Lifetime fitness is correlated more strongly with structural variant than SNP mutational load in a threatened bird species

Katarina C. Stuart[1,2*], Hui Zhen Tan[1], Annabel Whibley[1,3], Sarah Bailey[1], Patricia Brekke[4], John G. Ewen[4], Selina Patel[1], Anna W. Santure[1]

[1] Ecology and Evolutionary Biology Group, School of Biological Sciences, University of Auckland, Auckland, 1142, Aotearoa New Zealand

[2] University of New South Wales, Sydney, Australia

[3] Bragato Research Institute, Blenheim, Aotearoa New Zealand

[4] Institute of Zoology, Zoological Society of London, London, UK

*Correspondence: Katarina.stuart@auckland.ac.nz

Abstract

The field of conservation genomics is becoming increasingly interested in whether, and how, structural variant (SV) genotype information can be leveraged in the management of threatened species. The functional consequences of SVs are more complex than for single nucleotide polymorphisms (SNPs), as SVs typically impact a larger proportion of the genome due to their size, and thus may be more likely to contribute to load. While the impacts of SV-specific genetic load may be less consequential for large populations, the interplay between weakened selection and stochastic processes mean that smaller populations, such as those of the threatened Aotearoa hihi/New Zealand stitchbird (*Notiomystis cincta*), may harbour a high SV load. Hihi were once confined to a single remnant population, but have been reestablished into six sanctuaries and reserves, often via secondary bottlenecks, resulting in low genetic diversity, low adaptive potential and inbreeding depression. In this study, we use whole genome resequencing of 30 individuals from the Tiritiri Matangi population to identify the nature and distribution of both SNPs and SVs within this small avian population. We find that SNP and SV individual mutation load is only moderately correlated, likely because SVs arise in regions of high recombination and that are less evolutionarily conserved. Finally, we leverage a long-term monitoring dataset of pedigree and fitness data to assess the impact of SNP and SV mutation load on individual fitness, and demonstrate that SV load correlates more strongly than SNP load with lifetime fitness. The results of this study indicate that only examining SNPs neglects important aspects of intraspecific variation, and that studying SVs has direct implications for linking genetic diversity and genomic health to inform management decisions.

1 | Introduction

Inclusion of genomic data into conservation management is becoming more common practice. For wild unmanaged species and populations, this is being driven by the increasing recognition of the ability of genomic data to reveal many biologically relevant patterns within species, such as cryptic information about animal dispersal (Alves et al., 2023), demographics (Zhao et al., 2013), and local adaptation (Yang et al., 2016), which may provide valuable information to guide population delineations and management practices. For species undergoing more active management such as breeding programs, assisted translocation, and genetic rescue, genomic data is essential to facilitate optimal conservation outcomes. These data allow for quantification of inbreeding levels (Hedrick et al., 2016), assessing the impacts of deleterious mutations (Humble et al., 2023), and monitoring the outcomes of breeding programs (Frankham, 2008; Lynch & O'Hely, 2001), which may greatly impact the long-term survival of a population of interest.

As conservation managers increasingly incorporate genomic data into practice, there is growing discussion around the relative utility of different genomic sequencing approaches (Theissinger et al., 2023). Amongst these is whole genome resequencing (WGR), which involves sequencing of the entire genomic DNA present within an organism. Despite a steady decrease in sequencing costs over recent decades, WGR may still be prohibitively expensive for population-level sequencing in many systems (Fuentes-Pardo & Ruzzante, 2017). WGR has been primarily applied in the study of health (Turro et al., 2020), model organism systems (Cao et al., 2011) and for agriculturally important species (Usha et al., 2022), however its application in wild populations has been gaining popularity. Application of WGR data to conservation genomics has revealed that while WGR may reflect the same broad patterns seen in more economical approaches (Duckett et al., 2023), the high density of genetic markers is invaluable for some genomic analysis such as inferring the landscapes of inbreeding (Duntsch et al., 2021) and genomic divergence (Szarmach et al., 2021). An important question then remains of whether, given this extra cost, WGR has unique utility to conservation biology practice, and if so, under what set of circumstances (Fuentes-Pardo & Ruzzante, 2017; Haig et al., 2016; Kardos et al., 2021; McMahon et al., 2014).

One area in which WGR plays a vital role is in the profiling of complex genomic variants, including structural variants (SVs), alongside single nucleotide polymorphisms (SNPs) (Kardos

et al., 2021). SVs typically span tens of bases to multiple megabases (Scherer et al., 2007), and are increasingly being recognised as important components of intra-specific genetic variation (Catanach et al., 2019; Wellenreuther et al., 2019; Wold et al., 2023). The functional consequences of SVs may be more complex than for single nucleotide polymorphisms (SNPs), as SVs typically impact a larger proportion of the genome (Collins et al., 2020), may harbour and accumulate SNPs and other small genetic variants within their length (Mahmoud et al., 2019), and are more likely to be deleterious when compared to SNPs (Hämälä et al., 2021).

To this end, the field of conservation genomics is becoming increasingly interested how SV genotype information should be leveraged in the management of threatened species. Population-level profiling of genome-wide SVs require a minimum of moderate depth short-read WGR data, though they still remain technically challenging to identify (Mahmoud et al., 2019). The SVs may not be in strong linkage disequilibrium with the SNPs around them (Pang et al., 2010), so it is not guaranteed that focal structural variants can have their genotypes inferred from neighbouring or overlapping SNPs. Examining SNPs alongside SVs has revealed instances where patterns in one type of genetic variation are not reflected in the other (Catanach et al., 2019; Stuart et al., 2023). Ongoing research into intra-specific structural variants is continuing to demonstrate the important role they play in, for example, facilitating convergent adaptation (Kreiner et al., 2019), and revealing previously hidden genotype associations (e.g. environment (Y. Li et al., 2024) or phenotype (Ruigrok et al., 2022)). Thus, the use of WGR data is essential to provide a holistic representation of genome wide patterns of genetic diversity or to identify specific variants associated with an ecotype or deleterious/beneficial functional trait.

One pressing area of research is the relative impacts of SNP and SV variants within small populations, particularly in relation to concepts like inbreeding and genetic mutational load (the accumulation of deleterious variants) (Hohenlohe et al., 2021). In small populations, novel variants arise at initially higher frequencies and then may increase in frequency through drift. Conversely, in large populations novel variants arise at a low frequency and the change in frequency of the variants depend on the selection pressures present (Lohmueller, 2014). Because SVs are less likely to be neutral than SNPs, we may expect that the weakened selection within small populations (Edmonds et al., 2004) has the potential to produce differential patterns in these two different types of variants. Understanding how genetic mutational load, might differ when quantified using SNPs and SVs will be vital for linking

underlying patterns of genetic diversity to functional traits related to individual fitness, population persistence, and adaptive potential.

Characterising genetic load is particularly relevant in threatened species, such as those of the hihi/stitchbird (*Notiomystis cincta*) (Fig. 1a). This species once covered much of Aotearoa New Zealand's North Island, but their natural range had been reduced to a single offshore island, Te Hauturu-o-Toi by the 1880s (Fig. 1a; Brekke et al., 2011). This single remnant population has been used to establish a primary breeding colony on the offshore island of Tiritiri Matangi (Miskelly & Powlesland, 2013), from which most additional populations have been reestablished or supplemented (Brekke et al., 2011). However, the long-term success of established populations has been variable, with some experiencing limited population growth post-founding, and others failing to establish (Ewen et al., 2013; Parlato et al., 2021). While population success is dependent in part on extrinsic factors (e.g. Doerr et al., 2017), inbreeding and low adaptive potential within the species are also likely to be contributing to poor population health (de Villemereuil et al., 2019; Duntsch et al., 2023). The Tiritiri Matangi population is intensively monitored, including recording pedigree information and lifetime fitness, and so the hihi provides a means to better understand the differential impacts of SNP and SV genetic load on important conservation metrics such as fitness.

In this study, we use whole genome resequencing of hihi individuals from the Tiritiri Matangi population to identify the nature and distribution of both SNPs and SVs within this small avian population. We quantify mutation load and assess how well individual SNP load acts as a proxy for SV load. We also examine the genomic location of SVs with respect to underlying patterns of recombination and avian genomic evolutionary conservation, to better understand how SVs arise and are maintained within the genomes of a species. Finally, leveraging a long-term monitoring dataset of fitness data for each individual, we test the impact of SNP and SV mutation load on lifetime fitness. We hypothesise that relative patterns in these two types of variants will reveal that SVs have a greater impact on lifetime fitness, reflecting the more deleterious nature of SVs in comparison to SNPs.

2 | Methods

2.1 | Sampling

We selected a total of 31 hihi (*Notiomystis cincta*) individuals for inclusion in the study. These individuals comprise four hihi sampled from the remnant population on Te Hauturu-o-Toi / Little Barrier Island, Aotearoa / New Zealand (36°12' S, 175°05' E) and 27 sampled from the reintroduced population on the island of Tiritiri Matangi (36°36' S, 174°53' E) (Fig. 1a). Of the four Te Hauturu-o-Toi individuals, two were translocated to Tiritiri Matangi in 2010 as a genetic rescue attempt (Nichols et al., 2024), both of which successfully bred in the population. The remaining two were sampled in 2017 and 2018 as part of a population diversity survey on Te Hauturu-o-Toi and were subsequently selected as our reference male and reference female for genome assembly (Bailey et al., 2023). Individuals on Tiritiri Matangi were sampled in several cohorts from 2001-2011 and were chosen as individuals that successfully bred and therefore are likely to represent diversity present in the current population. We note that because many hihi individuals do not breed successfully, this study assesses how load impacts the level of reproduction in known breeding individuals, rather than whether load impacts the occurrence of successful breeding. Blood samples were taken at capture of all individuals via brachial venepuncture and stored in 95% ethanol. All individuals have been previously genotyped at a panel of 20 microsatellite markers (Brekke et al., 2009), which enables genetic sexing of the birds alongside inclusion in a genetically resolved pedigree for the entire Tiritiri Matangi population. The dataset includes 17 females, 14 males, and three close family groups; a parent-offspring duo, a half-sibling pair and their genotyped father, and eleven focal individuals connected in a multi generation family, which includes two parent-offspring trios and four further parent-offspring duos (family groups noted in Table S1).

We extracted gDNA from the blood samples of the 31 individuals with a Qiagen DNeasy Blood and Tissue Kit. Library preparation and whole genome resequencing (WGR) was conducted by AgResearch, New Zealand, in two experiments; in the first, libraries were prepared for sequencing with an Illumina kit, in the second with a Nextera XT library preparation that reduced adaptor contamination apparent in the first run. All samples (Table S1) were resequenced on the Illumina NovaSeq 6000 platform with S4 flowcells (150 bp paired end reads).

2.2 | Single nucleotide polymorphism calling

We merged raw reads across the two experiments. We then used TRIM_GALORE v0.6.7 (Krueger, 2021) to process the merged reads using default settings and the flags '--paired' for paired reads

and '--fastqc' to produce read quality reports that were then compiled and assessed with MULTIQC v1.13 (Ewels et al., 2016). We mapped these processed reads to the female *N. cincta* genome assembly version (Bailey et al., 2023) using the mapping algorithm BWA v0.7.17 (H. Li, 2013) *mem*. We used SAMTOOLS v1.16.1 (H. Li et al., 2009) to process the output files into sorted BAM files, before marking duplicate reads within each file using PICARD v2.26.10 (*Picard Toolkit*, 2019) *MarkDuplicates*. Variants were then called using the program BCFTOOLS v1.13 (Danecek et al., 2021), first using the *mpileup* (-C 50 -q 20 -Q 25) function to generate genotype likelihoods and coverage information, then the *call* (-f GQ) function to produce SNP calls, *view* (--exclude-types indels) function to retain only SNPs, and finally the *sort* function to order the SNPs into chromosome and positional order. We filtered the remaining variants using VCFTOOLS v0.1.15 (Danecek et al., 2011) to retain only bi-allelic SNPs (--max-alleles 2) with a minimum depth of 5 (--minDP 5) and a maximum mean depth of 100 (--max-meanDP 100), removing non-variant sites (--mac 1) and sites exceeding a missingness of 10% (--max-missing 0.9). We also filtered out all Z and W chromosome contigs, and excluded one individual that had very high heterozygosity estimates most likely due to sample contamination (individual 83318; contamination was apparent from both sequencing experiments). Our total dataset was therefore 30 birds, 26 of which were from Tiritiri Matangi and 4 of which were from Te Hauturu-o-Toi.

2.3 | Structural variant identification

In order to identify the SV content within the 30 hihi individuals, we reanalysed the duplicate-marked WGR data using the three short-read SV calling programs SMOOVE v0.2.8 (https://github.com/brentp/smoove), which is a wrapper for LUMPY-SV (Layer et al., 2014), DELLY v0.8.2 (Rausch et al., 2012), and MANTA v1.6.0 (X. Chen et al., 2016). Calling SVs is based on several lines of evidence produced by sequence data, including read depth and coverage, as well as the mapping patterns of read-pairs and split-reads to infer the existence of a SV relative to the reference genome. Identifying SVs from read data may contain many false positive calls (Chakraborty et al., 2019; Weissensteiner et al., 2020; Wold et al., 2023), and thus often a consensus approach is used where a particular variant is only retained if it is identified using two independent variant caller programs. Specifically for this study, we chose to retain only SVs that, after passing the individual program's quality filters, were found in a minimum of two out of the three variant calling programs used above. Briefly below we describe the three programs and the quality filtering criteria used to exclude low quality SV calls.

We identified SVs using the program SMOOVE by following the population-level calling workflow recommended in the program manual. This approach involves a per-individual variant calling step where SVs are identified within each sample independently. The individual's SV calls are then combined into one merged cross-sample SV list, which are then genotyped within each individual separately before being combined again into a dataset containing all individuals. We also included the optional annotation step, which uses the exon and untranslated region (UTR) information in the genome annotation matching the *N. cincta* reference assembly to assign quality scores to genotyped variants. Using these quality scores, we filtered deletions (DEL) to FMT/DHFFC < 0.7, duplications (DUP) to FMT/DHBFC > 1.3, and all genotypes to MSHQ>=3, following Wold et al. (2023), while also only retaining variants on autosomal chromosomes.

SVs were identified using the program DELLY by following the recommended germline SV calling workflow. The workflow is similar to that used by SMOOVE, and involves a per-individual SV calling step (DELLY *call*), a cross-sample SV calling merge step (DELLY *merge*), a per-sample genotyping step (DELLY *call*) and finally a cross-sample SV genotype merging step which combines all the individual SV calls into one variant file (BCFTOOLS *merge* -m id). We then filtered the resulting variant calls to retain only those where a minimum of 50% of individuals were assigned a per-individual 'PASS' score on their genotype, while also only retaining variants on autosomal chromosomes.

We identified SVs using the program MANTA by following the default pipeline, which involved using *configManta.py* to generate and run the MANTA workflow. We ran the initial variant file through the MANTA function *convertInversion.py* in order to extract inversions (INV) from breakend calls, which is not automatically done by the program. We then filtered variants using BCFTOOLS *view*, retaining only those variants that were assigned a per-variant 'PASS' score, while also only retaining variants on autosomal chromosomes.

We generated the consensus variant list using the program SURVIVOR v1.0.7 (Jeffares et al., 2017), which merged the SVs identified by each program using several criteria, including the maximum allowed distance between predicted breakends of 1 kb, two SV caller support needed to retain a call, that the type and strand of the SV must agree, and a minimum SV length of 30 bp (parameter values set as "1000 2 1 1 0 30"). Merging was done separately for each individual, and we also forced SURVIVOR to only merge SV calls that agreed on genotype by merging independently for homozygous reference, heterozygous, and homozygous alternate SV calls within samples. Finally, we merged the retained variants into a joint sample file using the

parameters “1000 1 1 1 0 30” (setting the minimum SV caller support to 1 to create a merged variant file).

2.4 | Structural variant curation and validation

The SURVIVOR consensus SV list was then filtered to only include deletion (DEL), duplication (DUP) and inversion (INV) calls, as these were the three variant types that all three SV calling programs could identify. We also removed large variants (>50,000 bp) and those with >50% missing data across the 31 individuals. We also removed any SV that had no heterozygous individuals, as it is unlikely that a polymorphic SV would exist without heterozygous individuals being present. Similarly to SNPs, we also removed a contaminated sample from the dataset at this step (individual 83318).

We manually curated these SVs to exclude false variants by visualising the raw sequence reads across the identified SV loci for two individuals of each genotype (homozygous reference, heterozygous, and homozygous alternate) where available. The visualisation was produced using SAMPLOT v1.3.0 (Belyeu, Chowdhury, et al., 2021) and then scored using PLOTCRITIC v1.0.1 (Belyeu et al., 2018) by one curator to check for signals of SVs in the sequencing read depth and insert sizes (Fig. S1-S6).

Because the Tiritiri Matangi population is small (~200 adults, Parlato et al., 2021), its pedigree is highly connected, and as noted above, our data contained several related individuals, with over half of the dataset organised into three family groups. Of these family groups, the large multi-generation family included two parent-offspring trios with all three individuals resequenced. We did not remove relatives from the dataset, but did use them to assess genotyping error rates and to validate our manual curation. The program PLINK v1.09b6.16 (Purcell et al., 2007) was employed to calculate Mendelian error rates using the two WGR trios using the ‘--mendel’ flag. We assessed Mendelian error rates before and after manual curation, using both the two true parent-offspring trios, as well as two fake trios. Mendelian error rates reported were the averages of the two offspring individual error rates, which were each calculated by counting the number of Mendelian errors present in the offspring divided by the total number of genotyped sites within the offspring.

2.5 | Structural variant filtering and profiling

Some of the inter-individual variants detected by the above structural variant workflow will be due to TE activity which has a different impact profile to SVs (Stuart et al., 2024). In this study we are seeking to examine SV signals only, and hence removed any SV variants that had likely

resulted from TE activity (for further discussion on this see section 4.4: Technical validation of SVs and exclusion of TEs). For this we ran REPEATMASKER v4.1.0 (Smit et al., 2013) on the SV breakends (+/- 30 bp) using the *N. cincta* repeat library (Bailey et al., 2023), as repeat identification in the genomic region around a breakpoint will allow for identification of variants that are really TEs, rather than true SVs that simply contain a TE signal within their length.

Variants that had been identified as a TE (classification: LINE, LTR, SINE, DNA) were removed from the SV data set, and then this data was filtered in VCFTOOLS to remove any non-variant sites (--mac 1) and variants with more than 50% missingness (--max-missing 0.5).

We used PLINK to run a pca (--pca) on both the SNP and SV data. In addition, we compared heterozygosity based inbreeding (VCFTOOLS --het), minor allele site frequency spectra (PLINK --freq), and per-variant genotype frequencies for SNPs and SVs. For these analyses and all others in the paper, the minor allele was used as the alternate allele.

2.6 | Quantifying mutational load

We sought to quantify per-individual genetic load for both SNPs and SVs, to examine the correlation between load calculated from these two different genetic marker types. We first needed to identify the positioning of variants relative to protein coding regions, as traditionally only variants close-to or overlapping coding regions are considered as contributing to load (Smeds et al., 2024). For this we used the variant effect predictor program VEP v107.0 (McLaren et al., 2016) to profile the impacts of the variants, and then quantified variants into one of four classes, based on their positioning relative to coding sequences. This included 'Gene' (variants occurring in exons), 'Intron' (variants occurring in introns), 'Up-/Down-stream' (variants <5kb up- and down-stream of a protein coding region), with all remaining variants being classed as 'None'. Variants of the three classes 'Gene', 'Intron', and 'Up-/Down-stream' were considered to be variants contributing to load.

We quantified per-individual mutational load as both masked load and realised load. Masked load was quantified as the proportion of load contributing variants that were heterozygous. Realised load was quantified as the proportion of load contributing variants that were homozygous for the minor allele. Both load calculations were done separately for SNPs and SVs. Unfortunately, no close species of the hihi has been whole genome resequenced, and additionally SVs often occur in more species-unique regions of the genome (see Fig. 4 below) which means the use of more distant relatives is likely to bias the polarisation of SVs compared to SNPs. Therefore, polarising these data was not possible, and so we use minor allele here in

place of derived allele. We compared how well correlated per-individual load was across SVs and SNPs, by performing two regressions between the two comparing masked, and then realised load using the *lm*() function from the TIDYR package in R.

2.7 | Recombination rates and linkage disequilibrium

We next sought to examine the genomic distribution of SVs within hihi, specifically in relation to underlying recombination rates. The relationship between SVs and recombination rates is a complex one, as recombination can give rise to SVs (Carvalho & Lupski, 2016), however the presence of large SVs may inhibit recombination (Morgan et al., 2017). As part of our quality control of SVs, acknowledging the challenges of identifying large SVs from short read data, we excluded all SVs larger than 50kb, so our SV dataset is on average small and therefore unlikely to inhibit recombination. The dataset is therefore most likely representative of the recombination landscape in which SVs are likely to arise. We therefore tested to see if regions of the genome with SVs present had higher recombination rates, with recombination rates summarised across 500Mb regions of the genome based on the hihi linkage map developed from pedigree (36 families) and SNP chip (33,890 autosomal SNPs) data (Tan et al., 2024). We analysed this separately across macro- (chromosomes 1-7,1A) and micro-chromosomes (8-29, 4A) because they have very different recombination rates (chromosomal classification as per Bailey et al. 2023). We identified that SVs were present in 450 of the 1,251 macrochromosome recombination windows, and for 221 of the 562 microchromosome windows.

Additionally, we examined patterns of linkage disequilibrium decay around SVs and contrasted these to those seen around SNPs. For this, we combined all SV and SNP markers into one file, and assessed the linkage disequilibrium for the SVs to the surrounding variants using the vcftools function --geno-r2-positions, specifying and assessing pairwise linkage disequilibrium up to 10 Mb away (--ld-window-bp 10000000). We also assessed SNP linkage disequilibrium decay using the same metrics, creating five random subset of 935 SNPs which were averaged. Finally, in order to better understand how processes of variant creation and selection may be differentially impacted by the genomic landscape of macro- and micro-chromosomes, we assessed the minor allele frequency of SNPs and SVs that do and do not contribute to load.

2.8 | Evolutionary conservation analysis on multiple whole-genome alignments

Next, we sought to identify where load contributing SNPs and SVs were positioned within the hihi genome, with respect to genomic conservation across other avian species. For this, we examined the hihi genome alongside additional passerine species downloaded from NCBI

including *Eopsaltria australis* (GCA_034509425.1), *Passer dom*esticus (GCF_036417665.1), *Hirundo rustica* (GCF_015227805.2), and *Taeniopygia guttata* (GCF_003957565.2), with *Gallus gallus* (GCF_016699485.2) as an additional species to serve as an outgroup. We used TimeTree v5 (Kumar et al., 2022) to create a Newick tree for these species, and used REPEATMASKER to soft mask all the genomes using the Aves lineage specific sequences. We then aligned these six genomes using the program PROGRESSIVE CACTUS v2.7.2 (Armstrong et al., 2020). We used the function *halAlignmentDepth* from the HAL v2.3 (Hickey et al., 2013) toolkit implemented in PROGRESSIVE CACTUS to calculate per base genome alignment coverage of the hihi genome. This reports the number of other species that successfully mapped to each base position of the hihi genome. A genome alignment value of 0 is indicative of a region that did not align across other species and is unique to the hihi genome relative to the other species examined. A genomic alignment value of 5 indicates that that all the other species' genomes aligned and is interpreted as a conserved region.

We examined the level of genomic alignment underlying the SNPs and SVs, across the four classes of 'Gene', 'Intron', 'Up-/Down-stream', and 'None' that were identified during load quantification. We used BEDTOOLS to identify the genomic alignment value for SNPs based on their positioning, while for SVs we used the average of their first and last base pairs so that the level of genomic alignment was based on their breakends and not their sequence identity. To quantify the genomic alignment values of SNPs and SVs relative to the whole genome, we also calculated average evolutionary conservation for all 'Gene', 'Intron', 'Up/Down-stream' and 'None' sites genome-wide.

2.9 Impacts of lifetime fitness

Finally, leveraging a long-term monitoring dataset of pedigree and lifetime fitness data, which was available for 28 out of the 30 individuals used in this study (because 2 of the Te Hauturu-o-Toi birds were translocated and bred on Tiritiri Matangi), we assessed the impact of SNP and SV mutation load on lifetime fitness. Lifetime fitness was measured as the total number of banded offspring of each individual across its lifetime, and therefore measures fitness across one life cycle, from banding to banding (de Villemereuil et al., 2019).

We set lifetime fitness as the dependent variable and ran three separate linear mixed models for SNPs and SVs each; one for realised load (i.e., minor allele homozygotes at load sites), one for masked load (i.e., heterozygous at load sites) and one for common allele homozygotes, quantified as the proportion of load contributing variants that were homozygous for the major allele (i.e., the remaining proportion of genotypes). In each model, the SNP / SV load was

included as a fixed effect along with individual sex, and sampling cohort was included as a random effect. To quantify the relative contribution of SNPs and SVs to load, we also ran a further two models where we included both variant types as fixed effects for each respective category of load (i.e., masked and realised). In both models, individual sex was again included as a fixed effect, and sampling cohort a random effect. Linear mixed models were run in R using the *lmer*() function in the package LME4 V 1.1-26 (Bates et al., 2015).

We sought to examine if the evolutionary conservation had an impact on load by following up significant results from the above models, of which only masked SV load was found to be significant. We partitioned the masked SV load contribution up into two components: those variants which fell into evolutionary conserved regions (where the average evolutionary conservation score of the two breakends averaged was four or larger), and the remaining variants. We ran a final linear mixed model as above, with just these two components of masked SV load, and with individual sex as a fixed effect and sampling cohort as a random effect.

3 | Results

3.1 | SV characterisation and validation

Across the three SV detection programs, DELLY identified the largest number of variants at 13,176 SVs, closely followed by MANTA AT 12,672 SVs, with SMOOVE identifying only 3,275 SVs (Table S2). From these, a total of 2,991 variants were identified in at least two out of the three programs and were retained after meeting consensus criteria.

After manual curation this number was reduced to 1,229. Mendelian inheritance error rates decreased from 1.6% to 0.9% after manual curation. We found that the Mendelian inheritance rates of the true parent-offspring trios were much higher than for the fake trios, which were 14.2% and 17.8% before and after manual curation respectively. Of these curated SVs, we identified 40.42% of them that were annotated as a type of repeat (Fig. 1b). Specifically, 293 of these were identified as SVs that were likely to be TEs, and these were removed from the final SV dataset, leaving a total of 936 variants. Of the identified TEs, LTRs were the most common TE class, followed by LINEs, with very few SINE or DNA transposons, reflecting the large proportion of LTRs and LINEs in the hihi genome (Bailey et al., 2023). Profiling of the SV length distributions pre- and post-TE filtering confirmed that TE associated peaks were removed during this filtering step (Fig. 1c) and identified that a majority of them fell in the range of 50-150 bp in length.

The cumulative length of the final filtered SV dataset was 420,304 bps, which was approximately one seventh of the coverage of the final filtered SNP data set that consisted of 3,111,629 SNPs. The PCA produced from SNPs and SVs were very similar to one another (Fig. S7), which both identified the Te Hauturu-o-Toi individuals distinct from those captured on Tiritiri Matangi. Additionally, individual heterozygosity based inbreeding metrics were well correlated across the two datasets (R-squared = 0.6291; Fig. S8), indicating that the SV dataset represented a collection of high confidence SVs.

The site frequency spectrum (SFS) of SNPs and SVs was very similar in showing very few rare alleles (Fig. 1d). However, the alternate/minor allele for SVs was more numerous than for SNPs, appearing as an increase in heterozygous genotypes compared to homozygous reference (Fig. 1e).

3.2 | Mutational load

The genome wide distribution of SNPs and SVs was overall fairly similar, and generally reflective of genome proportions of coding and non-coding regions, with the majority of variants either intergenic or within introns (Fig. S9). However, the proportion of variants of each type that were found to fall in coding or near coding regions was larger for SVs than SNPs (Fig. 2a, $F_{7,232}$=8056, p-value = <0.0001).

We calculated mutational loads for all 30 individuals in the study, and assessed how well load metrics were correlated with one another, as well as with individual inbreeding. From the total genome-wide SNPs and SVs, 1,694,577 (54.7%) and 548 (58.5%) respectively were identified as within or close to genes and therefore contributing to load (Figure 2a). SV load and SNP load is significantly correlated for both realised and masked load, with SV and SNP realised load being more highly correlated (Fig. 2c, $F_{1,28}$ = 51.29, p-value = <0.0001, R^2 = 0.6343) than SV and SNP masked load (Fig. 2b, $F_{1,24}$ = 23.59, p-value = <0.0001, R^2 = 0.4379). In general, levels of load in SVs and SNPs were more similar for realised load, while masked load values were slightly higher for SVs than SNPs.

3.3 | Recombination rates and structural variants

We identified that there was a positive association between recombination rate and the presence of SVs (Fig. 3a). We found that this difference was most pronounced on the microchromosomes, which in general had higher recombination rates even within regions without a SV present. Additionally, we found that pairwise linkage disequilibrium was slightly higher for SNPs than for SVs when examining linkage disequilibrium decay up to distances of 10

Mb (Fig. 3b). This pattern was more stochastic at smaller distances (<10kb), though the bin of variants closest to the SV breakends did see much lower linkage disequilibrium R^2 values than for SNPs (Fig. S10). We found that patterns of minor allele frequency within SVs and SNPs are reversed when split by load contribution and across chromosome type (macro- or micro-) (Fig. S12).

3.4 | Evolutionary conservation

The inter-species genome alignment was completed for five avian species alongside the hihi genome (Fig. 4a). We found that 61.1% of the hihi genome was evolutionary conserved across all the other avian species examined, as measured by at least four other species aligning to these sites, while 4.9% of the genome's length did not align to any other species and is unique to the hihi (Fig. S12). Patterns of genome conservation alignment were fairly consistent across all the hihi's autosomal chromosomes, though the proportion that was conserved across all species decreased as chromosome length also decreased (Fig. S12).

Genome wide patterns of genome alignment, which we use to benchmark patterns in SNPs and SVs, found that gene regions were the most evolutionarily conserved, up- and down-stream gene regions were the most unique, while intron and non-coding regions were similar (Fig. 4b). This pattern was consistent in SNPs and SVs, with the only deviation being that SNPs had higher levels of genomic alignment in the intron variants when compared to the variants that were not within or near coding regions ('None'). The average level of genomic alignment for both SNPs and SVs was lower than the genome-wide patterns (including both variant and non-variant sites), though this decrease in genomic alignment was more pronounced in SVs. This pattern is primarily due to an inflation of the number of SV breakends that occurred in evolutionarily unique regions of the hihi's genome, in comparison to SNPs (Fig. S13).

3.5 | Mutational load impact on lifetime fitness

We found a significant negative relationship between both SNP realised load (Fig. 5a) and SV (Fig. 5d) realised load with lifetime fitness (Table S3, β = -235.988, p-value = 0.0216 and β = -232.252, p-value = 0.0126 respectively). Contrastingly, we found a positive relationship between both SNP masked load (Fig. 5b) and SV masked load (Fig. 5e) with fitness, though SNPs had a p-value just above the significance threshold (Table S3, β = 168.152, p-value = 0.0622 for SNPs and β = 218.218, p-value = < 0.0001 for SVs). The contribution of common homozygote genotypes at load sites was less consistent in the SNP and SV datasets, with SNP homozygosity (Fig. 5c) not significantly correlating with fitness, while SV homozygosity (Fig. 5f) was negatively

and significantly correlated (Table S3, β = 0.7033, p-value = 0.9925 for SNPs and β = -137.680, p-value = 0.0124 for SVs).

To determine the relative contribution of SVs and SNPs to the load signals, we ran a further two linear mixed models, one fitting both SNP and SV masked mutational load, and one fitting both SNP and SV realised mutational load. Within the masked load model, we found that masked SV load was positively correlated with lifetime fitness (Table S3, β = 250.698, p-value < 0.0001), however this trend was neither significant nor in the same direction for masked SNP load (Table S3, β = -81.899, p-value = 0.3050). Within the realised load model, neither realised SV load (Table S3, β = -183.658, p-value = 0.1893) nor realised SNP load (Table S3, β = -65.136, p-value = 0.6680) was significantly associated with lifetime fitness, though both trends were consistently negatively associated with lifetime fitness.

We partitioned the SV masked load into evolutionary conserved and less conserved genomic regions and found that both these components of masked SV load were significantly and positively correlated with lifetime fitness (conserved regions: β = 233.937, SE= 67.190, t-value = 3.482, p-value = 0.003; less conserved regions: β = 201.120, SE= 70.017, t-value = 2.872, p-value = 0.009).

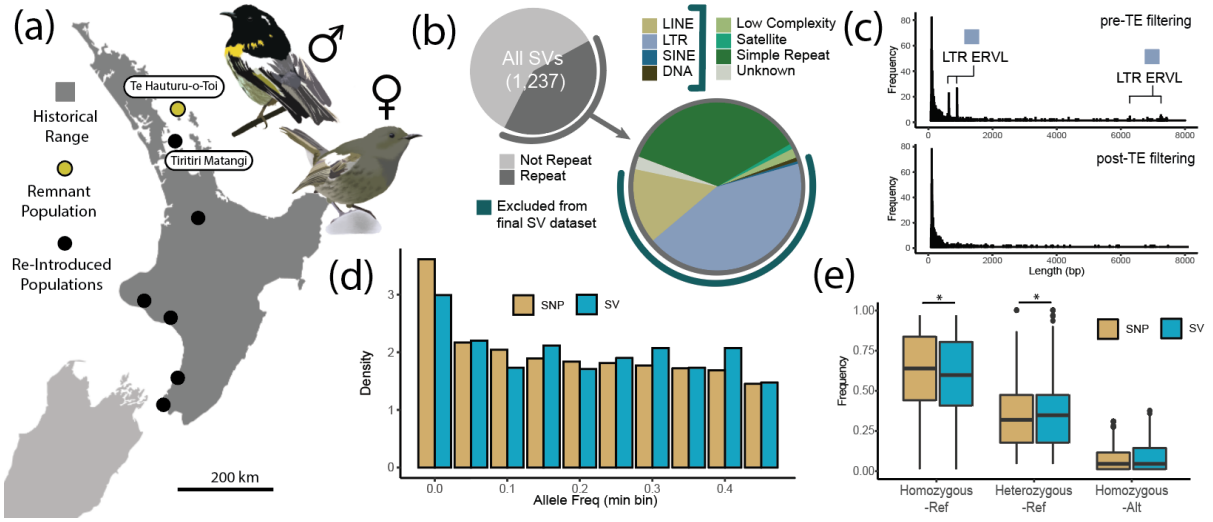
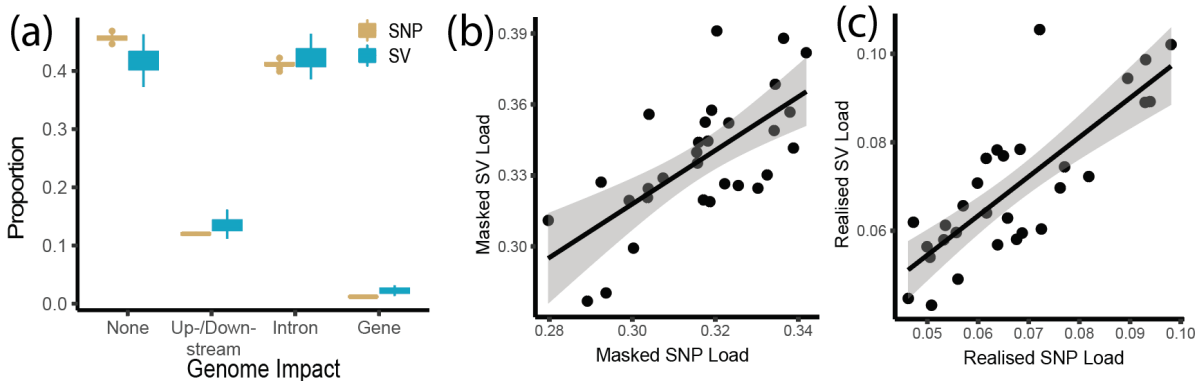


Figure 1 | Genetic diversity and genomic distribution profiles of single nucleotide polymorphisms (SNPs) and structural variants (SVs) for hihi (*Notiomystis cincta*) individuals. Panel **(a)** depicts the historic and present-day range of hihi on the North Island of Aotearoa New Zealand, with sampling locations from this study labelled with placenames. A depiction of the male and female hihi is also included (illustrations by Hui Zhen Tan). Panel **(b)** depicts the repeat annotations of the curated SVs. Panel **(c)** depicts the length profiles of the SVs pre- and post- transposable element (TE) filtering. The TE identity of the peaks in this plot are annotated with the primary TE responsible for the peak. Panel **(d)** is the site frequency spectrum of SNPs and SVs. Panel **(e)** is a boxplot of genotype frequencies across SNPs and SVs, averaged on a per-variant basis.



480

481 **Figure 2 | Patterns in variant impact and genetic load of single nucleotide polymorphisms**
482 **(SNPs) and structural variants (SVs).** Panel **(a)** is the proportion of SNP and SV minor allele
483 variants within individuals that has been categorised based on their overlap with gene regions**.**
484 Panels **(b)** and **(c)** show the significant correlation between SV load and SNP load for masked
485 and realised load respectively.

486

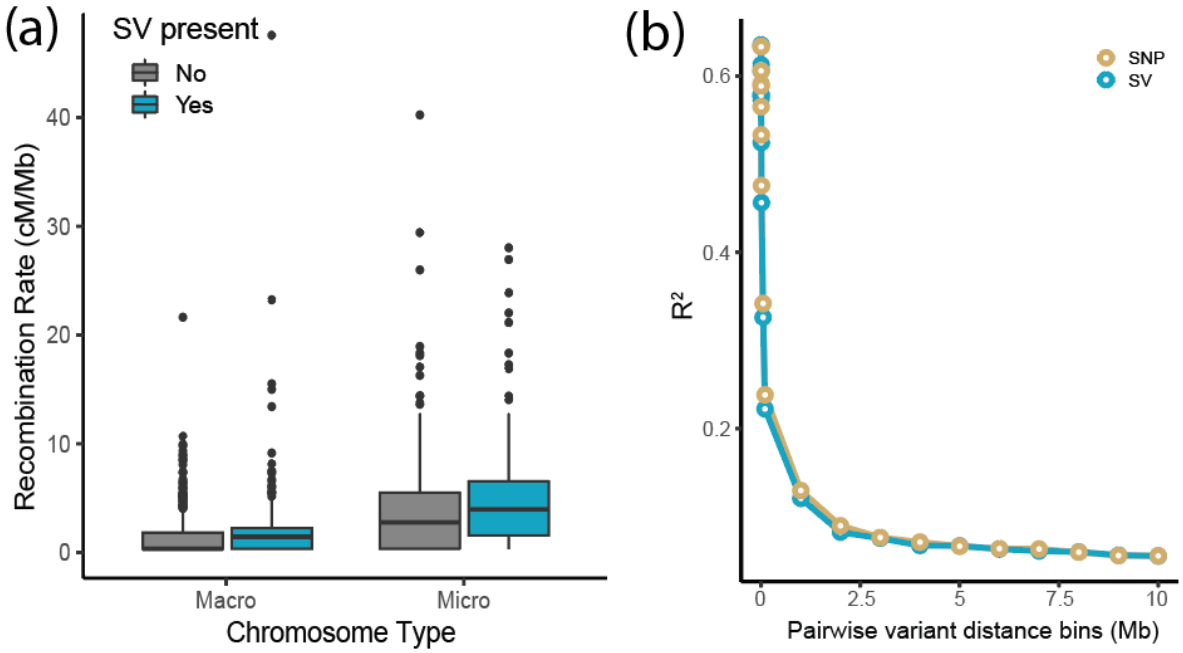
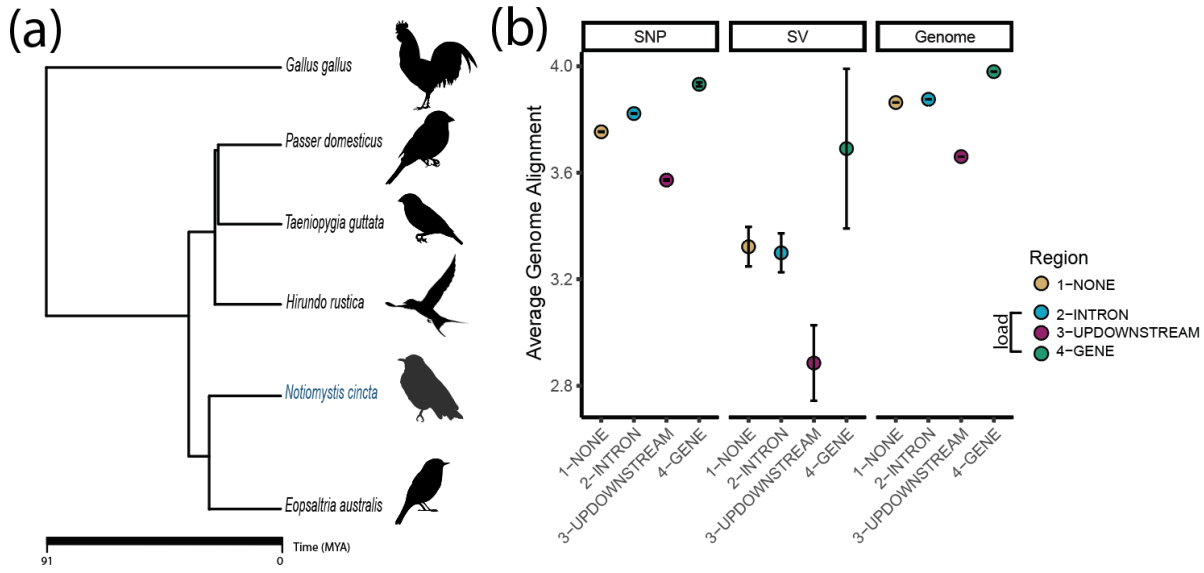


Figure 3 | Structural variants in hihi in relation to recombination rates, linkage disequilibrium, and minor allele frequency. Panel **(a)** displays average recombination rate across 500Mb regions of the genome, classified by whether at least one SV is present with that region or not, and across either macro or micro chromosomes. Panel **(b)** displays linkage disequilibrium decay up to 10 Mb away from the SV variant breakends (or the SNP position).



494

495 **Figure 4 | Evolutionary conservation and genome evolutionary conservation of the hihi**
496 **across genomic locations of structural variant (SVs) and single nucleotide polymorphism**
497 **(SNPs).** Panel (a) depicts the phylogenetic tree used in the PROGRESSIVE-CACTUS analysis, panel
498 **(b)** depicts the average cross species conservation of genomic regions where different variant
499 types and impact groups are located, where a value of 0 indicates a genomic sequence unique
500 to the hihi, and a value of 5 indicates a region that is conserved across all species analysed.
501 Pictures taken from https://www.phylopic.org/.

502

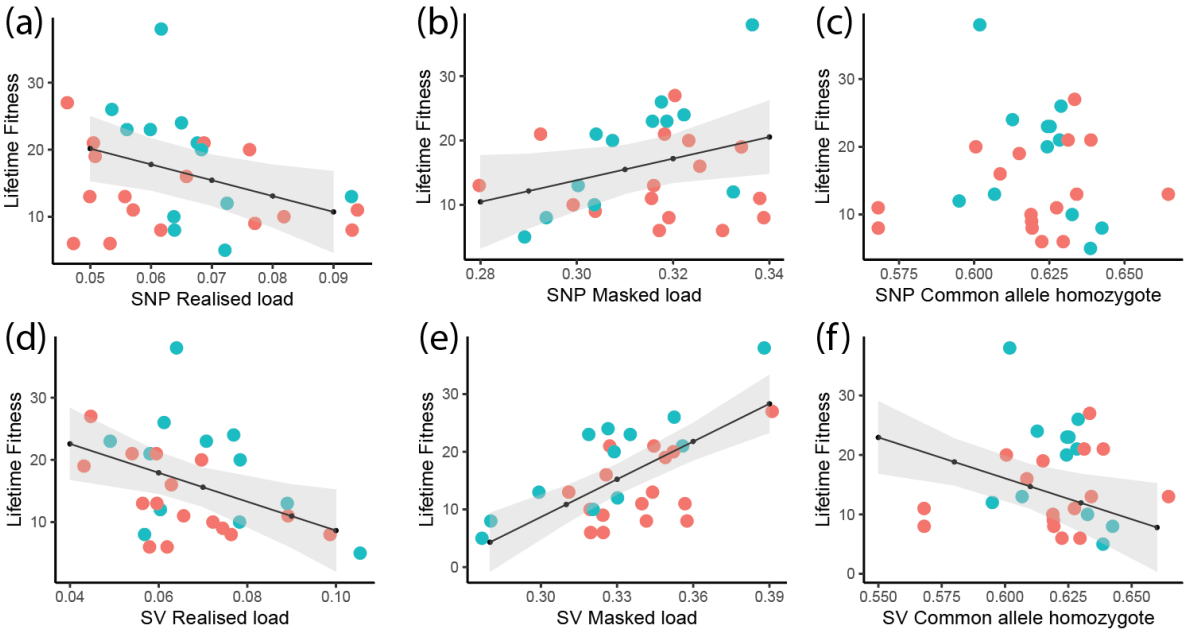


Figure 5 | Impact of structural variant (SVs) and single nucleotide polymorphism (SNPs) realised load, masked load, and common allele homozygotes proportion on lifetime fitness of hihi on Tiritiri Matangi. Scatterplot of the mutational load against the lifetime fitness, sex indicated by colour (female = red, male = blue), and mixed effect model fit lines indicated when the relationship with lifetime fitness was significant (or very near significance, in the case of SNP masked load). Two individuals from Te Hauturu-o-Toi were excluded from this analysis as no lifetime fitness values were available for them. Panel **(a)** shows SNP realised load, panel **(b)** shows SNP masked load, panel **(c)** shows SNP common allele homozygote proportion. Panel **(d)** shows SV realised load, panel **(e)** shows SV masked load, panel **(f)** shows SV common allele homozygote proportion.

4 | Discussion

Within small populations, conservation practitioners are particularly concerned with increased inbreeding, loss of genetic diversity, and greater mutational load burdens (Dussex et al., 2023). In order to best track and manage these, the development of genomics-informed best practices in conservation management requires that we understand how different components of the genome contribute to functional genetic diversity. Through whole genome resequencing approaches it is possible to quantify the more commonly studied single nucleotide polymorphisms (SNPs) alongside structural variants (SVs), allowing us to achieve a more holistic summary of inter-individual variation (Wellenreuther et al., 2019; Wold et al., 2021). In this study, we specifically focus on load contributing genetic variation in SNPs and SVs and find that masked and realised load is significantly but only moderately correlated across these two genetic variant types. When testing for a relationship between load and lifetime fitness, there was a significant negative relationship between realised load and fitness, and a significant positive relationship between masked load and fitness for both SNPs and SVs. Interestingly, when combining SNPs and SVs into the same model, SVs consistently had a stronger correlation with fitness. However, of the four terms tested, only masked SV load remained significantly correlated. This demonstrates that SVs are an important component of intra-specific variation in the highly inbred and endangered hihi population on Tiritiri Matangi, and that in order to optimise conservation genomic practice we may need to incorporate SV profiling.

4.1 | Mutational load may not be equivalent across SNPs and SVs in small populations

Our results demonstrate that patterns of reduced genetic diversity in the hihi are just as evident in SV as SNPs (Fig. S8) (Duntsch et al., 2021). Within the current study we have not identified SVs as hidden source of unexpected genome wide patterns (Fig S7) or rare variants (Fig. 1d) within the species, as has been found in some studies (Berdan et al., 2021; Stuart et al., 2023; Weissensteiner et al., 2020). The observed moderate correlation between SNP and SV load (44% for masked and 63% for realised load) indicates that within individuals, while one variant type's load may be somewhat predicted by the other, SNPs do not act as a perfect proxy for SV load (Fig. 2b & 2c). The moderate correlation could be explained by the differential heterozygosity patterns between the marker types, with a slightly higher prevalence of heterozygous SV than SNP genotypes in the population, alongside an enrichment of SVs in regions of the genome with higher recombination rates (Fig. 1e, Fig. 3a). The deviation in individual SNP and SV load contributes to differential patterns when linking individual load to

individual phenotype within the 28 assessed individuals. When testing both SNPs and SVs concurrently, we identified only masked SV load as statistically correlated with lifetime fitness within the successful hihi breeders used in this study (Fig. 5, Table S3) and found that this pattern was apparent in both evolutionary conserved and unique genomic regions. Further, though the joint SV and SNP realised load model was not significant, the magnitude of the effects (β) suggests that SVs once again contributed more strongly to the negative correlation with fitness than did SNPs (Table S3).

Genetic load in small, threatened populations is a result of the interaction between several factors. Reduced efficacy of selection in small populations means that purifying selection is less efficient because drift has a stronger effect (Brandvain & Wright, 2016; Dussex et al., 2023). Drift also causes stochastic loss of genetic diversity (Jamieson et al., 2008), and loss of genetic variants is further increased through elevated levels of inbreeding leading to effective purging of highly deleterious variants (Dussex et al., 2021; Robinson et al., 2022). Although hihi were once widespread across the North Island, it is likely that the remnant Te Hauturu-o-Toi island population has historically been isolated and small (Duntsch et al., 2023). Impacts of human settlement followed by a secondary extreme bottleneck of 21 individuals to found the Tiritiri Matangi population in 1995-1996 (Armstrong et al., 2002; Duntsch et al., 2023) is likely to have further impacted the landscape of genetic variation. Compounding on top of this is that supplementary feeding supports a high population growth rate on Tiritiri Matangi, with the island frequently used as a source of translocations (Ewen et al., 2013), which may also reduce the strength of natural selection and encourage retention of mildly deleterious variants (Fronhofer et al., 2024).

While this study only examined a small cohort of individuals the relationship between SV load and hihi fitness supports the hypothesis that, while purging may have effectively removed the most deleterious of variants, the strong effects of drift have led to mildly deleterious SVs becoming more common in the Tiritiri Matangi population. This has been seen in another recovered wild avian population, where highly deleterious variants were purged, but not moderately deleterious ones (Femerling et al., 2023), and is also supported by signals of genome-wide inbreeding depression in the population (Duntsch et al., 2023). Because SVs are more likely to be deleterious than SNPs (Hämälä et al., 2021) and thus have likely faced stronger purifying selection historically, this may mean that the weakened selection has resulted in a relatively larger increase in SV based load than SNPs when considering patterns pre- and post-the bottleneck imposed by reintroduction of hihi to Tiritiri Matangi. It is possible that more deleterious variants exist within the hihi individuals that are not successful breeders. Future

582 work on this species will look to profile SVs across this species more broadly, to understand the
583 role they play in, for example, different life stages and how the SVs themselves are impacted by
584 demographic processes.

585 **4.2 | The relationship between mutational load and fitness**

586 In our results, we demonstrate that (when fitted alongside SNP load) there is only a significant
587 correlation between SV load and fitness when considering SV masked load, and not SV realised
588 load. Traditionally realised load is considered to be the load that is currently impacting
589 individuals, whereas it is both masked and realised that will impact future generations (Dussex
590 et al., 2023). However, this assumes the negative impact of load is conferred by recessive and
591 deleterious derived variants, whereas phenotypic expression could be impacted by other
592 effects such as additive or overdominance expression (Li et al., 2008), and it is possible that the
593 relationship between variant and phenotypic impact is stronger and/or more complex for SVs
594 than in SNPs (Buchanan & Scherer, 2008).

595 It is somewhat unexpected that SNP based load metrics were significantly (or borderline
596 significantly) correlated with fitness when tested individually, but were no longer so when tested
597 jointly with SV load. This may be because the SNP dataset captures roughly 3 million SNP
598 markers containing proportionally more neutral variation than the ~1,000 variants captured in
599 the SV dataset (Fig. 2a) (Collins et al., 2020; Hämälä et al., 2021). Even when reduced to the
600 ~54.7% of variants in proximity to genes during load calculations, it is likely that contributions of
601 realised and masked SNP load to fitness are made less apparent due to the predominance of
602 neutral markers, compared to the higher impact but lower density markers like SVs. When both
603 SNPs and SVs were fitted jointly, the stronger relationship between SV load and fitness also
604 suggests that much of the relationship between SNP load and fitness may have been driven by
605 linkage disequilibrium between (neutral) SNPs and nearby (mildly deleterious) SVs. Exploring
606 the application of additional approaches such as selection analysis (e.g. Gautier, 2015) to
607 further refine which components of genetic diversity are functional and contribute to load may
608 help to better capture functional variation for load calculations when working with high density
609 SNP makers generated from WGS data.

610 Ultimately, the results of this study demonstrate that identifying SVs allows us to capture unique
611 components of inter-individual genetic variation which may be vital for linking genetic patterns
612 to important phenotypic traits such as fitness, especially within this threatened species. Where
613 possible, WGS data should be equally interrogated for SNP and SV patterns, to understand the
614 distribution of more neutral SNPs and more deleterious SVs within the population. That being

said, mildly deleterious variants are important sources of functional genetic diversity that may become beneficial under shifted section engines, and ultimately any reduction of genetic diversity reduces long-term adaptability to environmental shifts, i.e. adaptive potential (Femerling et al., 2023; Kardos et al., 2021). Thus, to best manage populations with highly reduced genetic diversity, the broad identification of genome wide variants is necessary, and management processes that maximise heterozygosity may be most useful for future proofing species of conservation concern.

4.3 | SVs arise in regions of high recombination and low genome conservation

The dynamic genomic recombination landscape that both causes and is impacted by SVs (Mérot et al., 2020; Morgan et al., 2017) may explain the slightly steeper linkage disequilibrium decay patterns seen in SVs compared to SNPs (Fig. 3b). The slightly weaker linkage disequilibrium between SVs and surrounding SNPs may be due to biological reasons such as recombination rate and differential selection impacts (Cumer et al., 2021; Pang et al., 2010), as well as technical reasons caused by imperfect variant calling of both the SV and surrounding SNPs (Audano & Beck, 2024; Geibel et al., 2022). The latter is likely to explain the highly stochastic patterns seen in the immediate vicinity of the SV breakends (Fig. S10), while the former may explain more distant pairwise patterns (Fig 3b).

We found that regions of high recombination are associated with the presence of SVs, and this pattern is more evident on the micro-chromosomes within the hihi (Fig. 3a). Microchromosomes are considered evolutionary building blocks in avians (Waters et al., 2021), and our results indicate that these adaptive hotspots may encourage SVs which contribute to adaptive diversity within a population (Fig. 3a). Higher levels of recombination on micro-chromosomes can both cause novel variation and break up linkage disequilibrium between variants, allowing adaptation to happen faster (Peischl et al., 2015) which in turn facilitates more efficient evolution (Haenel et al., 2018). We see this trend in the evolutionary conservation genome alignment results (Fig. 4) which demonstrate that SVs, regardless of their load contribution, on average arise within more unique regions of the hihi genome. Intriguingly, within micro-chromosomes load contributing SVs occur at higher frequencies than variants that do not contribute to load, a pattern reversed in SNPs (Fig. S12). This may reflect genomic region biases in which SVs and SNPs are likely to be created, as the former arise in high recombination regions which are often the regions in which genes occur also (Tan et al., 2024).

4.4 | Technical validation of SVs and exclusion of TEs

Because of the size and complex nature of SVs, it is more technically challenging to identify them in comparison to SNPs (Mahmoud et al., 2019; Wold et al., 2023). Through manual curation of SVs the false positives may be reduced (Belyeu et al., 2018; Belyeu, Chowdhury, et al., 2021), and the improvement in Mendelian inheritance errors in our data after manual curation (Fig. S1-S6) is indicative that the process improved the accuracy of the SV dataset. We also consider in our study the importance of excluding transposable element sequences from analysis (Fig. 1b). Some of the aims of this study were to examine patterns of recombination and evolutionary conservation near or at SVs and SNPs. TEs arise through self-replication and relocation (Belyeu, Brand, et al., 2021; Biémont & Vieira, 2006), and so their relationship to surrounding genomic features and patterns may interfere with identifying patterns pertinent to non-TE SVs which themselves arise through mechanisms of DNA recombination, replication, or repair (Carvalho & Lupski, 2016; Currall et al., 2013). It is also important to note that TE-mediated SVs are more likely to introduce technical challenges during both SV identification and genotyping in surrounding SNPs (Audano & Beck, 2024) and that targeted TE identification processes (e.g. Chen et al., 2023) would be more appropriate for assessing TE specific load.

We found targeting breakends to be an efficient approach to removing the offending repeat sequences from our final SV dataset (Fig. 1c), though note that profiling the impacts of TE load is a pressing future direction in conservation biology, particularly in small threatened populations that may experience greater stress and reduced TE suppression (Stapley et al., 2015).

4.5 | Future directions

The individuals used in this study were specifically successful breeders within the small Tiritiri Matangi population, and so arguably may not carry the most deleterious variants. Despite this, there was enough variation in slightly deleterious load across individuals, and enough variance in fitness, to detect an impact of genetic load on lifetime fitness. Future directions of study in this system will aim to profile SVs across a wider variety of individuals to identify correlations between load and the presence or absence of successful breeding, as this may better capture a wider range of deleterious homozygous minor alleles. We note also note that because of the limits of profiling SVs with short reads (Mahmoud et al., 2019) our methods are not appropriate for the detection of larger SVs. With the promise of increasingly affordable longer read sequencing technologies we can begin to accurately characterise larger SVs (Schloissnig et al., 2024) which are likely to have larger functional impacts which could impact fitness (Abel et al., 2020).

Further, we note that our designation of the minor allele as the derived allele in assessing load makes a number of the assumptions, namely that derived alleles are deleterious, that the deleterious, derived allele will be the rarest, and that the rarest allele in our small dataset is the minor allele in the population. Because hihi are phylogenetically distinct and even the closest species do not have well-assembled genomes, it would be challenging to accurately polarise even the SNP data. We also note that decreased evolutionary conservation for SVs relative to SNPs (Fig. S12) means that outgroup selection will be more prone to bias with increasing outgroup phylogenetic distance for the former variant type, indicating outgroup selection will be very important for correct SV polarisation. As discussed above, the small population size is likely to have led to some deleterious alleles reaching high frequencies by drift. Sampling a small number of individuals from the population is also likely to have mis-identified the 'correct' realised load genotype. Indeed, the negative correlation between fitness and the major allele homozygous genotype (Fig. 5c) may suggest that a number of realised load genotypes have been incorrectly classified. We expect therefore that if variants can be successfully polarised, the true realised load values will become be larger and the relationship with fitness stronger (Keightley & Jackson, 2018).

Finally, restricting load calculations to gene-proximate regions ignores the functional role non-coding regions of the genome may play in influencing phenotype. Sequence variation can influence phenotype through, for example, alterations to cis -regulatory elements and three dimensional structure DNA (Engreitz et al., 2016; Soto et al., 2023; Spielmann & Mundlos, 2016), which will undoubtedly be differentially impacted by SNP and in particular SV variants. Thus, limiting load estimates to coding regions of the genome may neglect important functional variation within a population, impacting our ability to accurately identify and predict the adaptive potential for a species (Kardos et al., 2021), and is another avenue for future research within this system and also more broadly within conservation genomics.

5 | Conclusion

In highly inbred species like the hihi, identifying all aspects of functional genetic diversity is extremely important. Previous research in the hihi has found fairly low adaptive potential as interpreted from low heritability of fitness and fitness-associated traits (de Villemereuil et al., 2019). However, quantifying a more comprehensive genome-wide profile of intra-individual variation has revealed previously hidden associations between fitness and underlying genetic variant patterns. This study demonstrates that the SV variants contain important aspects of variation that may be highly applicable to conservation decisions within a species. This study

joins the many that demonstrate the non-redundant genomic variation captured by structural variants may have important functional implications (e.g. Catanach et al., 2019; Li et al., 2024; Rafter et al., 2021), and reveals that our ability to assess genetic diversity and load in small populations will always be limited if we only investigate SNPs.

Author contributions

Project conceived by AWS and KCS. Sample coordination and extractions performed by SP. Analysis and manuscript drafting was completed by KCS, with additional data from HZT and AW. Additional resources were supplied by PB and JGE. All authors edited and contributed to the final version of this manuscript.

Data availability

Hihi are of cultural significance to the Indigenous People of Aotearoa New Zealand, the Māori and are considered a taonga (treasured) species whose whakapapa (genealogy) is intricately tied to that of Māori. The sequencing and variant data generated as part of this study will be made available on request though the recommendation of Ngāti Manuhiri, the iwi (extended kinship group) that affiliates as kaitiaki (guardians) for hihi. To obtain contact details for the iwi, please contact Dr Anna Santure: a.santure@auckland.ac.nz. All of the code associated with this project will be made available through GitHub at (https://github.com/katarinastuart/Nc3_HihiSV).

Acknowledgements

We acknowledge Ngāti Manuhiri as Mana Whenua and Kaitiaki of Te Hauturu-o-Toi and its taonga, including hihi. We extend many thanks to the volunteers, past students and Department of Conservation staff who have contributed to monitoring the Tiritiri Mātangi hihi population. We thank the Hihi Recovery Group, Department of Conservation, and Supporters of Tiritiri Mātangi for maintaining such a long-term vision in monitoring and management of this population, with a special mention of the current hihi conservation officer Mhairi McCready. We acknowledge the use of New Zealand eScience Infrastructure (NeSI) high-performance computing facilities, and thank the NeSI team, particularly Dinindu Senanayake, for all their support. We also thank the AgResearch Animal Genomics team, particularly Tracey Van Stijn, Rudiger Brauning and Shannon Clarke, for hihi Illumina sequencing. Many thanks to Jana Wold and Joseph Guhlin for helpful feedback and discussion on SVs and whole genome alignment. Open access publishing facilitated by The University of Auckland, as organised by the Council of Australian University

744 Librarians and its Member Institutions. All work was undertaken in accordance with Department
745 of Conservation permits 53614-FAU and 66751-FAU. The High Quality Genomes, and High
746 Quality Genomes and Population Genomics (HQG+PG) projects I and II of Genomics Aotearoa
747 supported AWS, AW and KS and funded the whole genome resequencing data and sequencing
748 of the reference genome utilised in this study. Funding from the George Mason Center for the
749 Natural Environment and the Little Barrier Island (Hauturu) Supporters Trust supported
750 fieldwork collection from Te Hauturu-o-Toi. A Marsden Grant (UOA1408) awarded to AWS from
751 the New Zealand Royal Society Te Aparangi supported AWS, PB, and JGE and funded
752 development of the SNP array genotypes and microsatellite genotyping. A Strategic Science
753 Investment Fund in Data Science from the Ministry of Business, Innovation and Employment
754 supports AWS and HZT. PB and JGE are supported by Research England.

755

REFERENCES:

Abel, H. J., Larson, D. E., Regier, A. A., Chiang, C., Das, I., Kanchi, K. L., Layer, R. M., Neale, B. M., Salerno, W. J., Reeves, C., Buyske, S., Matise, T. C., Muzny, D. M., Zody, M. C., Lander, E. S., Dutcher, S. K., Stitziel, N. O., & Hall, I. M. (2020). Mapping and characterization of structural variation in 17,795 human genomes. *Nature*, *583*(7814), 83–89. https://doi.org/10.1038/s41586-020-2371-0

Alves, F., Banks, S. C., Edworthy, M., Stojanovic, D., Langmore, N. E., & Heinsohn, R. (2023). Using conservation genetics to prioritise management options for an endangered songbird. *Heredity*, *130*(5), 289–301. https://doi.org/10.1038/s41437-023-00609-6

Armstrong, D. P., Davidson, R. S., Dimond, W. J., Perrott, J. K., Castro, I., Ewen, J. G., Griffiths, R., & Taylor, J. (2002). Population dynamics of reintroduced forest birds on New Zealand islands. *Journal of Biogeography*, *29*(5–6), 609–621. https://doi.org/10.1046/j.1365-2699.2002.00710.x

Armstrong, J., Hickey, G., Diekhans, M., Fiddes, I. T., Novak, A. M., Deran, A., Fang, Q., Xie, D., Feng, S., Stiller, J., Genereux, D., Johnson, J., Marinescu, V. D., Alföldi, J., Harris, R. S., Lindblad-Toh, K., Haussler, D., Karlsson, E., Jarvis, E. D., … Paten, B. (2020). Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature*, *587*(7833), 246–251. https://doi.org/10.1038/s41586-020-2871-y

Audano, P. A., & Beck, C. R. (2024). Small polymorphisms are a source of ancestral bias in structural variant breakpoint placement. *Genome Research*, *34*(1), 7–19. https://doi.org/10.1101/gr.278203.123

Bailey, S., Guhlin, J., Senanayake, D. S., Scherer, P., Brekke, P., Ewen, J. G., Santure, A. W., & Whibley, A. (2023). Assembly of female and male hihi genomes (stitchbird; Notiomystis cincta) enables characterization of the W chromosome and resources for conservation genomics. *Molecular Ecology Resources*, *n/a*(n/a). https://doi.org/10.1111/1755-0998.13823

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), Article 1. https://doi.org/10.18637/jss.v067.i01

Belyeu, J. R., Brand, H., Wang, H., Zhao, X., Pedersen, B. S., Feusier, J., Gupta, M., Nicholas, T. J., Brown, J., Baird, L., Devlin, B., Sanders, S. J., Jorde, L. B., Talkowski, M. E., & Quinlan, A. R. (2021). De novo structural mutation rates and gamete-of-origin biases revealed through genome sequencing of 2,396 families. *American Journal of Human Genetics*, *108*(4), 597–607. https://doi.org/10.1016/j.ajhg.2021.02.012

Belyeu, J. R., Chowdhury, M., Brown, J., Pedersen, B. S., Cormier, M. J., Quinlan, A. R., & Layer, R. M. (2021). Samplot: A platform for structural variant visual validation and automated filtering. *Genome Biology*, *22*(1), 161. https://doi.org/10.1186/s13059-021-02380-5

Belyeu, J. R., Nicholas, T. J., Pedersen, B. S., Sasani, T. A., Havrilla, J. M., Kravitz, S. N., Conway, M. E., Lohman, B. K., Quinlan, A. R., & Layer, R. M. (2018). SV-plaudit: A cloud-based framework for manually curating thousands of structural variants. *GigaScience*, *7*(7), giy064. https://doi.org/10.1093/gigascience/giy064

Berdan, E. L., Mérot, C., Pavia, H., Johannesson, K., Wellenreuther, M., & Butlin, R. K. (2021). A large chromosomal inversion shapes gene expression in seaweed flies (Coelopa frigida). *Evolution Letters*, *5*(6), 607–624. https://doi.org/10.1002/evl3.260

Biémont, C., & Vieira, C. (2006). Junk DNA as an evolutionary force. *Nature*, *443*(7111), Article 7111. https://doi.org/10.1038/443521a

Brandvain, Y., & Wright, S. I. (2016). The Limits of Natural Selection in a Nonequilibrium World. *Trends in Genetics*, *32*(4), 201–210. https://doi.org/10.1016/j.tig.2016.01.004

Brekke, P., Bennett, P. M., Santure, A. W., & Ewen, J. G. (2011). High genetic diversity in the remnant island population of hihi and the genetic consequences of re-introduction. *Molecular Ecology*, *20*(1), 29–45. https://doi.org/10.1111/j.1365-294X.2010.04923.x

807 Brekke, P., Dawson, D. A., Horsburgh, G. J., & Ewen, J. G. (2009). Characterization of
808 microsatellite loci In The hihi Notiomystis cincta (Notiomystidae, Aves). *Molecular*
809 *Ecology Resources*, *9*(4), 1255–1258. https://doi.org/10.1111/j.1755-0998.2009.02626.x
810 Buchanan, J. A., & Scherer, S. W. (2008). Contemplating effects of genomic structural variation.
811 *Genetics in Medicine*, *10*(9), 639–647. https://doi.org/10.1097/GIM.0b013e318183f848
812 Cao, J., Schneeberger, K., Ossowski, S., Günther, T., Bender, S., Fitz, J., Koenig, D., Lanz, C.,
813 Stegle, O., Lippert, C., Wang, X., Ott, F., Müller, J., Alonso-Blanco, C., Borgwardt, K.,
814 Schmid, K. J., & Weigel, D. (2011). Whole-genome sequencing of multiple Arabidopsis
815 thaliana populations. *Nature Genetics*, *43*(10), 956–963. https://doi.org/10.1038/ng.911
816 Carvalho, C. M. B., & Lupski, J. R. (2016). Mechanisms underlying structural variant formation in
817 genomic disorders. *Nature Reviews. Genetics*, *17*(4), 224–238.
818 https://doi.org/10.1038/nrg.2015.25
819 Catanach, A., Crowhurst, R., Deng, C., David, C., Bernatchez, L., & Wellenreuther, M. (2019).
820 The genomic pool of standing structural variation outnumbers single nucleotide
821 polymorphism by threefold in the marine teleost Chrysophrys auratus. *Molecular*
822 *Ecology*, *28*(6), 1210–1223. https://doi.org/10.1111/mec.15051
823 Chakraborty, M., Emerson, J. J., Macdonald, S. J., & Long, A. D. (2019). Structural variants exhibit
824 widespread allelic heterogeneity and shape variation in complex traits. *Nature*
825 *Communications*, *10*(1), 4872. https://doi.org/10.1038/s41467-019-12884-1
826 Chen, J., Basting, P. J., Han, S., Garfinkel, D. J., & Bergman, C. M. (2023). Reproducible
827 evaluation of transposable element detectors with McClintock 2 guides accurate
828 inference of Ty insertion patterns in yeast. *Mobile DNA*, *14*(1), 8.
829 https://doi.org/10.1186/s13100-023-00296-4
830 Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A. J.,
831 Kruglyak, S., & Saunders, C. T. (2016). Manta: Rapid detection of structural variants and

832 indels for germline and cancer sequencing applications. *Bioinformatics (Oxford,*
833 *England)*, *32*(8), 1220–1222. https://doi.org/10.1093/bioinformatics/btv710
834 Collins, R. L., Brand, H., Karczewski, K. J., Zhao, X., Alföldi, J., Francioli, L. C., Khera, A. V.,
835 Lowther, C., Gauthier, L. D., Wang, H., Watts, N. A., Solomonson, M., O'Donnell-Luria,
836 A., Baumann, A., Munshi, R., Walker, M., Whelan, C. W., Huang, Y., Brookings, T., ...
837 Talkowski, M. E. (2020). A structural variation reference for medical and population
838 genetics. *Nature*, *581*(7809), 444–451. https://doi.org/10.1038/s41586-020-2287-8
839 Cumer, T., Boyer, F., & Pompanon, F. (2021). Genome-Wide Detection of Structural Variations
840 Reveals New Regions Associated with Domestication in Small Ruminants. *Genome*
841 *Biology and Evolution*, *13*(8), evab165. https://doi.org/10.1093/gbe/evab165
842 Currall, B. B., Chiang, C., Talkowski, M. E., & Morton, C. C. (2013). Mechanisms for Structural
843 Variation in the Human Genome. *Current Genetic Medicine Reports*, *1*(2), 81–90.
844 https://doi.org/10.1007/s40142-013-0012-8
845 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E.,
846 Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call
847 format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158.
848 https://doi.org/10.1093/bioinformatics/btr330
849 Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A.,
850 Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and
851 BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008
852 de Villemereuil, P., Rutschmann, A., Lee, K. D., Ewen, J. G., Brekke, P., & Santure, A. W. (2019).
853 Little Adaptive Potential in a Threatened Passerine Bird. *Current Biology*, *29*(5), 889-
854 894.e3. https://doi.org/10.1016/j.cub.2019.01.072
855 Doerr, L. R., Richardson, K. M., Ewen, J. G., & Armstrong, D. P. (2017). Effect of supplementary
856 feeding on reproductive success of hihi (stitchbird, Notiomystis cincta) at a mature

857 forest reintroduction site. *New Zealand Journal of Ecology*, *41*(1), Article 1.
858 https://doi.org/10.20417/nzjecol.41.15
859 Duckett, D. J., Calder, K., Sullivan, J., Tank, D. C., & Carstens, B. C. (2023). Reduced
860 representation approaches produce similar results to whole genome sequencing for
861 some common phylogeographic analyses. *PLOS ONE*, *18*(11), e0291941.
862 https://doi.org/10.1371/journal.pone.0291941
863 Duntsch, L., Whibley, A., Brekke, P., Ewen, J. G., & Santure, A. W. (2021). Genomic data of
864 different resolutions reveal consistent inbreeding estimates but contrasting
865 homozygosity landscapes for the threatened Aotearoa New Zealand hihi. *Molecular*
866 *Ecology*, *30*(23), 6006–6020. https://doi.org/10.1111/mec.16068
867 Duntsch, L., Whibley, A., de Villemereuil, P., Brekke, P., Bailey, S., Ewen, J. G., & Santure, A. W.
868 (2023). Genomic signatures of inbreeding depression for a threatened Aotearoa New
869 Zealand passerine. *Molecular Ecology*, *32*(8), 1893–1907.
870 https://doi.org/10.1111/mec.16855
871 Dussex, N., Morales, H. E., Grossen, C., Dalén, L., & Oosterhout, C. van. (2023). Purging and
872 accumulation of genetic load in conservation. *Trends in Ecology & Evolution*, *38*(10),
873 961–969. https://doi.org/10.1016/j.tree.2023.05.008
874 Dussex, N., van der Valk, T., Morales, H. E., Wheat, C. W., Díez-del-Molino, D., von Seth, J.,
875 Foster, Y., Kutschera, V. E., Guschanski, K., Rhie, A., Phillippy, A. M., Korlach, J., Howe,
876 K., Chow, W., Pelan, S., Mendes Damas, J. D., Lewin, H. A., Hastie, A. R., Formenti, G., …
877 Dalén, L. (2021). Population genomics of the critically endangered kākāpō. *Cell*
878 *Genomics*, *1*(1), 100002. https://doi.org/10.1016/j.xgen.2021.100002
879 Edmonds, C. A., Lillie, A. S., & Cavalli-Sforza, L. L. (2004). Mutations arising in the wave front of
880 an expanding population. *Proceedings of the National Academy of Sciences*, *101*(4),
881 975–979. https://doi.org/10.1073/pnas.0308064100

882 Engreitz, J. M., Ollikainen, N., & Guttman, M. (2016). Long non-coding RNAs: Spatial amplifiers
883 that control nuclear structure and gene expression. *Nature Reviews Molecular Cell*
884 *Biology*, *17*(12), 756–770. https://doi.org/10.1038/nrm.2016.126
885 Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results
886 for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048.
887 https://doi.org/10.1093/bioinformatics/btw354
888 Ewen, J. G., Renwick, R., Adams, L., Armstrong, D. P., Parker, K. A., & North, N. Z. D. (2013).
889 1980-2012: 32 years of re-introduction efforts of the hihi (stitchbird) in New Zealand.
890 *Global Re-Introduction Perspectives: 2013. Further Case Studies from around the*
891 *Globe*, 68.
892 https://www.academia.edu/download/43117316/rsg_book_2013.pdf#page=83
893 Femerling, G., van Oosterhout, C., Feng, S., Bristol, R. M., Zhang, G., Groombridge, J., P. Gilbert,
894 M. T., & Morales, H. E. (2023). Genetic Load and Adaptive Potential of a Recovered Avian
895 Species that Narrowly Avoided Extinction. *Molecular Biology and Evolution*, *40*(12),
896 msad256. https://doi.org/10.1093/molbev/msad256
897 Frankham, R. (2008). Genetic adaptation to captivity in species conservation programs.
898 *Molecular Ecology*, *17*(1), 325–333. https://doi.org/10.1111/j.1365-294X.2007.03399.x
899 Fronhofer, E. A., Govaert, L., O'Connor, M. I., Schreiber, S. J., & Altermatt, F. (2024). The shape of
900 density dependence and the relationship between population growth, intraspecific
901 competition and equilibrium population density. *Oikos*, *2024*(2), e09824.
902 https://doi.org/10.1111/oik.09824
903 Fuentes-Pardo, A. P., & Ruzzante, D. E. (2017). Whole-genome sequencing approaches for
904 conservation biology: Advantages, limitations and practical recommendations.
905 *Molecular Ecology*, *26*(20), 5369–5406. https://doi.org/10.1111/mec.14264

Gautier, M. (2015). Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. *Genetics*, *201*(4), 1555–1579. https://doi.org/10.1534/genetics.115.181453

Geibel, J., Praefke, N. P., Weigend, S., Simianer, H., & Reimer, C. (2022). Assessment of linkage disequilibrium patterns between structural variants and single nucleotide polymorphisms in three commercial chicken populations. *BMC Genomics*, *23*, 193. https://doi.org/10.1186/s12864-022-08418-7

Haenel, Q., Laurentino, T. G., Roesti, M., & Berner, D. (2018). Meta-analysis of chromosome-scale crossover rate variation in eukaryotes and its significance to evolutionary genomics. *Molecular Ecology*, *27*(11), 2477–2497. https://doi.org/10.1111/mec.14699

Haig, S. M., Miller, Mark. P., Bellinger, R., Draheim, H. M., Mercer, D. M., & Mullins, T. D. (2016). The conservation genetics juggling act: Integrating genetics and ecology, science and policy. *Evolutionary Applications*, *9*(1), 181–195. https://doi.org/10.1111/eva.12337

Hämälä, T., Wafula, E. K., Guiltinan, M. J., Ralph, P. E., dePamphilis, C. W., & Tiffin, P. (2021). Genomic structural variants constrain and facilitate adaptation in natural populations of Theobroma cacao, the chocolate tree. *Proceedings of the National Academy of Sciences*, *118*(35). https://doi.org/10.1073/pnas.2102914118

Hedrick, P. W., Hellsten, U., & Grattapaglia, D. (2016). Examining the cause of high inbreeding depression: Analysis of whole-genome sequence data in 28 selfed progeny of Eucalyptus grandis. *New Phytologist*, *209*(2), 600–611. https://doi.org/10.1111/nph.13639

Hickey, G., Paten, B., Earl, D., Zerbino, D., & Haussler, D. (2013). HAL: A hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics*, *29*(10), 1341–1342. https://doi.org/10.1093/bioinformatics/btt128

Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology*, *30*(1), 62–82. https://doi.org/10.1111/mec.15720

Humble, E., Stoffel, M. A., Dicks, K., Ball, A. D., Gooley, R. M., Chuven, J., Pusey, R., Remeithi, M. A., Koepfli, K.-P., Pukazhenthi, B., Senn, H., & Ogden, R. (2023). Conservation management strategy impacts inbreeding and mutation load in scimitar-horned oryx. *Proceedings of the National Academy of Sciences*, *120*(18), e2210756120. https://doi.org/10.1073/pnas.2210756120

Jamieson, I. G., Grueber, C. E., Waters, J. M., & Gleeson, D. M. (2008). Managing genetic diversity in threatened populations: A New Zealand perspective. *New Zealand Journal of Ecology*, *32*(1), 130–137. https://www.jstor.org/stable/24058111

Jeffares, D. C., Jolly, C., Hoti, M., Speed, D., Shaw, L., Rallis, C., Balloux, F., Dessimoz, C., Bähler, J., & Sedlazeck, F. J. (2017). Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nature Communications*, *8*(1), 14061. https://doi.org/10.1038/ncomms14061

Kardos, M., Armstrong, E. E., Fitzpatrick, S. W., Hauser, S., Hedrick, P. W., Miller, J. M., Tallmon, D. A., & Funk, W. C. (2021). The crucial role of genome-wide genetic variation in conservation. *Proceedings of the National Academy of Sciences*, *118*(48), e2104642118. https://doi.org/10.1073/pnas.2104642118

Keightley, P. D., & Jackson, B. C. (2018). Inferring the Probability of the Derived vs. The Ancestral Allelic State at a Polymorphic Site. *Genetics*, *209*(3), 897–906. https://doi.org/10.1534/genetics.118.301120

Kreiner, J. M., Giacomini, D. A., Bemm, F., Waithaka, B., Regalado, J., Lanz, C., Hildebrandt, J., Sikkema, P. H., Tranel, P. J., Weigel, D., Stinchcombe, J. R., & Wright, S. I. (2019). Multiple modes of convergent adaptation in the spread of glyphosate-resistant Amaranthus

955 tuberculatus. *Proceedings of the National Academy of Sciences*, *116*(42), 21076–21084.
956 https://doi.org/10.1073/pnas.1900870116
957 Krueger, F. (2021). *Babraham Bioinformatics—Trim Galore!* (0.6.7) [Computer software].
958 https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
959 Kumar, S., Suleski, M., Craig, J. M., Kasprowicz, A. E., Sanderford, M., Li, M., Stecher, G., &
960 Hedges, S. B. (2022). TimeTree 5: An Expanded Resource for Species Divergence Times.
961 *Molecular Biology and Evolution*, *39*(8), msac174.
962 https://doi.org/10.1093/molbev/msac174
963 Layer, R. M., Chiang, C., Quinlan, A. R., & Hall, I. M. (2014). LUMPY: A probabilistic framework for
964 structural variant discovery. *Genome Biology*, *15*(6), R84. https://doi.org/10.1186/gb-
965 2014-15-6-r84
966 Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
967 *arXiv:1303.3997 [q-Bio]*. http://arxiv.org/abs/1303.3997
968 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
969 Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence
970 Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–
971 2079. https://doi.org/10.1093/bioinformatics/btp352
972 Li, L., Lu, K., Chen, Z., Mu, T., Hu, Z., & Li, X. (2008). Dominance, Overdominance and Epistasis
973 Condition the Heterosis in Two Heterotic Rice Hybrids. *Genetics*, *180*(3), 1725–1742.
974 https://doi.org/10.1534/genetics.108.091942
975 Li, Y., Yao, J., Sang, H., Wang, Q., Su, L., Zhao, X., Xia, Z., Wang, F., Wang, K., Lou, D., Wang, G.,
976 Waterhouse, R. M., Wang, H., Luo, S., & Sun, C. (2024). Pan-genome analysis highlights
977 the role of structural variation in the evolution and environmental adaptation of Asian
978 honeybees. *Molecular Ecology Resources*, *24*(2), e13905. https://doi.org/10.1111/1755-
979 0998.13905

Lohmueller, K. E. (2014). The Impact of Population Demography and Selection on the Genetic Architecture of Complex Traits. *PLoS Genetics*, *10*(5), e1004379. https://doi.org/10.1371/journal.pgen.1004379

Lynch, M., & O'Hely, M. (2001). Captive breeding and the genetic fitness of natural populations. *Conservation Genetics*, *2*(4), 363–378. https://doi.org/10.1023/A:1012550620717

Mahmoud, M., Gobet, N., Cruz-Dávalos, D. I., Mounier, N., Dessimoz, C., & Sedlazeck, F. J. (2019). Structural variant calling: The long and the short of it. *Genome Biology*, *20*(1), 246. https://doi.org/10.1186/s13059-019-1828-7

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, *17*(1), 122. https://doi.org/10.1186/s13059-016-0974-4

McMahon, B. J., Teeling, E. C., & Höglund, J. (2014). How and why should we implement genomics into conservation? *Evolutionary Applications*, *7*(9), 999–1007. https://doi.org/10.1111/eva.12193

Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020). A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation. *Trends in Ecology & Evolution*, *35*(7), 561–572. https://doi.org/10.1016/j.tree.2020.03.002

Miskelly, C. M., & Powlesland, R. G. (2013). Conservation translocations of New Zealand birds 1863-2012. *Notornis*, *60*, 3–28.

Morgan, A. P., Gatti, D. M., Najarian, M. L., Keane, T. M., Galante, R. J., Pack, A. I., Mott, R., Churchill, G. A., & de Villena, F. P.-M. (2017). Structural Variation Shapes the Landscape of Recombination in Mouse. *Genetics*, *206*(2), 603–619. https://doi.org/10.1534/genetics.116.197988

Nichols, S., Ewen, J. G., Gottelli, D., Grueber, C. E., Santure, A. W., Trask, A., & Brekke, P. (2024). Genetic rescue attempt in a small, inbred population of a wild endangered passerine. *Biological Conservation*, *290*, 110430. https://doi.org/10.1016/j.biocon.2023.110430

1006 Pang, A. W., MacDonald, J. R., Pinto, D., Wei, J., Rafiq, M. A., Conrad, D. F., Park, H., Hurles, M.
1007 E., Lee, C., Venter, J. C., Kirkness, E. F., Levy, S., Feuk, L., & Scherer, S. W. (2010).
1008 Towards a comprehensive structural variation map of an individual human genome.
1009 *Genome Biology*, *11*(5), R52. https://doi.org/10.1186/gb-2010-11-5-r52
1010 Parlato, E. H., Ewen, J. G., McCready, M., Parker, K. A., & Armstrong, D. P. (2021). A modelling
1011 framework for integrating reproduction, survival and count data when projecting the
1012 fates of threatened populations. *Oecologia*, *195*(3), 627–640.
1013 https://doi.org/10.1007/s00442-021-04871-5
1014 Peischl, S., Kirkpatrick, M., & Excoffier, L. (2015). Expansion Load and the Evolutionary
1015 Dynamics of a Species Range. *The American Naturalist*, *185*(4), E81–E93.
1016 https://doi.org/10.1086/680220
1017 *Picard toolkit*. (2019). [Computer software]. Broad Institute.
1018 http://broadinstitute.github.io/picard/
1019 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar,
1020 P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-
1021 Genome Association and Population-Based Linkage Analyses. *American Journal of*
1022 *Human Genetics*, *81*(3), 559–575.
1023 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1950838/
1024 Rafter, P., Gormley, I. C., Parnell, A. C., Naderi, S., & Berry, D. P. (2021). The Contribution of Copy
1025 Number Variants and Single Nucleotide Polymorphisms to the Additive Genetic
1026 Variance of Carcass Traits in Cattle. *Frontiers in Genetics*, *12*, 761503.
1027 https://doi.org/10.3389/fgene.2021.761503
1028 Rausch, T., Zichner, T., Schlattl, A., Stütz, A. M., Benes, V., & Korbel, J. O. (2012). DELLY:
1029 Structural variant discovery by integrated paired-end and split-read analysis.
1030 *Bioinformatics (Oxford, England)*, *28*(18), i333–i339.
1031 https://doi.org/10.1093/bioinformatics/bts378

1032 Robinson, J. A., Kyriazis, C. C., Nigenda-Morales, S. F., Beichman, A. C., Rojas-Bracho, L.,
1033 Robertson, K. M., Fontaine, M. C., Wayne, R. K., Lohmueller, K. E., Taylor, B. L., & Morin,
1034 P. A. (2022). The critically endangered vaquita is not doomed to extinction by inbreeding
1035 depression. *Science*, *376*(6593), 635–639. https://doi.org/10.1126/science.abm1742
1036 Ruigrok, M., Xue, B., Catanach, A., Zhang, M., Jesson, L., Davy, M., & Wellenreuther, M. (2022).
1037 The Relative Power of Structural Genomic Variation versus SNPs in Explaining the
1038 Quantitative Trait Growth in the Marine Teleost Chrysophrys auratus. *Genes*, *13*(7),
1039 1129. https://doi.org/10.3390/genes13071129
1040 Scherer, S. W., Lee, C., Birney, E., Altshuler, D. M., Eichler, E. E., Carter, N. P., Hurles, M. E., &
1041 Feuk, L. (2007). Challenges and standards in integrating surveys of structural variation.
1042 *Nature Genetics*, *39*(7), Article 7. https://doi.org/10.1038/ng2093
1043 Schloissnig, S., Pani, S., Rodriguez-Martin, B., Ebler, J., Hain, C., Tsapalou, V., Söylev, A., Hüther,
1044 P., Ashraf, H., Prodanov, T., Asparuhova, M., Hunt, S., Rausch, T., Marschall, T., & Korbel,
1045 J. O. (2024). *Long-read sequencing and structural variant characterization in 1,019*
1046 *samples from the 1000 Genomes Project* (p. 2024.04.18.590093). bioRxiv.
1047 https://doi.org/10.1101/2024.04.18.590093
1048 Smeds, L., Huson, L. S. A., & Ellegren, H. (2024). Structural genomic variation in the inbred
1049 Scandinavian wolf population contributes to the realized genetic load but is positively
1050 affected by immigration. *Evolutionary Applications*, *17*(2), e13652.
1051 https://doi.org/10.1111/eva.13652
1052 Smit, A., Hubley, R., & Green, P. (2013). *RepeatMasker Open-4.0* (http://www.repeatmasker.org)
1053 [Http://www.repeatmasker.org]. http://www.repeatmasker.org
1054 Soto, D. C., Uribe-Salazar, J. M., Shew, C. J., Sekar, A., McGinty, S. P., & Dennis, M. Y. (2023).
1055 Genomic structural variation: A complex but important driver of human evolution.
1056 *American Journal of Biological Anthropology*, *181 Suppl 76*(Suppl 76), 118–144.
1057 https://doi.org/10.1002/ajpa.24713

1058 Spielmann, M., & Mundlos, S. (2016). Looking beyond the genes: The role of non-coding variants
1059 in human disease. *Human Molecular Genetics*, *25*(R2), R157–R165.
1060 https://doi.org/10.1093/hmg/ddw205
1061 Stapley, J., Santure, A. W., & Dennis, S. R. (2015). Transposable elements as agents of rapid
1062 adaptation may explain the genetic paradox of invasive species. *Molecular Ecology*,
1063 *24*(9), 2241–2252. https://doi.org/10.1111/mec.13089
1064 Stuart, K. C., Edwards, R. J., Sherwin, W. B., & Rollins, L. A. (2023). Contrasting Patterns of Single
1065 Nucleotide Polymorphisms and Structural Variation Across Multiple Invasions.
1066 *Molecular Biology and Evolution*, *40*(3), msad046.
1067 https://doi.org/10.1093/molbev/msad046
1068 Stuart, K. C., Santure, A. W., & Rollins, L. A. (2024). STRUCTURAL VARIANTS AND
1069 TRANSPOSABLE ELEMENTS AS FACILITATORS OF RAPID EVOLUTIONARY CHANGE IN
1070 INVASIVE POPULATIONS. In *Invasion Genomics*. In Review.
1071 Szarmach, S. J., Brelsford, A., Witt, C. C., & Toews, D. P. L. (2021). Comparing divergence
1072 landscapes from reduced-representation and whole genome resequencing in the
1073 yellow-rumped warbler (Setophaga coronata) species complex. *Molecular Ecology*,
1074 *30*(23), 5994–6005. https://doi.org/10.1111/mec.15940
1075 Tan, H. Z., Scherer, P., Stuart, K. C., Bailey, S., Lee, K. D., Brekke, P., Ewen, J. G., Whibley, A., &
1076 Santure, A. W. (2024). A high-density linkage map reveals broad- and fine-scale sex
1077 differences in recombination in the hihi (stitchbird; Notiomystis cincta). *Under Review*.
1078 Theissinger, K., Fernandes, C., Formenti, G., Bista, I., Berg, P. R., Bleidorn, C., Bombarely, A.,
1079 Crottini, A., Gallo, G. R., Godoy, J. A., Jentoft, S., Malukiewicz, J., Mouton, A., Oomen, R.
1080 A., Paez, S., Palsbøll, P. J., Pampoulie, C., Ruiz-López, M. J., Secomandi, S., … Zammit,
1081 G. (2023). How genomics can help biodiversity conservation. *Trends in Genetics*, *39*(7),
1082 545–559. https://doi.org/10.1016/j.tig.2023.01.005

1083 Turro, E., Astle, W. J., Megy, K., Gräf, S., Greene, D., Shamardina, O., Allen, H. L., Sanchis-Juan,
1084 A., Frontini, M., Thys, C., Stephens, J., Mapeta, R., Burren, O. S., Downes, K., Haimel, M.,
1085 Tuna, S., Deevi, S. V. V., Aitman, T. J., Bennett, D. L., … Ouwehand, W. H. (2020). Whole-
1086 genome sequencing of patients with rare diseases in a national health system. *Nature*,
1087 *583*(7814), 96–102. https://doi.org/10.1038/s41586-020-2434-2
1088 Usha, T., Panda, P., Goyal, A. K., Kukanur, A. A., Kamala, A., Prasannakumar, M. K.,
1089 Sidhalinghamurthy, K., & Middha, S. K. (2022). Whole-Genome Sequencing of Plants:
1090 Past, Present, and Future. In R. L. Singh, S. Mondal, A. Parihar, & P. K. Singh (Eds.), *Plant*
1091 *Genomics for Sustainable Agriculture* (pp. 191–239). Springer Nature.
1092 https://doi.org/10.1007/978-981-16-6974-3_8
1093 Waters, P. D., Patel, H. R., Ruiz-Herrera, A., Álvarez-González, L., Lister, N. C., Simakov, O., Ezaz,
1094 T., Kaur, P., Frere, C., Grützner, F., Georges, A., & Graves, J. A. M. (2021).
1095 Microchromosomes are building blocks of bird, reptile, and mammal chromosomes.
1096 *Proceedings of the National Academy of Sciences*, *118*(45), e2112494118.
1097 https://doi.org/10.1073/pnas.2112494118
1098 Weissensteiner, M. H., Bunikis, I., Catalán, A., Francoijs, K.-J., Knief, U., Heim, W., Peona, V.,
1099 Pophaly, S. D., Sedlazeck, F. J., Suh, A., Warmuth, V. M., & Wolf, J. B. W. (2020). Discovery
1100 and population genomics of structural variation in a songbird genus. *Nature*
1101 *Communications*, *11*(1), 3403. https://doi.org/10.1038/s41467-020-17195-4
1102 Wellenreuther, M., Mérot, C., Berdan, E., & Bernatchez, L. (2019). Going beyond SNPs: The role
1103 of structural genomic variants in adaptive evolution and species diversification.
1104 *Molecular Ecology*, *28*(6), 1203–1209. https://doi.org/10.1111/mec.15066
1105 Wold, J. R., Guhlin, J. G., Dearden, P. K., Santure, A. W., & Steeves, T. E. (2023). The promise and
1106 challenges of characterizing genome-wide structural variants: A case study in a critically
1107 endangered parrot. *Molecular Ecology Resources*, *n/a*(n/a).
1108 https://doi.org/10.1111/1755-0998.13783

Wold, J. R., Koepfli, K.-P., Galla, S. J., Eccles, D., Hogg, C. J., Le Lec, M. F., Guhlin, J., Santure, A. W., & Steeves, T. E. (2021). Expanding the conservation genomics toolbox: Incorporating structural variants to enhance genomic studies for species of conservation concern. *Molecular Ecology*, *30*(23), 5949–5965. https://doi.org/10.1111/mec.16141

Yang, J., Li, W.-R., Lv, F.-H., He, S.-G., Tian, S.-L., Peng, W.-F., Sun, Y.-W., Zhao, Y.-X., Tu, X.-L., Zhang, M., Xie, X.-L., Wang, Y.-T., Li, J.-Q., Liu, Y.-G., Shen, Z.-Q., Wang, F., Liu, G.-J., Lu, H.-F., Kantanen, J., … Liu, M.-J. (2016). Whole-Genome Sequencing of Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments. *Molecular Biology and Evolution*, *33*(10), 2576–2592. https://doi.org/10.1093/molbev/msw129

Zhao, S., Zheng, P., Dong, S., Zhan, X., Wu, Q., Guo, X., Hu, Y., He, W., Zhang, S., Fan, W., Zhu, L., Li, D., Zhang, X., Chen, Q., Zhang, H., Zhang, Z., Jin, X., Zhang, J., Yang, H., … Wei, F. (2013). Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nature Genetics*, *45*(1), 67–71. https://doi.org/10.1038/ng.2494