Received: 8 March 2023 | Revised: 18 May 2023 | Accepted: 25 May 2023

DOI: 10.1111/1755-0998.13823

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

Assembly of female and male hihi genomes (stitchbird; *Notiomystis cincta*) enables characterization of the W chromosome and resources for conservation genomics

Sarah Bailey[1,2] | Joseph Guhlin[3] | Dinindu S. Senanayake[4] | Phoebe Scherer[1] | Patricia Brekke[5] | John G. Ewen[5] | Anna W. Santure[1,2] | Annabel Whibley[1]

[1]School of Biological Sciences, University of Auckland, Auckland, New Zealand

[2]Centre for Biodiversity and Biosecurity, School of Biological Sciences, University of Auckland, Auckland, New Zealand

[3]Biochemistry Department, School of Biomedical Sciences, University of Otago, Dunedin, New Zealand

[4]New Zealand eScience Infrastructure (NeSI), University of Auckland, Auckland, New Zealand

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

Correspondence
Anna W. Santure, School of Biological Sciences, University of Auckland, Auckland, New Zealand.
Email: a.santure@auckland.ac.nz

Handling Editor: Alison Nazareno

Abstract

A high-quality reference genome can be a valuable resource for threatened species by providing a foundation to assess their evolutionary potential to adapt to future pressures such as environmental change. We assembled the genome of a female hihi (*Notiomysits cincta*), a threatened passerine bird endemic to Aotearoa New Zealand. The assembled genome is 1.06 Gb, and is of high quality and highly contiguous, with a contig N50 of 7.0 Mb, estimated QV of 44 and a BUSCO completeness of 96.8%. A male assembly of comparable quality was generated in parallel. A population linkage map was used to scaffold the autosomal contigs into chromosomes. Female and male sequence coverage and comparative genomics analyses were used to identify Z-, and W-linked contigs. In total, 94.6% of the assembly length was assigned to putative nuclear chromosome scaffolds. Native DNA methylation was highly correlated between sexes, with the W chromosome contigs more highly methylated than autosomal chromosomes and Z contigs. 43 differentially methylated regions were identified, and these may represent interesting candidates for the establishment or maintenance of sex differences. By generating a high-quality reference assembly of the heterogametic sex, we have created a resource that enables characterization of genome-wide diversity and facilitates the investigation of female-specific evolutionary processes. The reference genomes will form the basis for fine-scale assessment of the impacts of low genetic diversity and inbreeding on the adaptive potential of the species and will therefore enable tailored and informed conservation management of this threatened taonga (treasured) species.

KEYWORDS
conservation genomics, differential methylation, genome assembly, *Notiomystis cincta*, ONT sequencing, sex chromosomes

1 | INTRODUCTION

Anthropogenic changes to the planet have had significant and ongoing detrimental effects for biodiversity (Ceballos et al., 2017). Conserving species has become a critical issue as the loss of complex ecosystems will impact the global food supply, access to biomaterials and medicines and is likely to exacerbate climate change (Lewin et al., 2018). Many species and ecosystems also have a high cultural value, which means there is a generational responsibility to preserve them (Hare et al., 2019). Multidisciplinary approaches are required to enable the preservation of biological diversity (Soulé, 2013) and genomic techniques are increasingly being applied as one 'tool' in the conservation biology 'toolbox' to inform conservation management of threatened species (Allendorf et al., 2010; Brandies et al., 2019; Supple & Shapiro, 2018). Genomic data and, in particular, a reference genome assembly, allow genetic diversity and adaptive variation to be studied at a more refined assessment scale than traditional methods that survey variation at a small number of regions in the genome using relatively few molecular markers (Allendorf et al., 2010; Shafer et al., 2015).

For threatened species, a genome assembly is a valuable resource because it allows for considerations of genetic processes in recovery plans to ensure a population's long-term viability (Brandies et al., 2019). Threatened species are characterized by isolation and small population sizes, which often leads to the reduction of genetic diversity and high levels of inbreeding (Brandies et al., 2019; Dussex et al., 2021; Kardos et al., 2021; Morin et al., 2020). Genome-wide data alongside an annotated reference genome can aid the fine-scale assessment of inbreeding depression, the reduced fitness of offspring from related parents, whereby genome annotation can help identify genes and the encoded biological pathways impacted by inbreeding depression (Allendorf et al., 2010; Keller & Waller, 2002; Mathur & DeWood, 2021). A genome assembly can also be used to help understand the sex-specific patterns of susceptibility to inbreeding (Duntsch et al., 2023). Genome-wide assessment of genetic diversity can be used to monitor populations and assess the adaptive potential of species to cope with future pressures such as environmental change caused by climate change (de Villemereuil, Rutschmann, Lee, et al., 2019; Supple & Shapiro, 2018).

Further, in managed populations with no pedigree information, genomic analysis can infer relatedness and inform breeding programs that seek to maximize genetic diversity and maintain adaptive potential (Galla et al., 2020). Genomic data can also be used to identify genomic regions under selection that might confer local adaptations, which may need to be maintained to ensure the fitness of a population in a particular habitat (Allendorf et al., 2010). It can also be used to track changes associated with conservation interventions (e.g. effects of captive breeding or genetic rescue attempts on long-term genetic health) in order to improve management interventions (Wright et al., 2021). Comparing genomic data of individuals can help define species boundaries and identify conservation units that are important when implementing policies and allocating resources for species conservation (Fuentes-Pardo & Ruzzante, 2017; Supple & Shapiro, 2018), or identify hybridisation between closely related species, which may negatively impact their fitness (Forsdick et al., 2021; Supple & Shapiro, 2018). Finally, a reference genome can give insight into the demographic history of a threatened species to infer previous population sizes compared to present-day populations (Bursell et al., 2022; Dussex et al., 2019; von Seth et al., 2021; Yan et al., 2023) and can also allow a comparison of genetic diversity over more recent timescales or across populations (Dussex et al., 2021; Feng et al., 2019; Mathur & DeWood, 2021; Wright et al., 2021).

With the development of sequencing technologies, the availability of genome assemblies for non-model species has increased due to the efforts of large genome sequencing consortia (Feng et al., 2020; Lewin et al., 2018; Rhie et al., 2020) as well as the contribution of many individual lab groups (e.g. Peñalba et al., 2020; Peona, Blom, et al., 2021; Prost et al., 2019; Robledo-Ruiz et al., 2022; Stuart et al., 2022). The optimal assembly approach is constantly updated as new sequencing technologies emerge with improved capabilities (Giani et al., 2020). Long-read sequencing technologies have transformed genome assembly as they can achieve high contiguity by spanning repetitive regions (Peona, Blom, et al., 2021; Whibley et al., 2021). Chromosome-level assemblies are increasingly achievable by exploiting scaffolding techniques such as chromatin conformation capture, linked reads, optical mapping or a linkage map to orient contigs and scaffolds in relation to each other (Whibley et al., 2021). A high-quality assembly then enables information about the structure and function of the DNA sequence to be added via gene annotation, including information for genes that are likely important for fitness, such as Toll-like receptor genes that have a role in immunity (Grueber et al., 2015).

For species with chromosomal sex determination (i.e. XX female vs. XY male or ZZ male vs. ZW female), an individual from the homogametic sex has commonly been chosen for genome sequencing because read coverage is equal across all regions of the genome, in contrast to the heterogametic sex where each sex chromosome is only present in one copy (Tomaszkiewicz et al., 2017; Xu et al., 2019). Further, while sex chromosomes often share a pseudoautosomal region, where chromosomes pair and recombine during meiosis (Zhou et al., 2014), the chromosome unique to the heterogametic sex is commonly degraded, meaning that it is highly repetitive, has reduced in size and has experienced gene loss relative to the homogametic chromosome (Bachtrog, 2013; Charlesworth & Charlesworth, 2000). The sex-limited chromosome is often highly condensed into heterochromatin, likely in order to repress the transposable element content (Peona, Palacios-Gimenez, et al., 2021). Assemblies of W and Y chromosomes are therefore underrepresented as it is difficult to assemble them, given that regions of the genome with high repeat and heterochromatin content are often under-sequenced (Peona, Blom, et al., 2021; Weissensteiner & Suh, 2019). Further, it has been assumed that the sex-limited chromosome contains mostly unimportant repetitive sequences (Peona, Palacios-Gimenez, et al., 2021), meaning that little emphasis has been placed on targeting W or Y chromosome sequencing assembly or attempting to identify W or Y chromosome content from unannotated regions of a genome

assembly. However, given the role of the sex-limited chromosome in reproductive isolation, fertility, disease and ageing (Peona, Palacios-Gimenez, et al., 2021), assembly and characterization of both sex chromosomes is likely to aid conservation genomics programmes aiming to conserve genome-wide diversity, identify the genetic basis of fitness traits, and understand differences in selection between the sexes.

In birds, the size reduction of the W chromosome is particularly dramatic in passerines, with the non-recombining portion of the W chromosome ranging from 1.33 to 7.24 Mb, corresponding to only 1.9%–8.5% of the Z chromosome length (Xu et al., 2019). Despite a rapidly increasing number of genome assemblies for birds (673 as of 22nd November 2022, https://www.ncbi.nlm.nih.gov/genome [Search details: "Aves"[Organism]]), only around 8.5% (57 out of 673) of assemblies include W chromosome assemblies (Table S1). However, this estimate fails to account for female assemblies where W-linked contigs have been identified (Xu et al., 2019) or there is a W-linked scaffold (Peona, Blom, et al., 2021).

Older techniques for sequencing the sex-limited chromosomes typically involved targeted sequencing, for example via chromosome microdissection, chromosome flow-sorting or targeted capture (Tomaszkiewicz et al., 2017). More current approaches for identifying sex-linked contigs use bioinformatics workflows which can be applied post de novo assembly (Palmer et al., 2019; Tomaszkiewicz et al., 2017). Long-read sequencing technologies have been pivotal in overcoming the challenging properties of W chromosomes, including the ability to sequence across large repeat regions. Still, twice the sequencing coverage of W chromosomes is needed to achieve equivalent coverage compared to autosomes. Following de novo assembly, sex-linked contigs can be identified via alignments to a high-quality reference genome of a closely related species, and/or comparisons of the male and female genomic depth of coverage, and/or comparisons between the genome content of a male and female genome assembly. In the case of utilizing the reference genome of a close relative, homology is used as evidence for determining the sex-linked contigs in the focal species, and this approach can be used to identify both Z and W contig (Xu et al., 2019; Zhou et al., 2014). The genomic depth of coverage approach is based on each sex chromosome being present at one copy in females and not in homologous ZZ pairs as they are in males; therefore, in females, sex-linked contigs should be present at half coverage compared to autosomes, Z-linked contigs will have twice the genomic depth of coverage in males compared to females and W-linked contigs should be mapped by very few male reads (Gan et al., 2019; Smeds et al., 2015; Xu et al., 2019; Zhou et al., 2014). Finally, mapping male and female genome assemblies against each other can reveal contigs that are unique to females and hence identify potential W-linked contigs. Each approach can be used on its own but combining multiple approaches, so that independent signals from each approach can be integrated, increases the success of identifying sex-linked contigs (Palmer et al., 2019).

In contrast to the wide availability of avian genomes (Bravo et al., 2021), studies of methylation and other epigenetic marks are rare, possibly due to the technical challenges that arise with sampling and measuring epigenetic marks (Lindner, Verhagen, et al., 2021; Yong et al., 2016). Methylation has been more extensively studied in humans and other mammals than in birds, but it is believed to serve the same function in both systems and is dependent on the sequence context it occurs in. Promoter methylation has been shown to downregulate gene expression (Laine et al., 2016), change chromatin structure and prevent transcription activation (Yong et al., 2016). Methylation of gene bodies downregulates gene expression (Laine et al., 2016) and has a role in gene splicing and the downregulation of repetitive elements within a gene body (Yong et al., 2016). The chicken (*Gallus gallus*) W chromosome has been observed to be highly methylated compared to Z and autosomes (Zhang et al., 2017), likely due to its repetitive content. In birds, DNA methylation, and the likely changes to gene expression as a result, has been associated with many processes including breeding (Lindner, Verhagen, et al., 2021), migration (Saino et al., 2017), and ageing (Sun et al., 2021). Despite methylation studies in a number of wild bird populations (e.g. Saino et al., 2017; Schrey et al., 2012), there is little work to-date linking methylation to phenotype in wild threatened species.

Methylation is most accurately assayed using bisulfite sequencing (Frommer et al., 1992). However, some third-generation sequencing technologies can distinguish between methylated and unmethylated bases (Yong et al., 2016). For Oxford Nanopore Technologies sequencing, modified bases can be detected due to a characteristic change in the electrical charge as a modified base passes through the pore (Rand et al., 2017). Extended basecalling can be used to specifically detect these modified bases, and has been shown to correlate well with bisulfite sequencing results (Liu et al., 2021). These recent findings offer promise that epigenetic modifications can be reliably detected from Nanopore data without the need for more expensive and laborious processes such as sodium bisulfite sequencing.

Hihi is a small, threatened passerine bird, endemic to the North Island of Aotearoa, New Zealand. Hihi is a forest dwelling species and sexually dimorphic, with males larger and more colourful than females (Ewen et al., 2006; Figure 1). Hihi were lost rapidly from the mainland of Aotearoa during European settlement, as this led to large-scale removal of native forests to turn to pasture and further mammalian predator introductions. One hihi population remained on an offshore island, Te Hauturu-o-Toi (36°12′S, 175°05′E) (Toy et al., 2018), until the 1980s when a reintroduction program was established (Brekke et al., 2011). Hihi have been successfully established in a further six predator-free sites across the North Island, that require active management. As part of ongoing management, two of these reintroduced populations, on the island of Tiritiri Matangi (36°36′S, 174°53′E) and Zealandia Sanctuary (41°17′S, 174°45′E), have been intensely monitored (Rutschmann et al., 2020; Thorogood et al., 2013). Individual-based longitudinal datasets containing morphological and life-history information as well as a large blood/tissue repository has enabled the development of genetically resolved pedigrees (Brekke et al., 2015; de Villemereuil, Rutschmann, Ewen, et al., 2019; de Villemereuil, Rutschmann, Lee,

FIGURE 1 Hihi (*Notiomystis cincta*) and their distribution. Right-hand side: A map showing the locations of all seven extant hihi populations across the North Island of Aotearoa New Zealand, including the only remnant population on Te Hauturu-o-Toi (in red). Top-left corner: picture of a male hihi. Bottom-left corner: picture of a female hihi. Both photographs taken by Charlotte Johnson and included with permission.

et al., 2019; Rutschmann et al., 2020) and other molecular resources (see below). Hihi reintroductions have reduced its extinction risk by increasing the number of individuals and populations, with approximately 2500 hihi across all sites (www.hihiconservation.com). However, the genetic bottlenecks associated with the establishment of new populations, as well as small population sizes, due to limited suitable habitat, have led to genetic erosion (Brekke et al., 2011; de Villemereuil, Rutschmann, Lee, et al., 2019), loss of adaptive potential (de Villemereuil, Rutschmann, Ewen, et al., 2019; de Villemereuil, Rutschmann, Lee, et al., 2019; Duntsch et al., 2020; Rutschmann et al., 2022), the accumulation of inbreeding and sex-biased inbreeding depression (Brekke et al., 2010, 2012; Duntsch et al., 2023) which risk the genetic health and long-term viability of these populations.

The blood/tissue repository for hihi has led to the development of several molecular resources. Twenty polymorphic microsatellite markers were characterized in hihi (Brekke et al., 2009), and used for assessing genetic structure and to genetically resolve the Tiritiri Matangi pedigree (Brekke et al., 2011, 2015). Two additional sex-linked microsatellite markers are used for assigning sex to embryos, nestlings and juveniles when sexual size and plumage dimorphism is not evident (Brekke et al., 2010; Dawson et al., 2015). Genetic diversity has been measured at a selection of toll-like receptor (TLR) loci (Grueber et al., 2015). Several whole-genome datasets have been generated to analyse genetic diversity and inbreeding. Low coverage whole genome sequencing (lcWGS) data was generated for 10 individuals using Illumina sequencing (de Villemereuil, Rutschmann, Lee, et al., 2019; Lee et al., 2021) and Restriction-site Associated DNA sequencing (RAD-seq) markers were also generated for 26 individuals (de Villemereuil, Rutschmann, Lee, et al., 2019; Lee et al., 2021), with identified polymorphism used to design a SNP array for hihi (Hihi 50K AXIOM 384HT array; Lee et al., 2021). The hihi SNP array data informed an autosomal linkage map with chromosomes named and identified using homology to the zebra finch (*Taeniopygia guttata*) reference genome (Scherer, 2017). The mitochondrial genome from a hihi from Tiritiri Matangi has been sequenced with Sanger sequencing and assembled to construct passerine phylogenies and improve timing estimates of the worldwide radiation of passerine birds (Gibb et al., 2015). Before commencing this project, draft versions of both the male and female assembly had been generated. The initial male assembly was used to compare RAD-seq, SNP array and lcWGS data to evaluate inbreeding measures (Duntsch et al., 2021) while the earlier female assembly was used to assess inbreeding depression (Duntsch et al., 2023).

A genome assembly is available of a hihi individual of unknown sex from the Bird 10,000 Genomes (B10K) consortia (Feng et al., 2020). However, this assembly is highly fragmented, limiting inferences about genetic diversity and inbreeding, which are important for conservation management of hihi (Rhie et al., 2020). A high-quality assembly will aid hihi conservation as a resource for assessing genetic variation in hihi and monitoring populations. Further, the B10K assembly was generated from an unknown individual that was not provided by a New Zealand institution and may not have had appropriate iwi (extended kinship group of Māori, the Indigenous People of Aotearoa) consultation for its use to generate genomic data.

Here, we present highly contiguous genome assemblies of a female (ZW) and male (ZZ) hihi to serve as a foundational resource for inferring the impacts of low genetic diversity and inbreeding on this threatened species. By identifying both W and Z chromosome sequences, we also enable future exploration of sex-linked gene expression and genomic architectures, and comparative analysis of this distinct evolutionary lineage.

2 | MATERIALS AND METHODS

2.1 | Biological samples

The two reference genome individuals are an adult male sampled in 2017 and a juvenile female sampled in 2018 from the remnant hihi population on Te Hauturu-o-Toi. Both individuals were caught

in mist nets and then blood sampled via venipuncture. Blood was preserved in 95% ethanol and stored at 4°C for the duration of field work before being stored long term at −30°C.

2.2 | Genomic sequencing

Genomic DNA was extracted using the Monarch® Genomic DNA Purification Kit or the Monarch® HMW DNA Extraction Kit (New England BioLabs) following the manufacturer's instructions. Some extracts were run through size selection prior to library preparation using the Blue Pippin High Pass Plus™ gel cassette (Sage Science) and targeting reads greater than 15 kb. Typically, 1.5 µg gDNA was used as input to the Oxford Nanopore Technologies (ONT) ligation sequencing protocol (LSK-109 or LSK-110) and following the manufacturer's protocol (including use of the Long Fragment Buffer in the final wash steps). Libraries were sequenced on a MinION device with R9.4.1 flow cells, with a target loading of 50–100 fmol. Multiple library loads were required for each flow cell, with intervening nuclease flushes (EXP-WSH004) to restore pore capacity and maximize sequence yield.

Illumina whole genome sequence libraries for the two genome individuals were prepared by AgResearch, Invermay, New Zealand using the Illumina DNA Prep library kit and sequenced on a NovaSeq 6000 in 150PE mode.

2.3 | Genome assembly

Oxford Nanopore Technologies read basecalling was performed using the super accurate model in GUPPY v5.0.7 with the min-qscore flag set to 7. GUPPY is a production basecalling tool developed and released by Oxford Nanopore Technologies and can be obtained by download from www.nanoporetech.com after registration. After basecalling, PORECHOP v0.2.4 was used to detect and remove residual sequencing adaptors (Wick et al., 2017) and the sequencing control strand was removed using NANOLYSE v1.2.0 (De Coster et al., 2018). Filtering thresholds and assembler performance and parameter optimisation were explored during initial phases of the project. The assemblies reported here imposed a minimum quality score of 10 and a minimum read length of 5 kb. Read filtering and summary statistics were performed using the NANOPACK suite of tools (De Coster et al., 2018). The properties of the read datasets are shown in Table S2. The primary assemblies were generated using FLYE v2.8.1 (Kolmogorov et al., 2019) in nano raw mode with the –keep_haplotypes flag enabled and an estimated genome size of 1.0 Gb.

Long read polishing was performed using two iterations of RACON v1.4.21 (Vaser et al., 2017). MINIMAP2 v2.20 (Li, 2018) was used to align all reads with a quality score greater than or equal to 10 back to the genome assembly before RACON polishing with the following settings: -m 8 -x -6 -g -8 -w 500, as recommended by MEDAKA (Oxford Nanopore Technologies, 2018) developers. MEDAKA v1.4.3 polishing was then conducted using default parameters and the r941_min_sup_g507 model.

Illumina short-read data from the two genome individuals was filtered using the TRIMGALORE wrapper to CUTADAPT (Martin, 2011) with default parameters plus the "–2-colour" flag enabled. The filtered Illumina reads were used for a final round of polishing with NEXTPOLISH v1.4.1 (Hu et al., 2020), as well as to obtain estimates of genome parameters using GENOMESCOPE (Vurture et al., 2017). For polishing, we performed two iterations of read-mapping to the draft assemblies using BWA mem (Li & Durbin, 2009), processing using SAMTOOLS (Li et al., 2009) to remove duplicates and running the NEXTPOLISH script with default parameters. The Illumina dataset was available late in the project cycle and so initial manual curation and transposable element library curation was performed prior to this final polishing; with repeat and TE detection performed on the final polished genomes. The genome assembly strategy is illustrated in Figure S1.

2.4 | Assembly evaluation

Summary statistics for the assemblies were generated using QUAST v5.2.0 (Gurevich et al., 2013), and BUSCO v5.3.2 with MetaEuk was used to assess assembly completeness with the ODB10_aves dataset (Manni et al., 2021). The contigs containing the mitochondrial genome were identified by blastn (Altschul et al., 1997) screening the assembly using the Sanger-sequenced GenBank accession KC545400.1 as the query. *K*-mer-based ($k=31$) assessments of assembly completeness and accuracy were made using MERQURY v1.3 (Rhie et al., 2020). Contamination in the assemblies was assessed using blastn searches to the BLAST nr nucleotide database (NCBI Resource Coordinators, 2014) and diamond blastx (Buchfink et al., 2014) searches to the Uniprot reference proteome (The UniProt Consortium, 2021). The collated hits from these searches were used to form a BLOBTOOLS database (blobDB) and visualized with a blobplot (Laetsch et al., 2017).

2.5 | Autosome scaffolding

Scaffolding using a genetic linkage map was performed with CHROMONOMER v1.14 (Catchen et al., 2020). The linkage map describes the relative ordering and location of markers across each linkage group in the genome, and by mapping the physical positions of these markers to the draft assembly, the contigs can be ordered and oriented into their linkage group positions. An autosomal linkage map had previously been constructed for hihi (Scherer, 2017) based on genotypes from our 50 K SNP array (Lee et al., 2021). In brief, following quality control, markers were first grouped into linkage groups based on homology of the SNP and its flanking sequence with the zebra finch genome. Those that aligned to the Z or W chromosomes were excluded. Within each of the autosomal linkage groups, marker order and positions were inferred using information from 436 individuals from the Tiritiri Matangi Island

population with confirmed family relationships. Markers were added only when their placement in the map was 1000 times more likely than any other position within the existing map, and local rearrangements of up to seven markers were assessed to select the order with the highest likelihood. Linkage map construction was performed with CRI-MAP v2.4 (Green et al., 1990), modified by Xuelu Liu (Monsanto), and CRI-MAP v2.507 (Evans & Maddox, 2015) to accommodate large numbers of markers and complex pedigree structures in constructing the linkage map. In total, the linkage map positioned 1773 SNP markers within 32 framework linkage groups representing 31 chromosomes, with a total map size of 2429.1 centi-Morgans (cM) (Scherer, 2017).

BWA mem was used to map the 50K SNP markers and their flanking sequence to the genome assemblies and this, along with the assembly contigs and the genetic map positions of the markers, were used as input to CHROMONOMER. In the initial round of scaffolding, a small number of chimeric contigs were identified in each of the assemblies due to SNP markers from two or, rarely, more linkage groups being located on the same assembly contig. Since the long-read contigs do not contain gaps, contig breaks were not introduced by CHROMONOMER automatically, and we instead used this information to inform manual curation. Starting from the misjoin interval identified by CHROMONOMER, we used depth of coverage, read-mapping positions and end-point information from the male and female draft assemblies to inform the placement of introduced contig breaks. The curated assembly contigs were then re-run through the CHROMONOMER pipeline to generate the final organization of the autosomal chromosome scaffolds, which consist of the contigs oriented and ordered with 100 bp "N" spacers introduced at the breakpoints. Scaffolded chromosomes were named based on homology with the zebra finch genome.

At this stage, small contigs (those under 2 kb) and contigs that showed weak support due to low coverage from short and/or long reads were also removed from the assembly. Contigs which demonstrated elevated coverage were tagged as likely collapsed repeats but retained in the assembly.

2.6 | Identification of sex-linked contigs

To identify sex-linked contigs and classify them as either W or Z in origin, we combined coverage-based analyses using both the Illumina and ONT read datasets with comparative genomics evidence from other avian genomes. ONT long reads were mapped back to the draft assembly using MINIMAP2 with the -x ont flag. Alignments were sorted with SAMTOOLS and reads with a mapping quality of <20 were excluded. Illumina reads were mapped to the assembly with BWA mem and processed with the same MQ >20 filter. We computed mean and median read coverage using MOSDEPTH v0.3.3 (Pedersen & Quinlan, 2018), normalized the coverage measures to account for the difference in sequencing effort between the two samples and assessed the female to male (F:M) coverage ratio for entire contigs and for non-overlapping 10 and 50 kb windows. Z-linked contigs have an expected F:M ratio of 0.5 compared to a ratio of 1 for the autosomes. Xu et al. (2019) employed a metric based on the ratio of mappable bases in males versus females and we used SAMTOOLS depth to recover counts and to calculate this, employing a coverage filter to retain reads with coverage <150 to exclude regions with a large contribution from collapsed repeats.

At the time of this study, we could identify only four published avian genomes with high-quality contiguous W chromosome assemblies (Table S3). We used RAGTAG v2.1.0 (Alonge et al., 2019) and an underlying assumption of chromosome synteny to link the hihi contigs to the Z or W of these references. We considered the placement of a contig >50 kb in length to the W or Z chromosome of one or more of these references with a confidence score of >0.95 to support the localisation, and evaluated this alongside contig coverage information. We also used this approach to cross-check the chromosomal assignment of the autosomal scaffolds and putatively assigned contigs >50 kb that were mapped to autosomal chromosomes with a confidence score of >0.95 as unplaced contigs of that chromosome. In the absence of linkage map information for the Z and W chromosomes, we were unable to scaffold the assigned sex-linked contigs into pseudo-chromosomes and so these are provided as annotated contigs in the final assembly.

2.7 | Genome repetitive content

Repeat libraries and landscapes for the male and female genome were generated with EARL GREY v1.2 (Baril et al., 2022). EARL GREY was developed as a TE discovery and annotation pipeline that also conducts automated curation of TE libraries, with the goal of generating improved TE consensus sequence lengths and reduced library redundancy compared to pipelines that omit TE curation. Although it has a focus on TE identification and classification, EARL GREY also identifies other repetitive sequences and summarizes these as satellites or as unclassified elements. The pipeline first uses REPEATMASKER to identify known repetitive elements, in our case sourced from the DFAM v3.2 database (Smit et al., 2013; Storer et al., 2021). Next, REPEATMODELER2 performs de novo TE discovery (Flynn et al., 2020). The pipeline-identified elements are processed to remove redundancy and an automated implementation of the "BLAST, extract, extend" approach is used to generate an optimal set of consensus sequences (Camacho et al., 2009; Capella-Gutiérrez et al., 2009; Katoh & Standley, 2013). The genome is then revisited with REPEATMASKER using the curated library and the output is combined with LTRFINDER results and post-processed with REPEATCRAFT (Ou & Jiang, 2019; Wong & Simakov, 2018).

2.8 | Gene annotation

The larger curated TE library from the female genome was used to soft-mask both genome assemblies before gene annotation using

GALBA v1.0.0. GALBA uses homology evidence, taking high-quality protein predictions from a single species (here, chicken NCBI annotation release 106 GCF_016699485.2-GRCg7b along with UniProt SwissProt Release 2022-10-12), to train AUGUSTUS models to generate gene predictions in the target genome (Hoff et al., 2019; Hoff & Stanke, 2019; Stanke et al., 2006). FUNANNOTATE v1.8.14 accessory scripts were used to clean and sort the assembly contigs prior to running the GALBA pipeline (Palmer & Stajich, 2020). Annotations were obtained by running GALBA in CRF-training mode (--crf) and with protein alignment using MiniProt (Li, 2023). The GALBA command also included the following flags: --softmasking --gff3 --species='Notiomystis cincta'. The female genome was annotated using this method, and the male genome was annotated using the female predictions to ensure similarity between the two annotations. Annotation quality was evaluated using BUSCO in protein mode with the aves ODB10 dataset. Functional annotation of the predicted proteins was performed using the EGGNOG-MAPPER v2.1.9 and the EGGNOG DATABASE v5.0.2 (Cantalapiedra et al., 2021; Huerta-Cepas et al., 2019). CpG islands were predicted using the EMBOSS v6.6.0 function cpgplot with default parameters (Rice et al., 2000). The mitochondrial contigs were annotated separately, using the MITOS2 webserver (http://mitos2.bioinf.uni-leipzig.de/; Donath et al., 2019). ORTHOFINDER v2.5.2 was used to identify orthologs between the male and female predicted protein datasets, and between the female dataset and both the chicken proteome used to support gene prediction and the zebra finch GCF_003957565.2 proteome (Emms & Kelly, 2019).

2.9 | Native DNA methylation analysis

The ONT raw fast5 data were revisited using a Remora-modified base calling model (dna_r9.4.1_450bps_modbases_5mc_cg_sup.cfg) in GUPPY v6.2.1 with the polished female genome assembly supplied as a reference. The model calls 5-methylcytosine (5mC) in a CpG context. The output bam files were processed with MODBAM2BED (https://github.com/epi2me-labs/modbam2bed) to generate a bed file with the counts of modified and non-modified bases at each CpG site, with signals from forward and reverse reads aggregated for downstream analyses.

In order to identify differentially methylated regions (DMRs) between the male and female individuals we first filtered the dataset to retain only sites that had a coverage depth of >5 reads and which were recovered in both individuals. We also excluded contigs that had been assigned to the W chromosome. The per-site modification data were analysed using the DMA function in NANOMETHPHASE v1.2, which calls the R package DSS with default parameters (Akbari et al., 2021; Park & Wu, 2016; R Core Team, 2021). DSS clusters differentially methylated sites into DMRs using thresholds of a minimum length of 100 bp, at least 15 CpG sites within the DMR and at least 50% of CpG sites having a significant *p*-value, with adjacent sites closer than 100 bp being automatically merged together. METHYLARTIST v1.2.6 and IGV v2.15.4 were used to generate visualizations of methylation signals (Cheetham et al., 2022; Thorvaldsdóttir et al., 2013).

TABLE 1 Global genome assembly statistics and quality assessment.

	Female genome	Male genome
Total assembly length (Mb)	1063.4	1045.8
Number of contigs	471	497
GC content (%)	42.88	42.81
Contiguity statistics		
Contig N50 (Mb)	7.00	6.87
Contig L50	47	51
Completeness statistics		
Ns per 100 kb	0.19	0.00
BUSCO complete *n* (%)	8077 (96.8)	8083 (96.9)
BUSCO complete and single copy *n* (%)	8033 (96.3)	8042 (96.4)
BUSCO complete and duplicated *n* (%)	44 (0.5)	41 (0.5)
BUSCO fragmented *n* (%)	46 (0.6)	44 (0.5)
BUSCO missing *n* (%)	215 (2.6)	211 (2.6)
MERQURY completeness	96.68	97.08
Accuracy statistics		
MERQURY QV	44.27	44.09
MERQURY estimated error rate	$3.74e^{-05}$	$3.90e^{-05}$

3 | RESULTS

Independent genome assemblies for the female and male individuals were generated using the FLYE assembler from reads basecalled with GUPPY version 5 using the super accurate model. The ONT sequence reads were used to polish these assemblies using RACON and then MEDAKA before introducing a final polishing step using NEXTPOLISH and a dataset of Illumina reads for the two individuals. The Illumina dataset was also used to generate *k*-mer based estimates of genome properties and for assembly evaluation using MERQURY.

The assembly properties are shown in Table 1. The male total assembly length of 1.046 Gb corresponded well to the *k*-mer base sized estimate produced by GENOMESCOPE of 1.038 Gb., The female assembly, which is larger than the male assembly due to the additional recovery of the W chromosome, was 1.063 Gb in length. In comparison, the GENOMESCOPE estimate of genome size of the female individual is 1.003 Gb due to the haploid contribution of the sex chromosomes in that context (Figure S2). Estimated heterozygosity was 0.189% and 0.371% in the male and female genomes, respectively, again with the heterogametic sex chromosomes likely inflating the estimate in the female. No contaminant non-avian sequences were detected in the genome assemblies as assessed by BLOBTOOLS (Figure S3).

The female assembly backbone was constructed from a dataset of 30.5 Gb ONT sequence, with a read N50 of 29.7 kb. The ONT contribution to the male assembly was 47.9 Gb sequence with a read N50 of 16.0 kb (Table S2). Despite the higher coverage of the male ONT dataset compared to the female, the assembly properties were broadly similar, with the genome represented by 471 contigs in the female and 497 contigs in the male, and N50 statistics of 7 and 6.9 Mb respectively (Table 1). Genome completeness as measured by BUSCO (gene-based) and MERQURY (*k*-mer-based) assessments were closely aligned and approached 97% (Table 1). Of the missing BUSCO genes, over 90% ($n = 196$) are absent from both genomes, suggesting these may be either genuinely absent or difficult to recover with current assembly approaches. Short-read polishing had minimal impact on structural features of the genome assembly but did improve the QV estimates of genome accuracy. Prior to short-read polishing the female and male genomes reported QV scores of 37.3 and 38.8, after short-read polishing these values increased to 44.3 and 44.1.

The mitochondrial genomes were recovered in a single contig for both the male and female assemblies (Figure S4). The two mitochondrial genomes recovered here are 99.98% identical to each other and 99.97% identical to the Sanger-sequenced accession reported in Gibb et al. (2015), which has been noted to contain an unusual gene order involving duplication of tRNA Pro-CR compared to other passeriforme genomes. Collectively these three 18.65 kb mitogenome accessions differ at just seven single nucleotide sites.

3.1 | Autosome scaffolding

Autosomal scaffolding of the draft assemblies was performed using a genetic linkage map generated from population genotyping data using our 50 K SNP array (Lee et al., 2021). Over 99.9% of the 50 K SNP array SNPs were successfully mapped back to the reference assemblies and 1773 map-informative markers were processed by CHROMONOMER to organize and orient the contigs into pseudo-chromosomes. These pseudo-chromosomes were then named based on homology with the zebra finch genome. When using the linkage map to scaffold the assembled contigs from the female assembly, 1696 markers were retained in the final map and 63 inconsistent markers were excluded. 914 Mb of assembly sequence, comprising 173 contigs and 14.2 kb of introduced gaps, was integrated into autosomal linkage groups. One of the contigs spanned the linkage groups for chromosome 1 and chromosome 1B, so the linkage map was updated to include this information by reassigning the 1B markers and adjusting the cM positions of the chromosome 1 markers to accommodate the terminal 1B marker dataset. Scaffolding was performed in the male genome in parallel, with 1698 markers retained in the map and 62 excluded. 913 Mb of assembly sequence, comprising 187 contigs and 15.6 kb of scaffold join gaps, was integrated into linkage groups. We did not initially recover chromosome 16, which is reported to contain the avian Major Histocompatibility Complex and is often not well-resolved in avian assemblies. tBLASTn was used to identify a single contig in both genomes that contained multiple hits to the MHC class I and MHC class II antigen proteins characterized in zebra finch, and these contigs were manually added to the assembled genomes as "chromosome 16". The highly syntenic arrangement of the autosomes in the female and male assembly is shown in Figure 2a. As expected from other avian genome analyses (e.g. Waters et al., 2021) and despite over 90 million years of separation, synteny between the chicken and hihi genome is broadly conserved with no major chromosome-scale rearrangements beyond well-characterized chromosome fission events (Figure S5).

3.2 | Identification of sex-linked contigs

Contigs were associated with the Z and W chromosomes using a combination of coverage-based assessments (using both long-read and short-read datasets) and homology to characterized sex chromosomes in high-quality avian assemblies. In general, these two lines of evidence were congruent.

In the female assembly, 19 contigs >50 kb in length, totalling 75.07 Mb, showed coverage profiles consistent with a single chromosome copy in the female genome and with two copies in the male genome. All 19 contigs were localized to the Z chromosome in the zebra finch and paradise crow assemblies, with one unplaced contig and 18 Z-chromosome-placed contigs in the New Caledonian crow assembly. In addition to the 19 contigs confidently assigned to the Z chromosome, a further 14 contigs (mean length 10.3 kb, 143 kb in total) had female:male coverage ratios consistent with a Z-chromosome localisation. One of these contigs (Ncf_contig_638, 44 kb) was linked to the zebra finch W chromosome by homology but had coverage consistent with Z-chromosome localisation in both long- and short- read datasets. In the female assembly, the pseudo-autosomal region (PAR) was recovered as a single 884 kb contig (Ncf_contig_1251) with equivalent coverage in the male and female datasets. This contig was excluded from analyses of the Z/W chromosome properties of the female genome.

In the male assembly, we identified 18 contigs >50 kb in length that showed coverage profiles consistent with Z-linkage and one further contig corresponding to the PAR (Ncm_contig_2631, 739 kb). Together, these contigs comprised 75.67 Mb sequence. Again, the assignment to the Z was supported by homology evidence. All 18 Z chromosome contigs were identified as such with reference to the zebra finch and paradise crow assemblies, and 17 of the 18 were placed on the New Caledonian crow Z assembly, with one unplaced contig. The PAR contig was identified as W-chromosomal with reference to all three assemblies. As with the female genome, there were several <50 kb contigs that showed coverage profiles consistent with Z-linkage and together these 10 contigs contributed 172 kb of sequence length and with a mean contig length of 17.2 kb.

A total of 24 contigs >50 kb in length in the female assembly, accounting for 14.81 Mb in total, were linked to the W chromosome by female–male coverage ratio. Twenty-two of these were mapped to the zebra finch W chromosome by RAGTAG2, with the remaining two contigs not placed in the zebra finch genome. Twenty-one

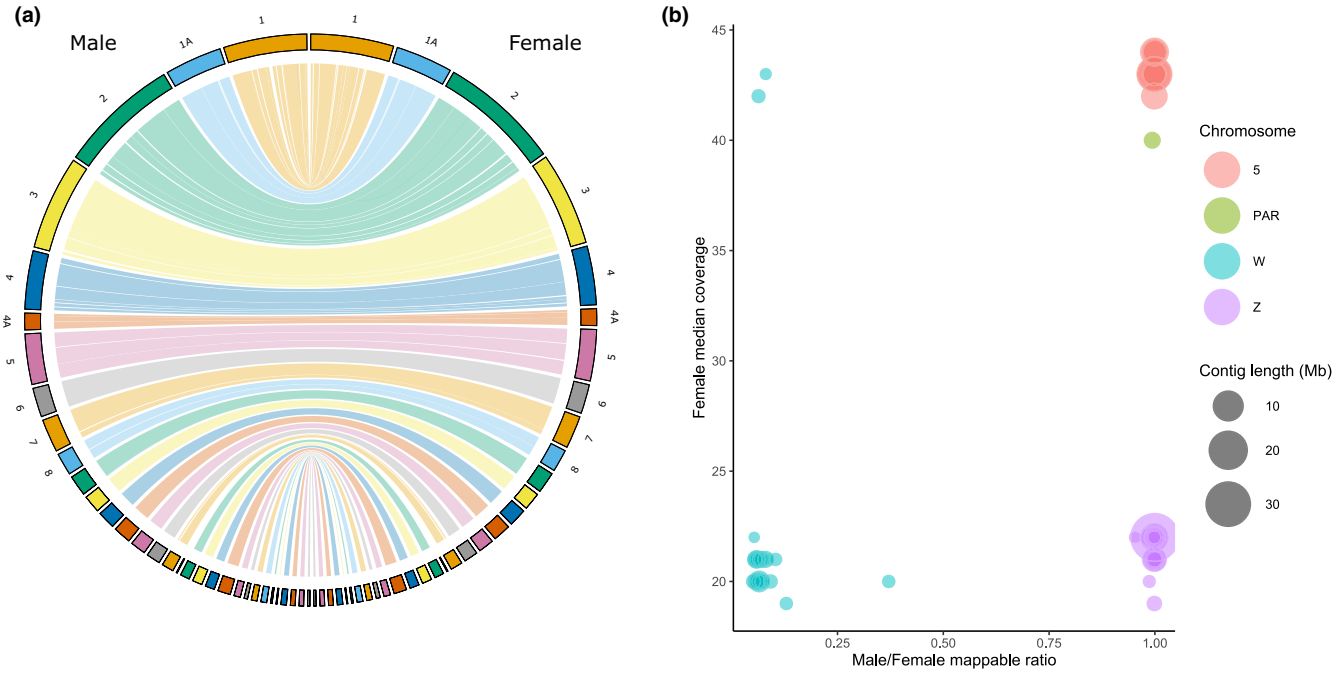

FIGURE 2 (a) Jupiter-type plot of the male and female autosomal hihi genomes. Chromosomes are named by homology to the zebra finch (*Taeniopygia guttata*) genome, with unlabelled chromosomes sequentially representing the microchromosomes 9–29, and including 25A and 25B; (b) Coverage plot showing Z, W and pseudoautosomal region (PAR) scaffold classification. A single autosomal chromosome, chromosome 5, is plotted for comparison. 'Male/Female mappable ratio' represents the proportion of male to female short reads mapping to each contig after correcting for the higher sequencing coverage for the female, 'Female median coverage' represents the median depth of female short read coverage for each contig.

contigs were linked to the W chromosome in the paradise and New Caledonian crow assemblies. Two contigs identified as W-linked likely represent collapsed duplications since despite near absence in the male assembly they both show coverage levels indicative of two genomic copies. These contigs appear as outliers in Figure 2b. Again, there were additional small contigs that could not be confidently ascribed to the W chromosome but which showed coverage profiles consistent with W-linkage. Together, these 8 contigs contributed 129 kb of sequence to the assembly, with a mean contig length of 16.1 kb.

After autosomal scaffolding, 1006 Mb (94.6%) of the female assembly and 989 Mb (94.6%) of the male assembly were assigned to chromosomes. Per-chromosome summary statistics, along with summary statistics for the sex-linked contigs, are provided in Tables S4 and S5 for the female and male genomes respectively. At this level, the contiguity statistics for the scaffolded assembly are dominated by the properties of the chromosomes themselves. In the female assembly, 12 microchromosome scaffolds are composed of a single contig and a further four contain just two contigs. The recovered N50 values for the two scaffolded assemblies are 61 Mb, and 90% of the assembly length is represented by 31 scaffolds in the female assembly and 30 scaffolds in the male assembly. We followed Waters et al. (2021) in classifying chromosomes as macro-chromosomes ($n=9$) or micro-chromosomes ($n=25$), with the smallest macro-chromosome being 38 Mb and the largest micro-chromosome 26 Mb. As is typical for avian genomes, the macro-chromosomes tend to have slightly lower GC proportions and gene densities than the micro-chromosomes. In the female assembly, mean gene density on the macro-chromosomes was 16.6 genes per Mb, whereas this was 45.0 genes per Mb on the micro-chromosomes. GC proportion averaged 41% on the macro-chromosomes and 50% on the micro-chromosomes (Tables S4 and S5).

3.3 | Genome repetitive content

Repeat annotation with EARL GREY suggested that the male and female hihi genomes had an overall repeat density of 10.00% and 11.15%, respectively (Table S6). The difference between the two genomes was largely explained by the long terminal repeat (LTR) retroviral content, which was annotated to encompass 2.9% of the male genome and 3.9% of the female genome. Both genomes had similar contributions from long interspersed nuclear elements (LINEs), which were the major family of elements identified and made up 4.7% of the male genome and 4.8% of the female genome. Contributions from DNA transposons (0.7% in both genomes), unclassified elements (1.0% male, 1.0% female) and satellite DNA (0.5% in both genomes) were broadly very similar. The most abundant element families are CR1 (LINE elements) and ERVL, ERV1 and ERVK (LTR elements).

We classified the genomes into autosomal, Z chromosome and W chromosome subsets and assessed the repetitive content of these partitions (Figure 3). Equivalent partitions for the male and female

datasets are closely aligned. The autosomal contigs numerically dominate the whole genome datasets and so show similar overall statistics (Figure 3a) and repeat landscapes (Figure 3b,c). In addition to reflecting the contribution of the sex chromosomes, the slightly lower densities of elements in the autosomes relative to the total genome also indicates that the unplaced contigs, which were excluded from these counts, are relatively enriched in TEs and satellites. The Z chromosome shows an enrichment of repetitive elements compared to the autosomes (an average density of 14.1% compared to 8.7%). In marked contrast to the properties of the other chromosomes, but in agreement with patterns on other avian chromosomes (Peona, Palacios-Gimenez, et al., 2021), the W has a high density of TEs (80.3%). The hihi W is dominated by LTR elements, particularly the ERVK and ERV families, with LTR elements accounting for 67.8% of the W contig sequence length (Table S6).

It is interesting to note a recent burst of transposon insertions in the genome, indicated by the peak of elements with a low genetic distance from the consensus seen in both the female and male genome repeat landscapes. This burst has contributions from both LTR and DNA transposons, with the latter being largely limited to autosomes since the low genetic-distance DNA transposon peak is seen only in the autosomal landscape (Figure 3c, and not the Z or W landscapes, Figure 3d,e). The DNA transposon families that dominate this recent burst are Ginger-2 elements and CMC-EnSpm, whereas the LTR contributors are, in order of decreasing coverage, ERVL, ERVK and ERV1.

3.4 | Gene annotation

Unfortunately due to collection permit restrictions it was not possible to obtain tissue RNAseq data to support gene annotation from available samples. Further, the blood samples utilized for our genomic sequencing failed to yield RNA of sufficient quality and quantity for us to generate a blood transcriptome. Instead, we generated a preliminary set of annotations using GALBA, which employs homology-evidence alone, opting to use the well-characterized chicken GRCg7b dataset as the reference proteome, as well as UniProt SwissProt. We recovered 22,857 protein-coding genes (35,264 total isoforms) in the female genome and 25,472 protein-coding genes (37,364 total isoforms) in the male. BUSCO evaluations of the annotation completeness were encouraging, each reporting ~97% completeness and low (<1%) duplication levels when the single, longest isoform per protein prediction was supplied (Table S7).

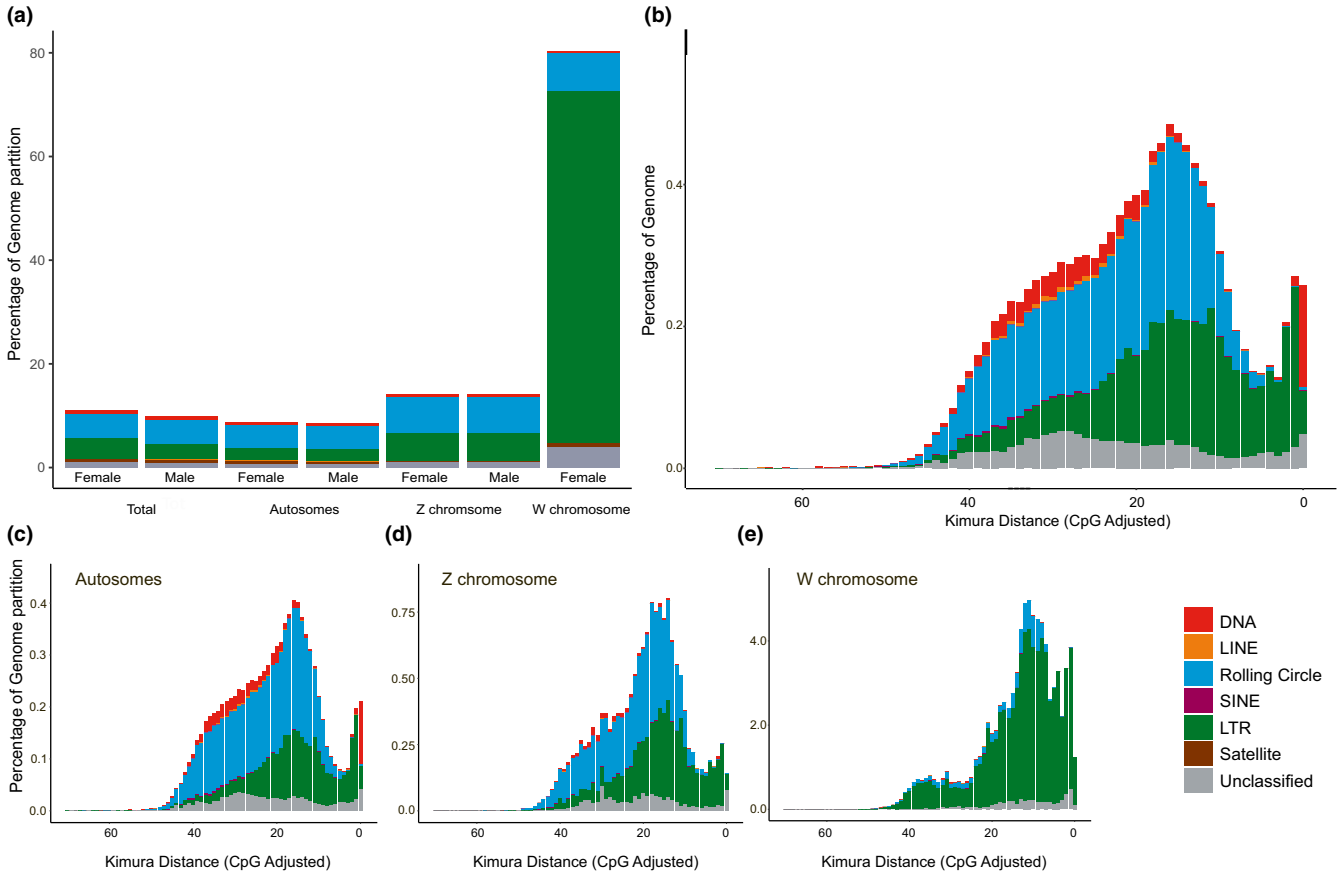

FIGURE 3 Genome repetitive content. (a) The repetitive element composition of genome partitions in the female and male assemblies. (b) Repeat landscape of the female genome. (c) Repeat landscape of female autosomal contigs. (d) Repeat landscape of female Z chromosome contigs. (e) Repeat landscape of female W chromosome contigs. The pseudoautosomal region (PAR) contig is excluded from the female Z and W chromosome datasets. Note the different scales on the *Y*-axis of panels (b–e).

We compared the female and male annotations and assessed the complement of annotated genes with respect to the zebra finch and chicken genomes using ORTHOFINDER (with a single isoform representing each protein). When comparing orthology properties between the two hihi genome assemblies, over 99% of orthogroups contained predicted proteins from both the female and male genome assemblies and a high percentage of the total genes were identified as orthologs between the two assemblies, with just 4.1% of female genes and 6.3% of male genes unassigned to orthogroups (Table S8). 20,214 genes were identified as one-to-one orthologues and these were used as links in the synteny comparisons of the two scaffolded genomes (Figure 2a). Comparisons of the hihi annotation to the zebra finch and chicken genomes indicate that there is some scope for further refinement of the hihi annotations (Table S9). Despite containing a number of predicted genes intermediate between zebra finch ($n=19{,}538$) and chicken ($n=14{,}429$), the percentage of hihi genes that could be assigned to orthogroups in this dataset was 74% (compared to >90% in both zebra finch and chicken). There were 10,097 one-to-one orthologs identified in the chicken-to-hihi comparison and 9657 in the zebra-finch-to-hihi comparison, which may reflect some degree of bias from the chicken dataset in gene prediction.

3.5 | Native DNA methylation analysis

The raw signal generated during ONT sequencing using unamplified templates can also provide information about native DNA base modifications. To investigate this, we re-called the raw fast5 datasets using a pre-developed model to detect 5-methylcytosine (5mC) in a CpG context. Both the male and female datasets were mapped to the female reference assembly for this analysis. A total of 10.28 M CpG dinuclotides were identified in the female dataset and 10.05 M in the male dataset, and these show similar distributions of methylation proportion (Figure S6) with a slight excess of fully unmethylated sites in the female. CpG islands were also independently identified from the reference sequence using EMBOSS cpgplot. This approach identified 59,320 CpG islands in the female genome and 56,622 CpG islands in the male genome. We partitioned the genome into autosomal, Z and W subsets to investigate the methylation status in parallel to the earlier analysis of repetitive content. Figure 4a shows the percent methylation of CpG sites averaged across 50 kb non-overlapping windows for different genome partitions. At this level of analysis, the autosomes and Z show similar distributions of methylation whereas the W chromosome is markedly hyper-methylated by comparison.

Since we had ONT datasets from a male and a female individual we were able to explore whether there were DMRs between these two individuals. In general, the methylation signals are strongly correlated between male and female samples: Figure 4b shows a representative snapshot 100 kb window. DMR potentially reflect sex-specific differences in epigenetic status, though they may also reflect non-sex-specific inter-individual differences. The methylation dataset for the two individuals was combined and filtered to exclude sites not represented in both individuals and sites with low read coverage. The data from the two DNA strands was aggregated resulting in a single, unphased set of methylation percentages per CpG site. After filtering, 9,973,798 sites were retained for analysis. The dma module in NANOMETHPHASE was used to identify differentially methylated sites and cluster these sites into regions. In total, 4097 DMR were identified with the analysis settings employed, of which 3276 could be localized to autosomes and 297 to the Z chromosome/PAR scaffold of the female genome. The DMRs contained between 16–466 CpG sites per locus (mean = 47.6) and ranged from 102 to 4.8 kb in size (mean = 552 bp). However, most of the identified DMR displayed quite minor differences in methylation levels (mean absolute difference of 0.14). We elected to focus on the subset of DMR showing an absolute difference in methylation of >0.4 (Table S10). These 43 DMR were generally found to be hyper-methylated in the male relative to the female, with only six of the 43 reporting hyper-methylation in the female individual. Eighteen of the 43 DMR were found to overlap a predicted gene (or genes) within the interval 1 kb upstream or downstream of the DMR boundaries. The 21 genes that are physically close to the DMRs are listed in Table S11. Not all predicted genes have functional annotations but we note the presence of genes implicated in transcription, post-translational modification, and replication, recombination and repair within this highly DMR subset. Of particular note is gene g22840 which is located on the pseudoautosomal contig and so has both a Z-linked and W-linked allele. This gene shows homology to the *TCF4* transcription factor, a key transcriptional mediator of the *Wnt* signalling pathway (reviewed in Liu et al., 2022). Jiang et al. (2021) recently reported that *TCF4* is negatively regulated by *Spindlin1-Z* (*SPIN1Z*), with the inhibition of *TCF4* promoting male sexual differentiation. A 2.3 kb view of the g22840 DMR interval is shown in Figure 4c. The male is homozygous for a ~ 380 bp deletion at this locus, and the female is heterozygous for this deletion, consistent with a Z-chromosomal localisation of the deleted allele and W-chromosome for the non-deleted allele. Note that the deleted bases were excluded from the analyses and so the DMR signal comes only from sites that are present in both the male and female samples and are flanking the deletion. The male is hyper-methylated in this region.

4 | DISCUSSION

A high-quality genome assembly for a threatened species is a valuable tool in the 'conservation toolbox', for example enabling inference of overall genetic variation and population structure that can be used to inform conservation management decisions. We undertook this project to provide a high-quality female genome assembly for hihi, a threatened avian species native to Aotearoa New Zealand. The genome assembly of a female complements the male genome assembly that was developed in parallel, as it contains the female-specific W chromosome. By identifying the hihi W chromosome, we have a more representative genome assembly for the females in the population, which can allow genetic diversity to be captured more

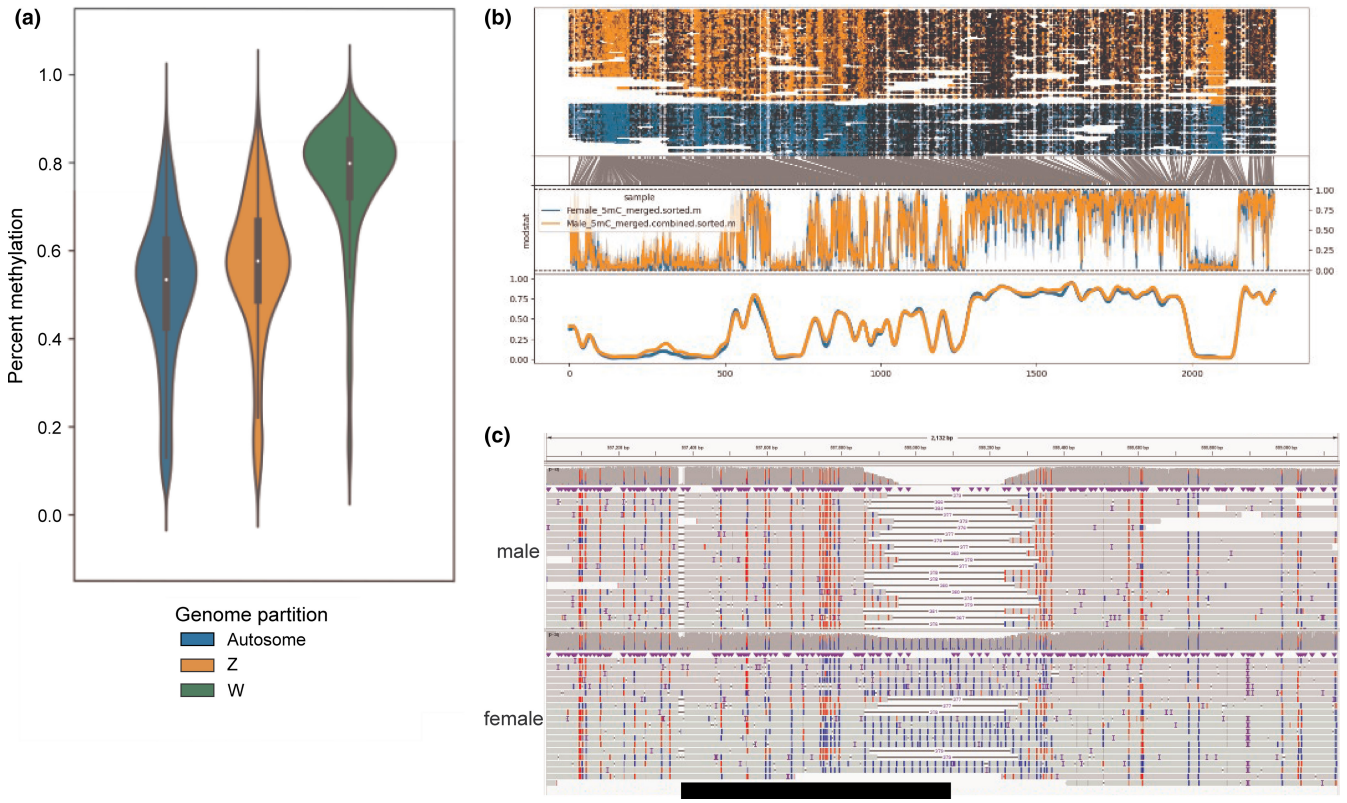

FIGURE 4 (a) Violin plot showing methylation levels for autosomal, Z and W scaffolds. (b) Representative locus plot showing local 5mC distribution. Plot generated using methylartist showing a representative 100 kb genomic segment (Ncf_contig_1251:360,000-460,000; PAR). The top panel shows read mappings and modifications across the region with the male in blue and female in orange. In the lower two panels, the reads are translated from genomic coordinates into CpG only coordinates and the raw log-likelihood ratios (centre) and smoothed methylation fraction (bottom) are displayed. (c) Integrated Genome Viewer (IGV) display of base modifications at the g22840 DMR locus. The region shown is intronic. The black bar demarcates the DMR bounds. Individual reads are mapped, with methylated CpG sites shown in red and unmethylated sites shown in blue. The region plotted is Ncf_contig_1251:557,006-559,136.

accurately in females and enable better understanding of sex differences in inbreeding and fitness. Furthermore, the two assemblies enable comparisons between the repeat content and methylomes of both sexes.

4.1 | Sequencing and assembly

We have constructed highly contiguous and complete genome assemblies from ONT reads polished with Illumina short read sequencing. We found avian samples to perform relatively poorly on the ONT flow cells, with steady decline in pore availability during the course of the experiment and typically only 10–16 h of run time before nuclease flushing of the library and reloading was required. We first assembled the male genome and were able to capitalize on learnings with this sample to obtain a comparable quality assembly of the female individual with lower coverage (the female read input to assembly was ~60% total length of the male input). The female reads had a larger fragment size, with a read N50 close to twice the length of the male dataset and this may have provided a degree of compensation for the lower input. In general, higher coverage datasets have typically been used in the construction of other high-quality avian assemblies (Friis et al., 2022; Peñalba et al., 2020; Peona, Blom, et al., 2021; Rhie et al., 2020). For hihi, it is likely that the low overall genetic diversity (de Villemereuil, Rutschmann, Lee, et al., 2019), and hence low heterozygosity, has positively contributed to the high assembly contiguity. In support of this hypothesis, although analyses were attempted, neither purge haplotigs (Roach et al., 2018) nor purge_dups (Guan et al., 2020) were able to recover both a primary and alternative assembly due to the low heterozygosity (data not shown).

The assembled female genome size of 1.06 Gb is typical for an avian genome (Zhang et al., 2014) while a contig N50 of 7.00 Mb and scaffolded N50 of 68.2 Mb are comparable to other long-read avian assemblies (Peona, Blom, et al., 2021; Rhie et al., 2020). The BUSCO completeness is also very typical for an avian assembly as in general, avian assemblies have a BUSCO completeness of over 95% (Peñalba et al., 2020). Scaffolding the assembly to a population linkage map with Chromonomer produced 31 autosomal scaffolds named for synteny with the zebra finch genome (1–15, 17–24, 26–29, 1A, 4A, 25A, and 25B). Chromosome 16 is thought to be small but highly repetitive, meaning it has been challenging to assemble in other passerines, including zebra finch (Ekblom et al., 2011). Although Chromosome 16 was missing from the linkage map, we were able to

identify part of this chromosome via homology to the major histocompatibility complex (MHC) genes. Passerine genomes are highly conserved across species, and most species have 40 chromosome pairs (Kretschmer et al., 2018), although the majority of these are microchromosomes (Waters et al., 2021). The haploid chromosome number is not known for hihi because no karyotyping has been performed, but it is likely that we are missing chromosome models of several microchromosomes. These micro-chromosomes are likely to be present in the unplaced contigs partition, which accounts for 57.4 Mb (less than 6%) of the final assembly and in the future it may be possible to link these contigs to specific micro-chromosomes.

Given the overall similarity in assembly statistics between the male and female genomes, including comparable levels of fragmentation for the same chromosome scaffolds (Tables 4 and 5), it is likely that extended repetitive regions remain the major impediment to remaining intra-chromosomal contig joins. We anticipate that an ultra-long read dataset (i.e. read N50 > 100 kb) would be required to substantially improve the assembly contiguity. The BUSCO and k-mer based estimates of completeness are both over 96% and, since functional elements (such as genes and regulatory elements) tend to be rare in repetitive regions, obtaining more a more contiguous assembly is likely to have a relatively small impact on our ability to draw biological conclusions from genomics datasets using these references (Peona, Blom, et al., 2021; Weissensteiner & Suh, 2019).

Our female assembly is more contiguous and complete than the B10K hihi assembly (Feng et al., 2020; ASM1339807v1; GenBank assembly accession: GCA_013398075.1) as reflected in a lower reported scaffold N50 for the B10K assembly (154.8 kb; compared to our female assembly of 68.23 Mb, Table 2) and a much higher number of scaffolds (38,987, compared to 288). The B10K assembly was assembled from Illumina HiSeq reads which contributes to the highly fragmented nature of the assembly. Although adults are sexually dimorphic, the sex of the hihi individual that was assembled by B10K was listed as unknown (Feng et al., 2020). Comparisons of assembly lengths show that the B10K assembly, at 1.025 Gb, is smaller than both male and female assemblies obtained here. BLASTn analysis with W-chromosome query sequences suggest that the B10K assembly does not contain W chromosome-derived contigs, suggesting that the individual was male.

TABLE 2 Scaffolded Genome assembly summary.

	Female genome	Male genome
Total assembly length (Mb)	1063.8	1045.9
Mb linked to chromosomes	1006.3	989.4
Mb unplaced	57.4	56.5
Total number of scaffolds	329	341
Number linked to chromosomes	77	52
Number unplaced	252	289
Contiguity statistics		
Scaffold N50 (Mb)	61.40	60.552
Scaffold L50	6	6
Scaffold L90	31	30
Completeness statistics		
Ns per 100 kb	1.52	1.40

4.2 | Local sequencing

Sequencing DNA in Aotearoa New Zealand was vital for our project due to our responsibility of working with a native species. As tauiwi (non-Indigenous) researchers working in Aotearoa, we are partners in Te Tiriti o Waitangi, the treaty that formalized a relationship between the British Crown and Māori. We acknowledge that all genetic research for the purposes of conservation within Aotearoa is of value and interest to Māori who are kaitiaki (guardians) for those species (Collier-Robinson et al., 2019; Hudson & Russell, 2009). Because hihi are a taonga (treasured) species, the samples and the data generated from the hihi individual used in this study are also taonga and need to be treated carefully (Caron et al., 2020; Hudson et al., 2021). Ngāti Manuhiri, the iwi who are kaitiaki for hihi, required the sequencing to be done in New Zealand to ensure the appropriate handling of the sample and genomic data, leading to selection of Oxford Nanopore Technologies MinION and Illumina HiSeq sequencing. Further, the limited quantity and quality of blood available for both the female and male individual meant that scaffolding options that could have been completed in-country, such as Hi-C, were unfeasible. However, we demonstrate the value of our genetic map, generated from SNP array data from the long-term studied population on Tiritiri Matangi, to scaffold the autosomal genome.

4.3 | Identification of sex-linked contigs

We used homology to the sex chromosomes of other passerine species and comparison of male versus female genomic depth of coverage to identify 24 W-linked and 19 candidate Z-linked contigs with a total length of 14.8 Mb (1.4% of the genome) and 75.1 Mb (7.1%), respectively. The PAR, which is found on both the Z and W chromosomes, was recovered as a single 0.88 Mb contig and retained as a separate contig in analyses. The total length of the W- and Z-linked contigs are within the ranges we expected based on the sizes of W (11–26 Mb) and Z (74–86 Mb) chromosomes in other avian species (Xu et al., 2019). Although avian genomes generally show high levels of synteny, this pattern does not appear to extend to the sex chromosomes (Zhou et al., 2014). The homology information provided good validation of the chromosome localisation of the putatively sex-linked contigs, however the relative ordering and orientation of these contigs was highly variable between the three reference genomes that we used, hence sex-linked contigs are provided as such and not further scaffolded.

The W is particularly challenging to assemble due to the high density of repetitive elements and proportionally lower coverage of this chromosome in the female sample. This is reflected in the lower

contiguity of the W chromosome compared to the Z: the N50 of the W contigs is 1.02 Mb, whereas the N50 of the Z is 8.0 Mb and close to the genome-wide contig N50 of 7.0 Mb. We also note that two W contigs have inflated coverage consistent with collapsed duplications. Overall, the female assembly is 17.9 Mb larger than the male assembly and we would expect that the bulk of any difference in assembly size would be explained by the contribution of the W chromosome to the female assembly, which potentially leaves 2.1 Mb unaccounted for. However, although we were quite conservative in assigning contigs to the W, less than 650 kb of unplaced contigs have a coverage ratio suggestive of W-linkage (and this using a permissive threshold of a female: male coverage ratio greater than 1.2:1). In avian genomes, highly repetitive and heterochromatic regions are systematically under-represented in assemblies, so some repetitive W-linked sequence could also be missing from our assembly altogether (Peona, Blom, et al., 2021; Weissensteiner & Suh, 2019). Both the male and female hihi genomes were scanned for non-avian contaminant sequences with BLOBTOOLS (Laetsch et al., 2017) and contamination was confirmed not to be present in the assembly, so it is unlikely that we have falsely identified contaminants as W-linked sequences (Smeds et al., 2015).

4.4 | Repeat identification

Repeat identification with EARL GREY, which combines and curates a custom library made up of de novo and previously annotated TE elements to identify TEs, as well as detecting satellite and unclassified repeats, revealed some variation in the repeat landscapes between the male and female genomes. This variation is to be expected due to the repeat-rich W chromosome in the female genome. The slightly elevated density of TEs and other repeats on the Z chromosome relative to autosomes are typical for avian genomes, as is the extreme enrichment of LTR sequences on the W chromosome, which has been hypothesised to be act as a refugium for endogenous retrovirus and a major source of genome-wide retrotransposition and genome instability (Peona, Palacios-Gimenez, et al., 2021). Unexpectedly, and in contrast to other avian genomes (Peona, Palacios-Gimenez, et al., 2021; Prost et al., 2019), there appears to have been a burst recent of DNA element transposon activity (driven largely by Ginger-2 DNA elements), along with recent LTR activity, which indicates recent expansions shaping the hihi genome. Possibly, the small effective population size of hihi means that selection cannot efficiently remove these insertions (Brekke et al., 2011). Taken together, this might indicate that female hihi is burdened by a toxic W chromosome which contains an abundance of active transposable elements that can disrupt genome stability and shorten the lifespan of the heterogametic sex (Brown et al., 2020; Peona, Palacios-Gimenez, et al., 2021). The observation data from the well-monitored population of Tiritiri Matangi suggests that female hihi have a slightly shorter lifespan than their male counterparts (2.9 vs. 3.3 years, Duntsch et al., 2023) and it would be interesting to explore what role that genomic factors may play in this.

4.5 | Methylation

Given the limited quantities of DNA available and the expense of other methods for assessing methylation, re-running the available ONT data with the extended basecalling model was an exciting opportunity to survey DNA methylation across the genome. There were no studies to our knowledge on avian species that used Nanopore data for studying methylation, with the majority of the studies utilizing bisulfite sequencing (Derks et al., 2016; Lindner, Laine, et al., 2021; Lindner, Verhagen, et al., 2021; Saino et al., 2017; Sun et al., 2019, 2021; Viitaniemi et al., 2019; Zhang et al., 2017), Methylated DNA immunoprecipitation (MeDIP) (Höglund et al., 2020; Taff et al., 2019) or Methylation-sensitive AFLP (MS-AFLP) (Schrey et al., 2012; Sheldon et al., 2022; Wenzel & Piertney, 2014). By using our ONT data, we found the overall methylome patterns consistent with results reported from studies in other avian species, with evidence for methylation at the vast majority of CpG sites, but limited differences in the mean level of methylation at these sites between the male and female genomes. The concordance with data from other bird species offers promise that methylation signals can be reliably detected from ONT data and used as a resource to explore further features of hihi genome architecture and phenotypic diversity. The methylation landscape also offers the opportunity to further explore the role of methylation in incomplete dosage compensation for hihi and birds more generally (for example, chromatin dynamics can explain incomplete dosage compensation in Eurasian crow, Catalán et al., 2021).

Despite the limited differences between the male and female methylome, on the autosomal, PAR and candidate Z-linked contigs there were 4097 regions that were differentially methylated at a significance threshold of $p < .05$ (Table S10). Of these, 43 regions were substantially differentially methylated between the female and male, showing an absolute difference in methylation proportion of >0.4. These DMRs most often showed hyper-methylation in the female relative to the male (37/43 DMRs). The differences between genomic depth of coverage for the male (45×) and female (25×) datasets, and between the autosomes and sex chromosomes, could be contributing to some of the differences seen between the male and female percentage of methylated sites, but GUPPY has been shown to stably predict methylation at different coverage levels (Yuen et al., 2021). Further, the candidate W-linked contigs had a higher mean level of methylation (62.8%) compared to autosomal, PAR and candidate Z-linked contigs (range 39.1%–44.8% across sexes) (Figure 4a), in agreement with previous avian studies. W chromosomes have a small effective population size and are often subjected to genetic drift causing W-linked genes to accumulate deleterious mutations (Sigeman et al., 2018). These mutations can disrupt the function of the genes and therefore methylation may be a mechanism to prevent expression of these non-functional gene sequences in females (Sigeman et al., 2018) as well as preventing the transcription of transposable elements (Laine et al., 2016; Yoder et al., 1997). In birds, only 5mC methylation has been studied but it is likely that other types of methylation that occur in vertebrates such as 5hmC also are present in the avian genome. 5hmC is not supported at this

stage in GUPPY (Liu et al., 2021) but if further basecalling models were added, the raw data used here could be re-basecalled in the future, or called using alternative software (e.g. SIGNALALIGN, Rand et al., 2017).

Dosage compensation via methylation-based repression, a common feature of mammalian genomes, does not occur in birds, which are described as having incomplete dosage compensation (Sigeman et al., 2018). The detection of relatively few DMR between the male and female samples supports this. The 43 DMR identified in this study highlight some preliminary candidates for further investigation of whether they reflect stable sex differences or are developmentally or individually variable. We are especially intrigued by the identification of a hyper-methylated DMR in the male at the *TCF4* transcription factor. Dosage of *Doublesex and Mab-3–Related Transcription factor 1* (*DMRT1*) is widely believed to be the primary sex determining factor in birds and recent targeted manipulation of this gene in chicken supports this (Ioannidis et al., 2021). Sex determination is, however, cell-autonomous and adult sex characteristics are independent of gonadal sex, with many of the molecular details still to be resolved. Jiang et al. (2021) have proposed a key role for *SPIN1-Z* in this process as an upstream regulator of *TCF4*. *SPIN1-Z* is a chromatin reader that recognizes multiple histone modification patterns (Zhao et al., 2007). Rose and Klose (2014) highlight how histone and DNA methylation can be mechanistically linked. Whether methylation at *TCF4* represents a useful biomarker or functionally relevant mark for sex differences remains to be determined, but a key role for *SPIN1-Z* in sex determination might support the observation that this locus serves as a highly reliable marker for avian sex across a range of taxa (Dawson et al., 2016).

In summary, the annotated genome we present here provides a foundational resource for future work to understand the adaptive potential and evolutionary distinctiveness of this taonga threatened species. These analyses include the analysis of gene families (Gemmell et al., 2020; Prost et al., 2019), analysis of genes under selection (Laine et al., 2016), a comparison of the gene content of chromosomes between species (Xu et al., 2019) and an analysis of the genetic diversity of genes that are likely important for hihi, such as Toll-like Receptor genes which have a role in immunity (Grueber et al., 2015). Moreover, being able to specify the impact of inbreeding depression at the gene level will inform management decisions and aid with conservation efforts (Duntsch et al., 2023; Shafer et al., 2015). Maintaining hihi populations contributes to allowing future generations to experience a world rich in biodiversity, and fulfils our cultural responsibility as Te Tiriti o Waitangi Treaty partners for preserving native species that are taonga to Māori.

AUTHOR CONTRIBUTIONS

A.W., S.B. and A.W.S. designed the research, A.W. conducted the ONT sequencing, and A.W. and S.B. led the hihi genome assembly and analyses, with support from D.S.S. J.G. performed the gene annotation. P.S. constructed the linkage map with support from A.W.S. J.G.E. supervised the data collection on Te Hauturu-o-Toi while J.G.E, P.B. and A.W.S. coordinate the long term data collection from Tiritiri Matangi and microsatellite genotyping that provided pedigree data for the linkage map. A.W. and S.B. led the writing of the paper, with input from A.W.S. and feedback from all authors. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

We acknowledge Ngāti Manuhiri as Mana Whenua and Kaitiaki of Te Hauturu-o-Toi and its taonga, including hihi. We are thankful the New Zealand eScience Infrastructure (NeSI) high-performance computing facilities. We thank Kate Lee for developing the hihi SNP array. We thank Stella Loke, Ross Crowhurst and Tobias Barril for advice on sequencing, assembly and analyses and Hui Zhen Tan for help with figures. We also thank the AgResearch Animal Genomicsteam, particularly Tracey Van Stijn, Rudiger Brauning and Shannon Clarke, for hihi Illumina sequencing, and Selina Patel for DNA extractions. We are thankful to the awesome team of Hihi Recovery Group members and friends who organized and attended hihi sampling trips to Te Hauturu-o-Toi, particularly Helen Taylor and Alex Knight along with Troy Makan, Kate Richardson, Neil Anderson, Fiona Gordon and Mhairi McCready. Funding from the George Mason Center for the Natural Environment and the Little Barrier Island (Hauturu) Supporters Trust supported the fieldwork collection. A Marsden Grant (UOA1408) awarded to A.W.S. from the New Zealand Royal Society Te Aparangi supported A.W.S., P.B. and J.G.E and funded sequencing, development and genotyping of the SNP array genotypes utilized in the linkage map. The High Quality Genomes and Population Genomics project of Genomics Aotearoa supported A.W.S. and A.W. and funded ONT and Illumina sequencing. ONT sequencing was also funded by a University of Auckland Digital Biology Interface Institute Seed Funding Grant, awarded to A.W. and A.W.S. A New Zealand National Science Challenge Biological Heritage Project Grant, Project 1.4, and a Strategic Science Investment Fund in Data Science from the Ministry of Business, Innovation and Employment also supported A.W.S. P.B and J.G.E are supported by Research England. S.B. was supported by a University of Auckland Centre for Biodiversity and Biosecurity writing stipend and a Faculty of Science Sustainability Award. P.S. was supported by a University of Auckland Summer Research Scholarship. All work was undertaken in accordance with Department of Conservation permits 53614-FAU and 66751-FAU. Open access publishing facilitated by The University of Auckland, as part of the Wiley - The University of Auckland agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Supporting figures and tables are provided in the Supplementary Material. 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. For this reason, the raw data, assemblies, SNP genotypes and associated metadata for hihi have been uploaded to the Aotearoa Genomic Data Repository, Project Identifier

TAONGA-AGDR00034 (https://doi.org/10.57748/ZD00-D451) and will be made available by request on 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.

BENEFIT SHARING

We consulted with the indigenous community, the iwi (extended kinship group) Ngāti Manuhiri, who are kaitiaki (guardians) for hihi. In the Acknowledgements, we acknowledge Ngāti Manuhiri as Mana Whenua and Kaitiaki of Te Hauturu-o-Toi and its taonga, including hihi. Regular hihi updates are provided to Ngāti Manuhiri via the Department of Conservation Hihi Recovery Group reports.

ORCID

Joseph Guhlin https://orcid.org/0000-0003-3264-7178
Patricia Brekke https://orcid.org/0000-0001-6298-3194
John G. Ewen https://orcid.org/0000-0001-6402-1378
Anna W. Santure https://orcid.org/0000-0001-8965-1042
Annabel Whibley https://orcid.org/0000-0003-1878-7705

REFERENCES

Akbari, V., Garant, J. M., O'Neill, K., Pandoh, P., Moore, R., Marra, M. A., Hirst, M., & Jones, S. J. M. (2021). Megabase-scale methylation phasing using nanopore long reads and NanoMethPhase. *Genome Biology*, *22*, 68.

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, *11*(10), 697–709. https://doi.org/10.1038/nrg2844

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., Lippman, Z. B., & Schatz, M. C. (2019). RaGOO: Fast and accurate reference-guided scaffolding of draft genomes. *Genome Biology*, *20*(1), 1–17. https://doi.org/10.1186/s13059-019-1829-6

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. https://doi.org/10.1093/nar/25.17.3389

Bachtrog, D. (2013). Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics*, *14*, 113–124. https://doi.org/10.1038/nrg3366

Baril, T., Imrie, R. M., & Hayward, A. (2022). Earl Grey: A fully automated user-friendly transposable element annotation and analysis pipeline. *bioRxiv*, 2022.06.30.498289. https://doi.org/10.1101/2022.06.30.498289

Brandies, B., Peel, E., Hogg, C. J., & Belov, K. (2019). The value of reference genomes in the conservation of threatened species. *Genes*, *10*(11), 846. https://doi.org/10.3390/genes10110846

Bravo, G. A., Schmitt, C. J., & Edwards, S. V. (2021). What have we learned from the first 500 avian genomes? *Annual Review of Ecology, Evolution, and Systematics*, *52*, 611–639. https://doi.org/10.1146/annurev-ecolsys-012121-085928

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

Brekke, P., Bennett, P. M., Wang, J., Pettorelli, N., & Ewen, J. G. (2010). Sensitive males: Inbreeding depression in an endangered bird. *Proceedings of the Royal Society B: Biological Sciences*, *277*, 3677–3684. https://doi.org/10.1098/rspb.2010.1144

Brekke, P., Dawson, D. A., Horsburgh, G. J., & Ewen, J. G. (2009). Characterization of microsatellite loci in the hihi *Notiomystis cincta* (Notiomystidae, Aves). *Molecular Ecology Resources*, *9*(4), 1255–1258. https://doi.org/10.1111/j.1755-0998.2009.02626.x

Brekke, P., Ewen, J. G., Clucas, G., & Santure, A. W. (2015). Determinants of male floating behaviour and floater reproduction in a threatened population of the hihi (*Notiomystis cincta*). *Evolutionary Applications*, *8*, 796–806. https://doi.org/10.1111/eva.12287

Brekke, P., Wang, J., Bennett, P. M., Cassey, P., Dawson, D. A., Horsburgh, G. J., & Ewen, J. G. (2012). Postcopulatory mechanisms of inbreeding avoidance in the Island endemic hihi (*Notiomystis cincta*). *Behavioral Ecology*, *23*(2), 278–284. https://doi.org/10.1093/beheco/arr183

Brown, E. J., Nguyen, A. H., & Bachtrog, D. (2020). The Y chromosome may contribute to sex-specific ageing in *Drosophila*. *Nature Ecology and Evolution*, *4*(6), 853–862. https://doi.org/10.1038/s41559-020-1179-5

Buchfink, B., Xie, C., & Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176

Bursell, M. G., Dikow, R. B., Figueiró, H. V., Dudchenko, O., Flanagan, J. P., Aiden, E. L., Goossens, B., Nathan, S. K. S. S., Johnson, W. E., Koepfli, K. P., & Frandsen, P. B. (2022). Whole genome analysis of clouded leopard species reveals an ancient divergence and distinct demographic histories. *iScience*, *25*(12), 105647. https://doi.org/10.1016/j.isci.2022.105647

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 1–9. https://doi.org/10.1186/1471-2105-10-421

Cantalapiedra, C. P., Hernandez-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution*, *38*, 5825–5829. https://doi.org/10.1093/molbev/msab293

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Caron, N. R., Chongo, M., Hudson, M., Arbour, L., Wasserman, W. W., Robertson, S., Correard, S., & Wilcox, P. (2020). Indigenous genomic databases: Pragmatic considerations and cultural contexts. *Frontiers in Public Health*, *8*, 111. https://doi.org/10.3389/fpubh.2020.00111

Catalán, A., Merondun, J., Knief, U., & Wolf, J. B. W. (2021). Epigenetic mechanisms of partial dosage compensation in an avian, female heterogametic system. *bioRxiv*, 2021.08.17.456618. https://doi.org/10.1101/2021.08.17.456618

Catchen, J., Amores, A., & Bassham, S. (2020). Chromonomer: A tool set for repairing and enhancing assembled genomes through integration of genetic maps and conserved synteny. *G3: Genes, Genomes Genetics*, *10*(11), 4115–4128. https://doi.org/10.1534/g3.120.401485

Ceballos, G., Ehrlich, P. R., & Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(30), E6089–E6096. https://doi.org/10.1073/pnas.1704949114

Charlesworth, B., & Charlesworth, D. (2000). The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *355*(1403), 1563–1572. https://doi.org/10.1098/rstb.2000.0717

Cheetham, S. W., Kindlova, M., & Ewing, A. D. (2022). Methylartist: Tools for visualizing modified bases from nanopore sequence data. *Bioinformatics*, *38*(11), 3109–3112. https://doi.org/10.1093/bioinformatics/btac292

Collier-Robinson, L., Rayne, A., Rupene, M., Thom, C., & Steeves, T. E. (2019). Embedding indigenous principles in genomic research of culturally significant species: A conservation genomics case study. *New Zealand Journal of Ecology*, *43*(3), 3389. https://doi.org/10.20417/nzjecol.43.36

Dawson, D. A., Brekke, P., Dos Remedios, N., & Horsburgh, G. J. (2015). A marker suitable for sex-typing birds from degraded samples. *Conservation Genetics Resources*, *7*(2), 337–343. https://doi.org/10.1007/s12686-015-0429-3

Dawson, D. A., Dos Remedios, N., & Horsburgh, G. J. (2016). A new marker based on the avian spindlin gene that is able to sex most birds, including species problematic to sex with CHD markers. *Zoo Biology*, *35*(6), 533–545. https://doi.org/10.1002/zoo.21326

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics*, *34*(15), 2666–2669. https://doi.org/10.1093/bioinformatics/bty149

de Villemereuil, P., Rutschmann, A., Ewen, J. G., Santure, A. W., & Brekke, P. (2019). Can threatened species adapt in a restored habitat? No expected evolutionary response in lay date for the New Zealand hihi. *Evolutionary Applications*, *12*(3), 482–497. https://doi.org/10.1111/eva.12727

de Villemereuil, P., Rutschmann, A., Lee, K. D., Ewen, J. G., Brekke, P., & Santure, A. W. (2019). Little adaptive potential in a threatened passerine bird. *Current Biology*, *29*, 889–894. https://doi.org/10.1016/j.cub.2019.01.072

Derks, M. F. L., Schachtschneider, K. M., Madsen, O., Schijlen, E., Verhoeven, K. J. F., & van Oers, K. (2016). Gene and transposable element methylation in great tit (*Parus major*) brain and blood. *BMC Genomics*, *17*(1), 332. https://doi.org/10.1186/s12864-016-2653-y

Donath, A., Jühling, F., Al-Arab, M., Bernhart, S. H., Reinhardt, F., Stadler, P. F., Middendorf, M., & Bernt, M. (2019). Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. *Nucleic Acids Research*, *47*(20), 10543–10552.

Duntsch, L., Tomotani, B. M., de Villemereuil, P., Brekke, P., Lee, K. D., Ewen, J. G., & Santure, A. W. (2020). Polygenic basis for adaptive morphological variation in a threatened Aotearoa | New Zealand bird, the hihi (*Notiomystis cincta*). *Proceedings of the Royal Society B: Biological Sciences*, *287*, 20200948. https://doi.org/10.1098/rspb.2020.0948/rspb20200948

Duntsch, L., Whibley, A., Brekke, P., Ewen, J. G., & Santure, A. W. (2021). Genomic data of different resolutions reveal consistent inbreeding estimates but contrasting homozygosity landscapes for the threatened Aotearoa New Zealand hihi. *Molecular Ecology*, *30*(23), 6006–6020. https://doi.org/10.1111/mec.16068

Duntsch, L., Whibley, A., de Villemereuil, P., Brekke, P., Bailey, S., Ewen, J. G., & Santure, A. W. (2023). Genomic signatures of inbreeding depression for a threatened Aotearoa New Zealand passerine. *Molecular Ecology*, *32*, 1893–1907. https://doi.org/10.1111/mec.16855

Dussex, N., van der Valk, T., Morales, H. E., Wheat, C. W., Díez-del-Molino, D., von Seth, J., Foster, Y., Kutschera, V. E., Guschanski, K., Rhie, A., Phillippy, A. M., Korlach, J., Howe, K., Chow, W., Pelan, S., Mendes Damas, J. D., Lewin, H. A., Hastie, A. R., Formenti, G., … Dalén, L. (2021). Population genomics of the critically endangered kākāpō. *Cell Genomics*, *1*, 100002. https://doi.org/10.1016/j.xgen.2021.100002

Dussex, N., von Seth, J., Knapp, M., Kardailsky, O., Robertson, B. C., & Dalén, L. (2019). Complete genomes of two extinct New Zealand passerines show responses to climate fluctuations but no evidence for genomic erosion prior to extinction. *Biology Letters*, *15*(9), 20190491. https://doi.org/10.1098/rsbl.2019.0491

Ekblom, R., Stapley, J., Ball, A. D., Birkhead, T., Burke, T., & Slate, J. (2011). Genetic mapping of the major histocompatibility complex in the zebra finch (*Taeniopygia guttata*). *Immunogenetics*, *63*(8), 523–530. https://doi.org/10.1007/s00251-011-0525-9

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*, 238. https://doi.org/10.1186/s13059-019-1832-y

Evans, I., & Maddox, J. (2015). *CRI-MAP*. https://www.animalgenome.org/tools/share/crimap/

Ewen, J. G., Surai, P., Stradi, R., Møller, A. P., Vittorio, B., Griffiths, R., & Armstrong, D. P. (2006). Carotenoids, colour and conservation in an endangered passerine, the hihi or stitchbird (*Notiomystis cincta*). *Animal Conservation*, *9*(2), 229–235. https://doi.org/10.1111/j.1469-1795.2006.00028.x

Feng, S., Fang, Q., Barnett, R., Li, C., Han, S., Kuhlwilm, M., Zhou, L., Pan, H., Deng, Y., Chen, G., Gamauf, A., Woog, F., Prys-Jones, R., Marques-Bonet, T., Gilbert, M. T. P., & Zhang, G. (2019). The genomic footprints of the fall and recovery of the crested ibis. *Current Biology*, *29*(2), 340–349.e7. https://doi.org/10.1016/j.cub.2018.12.008

Feng, S., Stiller, J., Deng, Y., Armstrong, J., Fang, Q., Reeve, A. H., Xie, D., Chen, G., Guo, C., Faircloth, B. C., Petersen, B., Wang, Z., Zhou, Q., Diekhans, M., Chen, W., Andreu-Sánchez, S., Margaryan, A., Howard, J. T., Parent, C., … Zhang, G. (2020). Dense sampling of bird diversity increases power of comparative genomics. *Nature*, *587*(7833), 252–257. https://doi.org/10.1038/s41586-020-2873-9

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, *17*, 9451–9457. https://doi.org/10.1073/pnas.1921046117

Forsdick, N. J., Martini, D., Brown, L., Cross, H. B., Maloney, R. F., Steeves, T. E., & Knapp, M. (2021). Genomic sequencing confirms absence of introgression despite past hybridisation between a critically endangered bird and its common congener. *Global Ecology and Conservation*, *28*, e01681. https://doi.org/10.1016/j.gecco.2021.e01681

Friis, G., Vizueta, J., Ketterson, E. D., & Milá, B. (2022). A high-quality genome assembly and annotation of the dark-eyed junco *Junco hyemalis*, a recently diversified songbird. *G3 Genes|Genomes|Genetics*, *12*(6), jkac083. https://doi.org/10.1093/g3journal/jkac083

Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L., & Paul, C. L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(5), 1827–1831. https://doi.org/10.1073/pnas.89.5.1827

Fuentes-Pardo, A. P., & Ruzzante, D. E. (2017). Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Molecular Ecology*, *26*(20), 5369–5406. https://doi.org/10.1111/mec.14264

Galla, S. J., Moraga, R., Brown, L., Cleland, S., Hoeppner, M. P., Maloney, R. F., Richardson, A., Slater, L., Santure, A. W., & Steeves, T. E. (2020). A comparison of pedigree, genetic and genomic estimates of relatedness for informing pairing decisions in two critically endangered birds: Implications for conservation breeding programmes worldwide. *Evolutionary Applications*, *13*(5), 991–1008. https://doi.org/10.1111/eva.12916

Gan, H. M., Falk, S., Morales, H. E., Austin, C. M., Sunnucks, P., & Pavlova, A. (2019). Genomic evidence of neo-sex chromosomes in the eastern yellow robin. *GigaScience*, *8*(9), 1–10. https://doi.org/10.1093/gigascience/giz111

Gemmell, N. J., Rutherford, K., Prost, S., Tollis, M., Winter, D., Macey, J. R., Adelson, D. L., Suh, A., Bertozzi, T., Grau, J. H., Organ, C., Gardner, P. P., Muffato, M., Patricio, M., Billis, K., Martin, F. J., Flicek, P., Petersen, B., Kang, L., … Edmonds, H. (2020). The tuatara genome reveals ancient features of amniote evolution. *Nature*, *584*(7821), 403–409. https://doi.org/10.1038/s41586-020-2561-9

Giani, A. M., Gallo, G. R., Gianfranceschi, L., & Formenti, G. (2020). Long walk to genomics: History and current approaches to genome sequencing and assembly. *Computational and Structural Biotechnology Journal*, *18*, 9–19. https://doi.org/10.1016/j.csbj.2019.11.002

Gibb, G. C., England, R., Hartig, G., McLenachan, P. A., Smith, B. L. T., McComish, B. J., Cooper, A., & Penny, D. (2015). New Zealand passerines help clarify the diversification of major songbird lineages during the oligocene. *Genome Biology and Evolution*, *7*(11), 2983–2995. https://doi.org/10.1093/gbe/evv196

Green, P., Falls, K., & Crooks, S. (1990). *Documentation for CRI-MAP, version 2.4*. https://www.animalgenome.org/hu/CRIMAPwkshp/crimap-doc.html

Grueber, C. E., Knafler, G. J., King, T. M., Senior, A. M., Grosser, S., Robertson, B., Weston, K. A., Brekke, P., Harris, C. L. W., & Jamieson, I. G. (2015). Toll-like receptor diversity in 10 threatened bird species: Relationship with microsatellite heterozygosity. *Conservation Genetics*, *16*(3), 595–611. https://doi.org/10.1007/s10592-014-0685-x

Guan, D., Guan, D., McCarthy, S. A., Wood, J., Howe, K., Wang, Y., Durbin, R., & Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, *36*(9), 2896–2898. https://doi.org/10.1093/bioinformatics/btaa025

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hare, K. M., Borrelle, S. B., Buckley, H. L., Collier, K. J., Constantine, R., Perrott, J. K., Watts, C. H., & Towns, D. R. (2019). Intractable: Species in New Zealand that continue to decline despite conservation efforts. *Journal of the Royal Society of new Zealand*, *49*(3), 301–319. https://doi.org/10.1080/03036758.2019.1599967

Hoff, K., & Stanke, M. (2019). Predicting genes in single genomes with AUGUSTUS. *Current Protocols in Bioinformatics*, *65*(1), e57.

Hoff, K. J., Lomsadze, A., Borodovsky, M., & Stanke, M. (2019). Whole–genome annotation with BRAKER. *Methods in Molecular Biology*, *1962*, 65–95. https://doi.org/10.1007/978-1-4939-9173-0_5

Höglund, A., Henriksen, R., Fogelholm, J., Churcher, A. M., Guerrero-Bosagna, C. M., Martinez-Barrio, A., Johnsson, M., Jensen, P., & Wright, D. (2020). The methylation landscape and its role in domestication and gene regulation in the chicken. *Nature Ecology and Evolution*, *4*(12), 1713–1724. https://doi.org/10.1038/s41559-020-01310-1

Hu, J., Fan, J., Sun, Z., & Liu, S. (2020). NextPolish: A fast and efficient genome polishing tool for long-read assembly. *Bioinformatics*, *36*(7), 2253–2255. https://doi.org/10.1093/bioinformatics/btz891

Hudson, M., & Russell, K. (2009). The treaty of Waitangi and research ethics in Aotearoa. *Bioethical Inquiry*, *6*, 61–68. https://doi.org/10.1007/s11673-008-9127-0

Hudson, M., Thompson, A., Wilcox, P., Mika, J., Battershill, C., Stott, M., Brooks, R. T., & Warbrick, L. (2021). *Te Nohonga Kaitiaki guidelines for genomic research on Taonga species*. Te Kotahi Research Institute, University of Waikato Hamilton, New Zealand.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*(Database issue), D309–D314. https://doi.org/10.1093/nar/gky1085

Ioannidis, J., Taylor, G., Zhao, D., Liu, L., Idoko-Akoh, A., Gong, D., Lovell-Badge, R., Guioli, S., McGrew, M. J., & Clinton, M. (2021). Primary sex determination in birds depends on DMRT1 dosage, but gonadal sex does not determine adult secondary sex characteristics. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(10), e2020909118. https://doi.org/10.1073/pnas.2020909118

Jiang, J., Zhang, C., Yuan, X., Li, J., Zhang, M., Shi, X., Jin, K., Zhang, Y., Zuo, Q., Chen, G., & Li, B. (2021). Spin1z induces the male pathway in the chicken by down-regulating Tcf4. *Gene*, *780*, 145521. https://doi.org/10.1016/j.gene.2021.145521

Kardos, M., Armstrong, E. E., Fitzpatrick, S. W., & Funk, W. C. (2021). The crucial role of genome-wide genetic variation in conservation. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(48), e2104642118. https://doi.org/10.1073/pnas.2104642118

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*, 772–780. https://doi.org/10.1093/molbev/mst010

Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in Ecology and Evolution*, *17*(5), 230–241. https://doi.org/10.1016/S0169-5347(02)02489-8

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, *37*(5), 540–546. https://doi.org/10.1038/s41587-019-0072-8

Kretschmer, R., Ferguson-Smith, M. A., & de Oliveira, E. H. C. (2018). Karyotype evolution in birds: From conventional staining to chromosome painting. *Genes*, *9*(4), 6–8. https://doi.org/10.3390/genes9040181

Laetsch, D. R., Blaxter, M. L., & Leggett, R. M. (2017). BlobTools: Interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. *F1000Research*, *6*, 1287. https://doi.org/10.12688/f1000research.12232.1

Laine, V. N., Gossmann, T. I., Schachtschneider, K. M., Garroway, C. J., Madsen, O., Verhoeven, K. J. F., de Jager, V., Megens, H. J., Warren, W. C., Minx, P., Crooijmans, R. P. M. A., Corcoran, P., The Great Tit HapMap Consortium, Adriaensen, F., Belda, E., Bushuev, A., Cichon, M., Charmantier, A., Dingemanse, N., … Groenen, M. A. M. (2016). Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nature Communications*, *7*, 10474. https://doi.org/10.1038/ncomms10474

Lee, K. D., Millar, C. D., Brekke, P., Whibley, A., Ewen, J. G., Hingston, M., Zhu, A., & Santure, A. W. (2021). The design and application of a 50 K SNP chip for a threatened Aotearoa New Zealand passerine, the hihi. *Molecular Ecology Resources*, *22*, 415–429. https://doi.org/10.1111/1755-0998.13480

Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., Durbin, R., Edwards, S. V., Forest, F., Gilbert, M. T. P., Goldstein, M. M., Grigoriev, I. V., Hackett, K. J., Haussler, D., Jarvis, E. D., Johnson, W. E., Patrinos, A., Richards, S., Castilla-Rubio, J. C., … Zhang, G. (2018). Earth BioGenome project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(17), 4325–4333. https://doi.org/10.1073/pnas.1720115115

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H. (2023). Protein-to-genome alignment with miniprot. *Bioinformatics*, *39*(1), btad014. https://doi.org/10.1093/bioinformatics/btad014

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lindner, M., Laine, V. N., Verhagen, I., Viitaniemi, H. M., Visser, M. E., van Oers, K., & Husby, A. (2021). Rapid changes in DNA methylation associated with the initiation of reproduction in a small songbird. *Molecular Ecology*, *30*(15), 3645–3659. https://doi.org/10.1111/mec.15803

Lindner, M., Verhagen, I., Viitaniemi, H. M., Laine, V. N., Visser, M. E., Husby, A., & van Oers, K. (2021). Temporal changes in DNA methylation and RNA expression in a small song bird: Within- and between-tissue comparisons. *BMC Genomics*, *22*, 36. https://doi.org/10.1186/s12864-020-07329-9

Liu, J., Xiao, Q., Xiao, J., Niu, C., Li, Y., Zhang, X., Zhou, A., Shu, G., & Yin, G. (2022). Wnt/β-catenin signalling: Function, biological mechanisms, and therapeutic opportunities. *Signal Transduction and Targeted Therapy*, *7*, 3. https://doi.org/10.1038/s41392-021-00762-6

Liu, Y., Rosikiewicz, W., Pan, Z., Jillette, N., Wang, P., Taghbalout, A., Foox, J., Mason, C., Carroll, M., Cheng, A., & Li, S. (2021). DNA methylation-calling tools for Oxford Nanopore sequencing: A survey and human epigenome-wide evaluation. *Genome Biology*, *22*, 295. https://doi.org/10.1186/s13059-021-02510-z

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution*, *38*(10), 4647–4654. https://doi.org/10.1093/molbev/msab199

Martin, M. (2011). Cutadapt removes adapter sequences from high–throughput sequencing read. *EMBnet Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

Mathur, S., & DeWood, J. A. (2021). Genetic load has potential in large populations but is realized in small inbred populations. *Evolutionary Applications*, *14*(6), 1540–1557. https://doi.org/10.1111/eva.13216

Morin, P. A., Archer, F. I., Avila, C. D., Balacco, J. R., Bukhman, Y. V., Chow, W., Fedrigo, O., Formenti, G., Fronczek, J. A., Fungtammasan, A., Gulland, F. M. D., Haase, B., Peter Heide-Jorgensen, M., Houck, M. L., Howe, K., Misuraca, A. C., Mountcastle, J., Musser, W., Paez, S., … Jarvis, E. D. (2020). Reference genome and demographic history of the most endangered marine mammal, the vaquita. *Molecular Ecology Resources*, *21*(4), 1008–1020. https://doi.org/10.1111/1755-0998.13284

NCBI Resource Coordinators. (2014). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *42*, 7–17. https://doi.org/10.1093/nar/gkt1146

Ou, S., & Jiang, N. (2019). LTR_FINDER_parallel: Parallelization of LTR_FINDER enabling rapid identification of long terminal repeat retrotransposons. *Mobile DNA*, *10*, 48. https://doi.org/10.1186/s13100-019-0193-0

Oxford Nanopore Technologies. (2018). *Medaka*. https://github.com/nanoporetech/medaka

Palmer, D. H., Rogers, T. F., Dean, R., & Wright, A. E. (2019). How to identify sex chromosomes and their turnover. *Molecular Ecology*, *28*(21), 4709–4724. https://doi.org/10.1111/mec.15245

Palmer, J. M., & Stajich, J. (2020). Funannotate v1.8.1: Eukaryotic genome annotation (v1.8.1). Zenodo https://doi.org/10.5281/zenodo.4054262

Park, Y., & Wu, H. (2016). Differential methylation analysis for BS-seq data under general experimental design. *Bioinformatics*, *32*(10), 1446–1453. https://doi.org/10.1093/bioinformatics/btw026

Pedersen, B. S., & Quinlan, A. R. (2018). Mosdepth: Quick coverage calculation for genomes and exomes. *Bioinformatics*, *34*(5), 867–868. https://doi.org/10.1093/bioinformatics/btx699

Peñalba, J. V., Deng, Y., Fang, Q., Joseph, L., Moritz, C., & Cockburn, A. (2020). Genome of an iconic Australian bird: High-quality assembly and linkage map of the superb fairy-wren (*Malurus cyaneus*). *Molecular Ecology Resources*, *20*(2), 560–578. https://doi.org/10.1111/1755-0998.13124

Peona, V., Blom, M. P. K., Xu, L., Burri, R., Sullivan, S., Bunikis, I., Liachko, I., Haryoko, T., Jønsson, K. A., Zhou, Q., Irestedt, M., & Suh, A. (2021). Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. *Molecular Ecology Resources*, *21*, 263–286. https://doi.org/10.1111/1755-0998.13252

Peona, V., Palacios-Gimenez, O. M., Blommaert, J., Liu, J., Haryoko, T., Jønsson, K. A., Irestedt, M., Zhou, Q., Jern, P., & Suh, A. (2021). The avian W chromosome is a refugium for endogenous retroviruses with likely effects on female-biased mutational load and genetic incompatibilities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*, 20200186. https://doi.org/10.1098/rstb.2020.0186

Prost, S., Armstrong, E. E., Nylander, J., Thomas, G. W. C., Suh, A., Petersen, B., Dalen, L., Benz, B. W., Blom, M. P. K., Palkopoulou, E., Ericson, P. G. P., & Irestedt, M. (2019). Comparative analyses identify genomic features potentially involved in the evolution of birds-of-paradise. *GigaScience*, *8*(5), 1–12. https://doi.org/10.1093/gigascience/giz003

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Rand, A. C., Jain, M., Eizenga, J. M., Musselman-Brown, A., Olsen, H. E., Akeson, M., & Paten, B. (2017). Mapping DNA methylation with high-throughput nanopore sequencing. *Nature Methods*, *14*(4), 411–413. https://doi.org/10.1038/nmeth.4189

Rhie, A., McCarthy, S. A., Fedrigo, O., et al. (2020). Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, *592*, 737–768. https://doi.org/10.1101/2020.05.22.110833

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics*, *16*(6), 276–277.

Roach, M. J., Schmidt, S., & Borneman, A. R. (2018). Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, *19*(460), 1–10. https://doi.org/10.1186/s12859-018-2485-7

Robledo-Ruiz, D. A., Gan, H. M., Kaur, P., Dudchenko, O., Weisz, D., Khan, R., Lieberman Aiden, E., Osipova, E., Hiller, M., Morales, H. E., Magrath, M. J. L., Clarke, R. H., Sunnucks, P., & Pavlova, A. (2022). Chromosome-length genome assembly and linkage map of a critically endangered Australian bird: The helmeted honeyeater. *GigaScience*, *11*, giac025. https://doi.org/10.1093/gigascience/giac025

Rose, N. R., & Klose, R. J. (2014). Understanding the relationship between DNA methylation and histone lysine methylation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1839*(12), 1362–1372. https://doi.org/10.1016/j.bbagrm.2014.02.007

Rutschmann, A., de Villemereuil, P., Brekke, P., Ewen, J. G., Anderson, N., & Santure, A. W. (2020). Consequences of space sharing on individual phenotypes in the New Zealand hihi. *Evolutionary Ecology*, *34*(5), 821–839. https://doi.org/10.1007/s10682-020-10063-z

Rutschmann, A., Santure, A. W., Brekke, P., Ewen, J. G., Shanahan, D., & de Villemereuil, P. (2022). Variation in shape and consistency of selection between populations of the threatened Hihi (*Notiomystis cincta*). *Journal of Evolutionary Biology*, *35*, 1378–1386. https://doi.org/10.1111/jeb.14088

Saino, N., Ambrosini, R., Albetti, B., Caprioli, M., De Giorgio, B., Gatti, E., Liechti, F., Parolini, M., Romano, A., Romano, M., Scandolara, C., Gianfranceschi, L., Bollati, V., & Rubolini, D. (2017). Migration phenology and breeding success are predicted by methylation of a photoperiodic gene in the barn swallow. *Scientific Reports*, *7*, 45412. https://doi.org/10.1038/srep45412

Scherer, P. (2017). *A high-density genetic linkage map provides insights into the sex-specific recombination landscape of the hihi* (Notiomystis cincta). [Honours thesis]. University of Auckland.

Schrey, A. W., Coon, C. A. C., Grispo, M. T., Awad, M., Imboma, T., McCoy, E. D., Mushinsky, H. R., Richards, C. L., & Martin, L. B. (2012). Epigenetic variation may compensate for decreased genetic variation with introductions: A case study using house sparrows (*Passer domesticus*) on two continents. *Genetics Research International*, *2012*, 979751. https://doi.org/10.1155/2012/979751

Shafer, A. B. A., Wolf, J. B. W., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., Colling, G., Dalén, L., de Meester, L., Ekblom, R., Fawcett, K. D., Fior, S., Hajibabaei, M., Hill, J. A., Hoezel, A. R., Höglund, J., Jensen, E. L., Krause, J., Kristensen, T. N., … Zieliński, P. (2015). Genomics and the challenging translation into conservation practice. *Trends in Ecology and Evolution*, *30*(2), 78–87. https://doi.org/10.1016/j.tree.2014.11.009

Sheldon, E. L., Ton, R., Boner, W., Monaghan, P., Raveh, S., Schrey, A. W., & Griffith, S. C. (2022). Associations between DNA methylation and telomere length during early life: Insight from wild zebra finches (*Taeniopygia guttata*). *Molecular Ecology*, *31*(23), 6261–6272. https://doi.org/10.1111/mec.16187

Sigeman, H., Ponnikas, S., Videvall, E., Zhang, H., Chauhan, P., Naurin, S., & Hansson, B. (2018). Insights into avian incomplete dosage compensation: Sex-biased gene expression coevolves with sex chromosome degeneration in the common whitethroat. *Genes*, *9*(8), 373. https://doi.org/10.3390/genes9080373

Smeds, L., Warmuth, V., Bolivar, P., Uebbing, S., Burri, R., Suh, A., Nater, A., Bureš, S., Garamszegi, L. Z., Hogner, S., Moreno, J., Qvarnström, A., Ruzic, M., Sæther, S. A., Sætre, G. P., Török, J., & Ellegren, H. (2015). Evolutionary analysis of the female-specific avian W chromosome. *Nature Communications*, *6*, 7330. https://doi.org/10.1038/ncomms8330

Smit, A. F. A., Hubley, R. R., & Green, P. R. (2013). *RepeatMasker Open-4.0*. http://RepeatmaskerOrg

Soulé, M. E. (2013). What is conservation biology? *The Future of Nature: Documents of Global Change*, *35*(11), 391–404. https://doi.org/10.11647/obp.0177.01

Stanke, M., Schöffmann, O., Morgenstern, B., & Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics*, *7*(1), 62.

Storer, J., Hubley, R., Rosen, J., Wheeler, T. J., & Smit, A. F. (2021). The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mobile DNA*, *12*, 2. https://doi.org/10.1186/s13100-020-00230-y

Stuart, K. C., Edwards, R. J., Cheng, Y., Warren, W. C., Burt, D. W., Sherwin, W. B., Hofmeister, N. R., Werner, S. J., Ball, G. F., Bateson, M., Brandley, M. C., Buchanan, K. L., Cassey, P., Clayton, D. F., de Meyer, T., Meddle, S. L., & Rollins, L. A. (2022). Transcript- and annotation-guided genome assembly of the European starling. *Molecular Ecology Resources*, *22*(8), 3141–3160. https://doi.org/10.1111/1755-0998.13679

Sun, D., Layman, T. S., Jeong, H., Chatterjee, P., Grogan, K., Merritt, J. R., Maney, D. L., & Yi, S. V. (2021). Genome-wide variation in DNA methylation linked to developmental stage and chromosomal suppression of recombination in white-throated sparrows. *Molecular Ecology*, *30*(14), 3453–3467. https://doi.org/10.1111/mec.15793

Sun, D., Maney, D. L., Layman, T. S., Chatterjee, P., & Yi, S. V. (2019). Regional epigenetic differentiation of the Z chromosome between sexes in a female heterogametic system. *Genome Research*, *29*(10), 1673–1684. https://doi.org/10.1101/gr.248641.119

Supple, M. A., & Shapiro, B. (2018). Conservation of biodiversity in the genomics era. *Genome Biology*, *19*, 131. https://doi.org/10.1186/s13059-018-1520-3

Taff, C. C., Campagna, L., & Vitousek, M. N. (2019). Genome-wide variation in DNA methylation is associated with stress resilience and plumage brightness in a wild bird. *Molecular Ecology*, *28*(16), 3722–3737. https://doi.org/10.1111/mec.15186

The UniProt Consortium. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(Database issue), D480–D489. https://doi.org/10.1093/nar/gkaa1100

Thorogood, R., Armstrong, D. P., Low, M., Brekke, P., & Ewen, J. G. (2013). The value of long-term ecological research: Integrating knowledge for conservation of hihi on Tiritiri Matangi Island. *New Zealand Journal of Ecology*, *37*(3), 298–306.

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative genomics viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*(2), 178–192. https://doi.org/10.1093/bib/bbs017

Tomaszkiewicz, M., Medvedev, P., & Makova, K. D. (2017). Y and W chromosome assemblies: Approaches and discoveries. *Trends in Genetics*, *33*(4), 266–282. https://doi.org/10.1016/j.tig.2017.01.008

Toy, R., Greene, T. C., Greene, B. S., Warren, A., & Griffiths, R. (2018). Changes in density of hihi (*Notiomystis cincta*), tīeke (*Philesturnus rufusater*) and tūī (*Prosthemadera novaeseelandiae*) on little barrier Island (Te Hauturu-o-Toi), Hauraki gulf, Auckland, 2005–2013. *New Zealand Journal of Ecology*, *42*(2), 149–157. https://doi.org/10.20417/nzjecol.42.24

Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, *27*(5), 737–746. https://doi.org/10.1101/gr.214270.116

Viitaniemi, H. M., Verhagen, I., Visser, M. E., Honkela, A., van Oers, K., Husby, A., & Meyer, M. (2019). Seasonal variation in genome-wide DNA methylation patterns and the onset of seasonal timing of reproduction in great tits. *Genome Biology and Evolution*, *11*(3), 970–983. https://doi.org/10.1093/gbe/evz044

von Seth, J., Dussex, N., Díez-del-Molino, D., van der Valk, T., Kutschera, V. E., Kierczak, M., Steiner, C. C., Liu, S., Gilbert, M. T. P., Sinding, M. H. S., Prost, S., Guschanski, K., Nathan, S. K. S. S., Brace, S., Chan, Y. L., Wheat, C. W., Skoglund, P., Ryder, O. A., Goossens, B., … Dalén, L. (2021). Genomic insights into the conservation status of the world's last remaining Sumatran rhinoceros populations. *Nature Communications*, *12*, 2393. https://doi.org/10.1038/s41467-021-22386-8

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics*, *33*(14), 2202–2204. https://doi.org/10.1093/bioinformatics/btx153

Waters, P. D., Patel, H. R., Ruiz-Herrera, A., Alvarez-Gonzalez, L., Lister, N. C., Simakov, O., Ezaz, T., Kaur, P., Frere, C., Grutzner, F., Georges, A., & Marshall Graves, J. A. (2021). Microchromosomes are building blocks of bird, reptile, and mammal chromosomes. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(45), e2112494118. https://doi.org/10.1073/pnas.2112494118

Weissensteiner, M. H., & Suh, A. (2019). Repetitive DNA: The dark matter of avian genomics. In R. Kraus, (Ed.), *Avian genomics in ecology and evolution: From the lab into the wild* (pp. 93–150). Springer International Publishing.

Wenzel, M. A., & Piertney, S. B. (2014). Fine-scale population epigenetic structure in relation to gastrointestinal parasite load in red grouse (*Lagopus lagopus scotica*). *Molecular Ecology*, *23*(17), 4256–4273. https://doi.org/10.1111/mec.12833

Whibley, A., Kelley, J. L., & Narum, S. R. (2021). The changing face of genome assemblies: Guidance on achieving high-quality reference genomes. *Molecular Ecology Resources*, *21*(3), 641–652. https://doi.org/10.1111/1755-0998.13312

Wick, R., Volkening, J., & Loman, N. (2017). *Porechop*. Github. https://github.com/rrwick/Porechop.

Wong, W. Y., & Simakov, O. (2018). RepeatCraft: A meta-pipeline for repetitive element de-fragmentation and annotation. *Bioinformatics*, *35*, 1051–1052. https://doi.org/10.1093/bioinformatics/bty745

Wright, B. R., Hogg, C. J., McLennan, E. A., Katherine Belov, K., & Grueber, C. E. (2021). Assessing evolutionary processes over time in a conservation breeding program: A combined approach using molecular data, simulations and pedigree analysis. *Biodiversity and Conservation*, *30*, 1011–1029. https://doi.org/10.1007/s10531-021-02128-4

Xu, L., Auer, G., Peona, V., Suh, A., Deng, Y., Feng, S., Zhang, G., Blom, M. P. K., Christidis, L., Prost, S., Irestedt, M., & Zhou, Q. (2019). Dynamic evolutionary history and gene content of sex chromosomes across

diverse songbirds. *Nature Ecology and Evolution*, *3*(5), 834–844. https://doi.org/10.1038/s41559-019-0850-1

Yan, C., Song, M.-H., Jiang, D., Ren, J.-L., Lv, Y., Chang, J., Huang, S., Zaher, H., & Li, J.-T. (2023). Genomic evidence reveals intraspecific divergence of the hot-spring snake (*Thermophis baileyi*), an endangered reptile endemic to the Qinghai-Tibet plateau. *Molecular Ecology*, *32*, 1335–1350. https://doi.org/10.1111/mec.16687

Yoder, J. A., Walsh, C. P., & Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics*, *13*(8), 335–340. https://doi.org/10.1016/S0168-9525(97)01181-5

Yong, W., Hsu, F., & Chen, P. (2016). Profiling genome-wide DNA methylation. *Epigenetics & Chromatin*, *9*, 26. https://doi.org/10.1186/s13072-016-0075-3

Yuen, Z. W. S., Srivastava, A., Daniel, R., McNevin, D., Jack, C., & Eyras, E. (2021). Systematic benchmarking of tools for CpG methylation detection from nanopore sequencing. *Nature Communications*, *12*, 3438. https://doi.org/10.1038/s41467-021-23778-6

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., Storz, J. F., Antunes, A., Greenwold, M. J., Meredith, R. W., Ödeen, A., Cui, J., Zhou, Q., Xu, L., Pan, H., Wang, Z., Jin, L., Zhang, P., Hu, H., … Froman, D. P. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, *346*(6215), 1311–1321. https://doi.org/10.1126/science.1251385

Zhang, M., Yan, F. B., Li, F., Jiang, K. R., Li, D. H., Han, R. L., Li, Z. J., Jiang, R. R., Liu, X. J., Kang, X. T., & Sun, G. R. (2017). Genome-wide DNA methylation profiles reveal novel candidate genes associated with meat quality at different age stages in hens. *Scientific Reports*, *7*, 45564. https://doi.org/10.1038/srep45564

Zhao, Q., Qin, L., Jiang, F., Wu, B., Yue, W., Xu, F., Rong, Z., Yuan, H., Xie, X., Gao, Y., Bai, C., Bartlam, M., Pei, X., & Rao, Z. (2007). Structure of human spindlin1. Tandem tudor-like domains for cell cycle regulation. *Journal of Biological Chemistry*, *282*, 647–656.

Zhou, Q., Zhang, J., Bachtrog, D., An, N., Huang, Q., Jarvis, E. D., Gilbert, M. T. P., & Zhang, G. (2014). Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science*, *346*(6215), 1246338. https://doi.org/10.1126/science.1246338

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: 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*, *00*, 1–21. https://doi.org/10.1111/1755-0998.13823