 2023.01.19.524190. 10.1101/2023.01.19.524190.
- 40.Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics 15, 356. 10.1186/s12859-014-0356-4.
- 41.Rumsby, G., and Cregeen, D.P. (1999). Identification and expression of a cDNA for human hydroxypyruvate/glyoxylate reductase. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression 1446, 383–388. 10.1016/S0167-4781(99)00105-0.
- 42.Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. Am J Hum Genet 88, 76–82. 10.1016/j.ajhg.2010.11.011.
- 43.Li, H., and Ralph, P.L. (2019). Local PCA Shows How the Effect of Population Structure Differs Along the Genome. Genetics 211, 289–304. 10.25386/GENETICS.7324526.
- 44.Mérot, C., Oomen, R.A., Tigano, A., and Wellenreuther, M. (2020). A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation. Trends in Ecology Evolution 35, 561–572. 10.1016/j.tree.2020.03.002.
- 45.Sendell-Price, A.T., Ruegg, K.C., Robertson, B.C., and Clegg, S.M. (2021). An island-hopping bird reveals how founder events shape genome-wide divergence. Molecular Ecology 30, 2495–2510. 10.1111/mec.15898.
- 46.Whitlock, M.C. (2000). FIXATION OF NEW ALLELES AND THE EXTINCTION OF SMALL POPULATIONS: DRIFT LOAD, BENEFICIAL ALLELES, AND SEXUAL SELECTION. Evolution 54, 1855–1861. 10.1111/j.0014-3820.2000.tb01232.x.
- 47.Clegg, S.M., Degnan, S.M., Moritz, C., Estoup, A., Kikkawa, J., and Owens, I.P.F. (2002). Microevolution in Island Forms: The Roles of Drift and Directional Selection in Morphological Divergence of a Passerine Bird. Evolution 56, 2090–2099. 10.1111/j.0014-3820.2002.tb00134.x.
- 48.Sendell-Price, A.T., Ruegg, K.C., and Clegg, S.M. (2020). Rapid morphological divergence following a human-mediated introduction: the role of drift and directional selection. Heredity 124, 535–549. 10.1038/s41437-020-0298-8.
- 49.Marques, D.A., Meier, J.I., and Seehausen, O. (2019). A Combinatorial View on Speciation and Adaptive Radiation. Trends in Ecology Evolution 34, 531–544. 10.1016/j.tree.2019.02.008.
- 50.Clegg, S.M., and Phillimore, A.B. (2010). The influence of gene flow and drift on genetic and phenotypic divergence in two species of Zosterops in Vanuatu. Philosophical Transactions of the Royal Society B: Biological Sciences 365, 1077–1092. 10.1098/rstb.2009.0281.
- 51.Mees, G.F. (1969). A systematic review of the Indo-Australian Zosteropidae, three parts.

- 52.Beauchamp, G. (2021). Do avian species survive better on islands? Biol Lett 17, 20200643. 10.1098/rsbl.2020.0643.

- 53.Zhang, G., Li, C., Li, Q., Li, B., Larkin, D.M., Lee, C., Storz, J.F., Antunes, A., Greenwold, M.J., Meredith, R.W., et al. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. Science 346, 1311–1320. 10.1126/science.1251385.

- 54.Sharp, J.D., White, D.L., Chiou, X.G., Goodson, T., Gamboa, G.C., McClure, D., Burgett, S., Hoskins, J., Skatrud, P.L., and Sportsman, J.R. (1991). Molecular cloning and expression of human Ca(2+)-sensitive cytosolic phospholipase A2. Journal of Biological Chemistry 266, 14850–14853. 10.1016/S0021-9258(18)98550-9.

- 55.Chen, H., Dou, Y., Tang, Y., Zheng, X., Niu, X., Yang, J., Yu, X., and Diao, Y. (2016). Experimental reproduction of beak atrophy and dwarfism syndrome by infection in cherry valley ducklings with a novel goose parvovirus-related parvovirus. Veterinary Microbiology 183, 16–20. 10.1016/j.vetmic.2015.11.034.

- 56.Chen, S., Xiao, S., Wang, S., Lin, F., Cheng, X., Zhu, X., Jiang, D., Chen, S., and Yu, F. (2019). Duckling Short Beak and Dwarfism Syndrome, A Newly Emerging Disease in China. HV 6. 10.17582/journal.hv/2019/6.6.141.146.

- 57.Scheffer, D.I., Shen, J., Corey, D.P., and Chen, Z.-Y. (2015). Gene Expression by Mouse Inner Ear Hair Cells during Development. J. Neurosci. 35, 6366–6380. 10.1523/JNEUROSCI.5126-14.2015.

- 58.Lovell, P.V., Clayton, D.F., Replogle, K.L., and Mello, C.V. (2008). Birdsong "Transcriptomics": Neurochemical Specializations of the Oscine Song System. PLoS One 3, e3440. 10.1371/journal.pone.0003440.

- 59.Derryberry, E.P., Seddon, N., Claramunt, S., Tobias, J.A., Baker, A., Aleixo, A., and Brumfield, R.T. (2012). Correlated Evolution of Beak Morphology and Song in the Neotropical Woodcreeper Radiation. Evolution 66, 2784–2797. 10.1111/j.1558-5646.2012.01642.x.

- 60.Potvin, D.A., and Clegg, S.M. (2015). The relative roles of cultural drift and acoustic adaptation in shaping syllable repertoires of island bird populations change with time since colonization. Evolution 69, 368–380. 10.1111/evo.12573.

- 61.Morales, H.E., Faria, R., Johannesson, K., Larsson, T., Panova, M., Westram, A.M., and Butlin, R.K. (2019). Genomic architecture of parallel ecological divergence: Beyond a single environmental contrast. Sci. Adv. 5, eaav9963. 10.1126/sciadv.aav9963.

- 62.Koch, E.L., Ravinet, M., Westram, A.M., Johannesson, K., and Butlin, R.K. (2022). Genetic architecture of repeated phenotypic divergence in Littorina saxatilis ecotype evolution. Evolution 76, 2332–2346. 10.1111/evo.14602.

- 63.Krist, M. (2011). Egg size and offspring quality: a meta-analysis in birds. Biological Reviews 86, 692–716. 10.1111/j.1469-185X.2010.00166.x.

- 64.Gao, G., Gao, D., Zhao, X., Xu, S., Zhang, K., Wu, R., Yin, C., Li, J., Xie, Y., Hu, S., et al. (2021). Genome-Wide Association Study-Based Identification of SNPs and Haplotypes Associated With Goose Reproductive Performance and Egg Quality. Frontiers in Genetics 12.

- 65.Akopyan, M., Tigano, A., Jacobs, A., Wilder, A.P., Baumann, H., and Therkildsen, N.O. (2022). Comparative linkage mapping uncovers recombination suppression across massive chromosomal inversions associated with local adaptation in Atlantic silversides. Molecular Ecology 31, 3323–3341. 10.1111/mec.16472.

- 66.Charlesworth, B., and Charlesworth, D. (1973). Selection of new inversions in multi-locus genetic systems. Genet. Res. 21, 167–183. 10.1017/S0016672300013343.

- 67.Ohta, T. (1971). Associative overdominance caused by linked detrimental mutations. Genetics Research 18, 277–286. 10.1017/S0016672300012684.

- 68.Leroy, T., Anselmetti, Y., Tilak, M.-K., Bérard, S., Csukonyi, L., Gabrielli, M., Scornavacca, C., Milá, B., Thébaud, C., and Nabholz, B. (2021). A bird's white-eye view on avian sex chromosome evolution. Peer Community Journal 1, e63. 10.24072/pcjournal.70.

- 69.Sigeman, H., Ponnikas, S., and Hansson, B. (2020). Whole-genome analysis across 10 songbird families within Sylvioidea reveals a novel autosome–sex chromosome fusion. Biology Letters 16, 20200082. 10.1098/rsbl.2020.0082.

- 70.Pala, I., Naurin, S., Stervander, M., Hasselquist, D., Bensch, S., and Hansson, B. (2012). Evidence of a neo-sex chromosome in birds. Heredity 108, 264–272. 10.1038/hdy.2011.70.

- 71.Kitano, J., Ross, J.A., Mori, S., Kume, M., Jones, F.C., Chan, Y.F., Absher, D.M., Grimwood, J., Schmutz, J., Myers, R.M., et al. (2009). A role for a neo-sex chromosome in stickleback speciation. Nature 461, 1079–1083. 10.1038/nature08441.

- 72.Liakath-Ali, K., Vancollie, V.E., Sequeira, I., Lelliott, C.J., and Watt, F.M. (2019). Myosin 10 is involved in murine pigmentation. Exp Dermatol 28, 391–394. 10.1111/exd.13528.

- 73.Bennett, R.D., Mauer, A.S., Pittelkow, M.R., and Strehler, E.E. (2009). Calmodulin-Like Protein Upregulates Myosin-10 in Human Keratinocytes and Is Regulated during Epidermal Wound Healing In Vivo. Journal of Investigative Dermatology 129, 765–769. 10.1038/jid.2008.288.
- 74.Liu, Y., Zhang, Y., Li, T., Han, J., and Wang, Y. (2020). The tight junction protein TJP1 regulates the feeding-modulated hepatic circadian clock. Nat Commun 11, 589. 10.1038/s41467-020-14470-2.
- 75.Mazerbourg, S., Sangkuhl, K., Luo, C.-W., Sudo, S., Klein, C., and Hsueh, A.J.W. (2005). Identification of Receptors and Signaling Pathways for Orphan Bone Morphogenetic Protein/Growth Differentiation Factor Ligands Based on Genomic Analyses*. Journal of Biological Chemistry 280, 32122–32132. 10.1074/jbc.M504629200.
- 76.Zhu, A.X., Zhao, Y., and Flier, J.S. (1994). Molecular Cloning of Two Small GTP-Binding Proteins from Human Skeletal Muscle. Biochemical and Biophysical Research Communications 205, 1875–1882. 10.1006/bbrc.1994.2889.
- 77.Seutin, G., White, B.N., and Boag, P.T. (1991). Preservation of avian blood and tissue samples for DNA analyses. Canadian Journal of Zoology 68.
- 78.Grabherr, M.G., Russell, P., Meyer, M., Mauceli, E., Alföldi, J., Di Palma, F., and Lindblad-Toh, K. (2010). Genome-wide synteny through highly sensitive sequence alignment: Satsuma. Bioinformatics 26, 1145–1151. 10.1093/bioinformatics/btq102.
- 79.Cornetti, L., Valente, L.M., Dunning, L.T., Quan, X., Black, R.A., Hébert, O., and Savolainen, V. (2015). The Genome of the "Great Speciator" Provides Insights into Bird Diversification. Genome Biol Evol 7, 2680–2691. 10.1093/gbe/evv168.
- 80.Rhie, A., McCarthy, S.A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., et al. (2021). Towards complete and error-free genome assemblies of all vertebrate species. Nature 592, 737–746. 10.1038/s41586-021-03451-0.
- 81.Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12. 10.14806/ej.17.1.200.
- 82.Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12, 357–360. 10.1038/nmeth.3317.
- 83.Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29, 644–652. 10.1038/nbt.1883.
- 84.Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Research 34, W435–W439. 10.1093/nar/gkl200.
- 85.Smit, A., Hubley, R., and Green, P. (2013). RepeatMasker Open.
- 86.Kent, W.J. (2002). BLAT—The BLAST-Like Alignment Tool. Genome Research 12, 656–664.
- 87.Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., and Zdobnov, E.M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Molecular Biology and Evolution 38, 4647–4654. 10.1093/molbev/msab199.
- 88.Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics 10, 421. 10.1186/1471-2105-10-421.
- 89.Griffiths, R., Double, M.C., Orr, K., and Dawson, R.J.G. (1998). A DNA test to sex most birds. Mol Ecol 7, 1071–1075. 10.1046/j.1365-294x.1998.00389.x.
- 90.Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890. 10.1093/bioinformatics/bty560.
- 91.Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26, 589–595. 10.1093/bioinformatics/btp698.
- 92.Johnson, P.L.F., and Slatkin, M. (2007). Accounting for Bias from Sequencing Error in Population Genetic Estimates. Molecular Biology and Evolution 25, 199–206. 10.1093/molbev/msm239.
- 93.Björklund, M. (2019). Be careful with your principal components. Evolution 73, 2151–2158. 10.1111/evo.13835.
- 94.Rising, J.D., and Somers, K.M. (1989). The Measurement of Overall Body Size in Birds. The Auk 106, 666–674.
- 95.Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software 80, 1–28. 10.18637/jss.v080.i01.

- 96.Pritchard, J.K., and Przeworski, M. (2001). Linkage Disequilibrium in Humans: Models and Data. The American Journal of Human Genetics 69, 1–14. 10.1086/321275.
- 97.Fox, E.A., Wright, A.E., Fumagalli, M., and Vieira, F.G. (2019). ngsLD: evaluating linkage disequilibrium using genotype likelihoods. Bioinformatics 35, 3855–3856. 10.1093/bioinformatics/btz200.
- 98.Jørsboe, E., and Albrechtsen, A. (2022). Efficient approaches for large-scale GWAS with genotype uncertainty. G3 Genes|Genomes|Genetics 12, jkab385. 10.1093/g3journal/jkab385.
- 99.Barton, N., Hermisson, J., and Nordborg, M. (2019). Why structure matters. eLife 8, e45380. 10.7554/eLife.45380.
- 100.Meisner, J., and Albrechtsen, A. (2018). Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. Genetics 210, 719–731. 10.1534/genetics.118.301336.
- 101.Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet 81, 559–575.
- 102.Kirkpatrick, M. (2010). How and Why Chromosome Inversions Evolve. PLOS Biology 8, e1000501. 10.1371/journal.pbio.1000501.
- 103.Mérot, C., Berdan, E.L., Cayuela, H., Djambazian, H., Ferchaud, A.-L., Laporte, M., Normandeau, E., Ragoussis, J., Wellenreuther, M., and Bernatchez, L. (2021). Locally Adaptive Inversions Modulate Genetic Variation at Different Geographic Scales in a Seaweed Fly. Molecular Biology and Evolution 38, 3953–3971. 10.1093/molbev/msab143.