Assessment of the impacts of anthropogenic hybridisation in a threatened non-model bird species through the development of genomic resources with implications for conservation

---

A thesis submitted in fulfilment
of the requirements for the degree of
Doctor of Philosophy
at the University of Otago,
New Zealand

Natalie Jane Forsdick
April 2020

Abstract

Interspecific hybridisation—the breeding between distinct species—can contribute to species extinction due to wasted reproductive potential, outbreeding depression, and introgression of genetic material mediated by backcrossing. Incomplete reproductive barriers can facilitate interspecific hybridisation as previously isolated species come into contact with one another. Interspecific hybridisation is relatively common among birds, but anthropogenic impacts that increase the incidence of such hybridisation between threatened native species and non-threatened species are of conservation concern due to the risks of genetic swamping, which at its most extreme may result in species extinction. While the impacts of interspecific hybridisation have previously been assessed using small numbers of genetic markers, new genomic sequencing developments now facilitate implementation of genome-wide reassessments providing greater resolution of analyses.

The critically endangered kakī (black stilt; *Himantopus novaezelandiae*) is one such species that can benefit from these new genomic data. Anthropogenic habitat change and introduction of mammalian predators resulted in the decline of this Aotearoa New Zealand endemic wading bird during the 1900s. An intense population bottleneck resulting in an ephemeral sex-bias among the remaining kakī contributed to hybridisation with the self-introduced poaka (the Aotearoa New Zealand population of the Australian pied stilt; *H. himantopus leucocephalus*), a congeneric species previously thought to have diverged from a common ancestor with kakī one million years ago. Intensive conservation management including captive breeding for translocation and predator control has increased kakī numbers from ~23 adults in 1981 to approximately 169 wild adults in 2020. Previous genetic studies identified minimal evidence of introgression of poaka genetic material into kakī, and

determined that moderate outbreeding depression in combination with stochastic processes likely limited introgression. These data informed the kakī captive breeding for translocation programme with the aim of maintaining genetic integrity. However, re-evaluation using genomic data was recommended for kakī.

Using high-throughput sequencing techniques, I sequenced and assembled the first reference genomes for kakī and Australian pied stilts as tools for use in analyses of introgression. The kakī mitochondrial genome was also assembled to facilitate comparisons of contemporary and historic stilt diversity, showing that conservation management aimed at maximising genetic diversity has largely maintained mitochondrial diversity despite kakī decline, identifying three mitochondrial haplotypes present among contemporary kakī. Kakī and poaka are well-differentiated, and are estimated to have diverged from a common ancestor approximately 750,000 years ago based on Bayesian analysis of mitochondrial data. In addition, the analysis of high-resolution genomic markers generated from approximately 65% of contemporary wild kakī detected no introgression from poaka to kakī despite past hybridisation. These findings confirm the results of previous genetic analysis of introgression and the success of past conservation management. As kakī recovery continues, these combined findings will be used by the New Zealand Department of Conservation's Kakī Recovery Programme to further maintain the genetic integrity of kakī. Overall, the genomic resources developed here have facilitated the transition from using genetic data to genomic data for kakī recovery, and contribute to our understanding of the impacts of anthropogenic hybridisation on a critically endangered taonga species.

Acknowledgements

This work has been a collaborative effort and I have many people to thank. First, I extend my thanks to the tangata whenua associated with kakī, namely Te Rūnanga o Ngāi Tahu, Te Ngāi Tūāhuriri Rūnanga, Te Rūnanga o Arowhenua, Te Rūnanga o Waihao, and Te Rūnanga o Moeraki.

This work would not have been possible without the ongoing efforts of the Department of Conservation's Kakī Recovery Team in collaboration with the Isaac Conservation and Wildlife Trust. Particular thanks go to Richard Maloney, Liz Brown, Simone Cleland, Cody Thyne, and Claudia Mischler for answering my many questions, sharing photos, and for their enthusiasm in supporting kakī recovery.

Special thanks go to my supervisors, Michael Knapp and Tammy Steeves. I am incredibly grateful to have had this opportunity and your ongoing support. I've been challenged, but also championed. In addition to producing research that contributes to the wider New Zealand conservation space, I've had the chance to be involved with a range of new opportunities, develop a long list of unanticipated skills, and contribute to a smart and kind research space, for which I am incredibly grateful.

I am grateful to my wider collaborative network, including Rod Hitchmough, Kevin Parker, Aisling Rayne, Anna Santure, Jana Wold, Stephanie Borelle, and Rebecca Jarvis. The bioinformatics duo of Roger Moraga and Marc Hoeppner were instrumental while I was finding my feet in the conservation genomics space.

Thank you to all those who provided samples for this project, including the Kakī Recovery Team, Richard Jakob-Hoff at Auckland Zoo, David McLelland at Adelaide Zoo, John Berry, Matt Rayner at Auckland Museum, Alan Tennyson and Colin Miskelly at Te Papa Tongarewa, and Paul Scofield at Canterbury Museum. I am also grateful to the University of Otago for funding my PhD, the Royal Society of New

Zealand for funding the bigger picture, and Birds New Zealand for additional research funding and for welcoming and encouraging my enthusiasm for birds.
For lab and bioinformatics assistance and advice, and all-round support and mentorship, there is a list of people that I cannot thank enough, including Denise Martini, Edana Lord, Anna Gosling, Alana Alexander, Hugh Cross, Olga Kardailsky, Eddy Dowle, George Pickerell, Sophie Fern, Helen Taylor, Graham Wallis, and Lisa Matisoo-Smith. Long chats, troubleshooting advice, and intense discussions have been hugely encouraging and motivating.
In more creative pursuits, I extend warm thanks and general awe to the creative talents of Brigitte Kammlein and Michelle Wilkinson, and special thanks to Madison Kelly, the artists with whom I collaborated on the Art + Genetics project. These collaborations were inspiring and helped alter my perspective regarding this research, and science communication more broadly.
LFB Room 333, 'lunch club', the University of Otago Department of Anatomy Emerging Researchers Group, and past and present members of the Matisoo-Smith/Knapp and ConSERTeam lab groups all merit thanks, especially Clare Adams, Susie Szakats, Rachel van Heugten, Ayla van Loenen, Hanna Rosinger, and Florence Glénnoec. To my Good Eggs, Giselle Clarkson, Edin Whitehead, and Jacqui Tizard, thank you so much for helping me develop into a fully-fledged bird nerd! Special thanks go to the inspirational Stephanie Galla, whose talent is only exceeded by her generosity, and to Ilina Cubrinovska, whose heavy application of logic and laughter have ensured the completion of this thesis.
Finally, I thank my family for their ongoing support through the madness that has been this thesis journey. One day we'll take a non-emergency visit to Dunedin and I'll show you all the places I've come to love.

Table of Contents

List of figures

List of tables

Abbreviations and terms

°C - degrees Celsius
µL - microlitre
µM - micromolar
µ - substitution rate
3’ - three prime
5’ - five prime
aDNA - ancient DNA
AIC - Akaike Information Criterion
Aotearoa - New Zealand (te reo Māori)
BAM - Binary Alignment Map
BIC - Bayesian Information Criterion
bp - base pairs
BP - before present
BWA - Burrows-Wheeler Aligner
BWT - Bind and Wash Buffer with Tween 20
cm - centimetres
contig - a contiguous DNA sequence produced through the assembly of short DNA fragments generated by genomic sequencing
DB - database
DNA - deoxyribonucleic acid
DOC - Department of Conservation
EDTA - Ethylenediaminetetraacetic acid
EtOH - ethanol
g - generation time
g - grams
GATK - Genome Analysis Toolkit
Gb - gigabase (one billion base pairs)
GB - gigabyte
GBS – genotyping-by-sequencing
gDNA - genomic DNA
GPU - graphics processing unit
GTR - Generalised Time Reversible
hapū - subtribal group (te reo Māori)
HPC - high performance computing
HPD - height posterior density
hr - hours
HTS - high-throughput sequencing
HWE - Hardy-Weinberg equilibrium
HWT - Hot Wash buffer with Tween 20
indel - insertion or deletion of base pairs
iwi - tribal group (te reo Māori)
K - thousand
kaitiaki - guardian (te reo Māori)
kakī - black stilt (te reo Māori)
kb - kilobase (one thousand base pairs)
kya - thousand years ago
LD - linkage disequilibrium
Māori - the indigenous people of Aotearoa New Zealand
Mb - megabase (one million base pairs)
mg - milligrams
$MgCl_2$ - magnesium chloride
min - minutes
mL - millilitres
mM - millimolar
mtDNA - mitochondrial DNA

Mya - million years ago

N50 - weighted median statistic such that 50% of the entire genome assembly is contained in contigs or scaffolds equal to or larger than this value

NaCl - sodium chloride

NCBI - National Centre for Biotechnology Information

NeSI - New Zealand eScience Infrastructure

ng - nanograms

nM - nanomolar

NZ - New Zealand

NZD - New Zealand dollars

PBS - phosphate-buffered saline

PCR - polymerase chain reaction

PE - paired-end

pM - picomolar

poaka - pied stilts in New Zealand (te reo Māori)

PSMC - pairwise sequential Markovian coalescent

qPCR - quantitative PCR

RAM - random access memory

RNA - ribonucleic acid

rpm - revolutions per minute

scaffold - a DNA sequence constructed from the assembly of contigs with known lengths and orientations

SD - standard deviation

SE - single-end

SNP – single-nucleotide polymorphism

s - seconds

taonga – treasured (te reo Māori)

tapu – sacred (te reo Māori)

TB - terabyte

TE - Tris-EDTA

Te Manahuna - the Mackenzie Basin (te reo Māori)

te reo Māori - the Māori language

TMRCA - time to most recent common ancestor

tRNA - transfer RNA

U - unit

VCF - Variant Call File

v - version

V - volts

W - watts

WGS - whole-genome sequencing

ya - years ago (years before present)

Preface

This thesis is composed of a series of inter-related yet stand-alone manuscripts, which are published or intended to be published, with overall Introduction and Discussion chapters. Thus there will be some unavoidable repetition, particularly among introductory sections. Here I describe the publication plans and statements of contribution for each research chapter.

- Chapter Two published as: Galla SJ, Forsdick NJ, Brown L, Hoeppner M, Knapp M, Maloney RF, Moraga R, Santure AW, Steeves TE. (2019). Reference genomes from distantly related species can be used for discovery of single-nucleotide polymorphisms to inform conservation management. *Genes* 10:1, doi:10.3390/genes10010009. This peer-reviewed manuscript is reproduced in full in Appendix A.

  All authors contributed to research conceptualisation. Genomes presented therein were prepared and assembled by NJF, RM, and MH. GBS and resequencing data was prepared and mapped by SJG with guidance from RM. Marker discovery and subsequent diversity analyses were performed by SJG. Original draft preparation and writing completed by SJG and NJF, with review and editing by all authors. The work presented in this chapter is my own, with the exception of whole-genome sequencing of prepared stilt libraries conducted at New Zealand Genomics Limited, Dunedin, New Zealand (NZ). Preliminary kakī genome annotation was conducted by MH (see Appendix B). The research was funded by the Brian Mason Scientific & Technical Trust (SJG, TES), the Mohua Charitable Trust (TES), New Zealand Ministry of Business, Innovation and Employment Endeavour Fund (TES, AWS), with specific funding for the research described in this chapter from the

Royal Society of New Zealand Rutherford Discovery Fellowship (MK). NJF was supported by a University of Otago Doctoral Scholarship.

- Chapter Three: Forsdick NJ, Brown L, Maloney RF, Steeves TE, Knapp M. Implementation of mitochondrial genomes for elucidating the phylogenetic relationship between congeneric stilts with relevance to conservation of kakī (*Himantopus novaezelandiae*). Proposed journal: Conservation Genetics or Animal Conservation.

  All authors contributed to research conceptualisation. Historic stilt samples were collected from Auckland War Memorial Museum by Matt Rayner (Auckland War Memorial Museum, NZ), Canterbury Museum by Denise Martini (University of Otago, NZ), and Te Papa Tongarewa Museum of New Zealand by NJF. Morphological identification of historic samples was confirmed by LB. Modern and historic samples were prepared for mitochondrial sequencing by NJF and sequenced at the Otago Genomics & Bioinformatics Facility, Dunedin, NZ. Whole-genome resequencing data was prepared by Stephanie Galla (University of Canterbury, NZ) as part of an aligned project, and sequenced at the Institute of Clinical Molecular Biology (IKMB), Kiel University, Germany. Additional samples were prepared and sequenced with Oxford Nanopore Technologies' MinION by NJF. Mitochondrial genome and MinION long-read assembly and mapping, and associated diversity and phylogenetic analyses were performed by NJF. BEAST phylogenetic analyses were implemented with assistance from Alana Alexander. Original draft preparation will be completed by NJF, with review and editing by all authors. The research was funded by a Royal Society of New Zealand Rutherford Discovery Fellowship (MK) and a Birds New Zealand

Research Grant (NJF). NJF was supported by a University of Otago Doctoral Scholarship.

- Chapter Four: Forsdick NJ, Maloney RF, Brown L, Steeves TE, Knapp M. Genomic sequencing confirms minimal introgression despite past hybridisation between kakī (*Himantopus novaezelandiae*) and poaka (*H. himantopus leucocephalus*). Proposed journal: Evolutionary Applications or Molecular Ecology.

  All authors contributed to research conceptualisation. Samples were prepared by NJF, and libraries for GBS were prepared, optimised, and sequenced at AgResearch Limited, Mosgiel, NZ. GBS mapping, marker discovery, and all subsequent analyses were performed by NJF. Original draft preparation will be completed by NJF, with review and editing by all authors. The research was funded by a Royal Society of New Zealand Rutherford Discovery Fellowship (MK), and NJF was supported by a University of Otago Doctoral Scholarship.

Chapter 1: Introduction

1.1 Objectives

The overall aim of this thesis is to leverage the power and resolution of new genomic resources to assess the impacts of hybridisation on a critically endangered wading bird in Aotearoa New Zealand. In this introductory chapter, I review the threats posed by interspecific hybridisation to global biodiversity, discuss whole-genome sequencing for assembly of reference genomes, population-level genotyping, and targeted mitochondrial sequencing and the use of these techniques for conservation management, and describe the demographic history and conservation genetic management of the focal species of this thesis, the critically endangered kakī (*Himantopus novaezelandiae*).

In Chapter Two I describe the process of sequencing and assembly of *de novo* reference genomes of kakī and Australian pied stilt (*Himantopus himantopus leucocephalus*)[1], which will be used in downstream analyses as references for marker discovery. Among the challenges for kakī conservation is past hybridisation with poaka, the New Zealand population of pied stilts. The development of a reference genome for kakī and its congener will facilitate the transition to a genomics approach for conservation of kakī (Appendix A, Chapter Four). These two genomes will be used to demonstrate the utility of more distantly-related genomes for

1 There is some disagreement over the taxonomic name, as pied stilts are sometimes recorded as a full species, *Himantopus leucocephalus*, and sometimes as a subspecies, *Himantopus himantopus leucocephalus* (see BirdLife Australia, 2019; Gill, 2010). There are also several common names by which the species is known, including Australasian pied stilt, black-winged stilt, and white-headed stilt. Here I use 'Australian pied stilt' to refer to the Australian population, and the te reo Māori name 'poaka' for the New Zealand population, and 'pied stilt' to refer to both collectively, with the scientific name *Himantopus himantopus leucocephalus* for both.

estimating genome-wide diversity and relatedness of threatened species with limited resources (Appendix A).

In Chapter Three, I assemble a mitochondrial genome (mitogenome) for kakī, and conduct sequencing and mapping of mitochondrial genomes for additional kakī, poaka and Australian pied stilts, and kakī-poaka hybrids to evaluate the evolution and phylogeography of these congeneric stilts. The use of mitogenomes allows incorporation of historic samples for analysis of temporal changes in diversity, where such samples may not be sufficiently well-preserved for analyses based on nuclear genomes. In the absence of a robust nuclear genome annotation, mitogenomes are used here for phylogenetic inference. Classification of samples included in sequencing and mapping is based on plumage following the results of Steeves et al. (2010) describing a correlation between plumage node and genetic assignment to kakī. Mitochondrial diversity is compared between poaka and kakī to assess the effects of their independent evolutionary trajectories and population histories. Historic samples are incorporated to assess the hypothesised decline in kakī genetic diversity following the peak of the population decline in the early 1980s. Bayesian phylogenetic analysis including mitochondrial sequence data from forty additional species within the Order Charadriiformes is used to estimate the timing of divergence of Australian pied stilts and kakī from a common ancestor. Following mapping of modern kakī resequencing data to the assembled kakī mitochondrial genome, an anomalous region with unexpectedly deep coverage was detected, which I hypothesised to represent a partial mitochondrial gene duplication, but which may also represent a nuclear pseudogene (see Appendix D).

To investigate the genome-wide impacts of hybridisation in kakī, in Chapter Four I implement a reduced-representation genotyping-by-sequencing approach for population-level genotyping. The kakī genome is used as a reference to guide the

discovery of single-nucleotide polymorphisms (SNPs) for assessment of introgression resulting from anthropogenic hybridisation with poaka during kakī decline, along with other conservation-relevant estimates. Using thousands of genomic markers, a population-level assessment of introgression into kakī confirms the extent of introgression resulting from past hybridisation and the effectiveness of conservation management aiming to minimise gene flow between kakī and poaka. In Chapter Five I synthesise the findings and implications of this thesis for kakī conservation management, and the role of conservation genomics for studies of hybridisation more broadly. I also illustrate the ways in which this thesis bridges the research-implementation gap and enhances kakī recovery, through project co-development, ongoing knowledge-sharing, and co-production of research outputs with conservation practitioners.

1.2 Hybridisation

In the face of the sixth mass extinction, understanding the interacting factors contributing to species extinctions is vital to minimise negative anthropogenic impacts on biodiversity. One such factor is interspecific hybridisation, here defined as the breeding between genetically distinct species. During the process of speciation, a range of barriers evolve that prevent gene flow between diverging populations (Coyne & Orr, 1989). Prezygotic barriers prevent fertilisation, and can include the evolution of distinct reproductive behaviours, and incompatible sexual morphologies. Such barriers evolve more rapidly when diverging species occur in sympatry, while postzygotic barriers typically evolve later in the evolutionary process, encompassing hybrid inviability, sterility, or breakdown (Coyne & Orr, 2004). Despite

the evolution of barriers to gene flow during speciation, interspecific hybridisation has been reported in up to 10% of animals and 25% of plants[2] (Mallet, 2005). Hybridisation is an important process in evolution and speciation, with a range of potential outcomes. It may improve individual fitness, population resilience and adaptive potential (Arnold, 1997; Dowling & Secor, 1997; Mallet, 2007; Seehausen, 2004), and as such hybridisation between closely related species or subspecies has been proposed as a conservation management tool to assist genetically depauperate threatened species by introducing novel genetic variation to increase genetic diversity and improve fitness (Arnold, 2016; Harrisson et al., 2016; Ingvarsson, 2001; Mallet, 2005). Studies exploring the effects of genetic rescue demonstrate the promise of this approach for conservation (e.g., Åkesson et al., 2016; Chan et al., 2019; Miller et al., 2012; Pimm et al., 2006; Quinzin et al., 2019; Rick et al., 2019), but implementation has thus far been limited due to concern around outbreeding depression (the reduction in fitness of outbred individuals; Edmands, 2007; Frankham et al., 2011). More recently, Ralls et al. (2017) have suggested that these concerns are inflated and have resulted in practitioner avoidance of genetic rescue as a management tool without full consideration of the costs and benefits of genetic rescue compared with a lack of action. Additional research to further support the suggested low risk of outbreeding depression will be beneficial in providing practitioners with the confidence to implement these management strategies where appropriate (such as is being implemented under the Genetic Rescue of Australian Wildlife project, Australian Research Council Linkage Project LP160100482, https://sites.google.com/monash.edu/geneticrescue).

[2] Although hybridisation is widespread among plants, and has been frequently harnessed for agricultural purposes, the complex genomic nature of these hybridisation events makes plant hybridisation beyond the scope of this thesis.

1.3 Hybridisation and conservation

While genetic rescue shows promise as a conservation management tool for genetically depauperate species, it can also have significant negative impacts on threatened species. Hybridisation is of conservation concern when it involves a threatened endemic species hybridising with a more common introduced species, the prevalence of which is increasing due to anthropogenic effects leading to changes in species distributions (Allendorf et al., 2001; Rhymer & Simberloff, 1996; Todesco et al., 2016; Vilà et al., 2000). Populations of threatened species are vulnerable to factors intrinsic to their small size, including increased inbreeding resulting in fitness reductions (Charlesworth & Charlesworth, 1999; Keller & Waller, 2002), loss of genetic diversity due to the strong impacts of genetic drift reducing adaptive potential (Lynch et al., 1995), along with greater risk of extinction due to stochasticity (Lande, 1993). Restricted mate choice in these small populations may promote hybridisation: when conspecifics are rare, hybrid pairs may form with a sympatric closely related species (the desperation principle; Hubbs, 1955; McCracken & Wilson, 2011; Steeves et al., 2010). Hybridisation can further increase the extinction risk for these small populations through population decline resulting from wasted reproductive effort (demographic swamping; (Allendorf et al., 2001; Wolf et al., 2001), outbreeding depression resulting from the replacement of locally adapted alleles and the breakdown of co-adapted gene complexes (Arnold, 1997; Edmands, 2007; Lynch, 1991), or through the introduction of maladaptive traits (Allendorf et al., 2013). Furthermore, hybridisation can directly cause species extinction, where introgression (gene flow between populations mediated by backcrossing) results in genetic swamping or homogenisation, to the extent that no genetically distinct individuals remain (Allendorf et al., 2001; Arnold, 1997; Quilodrán et al., 2018; Rhymer & Simberloff, 1996; Riley et al., 2003; Seehausen, 2004; Taylor

et al., 2006; Todesco et al., 2016; Yan et al., 2018). Conservation management has occasionally inadvertently promoted hybridisation which can have unintended negative impacts on population recovery (e.g., cross-fostering of Chatham Island black robin (*Petroica traversi)* eggs with Chatham Island tomtits (*Petroica macrocephala chathamensis*) to increase reproductive outputs of the critically endangered black robin resulted in mis-imprinting of robins on their tomtit hosts, and subsequent hybridisation; Butler & Merton, 1992).

Hybridisation also complicates species concepts and taxonomic delineations, creating additional challenges for conservation management and policy. This is particularly true when hybrids may have value for threatened species management but are not given the same protections as threatened species (Ellstrand et al., 2010; Grant & Grant, 1992; Haig & Allendorf, 2006). With the improved power of genomic data, perspectives regarding these challenges to policy and management may shift, particularly for species that have a long evolutionary history including hybridisation (e.g., North American *Canis* spp.; vonHoldt et al., 2016b, 2016a). Results of these genomic studies thus far suggest the primary consideration should be the impact of hybridisation on individual fitness among the species of concern, with adaptive management determining thresholds for introgression based on fitness impacts (vonHoldt et al., 2018). Galápagos tortoises represent an example of the potential value of hybrids for conservation and ecosystem restoration. Hybrids of the extinct *Chelonoidis niger* and extant *C. becki* have been identified as retaining introgressed material from the extinct species (Quinzin et al., 2019). A targeted captive breeding for translocation programme incorporating these hybrids has been proposed to produce individuals ecologically equivalent to the extinct *C. niger* for release and ecological restoration on Floreana Island.

Avian hybridisation is particularly well-documented (e.g., McCarthy, 2006; Ottenburghs et al., 2015), due in part to the popularity of bird-watching, and more recently the rise of wildlife- and eco-tourism (Connell, 2009; Rouche, 2003; Walther & White, 2018), and citizen science projects aimed at monitoring and documenting birds (e.g., iNaturalist, eBird). The high dispersal capabilities of many birds further promotes avian hybridisation. Anthropogenic effects such as climate change, habitat modification, and increased trade and travel are unintentionally promoting dispersal and contributing to altering species distributions, and new hybrid zones are likely to emerge (Chunco, 2014). With unpredictable, potentially significant negative impacts on small populations of threatened species, hybridisation thus represents a major challenge for conservation. Despite this, the toolbox of resources available to conservation practitioners continues to expand.

1.4 Conservation genomics

Genetic and genomic tools and principles are among the resources available to inform conservation management. Conservation genetics is currently in transition to a genomics approach (Allendorf et al., 2010; Garner et al., 2016; Ouborg et al., 2010; Primmer, 2009). Traditional conservation genetics uses population genetic tools including microsatellites and mitochondrial markers to understand and manage the underlying processes inherent to the small, isolated populations typically requiring conservation (Frankham et al., 2017). These tools can characterise genetic diversity and population differentiation (Boessenkool et al., 2007; Forsdick et al., 2017; Whitehouse & Harley, 2001), relatedness and inbreeding (Eldridge et al., 1999; Gómez-Sánchez et al., 2018; Oliehoek et al., 2006), and hybridisation and introgression (Goodman et al., 1999; Hänfling et al., 2005; Sant'Ana Sousa et al.,

2013; van Heugten et al., 2017). Indeed, when hybridisation and subsequent backcrossing are extensive, the production of cryptic hybrids indistinguishable from either parental type can present a challenge in identifying individuals with hybrid ancestry, and assessing the extent of introgression. In such cases, molecular tools including microsatellite markers and single-nucleotide polymorphisms (SNPs) are invaluable (Chan et al., 2006a; Cubrinovska et al., 2016; Goodman et al., 1999; Pierpaoli et al., 2003; Yan et al., 2018).

Conservation genetic tools benefit conservation by informing and assessing the implementation of strategies designed to reduce the adverse effects associated with small population size including inbreeding and loss of genetic diversity (Frankham et al., 2017). However, these genetic tools are limited by the small number of markers used relative to the size of the genome. Genome size varies markedly between and even among taxa, with genomes of birds and mammals averaging 1–3 Gb (Gregory, 2001). Conservation genetic analyses typically use fewer than twenty markers; thus these genetic markers may not be representative of genome-wide diversity (Allendorf et al., 2010; Ouborg et al., 2010). This may disproportionately impact genetically depauperate species where genetic markers may produce insufficient resolution for producing reliable estimates of diversity or relatedness, as is the case for many species of conservation concern (Galla et al., 2019, see Appendix A; Taylor, 2015).

Over the past decade, high-throughput sequencing (HTS) techniques have greatly improved, becoming faster, more accurate, and increasingly cost-effective (Check Hayden, 2014; Narum et al., 2013; Payseur & Rieseberg, 2016; van Dijk et al., 2014). The greater power and resolution of genomic data has already been shown to outperform previous genetic techniques (Galla et al., 2020; Hauser et al., 2011; Lemopoulos et al., 2019; Morin et al., 2009; Parejo et al., 2018; Puckett & Eggert, 2016; Santure et al., 2010), and genomic tools have been used for investigating

population genetic diversity (e.g., Ekblom et al., 2018), parentage and relatedness (e.g., Thrasher et al., 2018), evolution and adaptation (Zhang et al., 2014b), and hybridisation (e.g., Oswald et al., 2019). These improvements, combined with the development of HTS analysis pipelines reducing computational challenges (e.g., GATK, McKenna et al., 2010; dDocent, Puritz et al., 2014; TASSEL-GBS, Glaubitz et al., 2014; PyRAD, Eaton, 2014), have resulted in genomic methods becoming increasingly feasible for the study of non-model organisms (Andrews et al., 2016; Avise, 2010; Ekblom & Galindo, 2011; Grueber, 2015; Matz, 2018; McMahon et al., 2014; Wright et al., 2019). Despite a relatively slow initial uptake of genomics by conservation geneticists (Shafer et al., 2015; Taylor et al., 2017), HTS data is now being used to inform conservation management for a range of threatened species (e.g., Chinook salmon, *Onchorhynchus tshawytscha*, Larson et al., 2014; Przewalski's horse, *Equus ferus przewalskii*, Der Sarkissian et al., 2015; Visayan warty pig, *Sus cebifrons*, Nuijten et al., 2016; 'Alalā/Hawaiian crow *Corvus hawaiiensis*, Sutton et al., 2018; Tasmanian devil, *Sarcophilus harrisii*, Brandies et al., 2019; eastern tiger salamander, *Ambystoma tigrinum*, McCartney-Melstad et al., 2018).

With the development of HTS, increasingly complex questions can be addressed, including those relating to functional variation, evolutionary processes, gene activity, demographic history, fitness, and adaptive potential (Deakin et al., 2019; Primmer, 2009; Zhang et al., 2014b; Zhao et al., 2013). Along with whole-genome sequencing, HTS also simplifies the sequencing of mitochondrial genomes to explore the evolutionary history of species, enabling accurate taxonomic delineation, resolution of phylogenetic relationships and spatial structuring, and estimating species divergence times (Soares et al., 2016; Zhang et al., 2014a). Mitogenome sequencing and assembly is relatively straightforward compared with that of whole-genome

sequencing and assembly, and the small mitogenome size (typically < 20 kb) means such sequencing is more cost-efficient than whole-genome sequencing methods, although limited to inferences around maternal lineages. HTS also allows the incorporation of ancient DNA (aDNA) from fossils or museum specimens, which had previously yielded DNA too degraded to produce sufficiently informative data (Leonardi et al., 2017). With HTS methods, museum skins and fossil samples are widely used to provide information about past population processes and extinct species (e.g., Parks et al., 2015; Thomas et al., 2019), and can contribute to conservation management (e.g., Kearns et al. 2016).

Relatively cost-effective reduced-representation sequencing (e.g., genotyping-by-sequencing (GBS; Elshire et al., 2011), restriction-site associated DNA sequencing (RADseq; Davey & Blaxter, 2010; Peterson et al., 2012) is proving particularly useful for conservation projects, both allowing the sequencing of many individuals, and producing large SNP sets that can be used for various analyses, with or without existing genomic resources (e.g., Hohenlohe et al., 2011; Wright et al., 2019). These methods are facilitating the transition to conservation genomics approaches. As the available genomic resources (i.e., reference genomes) continue to increase, costs decline, and additional studies confirm the utility of these resources, genomic approaches will be increasingly implemented for conservation management. Nevertheless, genetic approaches are still effective for some aspects of population monitoring, such as genetic sexing and paternity assignment (e.g., Cremona et al., 2017; Rudnick et al., 2005).

HTS techniques are already being implemented to identify genome-wide impacts of introgression (Alexander et al., 2017; Russell et al., 2019; Seabra et al., 2019; Toews et al., 2016). These genomic methods using large numbers of SNPs have much greater power to identify hybrid origins of individuals than previous genetic

studies (McFarlane et al., 2020). Direct comparisons of microsatellite- and SNP-based estimates of individual-level admixture between invasive rainbow trout (*Onchorhynchus mykiss*) and westslope cutthroat trout (*O. clarkii lewisi*) confirmed the improved accuracy of SNP loci over microsatellites (Boyer et al., 2008; Hohenlohe et al., 2013). RADseq for SNP genotyping of 3,180 loci produced admixture estimates that correlated with those derived from seven microsatellite loci, but identified a low level of introgression previously undetected (Hohenlohe et al., 2013). Additionally, extensive investigation of hybridisation between invasive Sika deer (*Cervus nippon*) and native red deer (*C. elaphus*) in Kintyre, Scotland has used both genetics and genomics approaches to detect hybrids. A panel of 22 microsatellites and a region of the mitochondrial genome have been used for identifying hybrids. More recently, the use of a SNP chip comprising over 40,000 markers has been implemented, providing much greater diagnostic power for detecting deep backcrosses supporting individual reclassifications (McFarlane et al., 2020). These and similar studies provide promise for the improved elucidation of the impacts of hybridisation on threatened species.

1.5 The conservation genomics gap

Researchers and practitioners in the conservation space have observed a disconnect between conservation research and action, in both conservation practice and policy, known as the research-implementation gap (Knight et al., 2008). This gap emerges due to mismatches between researchers and practitioners, in the form of temporal, spatial, priority, communication, and institutional mismatches arising from the differing systems in which research and practice are enacted (Jarvis et al., 2015). The research-implementation gap sits within a wider space where the values and

perceptions of various publics, social interactions, political impacts, and the interactions between mismatches and these additional factors also come into play (Buschke et al., 2019; Toomey et al., 2017), with some calling for a paradigm shift to address this 'great divide' (Arlettaz et al., 2010). While progress has been made to address these mismatches (Jarvis et al., in review), there is still room for improvement, with researchers and practitioners making a range of recommendations that may bridge the gap within this shifting space (e.g., Dubois et al., 2020; Jarvis et al., in review, 2015; Knight et al., 2008; Laurance et al., 2012), improving the alignment of research and practice and enhancing conservation outcomes for threatened species.

While the research-implementation gap affects all aspects of conservation research and practice, and will require widespread change to adequately resolve, one particularly relevant facet of this disconnect is the conservation genomics gap. This refers to the relative paucity of genomic data implemented for conservation (Galla et al., 2016; Shafer et al., 2015), primarily resulting from the lag time associated with the transition from conservation genetic to genomic approaches. During the initial development of genomics, uptake of these tools among primary industry outstripped the uptake for conservation purposes, despite these two fields having aligned research interests (i.e., assessing inbreeding, relatedness, and genome-wide diversity; Galla et al., 2016). There are a number of interrelated factors contributing to this lag, with overlap between those factors associated with the research-implementation gap, largely centred on resource availability, whether that be in the form of time available to develop the skills and optimise techniques and workflows to implement conservation genomic approaches, or the (both real and perceived) costs associated with genomic sequencing (Galla et al., 2016; Shafer et al., 2015; Taylor et al., 2017). Another key contributor is the result of communication mismatch

associated with the research-implementation gap, where clear communication between researchers and practitioners is required to disseminate and discuss the utility and prioritisation of genomics approaches (Kadykalo et al., 2020), particularly for projects where genetic tools have already been applied and the need for a genomics approach may not have been made clear (Taylor et al., 2017). While the nature of conservation research and implementation may have resulted in the gap appearing wider than the actuality due to different methods of disseminating such work (see Garner et al., 2016; Shafer et al., 2015), this gap is steadily narrowing (Galla, 2019). There are a number of ways to bridge this gap, including the dissemination of evidence of the benefits of genomic research as implemented for conservation (e.g., Galla et al., 2020; McLennan et al., 2019; Ogden et al., 2013; Seabury et al., 2011), the development of workflows that can be readily adapted for widespread implementation (e.g., Wright et al., 2019), and through the co-development of multidisciplinary approaches to inform conservation management (Galla et al., 2016; Hogg et al., 2017).

1.6 Conservation in an Aoteraroa New Zealand context

New Zealand has a history of geographic isolation resulting in a diverse avian cohort occupying a wide array of ecological niches. As with many small islands, the endemic biodiversity has been adversely affected following a range of synergistic effects associated with human arrival, including the introduction of mammalian predators, extensive habitat modification, and genetic factors associated with small population size (Duncan & Blackburn, 2004, 2007; Frankham, 1997; Jamieson, 2009). Since the late 1970s, a range of innovative conservation efforts have been implemented to prevent further extinctions and assist population recovery, including

translocations of species to offshore islands (Elliott et al., 2001), predator-proof fencing of key mainland sites (Saunders & Norton, 2001), and more recently, the Predator Free 2050 programme aiming to eradicate introduced mammalian predators, specifically targeting mustelids (*Mustela* spp.), brush-tailed possums (*Trichosurus vulpecula*), and rodents (*Rattus* spp., *Mus musculus*) across the mainland of New Zealand (Russell et al., 2015). Conservation genetic principles and tools have also been widely used to inform species management (e.g., for takahe, *Porphyrio hochstetteri*, Wickes et al. 2009; kākāpō, *Strigops habroptilus*, White et al. 2015; and toutouwai/South Island robin, *Petroica australis*, Heber et al. 2013). Along with population declines resulting in bottlenecks, fragmentation and isolation, there are several examples of anthropogenic hybridisation among Aotearoa's biota. Studies of these hybridisation events largely focus on birds (but see Chapple et al. 2012, Banker et al. 2017, van Heugten et al. 2017), including ducks (Anatidae, Gillespie 1985), kiwi (Apterigidae, Ramstad and Dunning 2020), parakeets (Psittaculidae, Taylor 1975), gulls (Laridae, Gurr 1967), and stilts (Recurvirostridae, Pierce 1984b). Genetic analyses using allozymes (variant forms of enzymes), microsatellites (short sequence repeats), and/or mitochondrial markers (targeted regions within the mitochondrial genome) have been implemented for each of the above examples (Boon et al., 2000; Chan et al., 2006a; Mischler et al., 2018; Rhymer et al., 1994; Steeves et al., 2010). However, use of genomic tools has yet to be widely implemented for studies of hybridisation in New Zealand. Among the limited examples to date is the use of GBS alongside mitochondrial markers to clarify population structure among tarāpuka/black-billed gulls (*Larus bulleri*), which identified mitochondrial introgression from tarāpunga/red-billed gulls (*L. novaehollandiae scopulinus*) with no evidence of nuclear introgression (Mischler et al., 2018). Genomic and transcriptomic approaches are being employed to elucidate

divergence and adaptation among kiwi species, with hybridisation occurring between great spotted kiwi (*Apteryx haastii*), little spotted kiwi (*A. owenii*) and rowi (*A. rowi*). This hybridisation is of concern due to expected outbreeding depression resulting from the ~5 million years divergence between these species (Ramstad & Dunning, 2020).

Implementation of genomic approaches to questions pertaining to hybridisation are of benefit to species conservation, particularly when hybrids are difficult to distinguish from the parental taxa, complicating identification of the extent of hybridisation and the impact on the threatened species. Hybridisation between pārera (New Zealand grey duck, *Anas superciliosa*) and the introduced mallard (*A. platyrhynchos*) has been a key driver of pārera decline (Robertson et al., 2016). Accurate estimates of pārera numbers are confounded by observational uncertainty due to the challenges of correctly identifying pārera from morphologically similar hybrids and the sexually dimorphic mallard, obscuring the full impact of hybridisation (Williams, 2017). Mitochondrial analysis indicated bidirectional introgression had occurred between the species (Rhymer et al., 1994), there is concern that the extent of hybridisation has resulted in the formation of a hybrid swarm. The implementation of double-digest RADseq (ddRADseq) for SNP discovery and analysis of nuclear introgression indicates that there is extensive introgression from mallards into pārera, but that pārera on the west of the Southern Alps of the South Island remain genetically distinct, suggesting that the Southern Alps represent a geographic barrier to mallard dispersal (Brown et al., 2020).

1.7 Kakī, *Himantopus novaezelandiae* (Order Charadriiformes; Gould 1841)

The kakī is an endemic New Zealand wading bird, regarded by Māori, the indigenous people of New Zealand, as a taonga species (treasured species; (*Ngāi Tahu Claims Settlement Act*, 1998). Also known as black stilt, the kakī is a member of the Family Recurvirostridae comprising stilts and avocets, and is most closely related to the congeneric pied stilt (*Himantopus himantopus leucocephalus*). Kakī are thought to have evolved from a common ancestor with pied stilts following arrival in New Zealand, and are estimated to have been present in New Zealand for around one million years (Wallis, 1999), although estimates are yet to be confirmed. Kakī and pied stilts co-occur in New Zealand (where pied stilts are known as poaka), following a secondary invasion of the latter in the early 19th century (Pierce, 1984b). Kakī were once widely distributed across the country (Figure 1.1), but anthropogenic impacts including habitat alteration caused population decline. Today, the remaining population is limited to Te Manahuna/the Mackenzie Basin, in the central South Island (Figure 1.2), with occasional vagrants observed in coastal locations around the country (Pierce 1984; personal observation). The same anthropogenic effects that led to the decline of kakī facilitated the expansion of poaka southwards across the country (Pierce, 1984b), with the population estimated at over 30,000 individuals (Robertson & Heather, 2015), although anecdotal evidence suggests the population may be declining in the central South Island (DOC, pers. comm.).

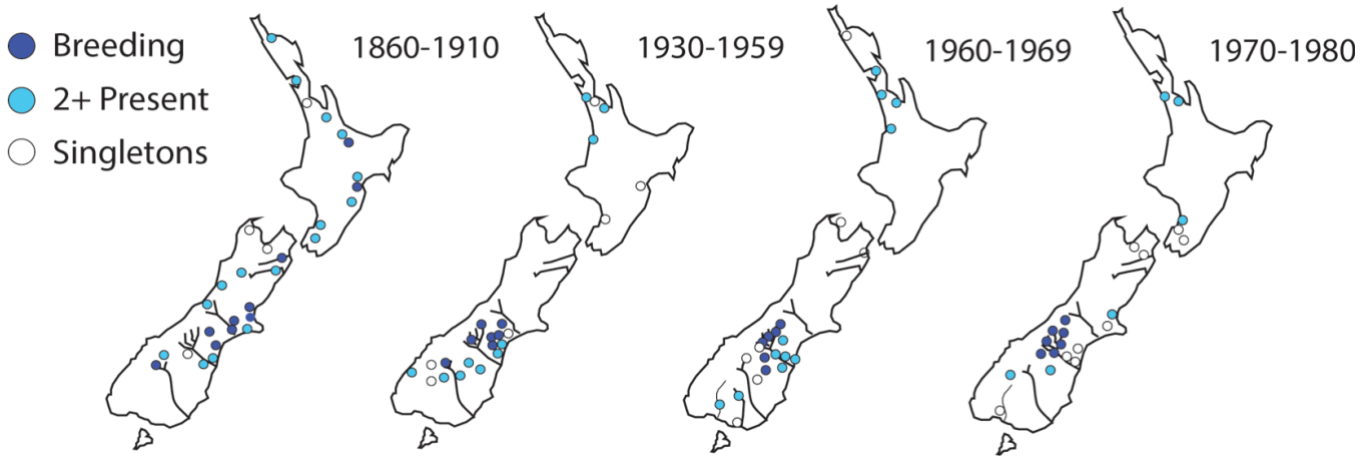


Figure 1.1: Historic kakī distribution across New Zealand. Modified from Pierce (1984a) by S. Galla, reproduced from Galla (2019) with permission.

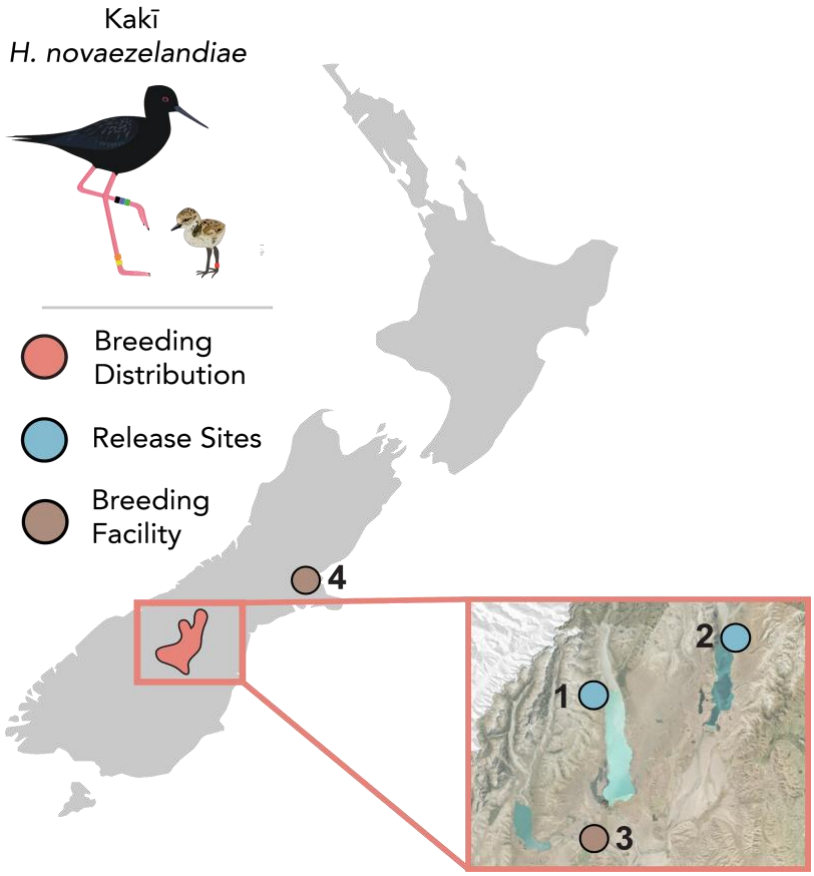


Figure 1.2: Current breeding distribution of kakī in Te Manahuna/Mackenzie Basin with current release sites for captive-reared kakī in 1) the Tasman River delta and 2) Mt. Godley. Site 3) marks the location of the Department of Conservation's kakī captive breeding and rearing facility in Twizel, and 4) marks the Isaac Conservation and Wildlife Trust's kakī captive breeding and rearing facility in Christchurch. Reproduced from Galla (2019) with permission.

1.7.1 Kakī biology

Kakī are medium-sized wading birds, around 40 cm in height and weighing around 220 g (Pierce, 1984a). They are braided river specialists, and forage for aquatic invertebrates and small fish. Kakī usually reach sexual maturity at two years and can live up to fifteen years. Breeding occurs annually, with females typically laying a single clutch of four eggs. After hatching, chicks are covered in brown-speckled down feathers, which acts to camouflage them in their stony riverbed habitat. Fledglings first leave the nest in their second month and reach independence at 3–6 months. As adults, they have completely black plumage with a green iridescence to the wings, long red legs, and a long narrow black bill (Figure 1.3). Juveniles have pale pink legs and mottled black and white plumage (Figure 1.4), becoming completely black by their second winter (Pierce, 1984a).

Figure 1.3: Banded adult (left) and subadult kakī (right). Photo: L. Brown, reproduced with permission.

Figure 1.4: Four juvenile kakī. Photo: C. Forsdick, reproduced with permission.

1.7.2 Kakī demographic history and conservation management

Kakī declined during the 1900s in both numbers and range due to anthropogenic habitat modification and predation by introduced species (Pierce, 1984b). The species is currently listed as Critically Endangered in the IUCN Red List of Threatened Species (BirdLife International, 2018), and classified as Nationally Critical by the New Zealand Department of Conservation (DOC; Robertson et al. 2016). In 1981, the population had declined to approximately 23 birds (Steeves et al., 2010), and a conservation management programme was initiated. Since its inception, management has incorporated predator control, habitat management, and a programme of captive breeding for translocation (Cruz et al., 2013; Maloney & Murray, 2001), which has resulted in the current wild adult population increasing to 169 adults at the end of the 2019/20 breeding season (DOC, personal communication).

Key threats to species survival include predation, and habitat loss and modification (Maloney & Murray, 2001). Introduced predators include feral cats (*Felis catus*),

hedgehogs (*Erinaceus europaeus*), stoats (*Mustela erminea*) and ferrets (*Mustela furo*), with disproportionate impacts on chicks and juveniles (Pierce, 1986b; Sanders & Maloney, 2002). Since 1850, 40% of Te Manahuna wetlands have been drained (Wilson, 2001), with similar land conversion across the country. This land conversion has reduced floodplains and wetlands, and increased open water habitat and lake shorelines (Wilson, 2001), and the loss of the major disturbance mechanism has had a range of flow-on effects to kakī, including reduced nesting and feeding areas (Maloney et al., 1997; Maloney & Murray, 2001). More recently, primary industry and recreational use of the land and waterways in these areas has also had significant impacts on the habitat available to kakī (Caruso, 2006). In addition, limited mate choice during the decline resulted in kakī hybridising with the more numerous poaka (Pierce, 1984a), which may have exacerbated the decline (Steeves et al., 2010). Current conservation management includes extensive predator control around key nesting areas (Keedwell et al., 2002; Maloney & Murray, 2001). DOC's Project River Recovery is acting in tandem with the newly established Te Manahuna Aoraki landscape restoration project, substantially increasing the area and intensity of predator control in the river deltas and valleys to the north of lakes Pūkaki and Tekapō (Te Manahuna Aoraki Project Annual Report 2019). These projects aim to ensure Te Manahuna remains a key New Zealand braided river habitat for birds, invertebrates, and fish (Caruso, 2006; Maloney et al., 1997).

A captive breeding programme was established as part of the Kakī Recovery Programme in 1987 to increase population numbers and prevent extinction (Pierce, 1996). As part of this programme, all individuals are banded for identification, and data is collected in the 'Alpha List' tracking individual breeding, release, and survival data (DOC, pers. comm.). These data, along with paper pedigree records, have been used to build a full studbook for kakī (Galla et al., 2020). A small number of

individuals are retained in captivity each year to form managed breeding pairs. Small population size has resulted in genetic challenges such as inbreeding depression, where closely related breeding pairs have lower hatching success than unrelated pairs (Hagen et al., 2011). Until recently, individuals in the captive breeding programme have been genotyped at eight microsatellite loci and relatedness estimates combined with pedigree information, to select individuals with the lowest relatedness estimates as potential pairs (Hagen et al., 2011). However, comparison of estimates of relatedness obtained from pedigrees, microsatellite and SNP markers derived from whole-genome resequencing data revealed that the use of SNPs alongside pedigree data produced more precise and accurate estimates than those from microsatellites (Galla et al., 2020). Moving forwards, this approach will be implemented by the Kakī Recovery Programme to minimise inbreeding.

Alongside captive breeding, captive rearing is conducted, whereby eggs are collected from both captive and wild kakī pairs and artificially incubated. Egg-pulling encourages pairs to produce multiple clutches, increasing the number of eggs produced beyond the potential of the wild population (Heezik et al., 2005). High rates of predation in the wild are the main driver of low survival to adulthood (Keedwell et al., 2002; Maloney & Murray, 2001; Pierce, 1986b, 1996; Sanders & Maloney, 2002) and so chicks are reared in captivity. Juveniles reaching the age of independence are released back into the wild to supplement subpopulations (Maloney & Murray, 2001). This is intended to increase effective population size, and reduce demographic stochasticity of these subpopulations (Heezik et al., 2009). As hybridisation is more likely to occur when conspecifics are rare, and particularly when there is a bias in the sex ratio (Hubbs, 1955; McCracken & Wilson, 2011; Randler, 2002; Steeves et al., 2010), wild release also aims to reduce the likelihood of hybridisation (Maloney & Murray, 2001).

1.7.3 Hybridisation between kakī and poaka

Hybridisation between kakī and poaka occurred during the period of kakī decline, resulting in the production of viable hybrid offspring with intermediate plumage between completely black kakī and black and white poaka (Figure 1.5 and Figure 1.6). Hybrids have been observed since the late 1800s, though these observations led to the initial categorisation of a number of different stilt species, rather than the range of plumage nodes comprising the two species and their hybrids that are recognised today (Pierce, 1984a; Figure 1.5). Additional confusion over species delimitation may have arisen due to the black and white plumage of juvenile kakī. The nodes described vary in the extent of white plumage and are used by DOC for kakī management, where nodes A–C2 represent poaka, node J is kakī, and the intermediate nodes D1–E are light hybrids and F–I/J are dark hybrids (Pierce, 1984a; Steeves et al., 2010). Other differences in morphology and behaviour exist between the species. Kakī have a longer bill and shorter tarsus, a higher-pitched call, different foraging and predator-avoidance behaviours, and are non-migratory, whereas poaka in Te Manahuna migrate out of the basin over winter (Pierce, 1984a). The observation of poaka having a greater extent of black plumage (nodes B–C2) compared with Australian conspecifics (predominantly node A) indicates that the impacts of hybridisation have been bidirectional, leading to the suggestion that genetically distinct poaka and kakī may no longer exist, and instead represent a hybrid swarm (Pierce, 1984a).

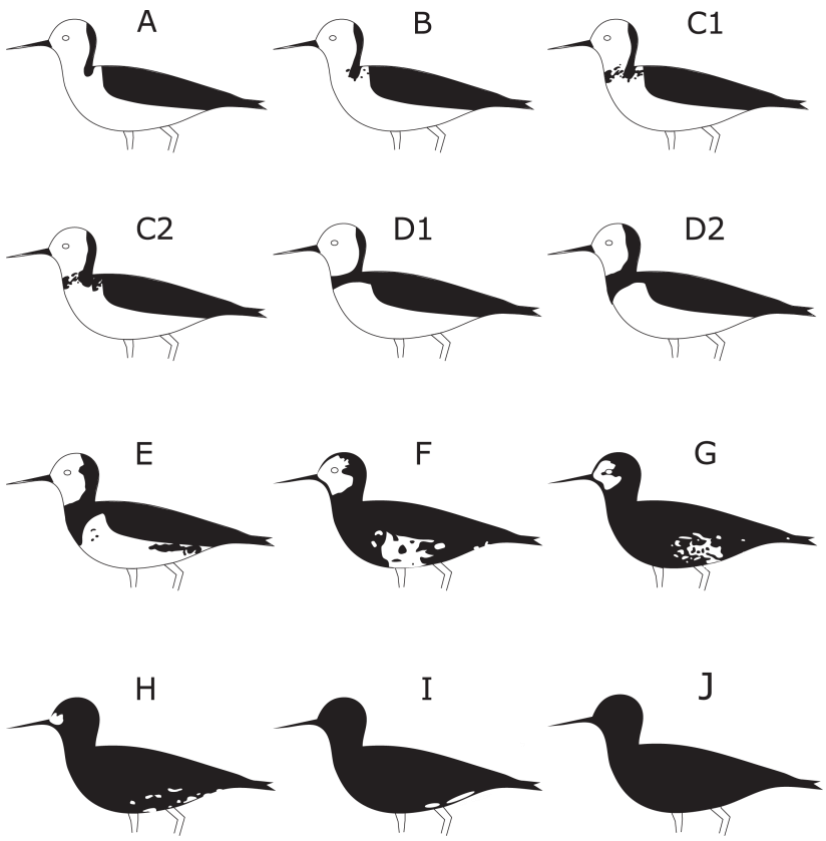


Figure 1.5: Plumage nodes of kakī (J), poaka (A–C2), and kakī poaka hybrids (D–I). These node scores are used by the Department of Conservation to identify kakī individuals for management, with non-kakī individuals excluded. Modified from Pierce (1984b) by S. Galla, reproduced with permission.

Figure 1.6: Two adult stilts foraging. The individual on the left has hybrid node D plumage, while the individual on the right is a node B poaka. Photo: E. Whitehead, reproduced with permission.

In 1981, hybrids were estimated to comprise 17.5% of the adult stilt population in Te Manahuna (Pierce, 1984a). Between 1981 and 1987, management included cross-fostering of kakī eggs to pairs including poaka and hybrids (Reed et al., 1993b). This was in an effort to stimulate multiple clutching by kakī to rapidly boost population numbers. However, there is some evidence kakī that were cross-fostered had learned poaka-type behaviours, including migrating out of Te Manahuna with poaka over the winter. As kakī form breeding pairs in the wild prior to the return of poaka to the basin, cross-fostered migratory kakī were less likely to pair appropriately on return. This behaviour limited the success of such cross-fostered individuals in contributing to the gene pool, and so cross-fostering was restricted to kakī pairs (DOC, pers. comm.).

Dark hybrids (nodes G–I) were actively managed alongside node J kakī as part of the captive breeding programme between 1988–99, as it was thought that the potential risks of dark hybrids backcrossing with kakī was better than kakī individuals not breeding at all (Greene, 1999). As kakī began to recover, hybridisation became less frequent, and minimising gene flow between the species became a key objective in the recovery plan in response to data from Wallis (1999; see Steeves et al. 2010) and thus dark hybrids were actively excluded from conservation management (Maloney & Murray, 2001).

The extent of hybridisation created concern that genetic swamping may have resulted in kakī no longer existing as a distinct species, but merely as a hybrid swarm, where all node J individuals were thought to be cryptic hybrids (Steeves et al., 2010). However, genetic analyses have determined the genetic integrity of kakī, indicating that introgression has had a lower impact on the genetic composition of the recovering kakī population than predicted (Steeves et al., 2010; Wallis, 1999). Reduced fitness of hybrids and strong positive assortative mating by kakī, and

stochasticity appeared to have limited introgression (Steeves et al., 2010). As the population continues to increase, the kakī range is likely to expand into areas with larger poaka populations, and hybridisation may once again present a challenge for kakī due to limited mate choice and potential sex ratio imbalance at the expansion front. Furthermore, genetic studies using small marker sets (e.g., < 20 microsatellite markers) assumed to be selectively neutral, while confirming genetic distinctiveness of the two species, may not be representative of genome-wide introgression (Gómez-Sánchez et al., 2018; Hohenlohe et al., 2013; Thongda et al., 2019). Thus far, there is limited data to predict the future potential for hybridisation and the subsequent outcomes if it were to again increase in prevalence, complicating future management decisions. As costs associated with genomic studies are declining, and such approaches more robustly identify admixture and introgression through advanced backcrosses than previous genetic methods, a genomic approach to investigate the impacts of hybridisation in kakī is now feasible.

In this thesis, I will develop genomic resources to reassess the impacts of hybridisation between the critically endangered kakī, and the non-threatened congeneric poaka, providing information to enhance current and future conservation management for kakī. Reference genomes for kakī and Australian pied stilts will be produced in Chapter Two to assist the conservation genomics transition for kakī. To provide context to the evolutionary history of these stilt species, mitochondrial genomes for kakī and Australian pied stilts will be used to estimate the timing of divergence of these species from a common ancestor, and to compare pre- and post-decline mitochondrial diversity for kakī (Chapter Three). Population-level genotyping-by-sequencing and reference-guided SNP discovery will be used to assess the extent of introgression between the two species (Chapter Four). Conservation implications of these results will be synthesised in Chapter Five, along

with discussion of potential future research questions. To minimise the research-implementation gap, and maximise conservation outcomes for kakī, this research has been developed in partnership with conservation practitioners (DOC's Kakī Recovery Programme), and collaboration with practitioners involving ongoing knowledge exchange has been an essential component of the project. Practitioners are included as co-authors on published manuscripts and those in preparation. Thus, the results from this thesis are already in use by the Kakī Recovery Programme to inform future management strategies to enhance kakī recovery. Furthermore, this thesis presents tools and workflows that may assist other threatened species recovery projects making the transition to a conservation genomics approach, thereby contributing to closing the conservation genomics gap.

Chapter 2: Whole-genome sequencing and *de novo* genome assembly for kakī (*Himantopus novaezelandiae*) and Australian pied stilts (*Himantopus himantopus leucocephalus*) to inform conservation

2.1 Abstract

Threatened species management benefits from the integration of genetic tools to assess factors of conservation interest including genetic diversity, population differentiation, inbreeding, relatedness, and hybridisation and introgression. Conservation is currently in transition to implementing a genomics approach, whereby many thousands of genome-wide markers are used to provide greater power and accuracy to estimate relevant metrics. Kakī (black stilt; *Himantopus novaezelandiae*)—a critically endangered wading bird endemic to New Zealand—is one such threatened species that has greatly benefitted from the use of conservation genetics to inform current management including a conservation breeding for translocation programme. In addition, genetic tools have been used to assess the impacts of anthropogenic hybridisation with the self-introduced pied stilt (*Himantopus himantopus leucocephalus*). Here, I use cost-effective Illumina short-read sequencing and a *de novo* approach to assemble the ~1.1 Gb reference genomes for both kakī and the congeneric Australian pied stilt. This is the first step in facilitating the transition to a genomics approach to conservation management for kakī, enabling subsequent genomic investigation of hybridisation between these closely-related species.

2.2 Introduction

In an era where global biodiversity is in decline (Butchart et al., 2010), threatened species management can benefit from the inclusion of genetic data to inform decision-making (Allendorf et al., 2013; Brandies et al., 2019; Caballero et al., 2010; Hoban et al., 2013). With the development of high-throughput genomic sequencing techniques, whole-genome sequencing and *de novo* genome assembly is becoming increasingly accessible for such non-model species, resulting in a transition from conservation genetics to conservation genomics (Allendorf et al., 2010). Using these approaches, genomic markers such as single-nucleotide polymorphisms (SNPs) provide greater power and resolution than genetic approaches using limited marker sets, producing greater accuracy than microsatellite-based estimates in a range of studies (Galla et al., 2020; Hauser et al., 2011; Hohenlohe et al., 2013; Santure et al., 2010; Weinman et al., 2015). While high-throughput sequencing costs continue to decline (Check Hayden 2014, see NIH for human genome sequencing costs), whole-genome sequencing remains a substantial investment for conservation projects (Galla et al., 2019), and genome assembly remains bioinformatically challenging and computationally intensive, contributing to the conservation genomics gap (defined by most as the lag in uptake of genomic tools for conservation management; Knight et al. 2008; Shafer et al. 2015; Garner et al. 2016; Galla et al. 2016; Taylor et al. 2017). Despite these challenges, an increasing number of genomic resources (i.e., reference genomes) are becoming available for non-model species (Figure 2.1), along with a range of proof-of-concept studies demonstrating the utility of genomic approaches for conservation (Galla et al., 2019; Larson et al., 2014; Thrasher et al., 2018; Torkamaneh et al., 2016), thus assisting the transition to conservation genomics approaches for threatened species management.

One of the most valuable resources for genomic studies using thousands of genome-wide markers for estimates of nucleotide diversity, inbreeding, and relatedness are reference genomes. Reference genomes can support conservation management by enabling more accurate reference-guided variant discovery for population genomic analyses (Brandies et al., 2019; Wright et al., 2019), and allow structural, functional, and adaptive comparisons between species (Lamichhaney et al., 2016; Yu et al., 2016; Zhang et al., 2014b; Zhao et al., 2013). For many threatened species, reference genomes are yet to be developed, but genomes may be available for closely related species that may be sufficient to use as references to produce estimates of conservation-relevant metrics. Among the first sequenced genomes were those of model organisms, including those for mice (*Mus musculus*, Waterston et al. 2002), rats (*Rattus norvegicus*, Rat Genome Sequencing Project Consortium 2004), cattle (Schibler et al., 2004), and chickens (*Gallus gallus*, International Chicken Genome Sequencing Consortium 2004). Consortia have been particularly prolific in producing reference genomes for non-model organisms at species-specific scale to subphylum-level and beyond (e.g., Vertebrate Genomes Project (VGP (Genome 10K Community of Scientists 2009)), Bird 10,000 Genomes Project (B10K (Zhang et al. 2014)), 5,000 Insect Genome Project (i5K, Robinson et al. 2011), 1,000 Plants Project (1KP (Matasci et al. 2014)), Oz Mammalian Genomics (Duchêne et al., 2018), Earth BioGenome Project (Lewin et al., 2018)). B10K has contributed to the sequencing and assembly of reference genomes for one member of every modern avian order, with the second phase of the project underway, aiming to produce representative genomes for every modern avian family (Zhang, 2015). Such large-scale efforts make genomics an evolving field with new resources emerging rapidly (Figure 2.1). Despite these large projects, threatened species recovery programmes may be disadvantaged by limited resources, where

urgent conservation actions may be prioritised over the development of species-specific genomic resources. Despite a more than two-fold increase in the available avian genomic resources since the beginning of this study in September 2016, only 52 (24.07%) represent threatened species (VU, EN, CR, EW status according to the IUCN Red List, April 1st 2020; Figure 2.2, Appendix B). With an estimated 13.4% of recognised bird species threatened with extinction (BirdLife International, 2017), and anthropogenic impacts resulting in population trends declining for many more, there is an increasing need for genomic resources as a growing number of species require conservation management.

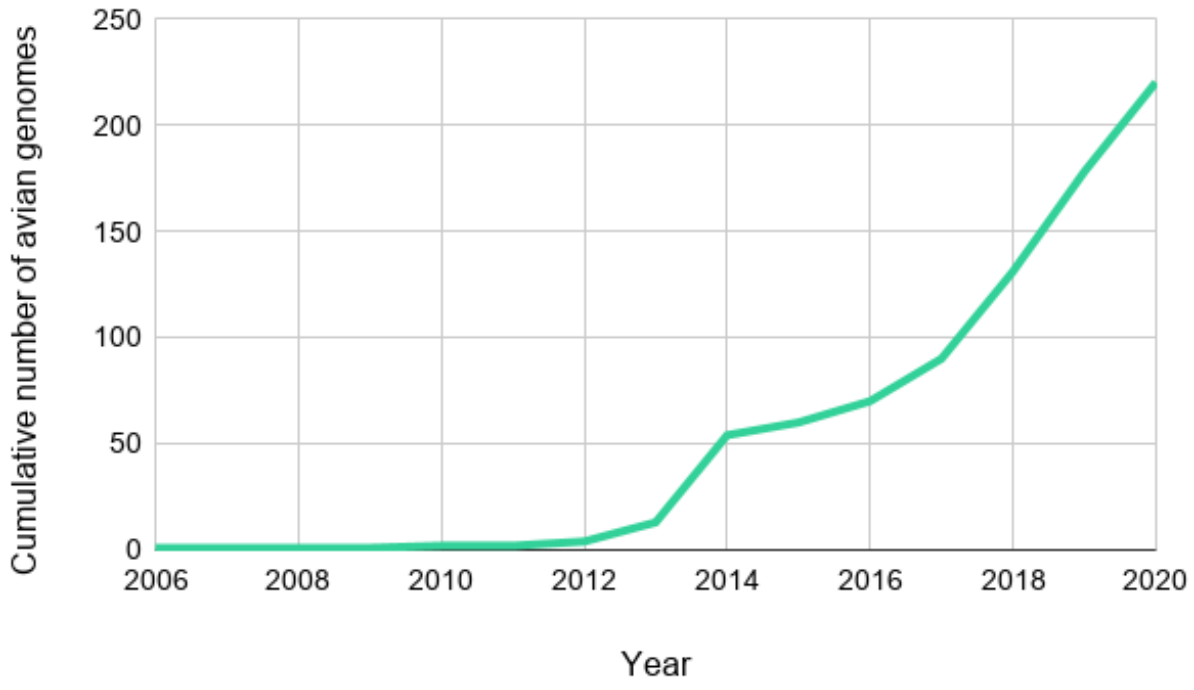


Figure 2.1: The cumulative increase in avian genomes available via NCBI as assessed on April 1st 2020, based on the first date of genome availability in the database, with one representative genome per species. See Appendix B for additional information.

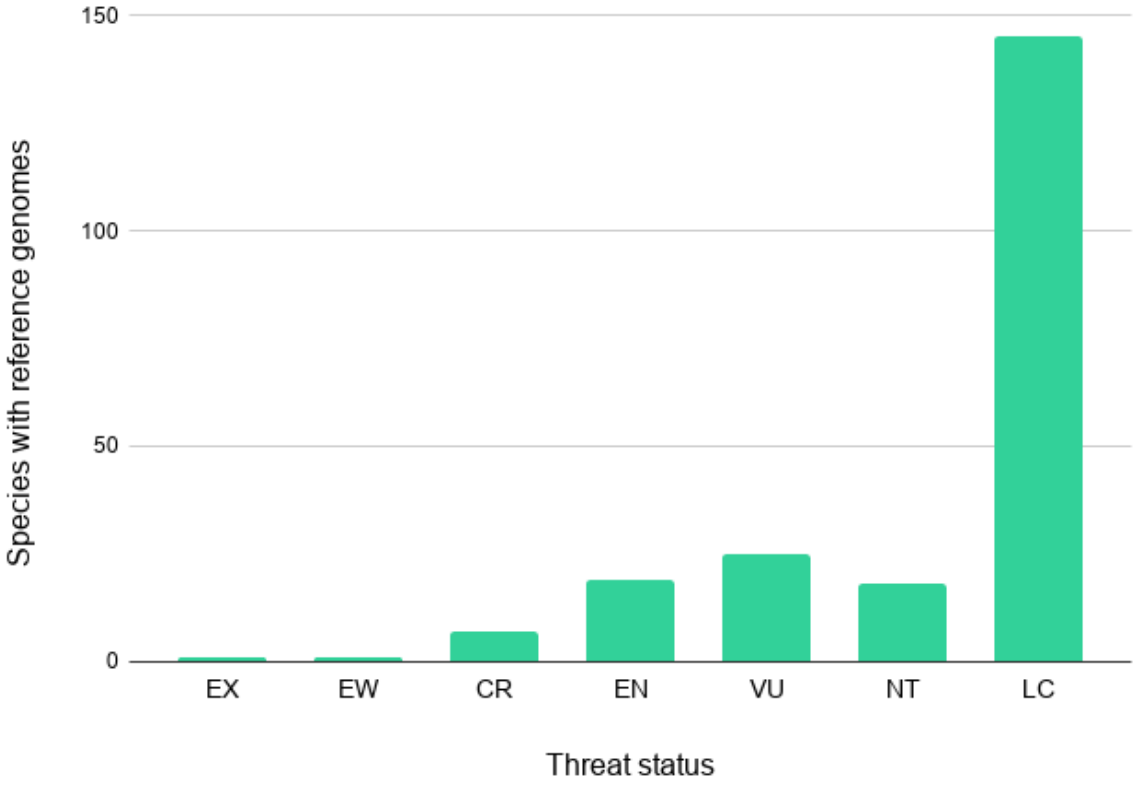


Figure 2.2: IUCN Red List threat status for the 220 avian species with reference genomes available on NCBI, as assessed on April 1st 2020. Threat categories are as listed by the IUCN: EX = Extinct, EW = extinct in the wild, CR = Critically endangered, EN = Endangered, VU = Vulnerable, NT = Near threatened, LC = Least concern. See Appendix B for supporting information.

2.2.1 Study system, kakī (*Himantopus novaezelandiae*) and Australian pied stilts (*Himantopus himantopus leucocephalus*) (Order Charadriiformes)

The kakī (black stilt, *Himantopus novaezelandiae*, Order Charadriiformes) is a New Zealand endemic wading bird. Anthropogenic impacts including habitat modification and loss, and the introduction of invasive predators resulted in population decline throughout the 1900s, until by 1981 only ~23 birds remained in the river systems of the Mackenzie Basin/Te Manahuna (Pierce, 1984b; Steeves et al., 2010). Intensive conservation management has resulted in kakī numbers increasing to 169 wild adults in the 2019/20 summer season (DOC, personal communication). Although the

population has been increasing, kakī remain critically endangered (BirdLife International, 2018; Robertson et al., 2016). Kakī have benefitted from the use of conservation genetics over the past 25 years. Allozymes, microsatellites, and mitochondrial markers have all been used to understand the underlying genetic impacts of the population bottleneck they have experienced, impacts of past hybridisation with poaka (the New Zealand population of Australian pied stilts, *Himantopus himantopus leucocephalus*), and to inform captive breeding for translocation (Greene, 1999; Hagen et al., 2011; Steeves et al., 2010). In comparison, despite the wide distribution of Australian pied stilts throughout the Pacific (e.g., Iqbal 2008; Minton et al. 2017), little is known about the species, with the majority of research in relation to kakī in New Zealand (e.g., Pierce 1984a; Pierce 1984b; Pierce 1986a; Pierce 1986b). In New Zealand, poaka outnumber kakī with an estimated population size of 30,000 individuals (Robertson & Heather, 2015). At the inception of this project, reference genomes were available for only two species at the conordinal level (Order Charadriiformes), those for killdeer (*Charadrius vociferus*, Family Charadriidae, NCBI Accession No.: GCA_000708025.2; Zhang et al. 2014), and the ruff (*Calidris pugnax*, Family Scolopacidae NCBI Accession No.: GCA_001431845.1 and GCA_001458055.1; Küpper et al. 2016; Lamichhaney et al. 2016). The genome of killdeer represents the most closely related genome available to stilts (Family Recurvirostridae). The two families are estimated to have diverged from one another approximately 69 million years ago (Baker et al. 2007, but see Jarvis et al. 2014). To date (April 1st 2020), the number of species with reference genomes available on NCBI within the order Charadriiformes has increased to eleven (not including the two genomes produced in this chapter). This includes genomes for the Amami woodcock (*Scolopax mira*, GCA_004320125.1, unpublished), bar-tailed godwit (*Limosa lapponica baueri*,

GCA_002844005.1, Parody Merino 2018), common tern (*Sterna hirundo*, GCA_009819605.1, unpub.), Eurasian thick-knee (*Burhinus oedicnemus*, GCA_008921705.1, unpub.), Kentish plover (*Charadrius alexandrinus*, GCA_008711295.1, Wang et al. 2019), pied avocet (*Recurvirostra avosetta*, GCA_004023745.1, sequenced and assembled as part of an aligned project, Galla et al. 2019, Appendix A), razorbill (*Alca torda*, GCA_008658365.1, unpub.), spoon-billed sandpiper (*Calidris pygmaea*, GCA_003697955.1, unpub.), and thick-billed guillemot (*Uria lomvia*, GCA_002289315.1, Tigano et al. 2018).

2.2.2 Genome sequencing and assembly

Numerous genome sequencing platforms (e.g., Illumina, Oxford Nanopore Technologies, 10x Genomics) and assembly methods (e.g., SOAPdenovo2, Luo et al. 2012; Meraculous2, Chapman et al. 2017; ABySS, Simpson et al. 2009; Jackman et al. 2017; Canu, Koren et al. 2017) have been developed to facilitate genome assembly using high-throughput sequencing data. While these sequencing platforms differ in the length of sequence reads produced, the assembly methods utilise varying algorithms to assemble the data, each with differing strengths and weaknesses (see Bradnam et al. 2013). For assembly of genomes from short reads, high-depth sequencing to the point of redundancy is required to produce sufficient sequence coverage to confidently assemble a consensus sequence (although high-depth may not be a requirement of long-read sequencing platforms). *De novo* genome assembly of short reads methods involves merging overlapping short paired-end sequences into consensus sequences known as contigs, often through the use of de Bruijn graph approaches. Overlapping contigs can then be assembled into long scaffolds using paired-end sequence information to determine read orientation and distances between contigs. Improvements in accuracy and contiguity

can be achieved through the use of long mate-pair libraries (1–20 kb; Chakraborty et al., 2016; Ekblom & Wolf, 2014; Haridas et al., 2011), or with a reference-assisted approach using a genome of a closely related species to guide the alignment of sequenced reads (Bao et al., 2014; Kim et al., 2013; Lischer & Shimizu, 2017; Pop et al., 2004; Vezzi et al., 2011; Wang et al., 2014). While computationally less demanding, this approach may not capture novel variation in the genome of interest. Hybrid assembly incorporating both Illumina short-read sequencing and third-generation sequencing (e.g., reads 10–100 kb or longer using platforms including Oxford Nanopore Technologies or Pacific Biosciences single-molecule real-time sequencing, reviewed in Laver et al. 2015; Rhoads and Au 2015; Lu et al. 2016) greatly improve scaffolding, assembly of repetitive regions, and span gaps in short-read assemblies (Gordon et al., 2016; Tan et al., 2018). However, these new long-read technologies have higher sequencing error rates than short-read Illumina sequencing (error rates of 5–25% with Oxford Nanopore sequencing (Wick et al., 2018), 13–15% with PacBio platforms and ~0.1% with Illumina platforms (Ardui et al., 2018)) and thus assembly using these techniques may be more computationally challenging, but can be very effective when combined with high-coverage short-read sequencing to resolve errors (e.g., Walker et al. 2014). Genome size and the extent of repetitive regions and heterozygosity can increase the challenge of assembly, and with new sequencing techniques and assembly algorithms, there is potential for continual improvement (see the multiple versions of the model chicken genome since first assembly, International Chicken Genome Sequencing Consortium 2004; Warren et al. 2017). However, researchers should consider the downstream use of the genome to determine the assembly quality required to answer the questions of interest and avoid becoming entangled in ongoing assembly improvements with diminishing returns. Low coverage sequencing and reference-guided assembly may

represent an efficient, economical pathway to produce assembled genomes with sufficient quality for downstream analyses (Card et al., 2014), such as for producing robust estimates of metrics informative for conservation management.

The combined challenges of genome assembly and both perceived and actual costs (time, money, expertise) are limiting factors in the transition to a genomics-informed management approach for many threatened species as conservation programmes are typically resource-limited, and on-the-ground management may be prioritised over developing genomic resources (Allendorf et al., 2010; Primmer, 2009; Shafer et al., 2015). However, the costs of genome assembly have greatly declined (Check Hayden, 2014; Lewin et al., 2018) and bioinformatics expertise and genomic resources are expanding, providing increasing opportunities for conservation projects to harness these new approaches to assist in decision-making (McMahon et al. 2014; Ekblom et al. 2018; Sutton et al. 2018; Brandies et al. 2019; Wright et al. 2019; Galla et al. 2020; see Chapter Four and Appendix A). Nevertheless, there is a critical need for strong relationships between researchers and practitioners to ensure such genomic tools are implemented according to existing conservation needs.

2.2.3 Aims

In this chapter I will produce a high-quality *de novo* genome assembly for kakī, thus facilitating the transition from conservation genetic to conservation genomic management for this threatened species. In addition, I will produce a *de novo* genome assembly for Australian pied stilts that can be used for comparative analyses between these congeners. Genome sequencing will use the relatively cost-effective Illumina sequencing of a single paired-end library for each individual, representing an accessible approach for resource-limited conservation projects. The final assemblies will be used for estimating conservation-relevant metrics and to

make comparisons between these two closely related species, laying the foundation for understanding the underlying genomic impacts of interspecific hybridisation. As such, the assembly goal is for high-quality in terms of assembly contiguity and completeness, but a chromosomal-level assembly is not required. Furthermore, Galla et al. (2019, see Appendix A) used both genomes to assess the utility of genomes of closely related species for generating conservation genomic estimates of diversity, heterozygosity, and inbreeding. This aligned study aims to demonstrate that by using existing reference genomes, threatened species recovery programmes can more easily transition to implementing genomics in conservation management. This chapter describes the process of whole-genome sequencing and assembly for stilts, along with subsequent preliminary genome annotation and estimation of past demographic patterns.

2.3 Methods

2.3.1 Sample collection and DNA extraction

Blood samples from two female kakī (DNA IDs DNA1914 and DNA1929) were provided by the DOC's Kakī Recovery Programme captive breeding facility, Twizel, New Zealand (Figure 2.3, Table 2.1), collected by approval of the DOC Animal Ethics Committee (AEC #283). Based on kakī pedigree data, these individuals have no recorded history of hybrid ancestry in their pedigree data (four generations spanning up to 36 years). Female birds were preferentially selected for genome sequencing and assembly as females are heterogametic among birds, thus these samples should contain both Z and W sex chromosomes. Blood samples from Australian pied stilts were collected via brachial venepuncture during routine veterinary checks from a one-year old female (DNA ID B60406) and a six-year old

male (DNA ID B60480) held at Adelaide Zoo, South Australia, with samples provided under the Royal Zoological Society of South Australia Specimen Licence Agreement (Import Permit: 2016061954). Both individuals were offspring of birds wild-caught in the MacDonnell ranges, Northern Territories, Australia (Figure 2.4). A tissue sample was collected from a recently deceased poaka (DNA ID Poaka1) brought to an avian rehabilitation facility from the Clive River, Hawke's Bay, New Zealand (Figure 2.3), following an injury. I performed the following protocols for each of the five sampled individuals to produce sequence data for genome assembly and downstream analyses.

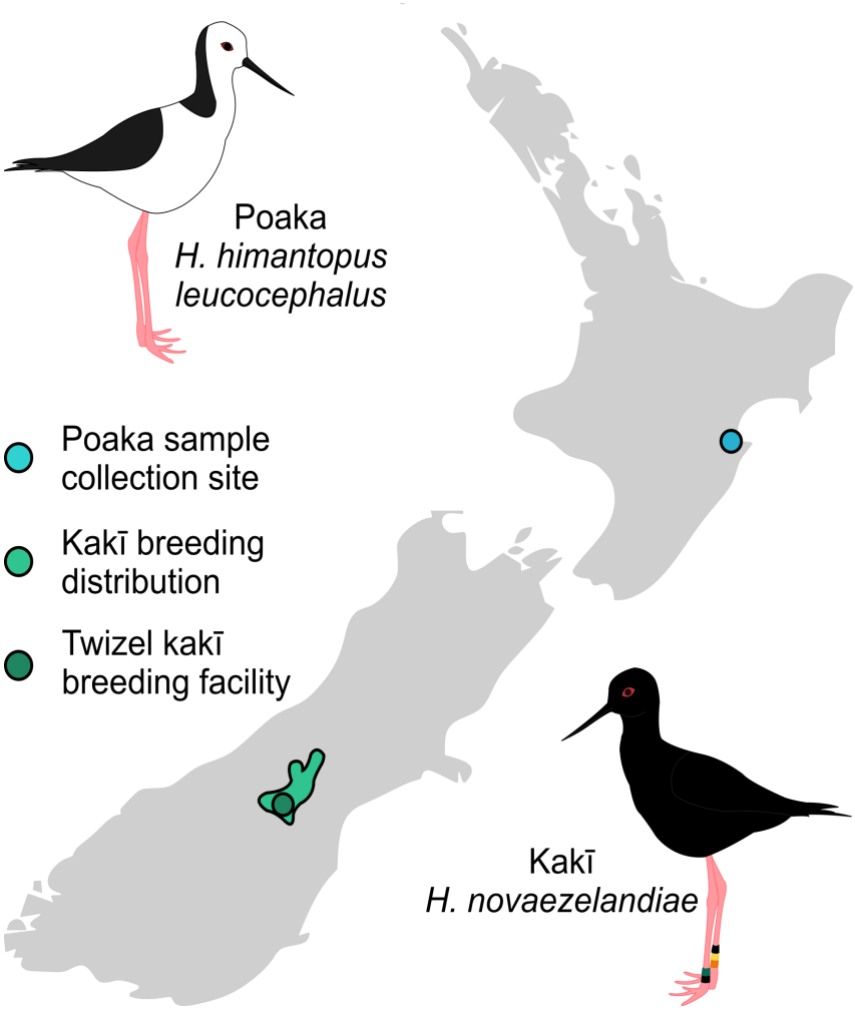


Figure 2.3: Location of kakī and poaka sampling sites in New Zealand and kakī breeding distribution in Te Manahuna/Mackenzie Basin. Kakī blood samples were provided by the

Department of Conservation's Kakī Recovery Programme, based at the Twizel breeding facility. The poaka individual was collected from the Clive River, Hawke's Bay.

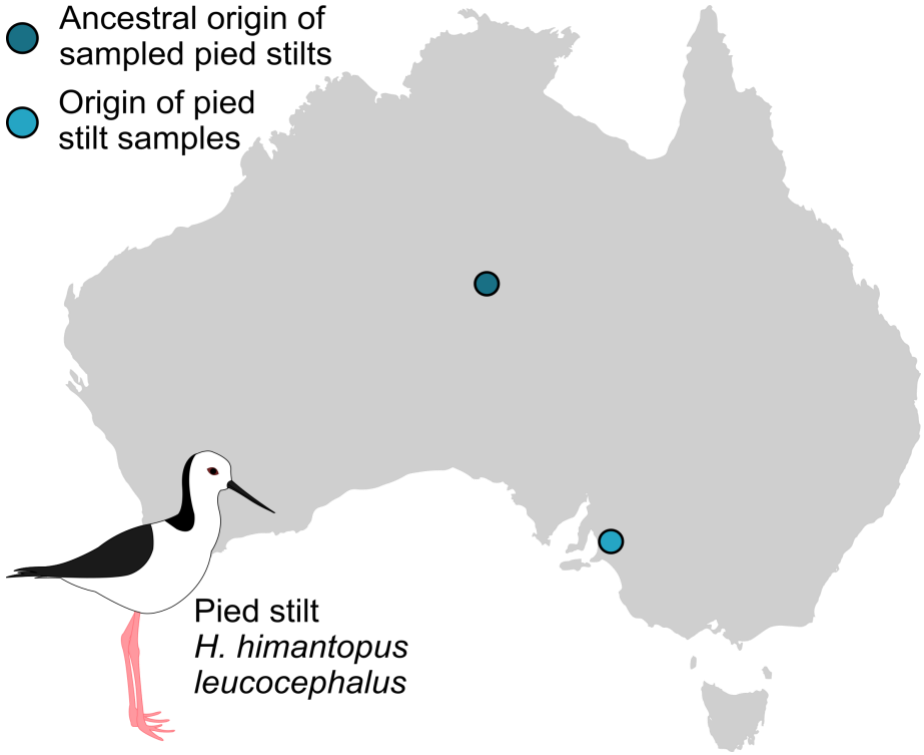


Figure 2.4: Origin of Australian pied stilt samples, indicating the origins of sampled individuals in the MacDonnell Ranges, Northern Territories, and the current location of individuals at Adelaide Zoo, South Australia.

Table 2.1: Details of five stilts sampled for whole-genome sequencing. APS = Australian pied stilt, M = male, F = female, U = unknown.

DNA ID	Species (common name)	Sampling location	Sex
B60480	*Himantopus himantopus leucocephalus* (APS)	Adelaide Zoo, Australia	M
B60406	*H. h. leucocephalus* (APS)	Adelaide Zoo, Australia	F
DNA1914	*H. novaezelandiae* (Kakī)	Twizel, New Zealand	F
DNA1929	*H. novaezelandiae* (Kakī)	Twizel, New Zealand	F
Poaka1	*H. h. leucocephalus* (Poaka)	Hawke's Bay, New Zealand	U

I used a DNeasy® Blood and Tissue kit (QIAGEN) for initial extractions of gDNA for B60406 and B60480, with the whole blood extraction protocol. Briefly, I added approximately 20 µL of blood to a 2 mL microcentrifuge tube containing 20 µL Proteinase K and 180 µL PBS, before adding 200 µL Buffer AL. I vortexed samples

and incubated them at 56°C for ten minutes, vortexing occasionally, before adding 200 µL EtOH (100%). I then transferred samples to DNeasy Mini spin columns placed in collection tubes and centrifuged them at 8000 rpm for one minute. I discarded the flow-through and added 500 µL Buffer AW1 to the spin columns before centrifuging and discarding the flow-through again. I then added 500 µL Buffer AW2 to the samples and centrifuged them at 14,000 rpm for three minutes to allow the spin column membrane to dry. I then placed the spin column in a new 2 mL microcentrifuge tube, and added 200 µL Buffer AE to the membrane, before incubating the samples at room temperature for one minute. Samples were centrifuged at 8000 rpm for one minute to elute the purified DNA, and repeated these last elution steps. I quantified DNA extractions with a NanoDrop™ 8000 Spectrophotometer. Initial extractions with this method resulted in low DNA concentrations insufficient for the library preparation protocols for high-throughput sequencing.

I then tested a Thermo Scientific™ MagJET™ Genomic DNA kit, following Protocol E (manual genomic DNA purification from up to 20 mg tissue). For this DNA extraction protocol, I suspended whole blood in 200 µL of digestion solution, added 20 µL Proteinase K, and vortexed briefly before incubating the samples at 56°C for 10 minutes, with occasional vortexing. Following incubation, I added 300 µL lysis buffer and vortexed samples briefly before transferring the lysate to a tube pre-filled with 400 µL isopropanol and 25 µL of magnetic bead suspension. After vortexing, I placed the tubes on the magnetic rack to allow the magnetic beads to collect against the magnet for three minutes. I removed the supernatant and added 800 µL wash buffer 1, mixing gently before returning the sample to the magnetic rack. After two minutes, I removed the supernatant was removed, and added 800 µL wash buffer 2, mixing gently before returning the sample to the magnetic rack, where after two

minutes, I removed the supernatant. This second wash step was repeated, and with all remaining supernatant carefully removed. I then added 200 µL elution buffer and warmed the samples for five minutes at 72°C. I returned the samples to the magnetic rack for two minutes before transferring the eluate containing purified DNA to a new clean tube. This elution step was repeated to produce two elutions from each sample. I assessed the quality of extracted gDNA by gel electrophoresis of 5 µL of DNA on a 1% agarose gel run at 100 V for 45 minutes. A distinct band of high molecular weight was obtained for each sample. I then quantified samples on both a NanoDrop™ 8000 Spectrophotometer and via a Qubit dsDNA High Sensitivity Assay kit on a Qubit® 2.0 Fluorometer (Invitrogen). DNA concentrations extracted with the MagJET protocol were higher than for the DNeasy protocol, and so the MagJET protocol was used for all subsequent extractions.

I prepared Illumina sequencing libraries for the five samples using the TruSeq DNA PCR-free library prep protocol, according to manufacturer specifications. This involved normalisation of gDNA to a concentration of 20 ng/µL, and fragmentation to ~350 bp using a Covaris S220 focused-ultrasonicator (duty factor of 5%, 175 W peak power, 200 cycles/burst for 50 s at 6°C). I conducted PCR amplifications for end repair, universal adapter ligation for sequence barcoding, and indexing with a unique 6 bp sequence.

I confirmed the quantity and quality of the produced libraries via Qubit and ran them on an Agilent 2100 BioAnalyzer high sensitivity gDNA chip to estimate average fragment size. I used qPCR for further quantification following 1:10,000 and 1:20,000 dilutions compared against six DNA standard dilutions. The qPCR mix consisted of 10 µL SYBR Master Mix, 0.2 µL of each 10 nM primer, 7.6 µL water, and 2 µL template DNA. For each sample, I included three replicate 1:10,000 and 1:20,000 dilutions of DNA as templates. I loaded three replicates of the KAPA library

standards diluted to 20, 2, 0.2, 0.02, 0.002, and 0.0002 pM, along with three replicate no-template controls. qPCR thermocycling was conducted on an MxPro 3000P (Stratagene) and consisted of a hot start of 95°C for 10 s, forty cycles of denaturation at 95°C for 15 s and annealing at 60°C for one minute, and melt-curve analysis of 65–96°C. I then pooled sample libraries, which were then split across five lanes of a flow cell for paired-end 2 x 125 bp sequencing on an Illumina HiSeq 2500 with V4 chemistry, following a preliminary MiSeq test run, at the Otago Genomics and Bioinformatics Facility (OGBF; formerly New Zealand Genomics Limited).

2.3.2 Data quality control and estimation of assembly metrics

Sequenced paired-end reads were demultiplexed by OGBF using their own in-house pipeline. Scripts associated with the following workflows are detailed in my GitHub 'Himantopus' repository at https://github.com/natforsdick/Himantopus. I assessed raw paired-end sequence quality using FastQC version 0.11.5 (Andrews, 2010). To test for exogenous contamination, I randomly subsampled 5,000 reads from each library and performed BLAST (Altschul et al., 1990) searches against the NCBI nucleotide database to identify non-avian sequences. I conducted sequence cleaning in three steps. First, I removed the Illumina adapters used for sequence barcoding using Trimmomatic v0.35 (Bolger et al., 2014). I trimmed low quality bases using ConDeTri v2.3 (Smeds & Künstner, 2011) with default settings (high quality threshold = 25, low quality threshold = 10, fraction of read that must exceed high quality threshold after trimming = 0.8, maximum fraction of bases with quality below the low quality threshold after quality trimming = 0, trimming stops when five consecutive high quality bases are reached, one low quality base allowed within a stretch of high quality bases, minimum sequence length = 50). I then deduplicated reads with the ConDeTri Perl script *filterPCRdupl.pl*, using the first 50 bp of both

reads in a pair for comparisons. I assessed the quality of cleaned sequences with FastQC for comparison with raw sequence quality. In an additional contamination check, I mapped a subset of 5,000 cleaned sequences from each individual against mitochondrial genomes of potential lab contaminants (reagent contaminants and species previously handled in the lab), namely human (*Homo sapiens*), wolf (*Canis lupus*), chicken (*Gallus gallus*), wild boar (*Sus scrofa*), domestic cattle (*Bos taurus*), black rat (*Rattus rattus*), kea (*Nestor notabilis*), and guinea pig (*Cavia porcellus*), with an in-house pipeline.

To assess *k*-mer content and estimate genomic heterozygosity, I passed trimmed paired-end reads to KmerGenie (Chikhi & Medvedev, 2014) and Jellyfish v2.2.7 (Marçais & Kingsford, 2011). *k*-mers are unique sequences with length *k*, and estimates of *k*-mer value that maximise the number of unique *k*-mers are used in de Bruijn graph-based assemblies. KmerGenie was run in diploid mode across *k*-mers 19–99. Jellyfish was used to count *k*-mers, with outputs passed to the web application GenomeScope (Vurture et al., 2017) to visualise estimated genome assembly metrics including genome size, heterozygosity, and sequence duplication levels. I visualised additional assembly metrics with the SGA-preqc v0.9.4 pipeline (Simpson, 2014).

2.3.3 Genome assembly

Using the whole-genome short-read paired-end sequence data generated, I tested four *de novo* genome assembly pipelines built for assembling short-read paired-end sequences: ABySS v2.0.2 (Simpson et al., 2009), DiscovarDeNovo (Weisenfeld et al., 2014), Meraculous v2.2.4 (Chapman et al., 2017), and SOAPdenovo2 (Luo et al., 2012). All assemblies were performed using a local (University of Otago) computing cluster with 32 cores and 1 TB RAM. Preliminary assembler testing used the

sequenced reads for DNA1914 as input. Parameters were optimised over multiple assembler tests to generate improved assemblies. Trimmed sequence reads were passed as input to all assemblers except DiscovarDeNovo which takes raw sequence reads as input and performs sequence cleaning independently.

Assembly tests with ABySS were conducted in paired-end mode, with three tests using *k*-mers of 31, 35, and 49. DiscovarDeNovo required no additional parameters. Kakī assembly with Meraculous specified an average insert size of 375 bp, estimated standard deviation in insert size of 75 bp, average read length of 125 bp, standard orientation of sequence reads, all reads specified for use in contiging, scaffolding and gap closing, adapter- and quality-trimmed reads specified as input, estimated genome size of 1.2 Gb (based on SGA-preqc results, genome size of other Charadriiformes, and estimates from Gregory (2001), *k* = 49, diploid mode, with no minimum depth cut-off. I tested a range of *k*-mer values in SOAPdenovo2, but other parameters remained the same throughout assembly testing.

The final parameters for SOAPdenovo2 assembly of kakī DNA1914 used paired-end mode with *k* = 49, minimum contig length for scaffolding of 200 bp, expected genome size of 1.2 Gb, maximum read length of 125 bp, an average insert size of 385 bp, pair number cut-off value of 3 for accurate connections, reads were specified for use in both contig and scaffold assembly, and a minimum alignment length of 32 bp. I tested an alternative approach for assembling B60406, B60480, and Poaka1 to handle the higher levels of heterozygosity observed in preliminary quality control assessment for these genomes. I interleaved forward and reverse reads with Khmer v2.0 (Crusoe et al., 2015), and conducted digital normalisation of the interleaved reads (Brown et al., 2012). I then filtered reads by abundance (Zhang et al., 2017) with a minimum *k*-mer coverage of 3. The normalised, filtered reads were then passed to VelvetOPTIMISER v2.2.6 (Zerbino & Birney, 2008) to test *k* = 21–121 with

a size step of *k* = 6. The B60406 interleaved reads were then passed to the Velvet v1.2.08 (Zerbino & Birney, 2008) assembly pipeline, with *k* = 15–29 and a step-size of *k* = 6, expected coverage of 12×, minimum coverage of 4×, maximum coverage of 50×, and an insert-size of 350 bp. For Poaka1, *k* = 21–61 was tested with step-size of *k* = 4 in VelvetOPTIMISER, and interleaved reads were passed to Velvet, where *k* = 11–35 were tested with a step-size of *k* = 4, with other parameters the same as for B60406.

The final, highest-quality assembly for Australian pied stilt B60406 was produced with SOAPdenovo2, using paired-end mode with the following parameters: *k* = 43, minimum contig length for scaffolding of 200 bp, estimated genome size of 1.2 Gb, average insert size of 370 bp, pair number cut-off value of 3 for accurate connections, reads were used in both contig and scaffold assembly, and minimum alignment length of 32 bp. Despite several attempts to produce a high-quality assembly for Poaka1, all approaches produced very fragmented assemblies compared with those of kakī DNA1914 and Australian pied stilt B60480, likely due to sequence error or high heterozygosity. For the remainder of the thesis the Australian pied stilt genome assembly is used as a proxy for poaka.

2.3.4 Assembly quality assessment

I assessed all assemblies with the Assemblathon2 Perl script developed for the Assemblathon2 genome assembly challenge comparing current genome assembly methods (Bradnam et al., 2013). Although most assemblers produce details of assembly metrics, using one independent script to assess all genomes provides greater confidence in robust, unbiased assessments. To validate completeness of the assembly outputs, I assessed assemblies with BUSCO v3.0.1 (Simão et al. 2015; Waterhouse et al. 2018) assessment based on expected gene content from

the database of 2,586 orthologous protein-coding genes for vertebrates derived from OrthoDB v9 (Zdobnov et al., 2017). Final assemblies were benchmarked against the 4,915 orthologous protein-coding genes for birds. I considered both Assemblathon2 metrics and BUSCO results when comparing assembly quality.

2.3.5 Assembly enhancement

To increase assembly contiguity, I passed the highest quality assembly from initial testing with DNA1914 to SOAPdenovo2's GapCloser v1.12 along with the cleaned sequence reads. Following gap-closing, I removed contigs shorter than 5 kb, and syntenically aligned the remaining scaffolds against the chicken genome (*Gallus gallus* v5.0, GenBank Assembly ID GCA_000002315.3) using the Satsuma2 *Chromosemble* pipeline (Grabherr et al., 2010) to generate a superscaffolded assembly. Chromosemble aligns two genomes, and only retains orthologous regions. I then assessed the superscaffolded assemblies with the Assemblathon2 script and BUSCO using the set of 4,915 avian-specific orthologs. This superscaffolded assembly was superior at all metrics to the initial assembly, and so I repeated the process for the highest quality Australian pied stilt B60480 assembly. I removed contigs mapping to the chicken mitochondrial genome from the final assemblies as these sequences were thought to represent true mitochondrial sequences. I soft-masked repeat sequences identified with RepeatMasker v4.0.7 (Smit et al., 1996) using the chicken as reference. I identified putative sex chromosomes (Z and W in birds, where males are ZZ and females are ZW) through BLAST alignments of all scaffolds against the chicken sex chromosomes. The BLAST+ v2.7.1 (Altschul et al., 1990) *blastn* search used a word size of 15, an e-value of $1x10^{-10}$, and outputs were filtered for ≥ 80% identity across a minimum 1 kb.

2.3.6 Enhanced assembly quality assessment

I performed additional quality assessment with QUAST v4.0 (Gurevich et al., 2013) and REAPR v1.0.16 (Hunt et al., 2013), and conducted further validation by mapping species-specific trimmed sequence reads to the final genome assemblies (e.g., DNA1914 and DNA1929 reads were mapped to the kakī genome assembly) with the Burrows-Wheeler Aligner v0.7.17 (BWA; Bayat et al., 2017; Li & Durbin, 2009) and Picard v2.1.0 (Picard Toolkit, 2019) in a custom pipeline. This pipeline aligned paired-end reads to the reference genome assembly with BWA-MEM. The output was then sorted with Picard's *SortSam*, and indexed with SAMtools v1.9 (Li et al., 2009). Sequence and alignment metrics were extracted with Picard *CollectInsertSizeMetrics*, and SAMtools *stats* and *idxstats*. The resulting BAM files were then cleaned to fix errant MAPQ scores and soft-clip overhanging reads with Picard *CleanSam*. Unmapped reads were filtered with SAMtools *view*, and duplicates were removed with Picard *MarkDuplicates*. Mean mapped genome coverage depth and proportions for the output BAM files were extracted with SAMtools *depth* and *flagstat* tools. I then conducted variant discovery for each species with a consensus file first generated from the species-specific BAM files with BCFtools v1.9 *mpileup* and *call* pipelines to produce a Variant Call File (VCF) for each species. Variants were subsequently filtered to exclude indels and non-biallelic SNPs to produce comparable data sets.

2.3.7 Estimating demographic history

I conducted preliminary assessment of demographic change over time through pairwise sequentially Markovian coalescent (PSMC) analysis (Li & Durbin, 2011) for all samples excluding Poaka1. PSMC estimates effective population size (N_e) over time, using rates of coalescent events for a single diploid genome. The use of PSMC

was considered best-practice when these analyses were conducted, although MSMC (multiple sequentially Markovian coalescent; Schiffels and Durbin 2014) is now recommended, however this approach requires producing phased haplotypes for the individuals of interest. This is beyond the scope of this thesis, where the goal is to produce genomes sufficient for use as references to generate conservation-relevant metrics.

To generate high coverage whole-genome diploid consensus files for each of the four individuals based on the species-specific BAM files generated in the previous mapping steps, I used SAMtools *mpileup* and BCFtools *call* tools, followed by SAMtools *vcfutils* to filter and convert the consensus to a FASTQ file with a minimum read depth of 10 and a maximum of 100. These outputs were then passed to PSMC independently, with parameters based on those used for the congeneric killdeer in (Nadachowska-Brzyska et al., 2015) of -N30 (number of rounds) -t5 (upper limit of time to most recent common ancestor) -r5 with 84 atomic time intervals and 34 free intervals (-p "4+30∗2+4+6+10"). I manually confirmed results to ensure that at least ten recombinations were counted in each set of intervals from round twenty onwards. To assess variance in the estimates of N_e produced, I performed 100 bootstrap tests for each individual. I visualised results with the PSMC in-built Perl script *psmc_plot.pl*, excluding results more recent than 10,000 ya during which estimates become unreliable (Li & Durbin, 2011). The estimated generation time for kakī was used for all individuals (6 years, DOC, pers. comm.), and a mutation rate of 7.38 x 10^{-9} substitutions per site per generation, based on the lower avian mutation rate of 1.23 x 10^{-9} substitutions per site per year estimated in (Nam et al., 2010), where the mean avian mutation rate is estimated at 1.23×10^{-9}–2.21×10^{-9} substitutions per site per year (although both rates were compared with little difference in final outcomes; results not shown).

2.4 Results

2.4.1 DNA extraction and whole-genome sequencing

The MagJET DNA extraction protocol produced higher quantities of gDNA extracted than the DNeasy extraction kit, with extractions ranging in concentration from 7.1–57.5 ng/μL from the DNeasy kit compared with 18.7–254.3 ng/μL from the MagJET protocol. TruSeq library preparation produced a pool of the five libraries with a gDNA concentration of approximately 9.1 nM. Illumina HiSeq 2500 sequencing combined with data produced from the MiSeq test run produced a total of 2.46 billion paired-end reads, with 452.98–534.41 million sequences per individual, for an average of 61.55 ± SD 3.80 Gb sequence data per individual (Table 2.2). With an estimated size of 1.2 Gb based on the genome assembly of the conordinal killdeer (1.22 Gb; Zhang et al. 2014) and confamilial pied avocet (*Recurvirostra avosetta*, 1.02 Gb; Galla et al. 2019), this equates to approximately 47–56× sequencing coverage per individual.

Table 2.2: Raw and cleaned sequence quality metrics for five stilt samples. bp = base pairs fq1 = forward reads, fq2 = reverse reads, Q = base quality Phred score.

DNA ID	Sequence ID	Total sequences	Length (bp)	% GC content	% ≥ Q30	Total cleaned sequences	Cleaned % remaining	Cleaned sequence length (bp)	Cleaned % GC content	Cleaned % ≥Q30
B60480	Aus1_all_fq1	226,490,412	125	42	94.07	194,635,913	85.94	50–125	42	99.95
	Aus1_all_fq2	226,490,412	125	42	94.07	194,635,913	85.94	50–125	41	99.87
B60406	Aus2_all_fq1	253,473,713	125	42	93.64	218,219,053	86.09	50–125	42	99.94
	Aus2_all_fq2	253,473,713	125	42	89.74	218,219,053	86.09	50–125	41	99.89
DNA1914	Kaki1_all_fq1	243,342,517	125	42	93.90	210,489,519	86.50	50–125	42	99.95
	Kaki1_all_fq2	243,342,517	125	42	89.50	210,489,519	86.50	50–125	42	99.88
DNA1929	Kaki2_all_fq1	240,540,638	125	42	94.43	208,766,541	86.79	50–125	41	99.95
	Kaki2_all_fq2	240,540,638	125	41	89.64	208,766,541	86.79	50–125	41	99.87
Poaka1	Poaka_all_fq1	267,207,157	125	45	91.63	215,057,780	80.48	50–125	45	99.94
	Poaka_all_fq2	267,207,157	125	45	84.94	215,057,780	80.48	50–125	45	99.87

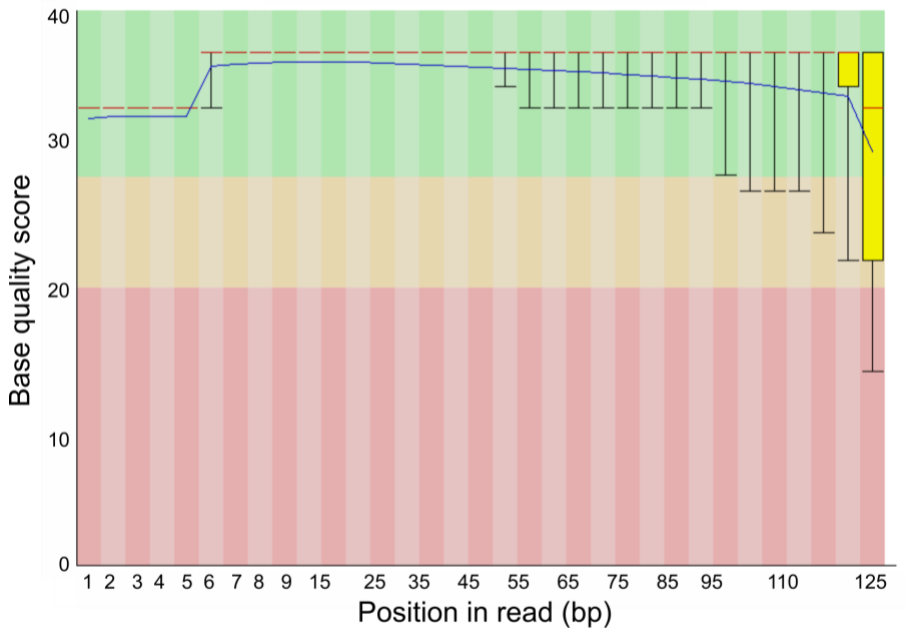
2.4.2 Data quality control and estimation of assembly metrics

FastQC statistics for the raw reads found an average 92.35% of all sequences achieved quality scores ≥ 30 (Figure 2.5, Table 2.2, Table 2.3; Appendix B for additional QC information). GC content displayed the typical skewed pattern observed in avian genomes (Figure 2.6; Nabholz et al. 2011; Hartono et al. 2015). Following adapter removal, quality trimming, and read deduplication, on average 85.16% of sequences were retained (Figure 2.7, Table 2.2), with 99.91% of cleaned sequences achieving quality scores ≥ 30. The contamination checks for raw reads produced no hits to the NCBI database. Mapping of cleaned reads to mitochondrial genomes of potential lab contaminants produced only two sequence matches to any mitochondrial genomes—one to chicken and one to guinea pig, both from DNA1914.

Table 2.3: Estimated genome assembly metrics based on raw sequence data with SGA-preqc and Jellyfish/GenomeScope. Gb = gigabase pairs.

DNA ID	Genome size (SGA; Gb)	Genome size (Jellyfish; Gb)	Heterozygosity (%)	Read error (%)	Duplication (%)
B60480	1.181	1.149	0.646	0.120	0.208
B60406	1.250				
DNA1914	1.201	1.147	0.349	0.121	0.234
DNA1929	1.187				
Poaka1	0.785	0.631	2.36	0.821	0.929

a)



b)

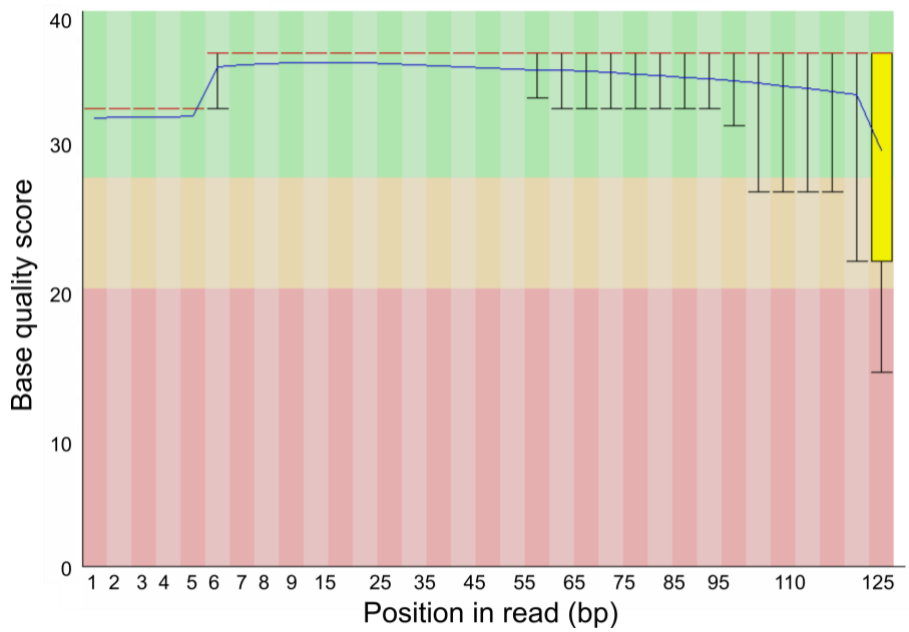


Figure 2.5: Examples of the raw per-base sequencing quality scores for a) kakī DNA1914 forward reads and b) Australian pied stilt B60480 forward reads. Sanger/Illumina 1.9 encoding.

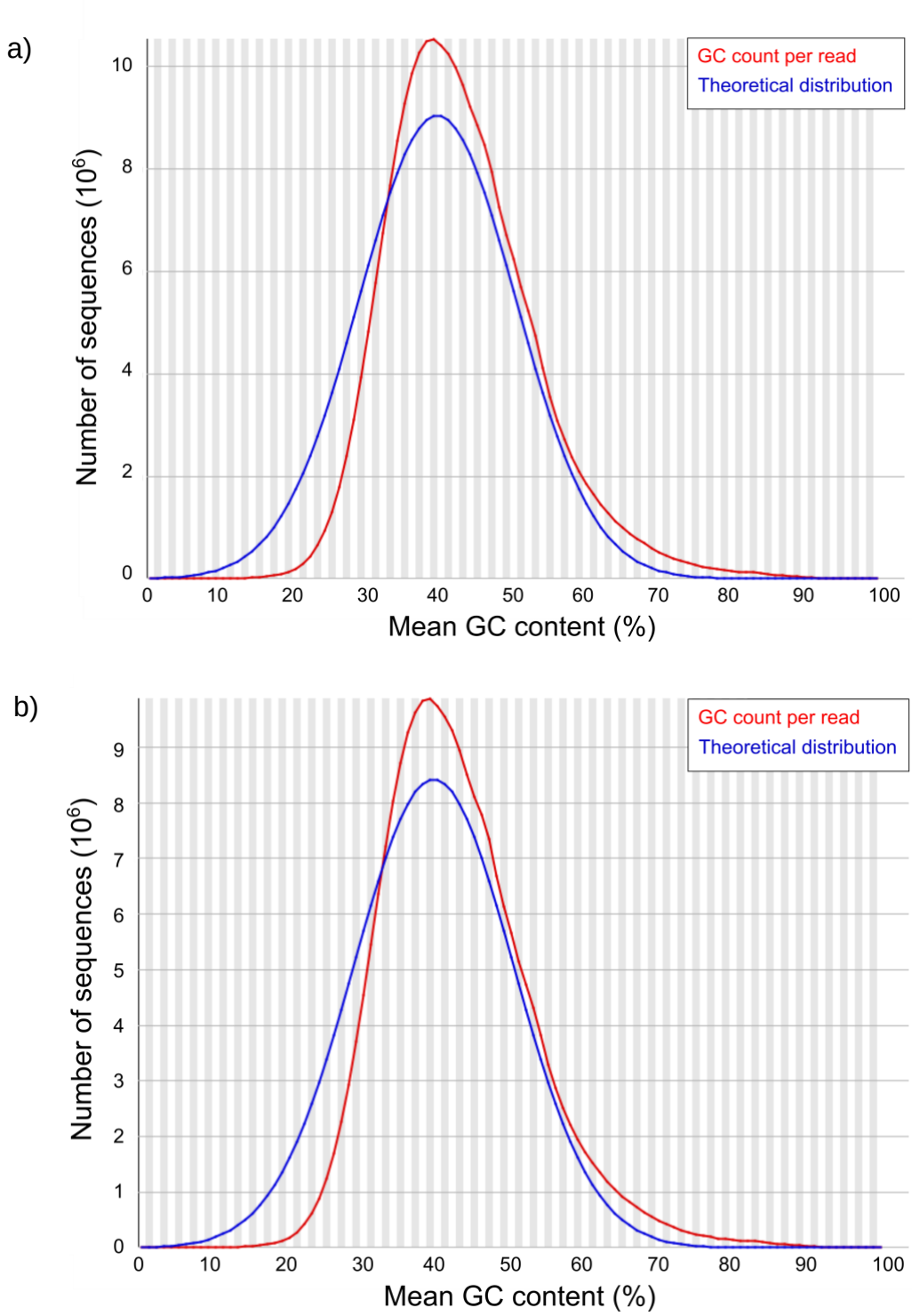


Figure 2.6: GC distribution over all raw sequences for a) kakī DNA1914 forward reads and b) Australian pied stilt B60480 forward reads.

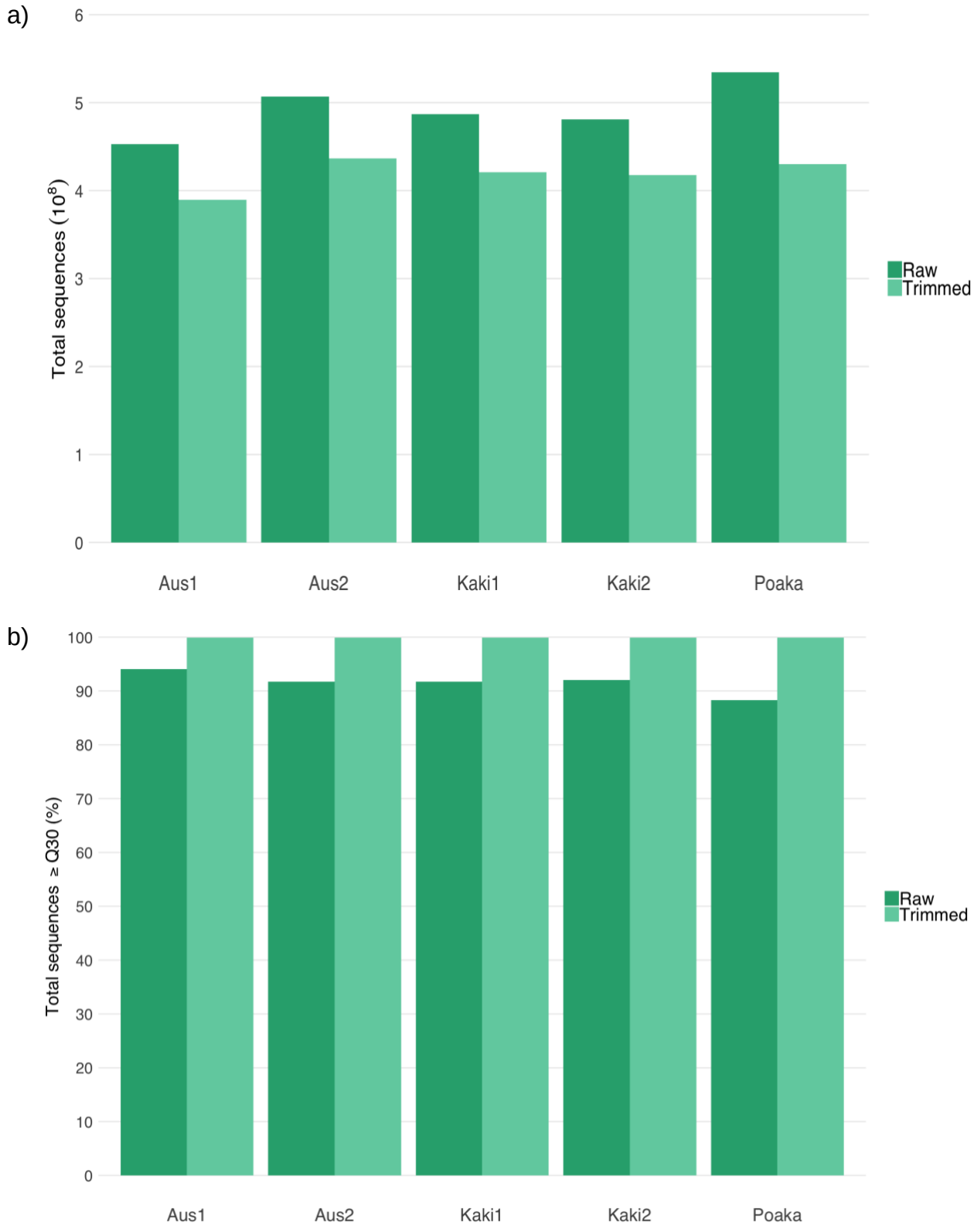
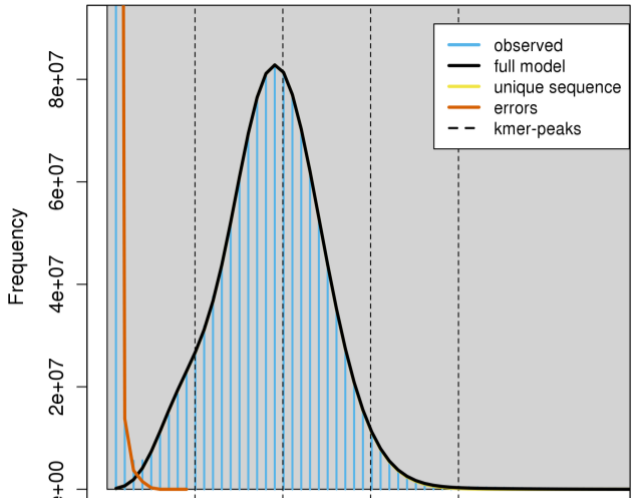


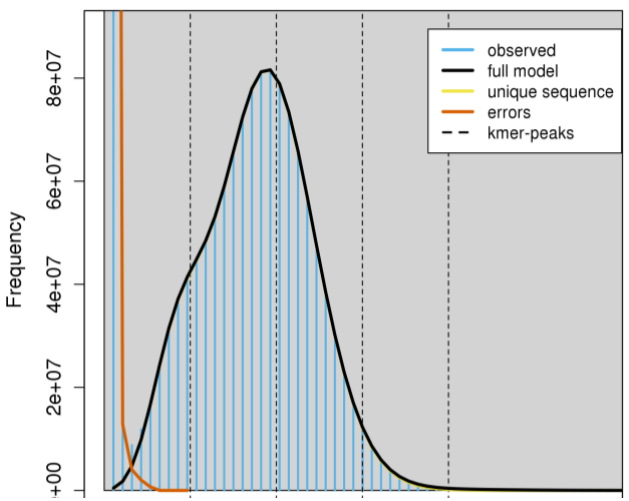
Figure 2.7: Comparison of raw and trimmed a) sequence number and b) average sequence quality for five stilt individuals: Aus1 = Australian pied stilt B60480, Aus2 = Australian pied stilt B60406, Kaki1 = kakī DNA1914, Kaki2 = kakī DNA1929, Poaka = poaka Poaka1. Q = Phred based quality score.

GenomeScope visualisations of Jellyfish results for DNA1914 and B60480 were similar, with an estimated genome size of approximately 1.14 Gb (Figure 2.8, Table 2.3). Heterozygosity was estimated at 0.349% for DNA1914 and 0.646% for B60406. Duplication rates were similar between these two individuals, estimated at 0.23% for DNA1914 and 0.21% for B60406. Error rates for both DNA1914 and B60406 were estimated at 0.12%. SGA-preqc reports for raw data produced no obvious differences with those produced from trimmed data; all results reported here are from raw data (see Appendix B for pre-assembly reports). SGA-preqc assessment predicted a genome size of ~1.2 Gb for kakī and Australian pied stilts (Table 2.3), similar to that of the killdeer genome. Data from all individuals exhibited the expected avian GC content. Estimated read duplication was below 0.04% for all individuals. The frequency of variant branches indicated lower heterozygosity among kakī than Australian pied stilts, likely due to the population decline kakī have experienced. Results from Poaka1 indicated low sequence coverage, and underestimated the genome size at 0.8 Gb, likely due to the high levels of heterozygosity or sequence error combined with a high frequency of error branches identified by SGA-preqc and GenomeScope visualisation of Jellyfish *k*-mer counts (Figure 2.8). Kakī DNA1914 displayed the lowest estimates of heterozygosity of all individuals, and so was selected for preliminary genome assembly testing, as greater heterozygosity increases complexity of assembly (e.g., Tigano et al. 2018). Of the two Australian pied stilt stilts, B60406 was female, and displayed lower estimates of heterozygosity, error frequencies, and duplication, and higher read quality, and so was selected for representative genome assembly.

a)



b)



c)

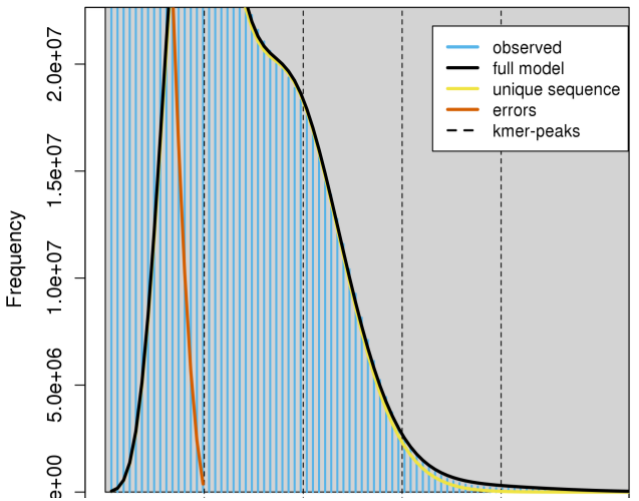


Figure 2.8: GenomeScope visualisation of Jellyfish results describing characteristics of whole-genome sequencing data for a) kakī DNA1914, b) Australian pied stilt B60480, and c) poaka Poaka1. Len = length, bp = base pairs, dup = duplication, err = error, het = heterozygosity, kcov = k-mer coverage, k = k-mer.

2.4.3 Genome assembly and assessment

The four assemblers tested had run-times ranging from 11 (SOAPdenovo2) to > 90 real-time hours (ABySS) per assembly as implemented on the local computing cluster. In total, 34 assembly runs with varying parameters were conducted across the four assemblers using DNA1914 sequence data, most testing SOAPdenovo2 with varying *k*-mer values. Preliminary results from assembler tests indicated that while all four assemblers produced assemblies around the target genome size, quality was low. Here, quality encompasses overall genome size, contiguity of assembly (a lower number of scaffolds, higher scaffold N50, and a lower proportion of unknown (N) bases are superior), and presence of complete single-copy gene orthologs (Figure 2.9). In addition, eight assembly trials were conducted for each of B60480 and Poaka1 across different assemblers.

Through testing DiscovarDeNovo with both raw and trimmed DNA1914 sequencing reads, I confirmed that it performed better with raw sequence reads than cleaned reads. DiscovarDeNovo was the most 'black box'-like assembler, providing no opportunity to adjust parameters or assess the read trimming included in the pipeline. While assembly was computationally efficient (average run time was 18 hrs), contiguity was poor (scaffold N50 of 43,607 bp, mean scaffold length of 3,875 bp, and 325,817 scaffolds), a low proportion (13.27%) of contigs were captured in scaffolds, and BUSCO analysis indicated only 74% of gene orthologs were present when the raw sequences were used as input (Table 2.4).

Assembly tests with Meraculous produced long contigs and scaffolds (longest contig 437,816 bp, longest scaffold 519,129 bp). However, there were large gaps in the assembly (average length of breaks between scaffolded contigs was 167 bp), only 32% of contigs were incorporated in scaffolds, and the final assembly spanned only 1.075 Gb. Complete, single-copy gene orthologs were present for 82.6% of the

vertebrate database. Assembly with ABySS was computationally slow, running for upwards of 90 hours, producing highly fragmented assemblies (up to 1,972,559 scaffolds with mean length 628 bp) that included a maximum of only 80.7% of the expected single-copy vertebrate orthologs.

Overall, SOAPdenovo2 produced the highest quality assemblies of those tested. The highest quality SOAPdenovo2 assembly for kakī DNA1914 (using *k* = 49) comprised 148,989 scaffolds with mean length 8,315 bp, a longest scaffold of 1,028,732 bp, and scaffold N50 of 119,990 bp, spanning a total length of 1.239 Gb (Table 2.4). Breaks between scaffolded contigs were relatively short (average 85 bp), and 97.8% of contigs were captured in scaffolds. Of the vertebrate set of orthologs, 92% were present, and in the final assessment, 88.3% of the 4,915 avian orthologs were present, compared with 81.3% in the highest quality Meraculous assembly, and 74.0% in the highest quality DiscovarDeNovo assembly. Following similar testing with B60480 sequence reads, the final SOAPdenovo2 assembly for Australian pied stilt B60480 consisted of 816,713 scaffolds totalling 1.28 Gb in length, with a scaffold N50 of 34,001 bp. BUSCO analysis detected 75.9% of the expected complete single copy orthologs in the avian database. Despite the additional steps taken, the Velvet assembly pipeline trialled for B60480 and Poaka1 did not produce results that improved on the best SOAPdenovo2 assembly for B60480, and could not resolve an assembly for Poaka1, thus the Australian pied stilt assembly for B60480 is used as a proxy reference genome for conspecific poaka.

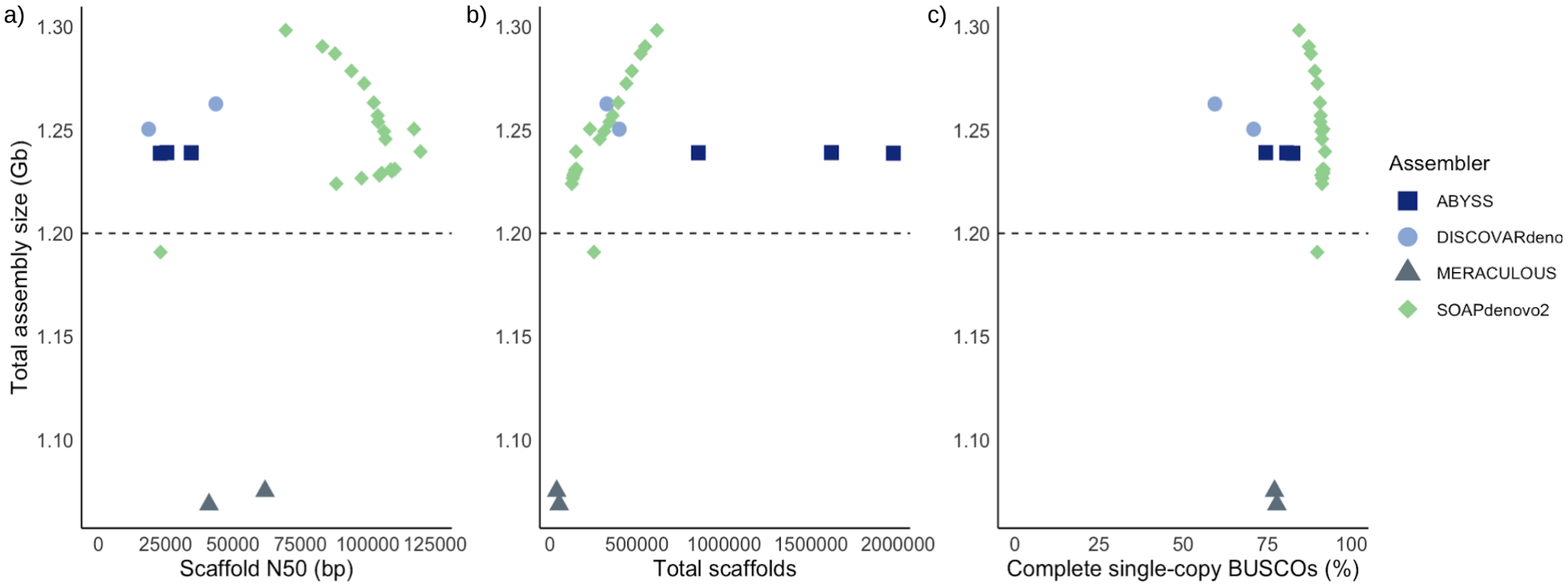


Figure 2.9: Preliminary results of four different genome assemblers in output assembly size and a) scaffold N50 length, b) total number of scaffolds, and c) presence of complete single-copy gene orthologs from the avian database. The dashed horizontal line indicates the estimated genome size of 1.2 Gb.

Table 2.4: Metrics of genome assemblies produced with various assemblers during the testing phase for kakī DNA1914 and Australian pied stilt B60480. Only the highest quality assembly produced with each assembler for each individual is reproduced here. BUSCO analysis is against the vertebrate database (VDB) or avian database (ADB). When assembly contiguity was deemed poor, BUSCO analysis was not conducted due to the required run-time. Gbp = gigabase pairs, bp = base pairs.

	Kakī (DNA1914)				Australian pied stilt (B60480)			
Assembler	**ABySS**	**DiscovarDeNovo**	**Meraculous**	**SOAPdenovo2**	**DiscovarDeNovo**	**Meraculous**	**SOAPdenovo2**	**Velvet**
Total scaffolds	852,124	325,817	37,769	148,989	1,369,404	270,899	816,713	673,989
Assembly size (Gbp)	1.239	1.2627	1.0754	1.2389	1.4006	1.1097	1.2757	1.2419
Longest scaffold (bp)	355,052	510,656	519,142	1,028,732	76,641	349,000	743,085	500,488
Mean scaffold length (bp)	1,454	3,875	28,472	8,315	1,023	4,096	1,563	1,843
Scaffold N50 (bp)	34,465	43,607	62,404	119,990	2,575	5,358	34,001	35,112
Scaffold L50	9,707	7576	5,005	2,821	109,642	42,166	9,650	9,782
Scaffold %N	0.16	0.26	0.15	3.12	0	0.21	3.12	0.75
Proportion of assembly in scaffolded contigs (%)	52.2	63.7	32.5	97.8	0	11.6	89.4	79.3
Mean break between contigs in scaffolds (> 25 Ns; bp)	58	100	170	84	0	125	63	82
Total contigs	887,800	358,821	46,583	594,827	1,369,404	288,142	1,415,249	766,485
Longest contig (bp)	260,003	196,825	437,821	122,633	746,641	296,798	175,002	239,595
Mean contig length (bp)	1,403	3,510	23,053	2,020	1,023	3,844	875	1,610
Contig N50 (bp)	22,540	16,628	51,223	10,746	2,575	5,145	3,869	14,881
Contig L50	15,184	21,213	6,059	31,271	109,642	44,493	69,832	23,549
Complete single-copy BUSCOs (%, VDB)	80.7	75.5	82.9	92				
Complete duplicated BUSCOs (%, VDB)	0.5	0.5	0.5	0.5				
Missing BUSCOs (%, VDB)	3	5.5	6.1	1.4				
Complete single-copy BUSCOs (%, ADB)		74	81.3	88.3			78.7	
Complete duplicated BUSCOs (%, ADB)			0.8	1.2			0.8	
Missing BUSCOs (%, ADB)		11.1	7.7	3.5			8.1	

2.4.4 Genome assembly enhancement and additional quality assessment

The initial DNA1914 assembly gap count was 552,385 gaps, covering a total 38,693,531 bp (3.12% of the genome). After gap filling, 438,757 gaps remained, spanning 25,008,481 bp (64.63% of the prior total, representing 2.13% of the genome). For B60406, the initial gap count was 758,472 gaps spanning 39,849,765 bp (3.12%). After filling, 637,159 gaps remained, covering 38,997,858 bp (3.06% of the genome). Superscaffolding of genome assemblies with the chicken genome as reference produced genomes with improved contiguity metrics and higher BUSCO scores (Table 2.5). Genome size was slightly reduced, but remained within the expected range (1–1.2 Gb). Following assessment, one contig from each assembly that was identified as analogous to the chicken mitochondrial genome was removed from the final assemblies. BLAST searches to identify putative sex chromosomes confirmed the Z and W pseudochromosomes superscaffolded to the chicken assembly as Z and W chromosomes for both genome assemblies.

Repeat-masking of the superscaffolded kakī genome assembly resulted in 6.58% of the genome soft-masked (Table 2.6). Of these repeats, 5.06% were identified as repetitive retroelements (SINEs (short interspersed elements), LINEs (long interspersed elements) and LTRs (long terminal repeats)), 0.44% as repetitive DNA transposon repeats, 0.07% as unclassified repeats, and 1.0% as simple low complexity repeats. Similarly, 5.25% of the superscaffolded Australian pied stilt genome was soft-masked, with 3.76% comprising repetitive retroelements, 0.45% transposons, and 0.08% in unclassified repeats (Table 2.6). Simple low complexity repeats comprised a further 0.95% of the genome. Assessment of genomes with REAPR identified 94.08% of the kakī genome and 92.24% of the Australian pied stilt genome as containing error-free bases (Table 2.7). The main sources of errors were

gap-spanning regions, with warnings mainly due to incorrect orientation of sequences.

Mapping of trimmed reads to the respective species' genomes resulted in 95.49% of B60480 and 95.14% of B60406 reads mapping to cover 98.97% and 98.32% of the Australian pied stilt genome respectively (Table 2.8). Average coverage depth for B60480 was 38×, and 42× for B60406. For kakī, 97.53% of both DNA1914 and DNA1929 reads were successfully mapped to the kakī genome, covering 99.51% and 99.46% of the assembly, with depth of coverage of 40× for DNA1914 and 29× for DNA1929.

A total of 1,487,200 variants were discovered using the mpileup pipeline with the cleaned kakī reads mapped to the kakī genome. Following exclusion of indels and non-biallelic variants, a total of 1,374,045 SNPs were discovered. Variant discovery with Australian pied stilt cleaned reads mapped to the Australian pied stilt genome produced over six times more variants than for kakī, with a total of 9,116,038 variants discovered, and 8,386,907 SNPs after filtering. Conversion of VCFs to PED files identified a total of 1,352,296 variant sites among kakī and 8,371,786 sites among Australian pied stilts (1.147 SNPs/kb for kakī and 7.496 SNPs/kb for Australian pied stilts). There was very little missing data, with a genotyping rate of 0.9998 for kakī and 0.9993 for Australian pied stilts. Nucleotide diversity (π) was similar between species (kakī π = 0.5922, Australian pied stilts π = 0.5498), although these estimates were obtained from only two individuals of each species, so are unlikely to be representative of species-wide diversity.

Table 2.5: Assembly metrics for the gap closing and superscaffolding process for kakī and Australian pied stilt genome assemblies. bp = base pairs, Gbp = gigabase pairs. BUSCO analysis was against the database of vertebrate orthologs. Grey boxes indicate BUSCO assessment was not completed.

	Kakī (DNA1914)			**Australian pied stilt (DNA B60480)**		
	SOAPdenovo2	**Gap closed**	**Superscaffolded**	**SOAPdenovo2**	**Gap closed**	**Superscaffolded**
Scaffold #	148,989	19,214	522	816,713	816,713	1,443
Assembly size (Gbp)	1.2389	1.174	1.1792	1.2762	1.2757	1.1168
Longest scaffold (bp)	1,028,732	1,022,492	238,324,410	743,085	743,035	221,521,436
Mean scaffold length (bp)	8,315	31,104	2,258,948	1,563	1,562	773,955
Scaffold N50 (bp)	119,990	146,413	105,710,992	34,001	34,038	99,457,149
Scaffold L50	2,821	2,587	4	9,650	9,642	4
% GC content	41.12	42.27	42.09	41.14	42.25	41.64
Scaffold %N	3.12	0.04	0.48	3.12	0.04	1.04
Proportion of assembly in scaffolded contigs (%)	97.8	31.3	99.7	89.4	12.1	99
Mean break between contigs in scaffolds (> 25 Ns; bp)	84	66	254	63	58	305
Contig #	594,827	24,092	21,843	1,415,249	822,989	38,620
Longest contig (bp)	122,633	917,245	917,245	175,002	629,818	629,818
Mean contig length (bp)	2020	48,718	53,735	875	1,550	28,624
Contig N50 (bp)	10,746	107,864	11,346	3,869	32,189	43,536
Contig L50	31,271	3,068	3,034	69,832	10,266	7,287
Complete single-copy BUSCOs (%)	92		91	75.9	77.8	87.4
Missing BUSCOs (%)	1.4		3.1	8.8	7.2	8.2

Table 2.6: Results of repeat-masking of superscaffolded kakī and Australian pied stilt genomes with RepeatMasker v4.0.7, where query species was set to chicken. bp = base pairs, LINEs = long interspersed elements, LTR = long terminal repeats, SINEs = short interspersed elements.

	Kakī (DNA1914)			Australian pied stilt (DNA B60480)		
	Number of elements	**Length (bp)**	**% sequence**	**Number of elements**	**Length (bp)**	**% sequence**
Retroelements	119,661	59,647,666	5.06	95,631	42,046,355	3.76
SINEs	12,206	2,656,224	0.23	11,807	2,571,124	0.23
Penelope	164	49,916	0	156	46,682	0
LINEs	107,198	56,931,036	4.83	83,570	39,416,123	3.53
CRE/SLACS	0	0	0	0	0	0
L2/CR1/Rex	106,916	56,841,843	4.82	83,301	39,331,866	3.52
R1/LOA/Jockey	0	0	0	0	0	0
R2/R4/NeSL	0	0	0	0	0	0
RTE/Bov-B	0	0	0	0	0	0
L1/CIN4	118	39,277	0	113	37,575	0
LTR elements	257	60,406	0.01	254	59,108	0.01
BEL/Pao	0	0	0	0	0	0
Ty1/Copia	0	0	0	0	0	0
Gypsy/DIRS1	0	0	0	0	0	0
Retroviral	0	0	0	0	0	0
DNA transposons	20,429	5,223,101	0.44	19,732	5,026,674	0.45
hobo-Activator	3,179	745,617	0.06	3,085	719,919	0.06
Tc1-IS630-Pogo	898	203,169	0.02	878	196,836	0.02
En-Spm	0	0	0	0	0	0
MuDR-IS905	0	0	0	0	0	0
PiggyBac	0	0	0	0	0	0
Tourist/Harbinger	7,306	2,354,311	0.2	6,976	2,244,614	0.2
Other (Mirage, P-element, Transib)	0	0	0	0	0	0
Rolling-circles	0	0	0	0	0	0
Unclassified	3,569	880,281	0.07	3463	854,848	0.08
Total interspersed repeats		65,751,048	5.58		47,927,877	4.29
Small RNA	4,426	1,011,644	0.09	4226	972,294	0.09
Satellites	0	0	0	0	0	0
Simple repeats	247,305	9,586,939	0.81	226510	8,652,423	0.77
Low complexity	48,071	2,255,105	0.19	44414	2,056,826	0.18
Total bases masked		77,639,702	6.58		58,676,607	5.25

Table 2.7: Reapr results of final scaffolded assemblies for kakī and Australian pied stilt. bp = base pairs, FCD = fragment coverage distribution, n = number of reads.

	Kakī (DNA1914)	Australian pied stilt (DNA B60480)
Total length (bp)	1,179,170,728	1,116,816,459
Total sequences	522	1,443
Mean sequence length (bp)	2,258,947.75	773,954.58
Length of longest sequence	238,324,410	221,521,436
N50 (n)	105,710,992 (4)	99,457,149 (4)
N60 (n)	70,199,135 (6)	65,902,221 (6)
N70 (n)	42,018,823 (8)	39,316,847 (8)
N80 (n)	25,157,471 (11)	23,457,286 (11)
N90 (n)	17,024,195 (16)	14,219,872 (17)
N100 (n)	5,013 (522)	5,001 (1,443)
Number of gaps	106,779	136,444
Total gap length (bp)	5,602,890	11,591,932
Error free bases	94.08%	92.24%
Total errors	47,904	62,915
FCD errors within a contig	6,026	4,610
FCD errors over a gap	23,037	22,713
Low fragment coverage within a contig	6,703	8,288
Low fragment coverage over a gap	12,138	27,304
Total warnings	64,714	111,317
Low score regions	18	0
Links	4,208	4,869
Soft clip	7,268	8,055
Collapsed repeats	1,420	882
Low read coverage	189	103
Low perfect coverage	0	0
Wrong read orientation	51,611	97,408

Table 2.8: Results of mapping cleaned reads from kakī and Australian pied stilts to the respective species' reference genome assembly.

Reference genome	DNA ID of individual mapped	Mean depth (×)	Genome covered (%)	Reads mapped (%)
Kakī	DNA1914	40.57	99.51	97.53
	DNA1929	28.58	99.46	97.53
Australian pied stilt	B60480	37.81	98.97	95.49
	B60406	42.11	98.32	95.14

2.4.5 Inferring demographic history

Preliminary pairwise sequential Markovian coalescent (PSMC) model results indicate distinct population trajectories for kakī and Australian pied stilts (Figure 2.10), although resolution of the data was particularly poor for Australian pied stilt individual B60480. According to these results, kakī have experienced population decline from a maximum estimated effective population size of approximately 10,000 individuals simultaneously with the occurrence of the Waimea glacial period 125–180 kya, and have remained at small population size since. These results indicate Australian pied stilts have remained at a relatively larger population size, but estimates produced here appear low given their current wide distribution.

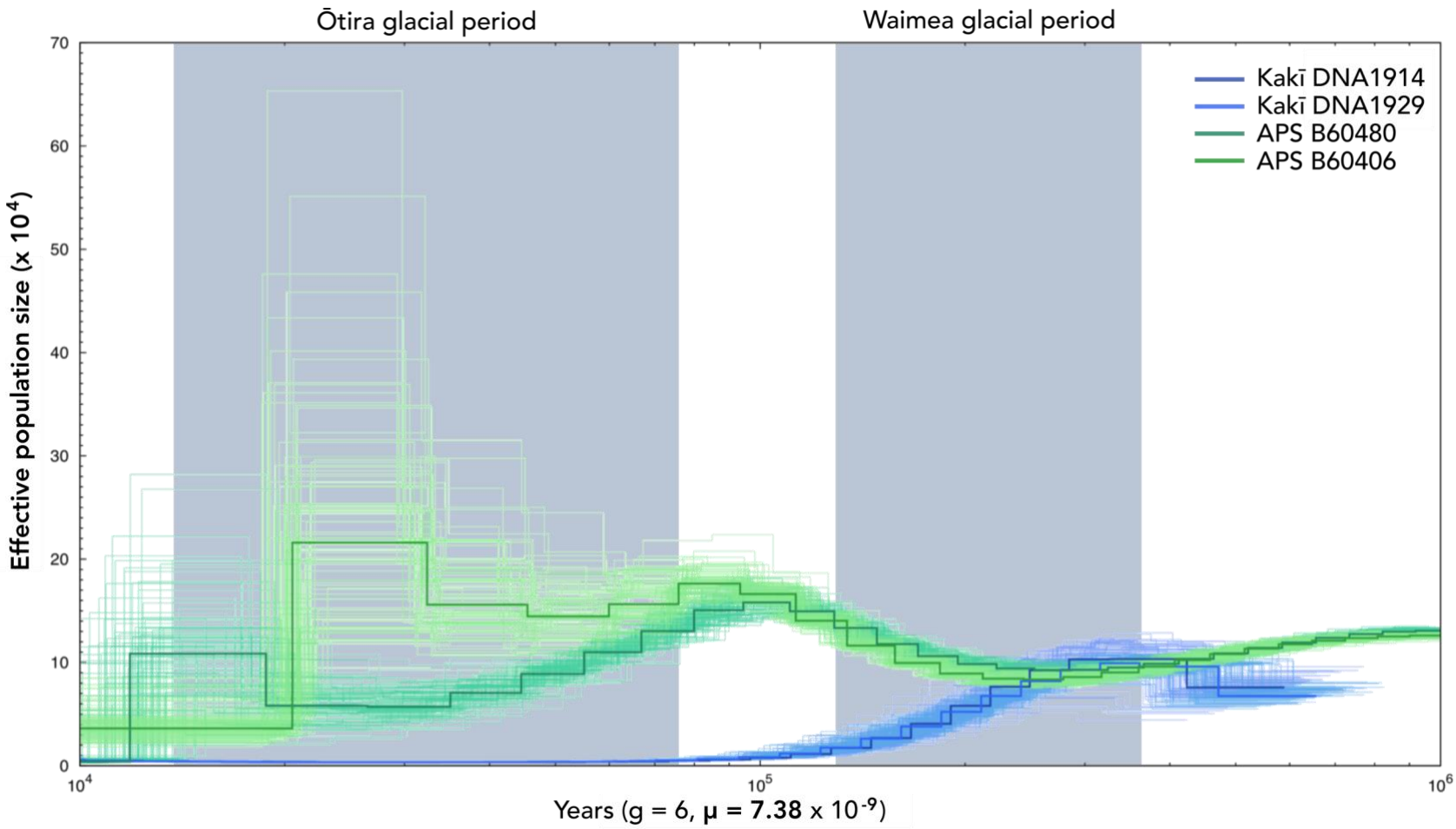


Figure 2.10: Pairwise sequentially Markovian coalescent (PSMC) model plot for two kakī (DNA1914 and DNA1929) in blue and two Australian pied stilts (APS; B60480 and B60406) in green. The y-axis represents effective population size. The x-axis represents a log-scale of time in years before the present, calibrated assuming a generation time (g) of six years for both species, and a substitution rate (μ) of 7.38 x 10^{-9} substitutions/site/generation, inferred from the lower mean avian mutation rate estimated by Nam et al. (2010). Lighter hued lines represent 100 bootstrap replicates for each individual. New Zealand glaciation events are indicated in grey, although these events may have had different impacts on Australian pied stilts.

2.5 Discussion

Here I produced short-read sequencing data for five stilt individuals that was then used to assemble draft genomes for each of two stilt species, kakī and Australian pied stilts. The small size, low complexity, and high synteny among bird genomes make avian genomes relatively straightforward to resolve (Zhang et al., 2014b) compared with those for species with large and/or highly repetitive genomes (e.g., amphibians (Rogers et al., 2018; Smith et al., 2019), cartilaginous fishes (Hara et al., 2018; Marra et al., 2019)), although reference-assisted assembly may result in under-reporting of exceptions to this perception of a high degree of synteny (e.g., Gan et al., 2019; Dierickx et al., 2020). A range of quality metrics were used to compare the quality of assemblies produced using different assemblers. Ideally, a complete genome assembly would span the expected genome size, contain one scaffold representing each chromosome, and the full complement of expected genes. However, highly repetitive regions and highly-heterozygous genomes present a challenge to both sequencing and assembly, particularly when using only short paired-end reads (Pryszcz & Gabaldón, 2016). Thus, a genome assembly with a low number of highly contiguous scaffolds covering the expected genome size (i.e., low scaffold number, high scaffold N50, high mean scaffold length, low proportion of gaps, total genome size ~1.1 Gb) and containing a high proportion of expected genes is considered optimal. This approach indicates that a simple, cost-effective sequencing approach is sufficient to produce high-quality draft avian genome assemblies (see Appendix A), though I acknowledge that the genomes presented here do not represent ‘complete’ chromosomally-assembled genomes (Peona et al., 2018). Even the very high-quality fifth version of the chicken genome assembly represents an incomplete assembly (Warren et al., 2017). Nevertheless, these

genomes are of the required quality for downstream analyses (see Galla et al. 2019, Appendix A, and Chapter Four). For future research, the assemblies presented here would benefit from additional long-read sequencing and assembly to isolate complete chromosomes, accurately cover repetitive regions, and span assembly gaps, all of which would contribute to improving genome annotation (see Appendix B).

For example, the low level of error identified in the genomes with REAPR quality assessment cannot be resolved here, as this requires correction with long-read sequencing to improve scaffolding across gaps and correct sequence orientation errors. Although assembly of Australian pied stilt B60480 produced a lower quality assembly than that of kakī DNA1914, likely due to the higher levels of heterozygosity present as indicated in pre-assembly assessment, both draft genomes are of similar quality across all metrics to those of genomes for other Charadriiformes species produced via similar sequencing and assembly methods (Table 2.9). The additional steps taken to conduct superscaffolding of the assembly by leveraging the high level of synteny between avian species proved effective despite the lack of long mate-pair reads, improving the contiguity of the assembled genomes (in terms of reduced number of scaffolds and increased mean scaffold lengths).

Table 2.9: Comparison of Australian pied stilt and kakī genome assemblies (green) with genome assemblies of species in the Order Charadriiformes with genomes available on NCBI as at April 1st 2020. Gb = gigabase pairs, bp = base pairs, BUSCOs = complete single-copy orthologs from the avian database, MP = mate-pair, PE = paired-end, U = unknown. Metrics and assembly details were derived from GenBank and the associated publications, except for BUSCO metrics, which I assessed independently against the set of avian ortholog database.

Species	Library	Sequencing platform	Assembly method	Depth	Genome size (Gb)	Scaffold N50 (bp)	Scaffolds	BUSCOs (%)
Alca torda, Razorbill[1]	U	Illumina NovaSeq, PacBio Sequel, 10X Genome, Arima Hi-C	FALCON, purge_haplotigs, scaff10x, Bionano 2 enzyme, Salsa2, longranger align, freebayes, gEVAL manual curation	65×	1.178	1,174,931,441	95	
Burhinus oedicnemus, Eurasian thick-knee[2]	U	Illumina HiSeq	U	137×	1.228	1,191,749,133	9,229	
Calidris pugnax, Ruff[3]	PE + MP	Illumina HiSeq 2000	SOAPdenovo2	113×	1.229	10,060,041	3,753	93.1
Calidris pugnax, Ruff[4]	PE + MP	Illumina HiSeq 2500, PacBio RS II	ABySS, SSPACE, PBJelly	110×	1.173	868,209	47,740	92.0
Calidris pygmaea, Spoon-billed sandpiper[5]	U	Illumina HiSeq	DiscovarDeNovo, SSPACE-SR, FAST-SG	38×	1.178	3,136,279	29,819	94.0
Charadrius alexandrinus, Kentish plover[6]	PE + MP	Illumina HiSeq	SOAPdenovo2	134×	1.232	3,290,793	17,135	
Charadrius vociferus, Killdeer[7]	PE + MP	Illumina HiSeq	SOAPdenovo	100×	1.220	3,657,050	15,167	92.6
Himantopus himantopus leucocephalus, Australian pied stilt[8]	PE	Illumina HiSeq 2500	SOAPdenovo2, Satsuma	51×	1.117	99,457,149	1,442	87.4
Himantopus novaezelandiae, Kakī[9]	PE	Illumina HiSeq 2500	SOAPdenovo2, Satsuma	52×	1.179	105,710,992	522	91.0
Limosa lapponica baueri, Bar-tailed godwit[10]	PE + MP	Illumina HiSeq 2000	AllPaths-LG	71×	1.035	283,007	32,319	61.8
Recurvirostra avosetta, Pied avocet[11]	PE	Illumina HiSeq 4000	Velvet	78×	1.193	96,778	35,873	82.4
Scolopax mira, Amami woodcock[12]	PE	Illumina HiSeq X	Genomic Workbench	55×	1.091	124,759	40,750	86.0
Sterna hirundo, Common tern[13]	U	PacBio Sequel, Illumina NovaSeq, Arima Hi-C, Bionano Genomics	FALCON, purge_haplotigs, scaff10x, Bionano Solve DLS, Salsa HiC, Arrow, longranger align	68×	1.230	1,219,607,931	123	
Uria lomvia, Thick-billed guillemot[14]	PE + MP	Illumina HiSeq 2500	Platanus	68×	1.179	15,847,591	9,327	93.6

1) NCBI Accession No.: GCA_008658365.1 (Unpublished), 2) GCA_008921705.1 (Unpub.), 3) GCA_001431845.1 (Lamichhaney et al., 2016), 4) GCA_001458055.1 (Küpper et al., 2016), 5) GCA_003697955.1 (unpub.), 6) GCA_008711295.1 (Wang et al., 2019), 7) GCA_000708025.2 (Zhang et al., 2014b), 8) GCA_003993805.1 (Galla et al., 2019), 9) Not available via NCBI, see Data Accessibility, this chapter (Galla et al., 2019), 10) GCA_002844005.1 (Parody Merino, 2018), 11) GCA_004023745.1 (Galla et al. 2019), 12) GCA_004320125.1 (unpub.), 13) GCA_009819605.1 (unpub.), 14) GCA_002289315.1 (Tigano et al., 2018).

The development of long-read sequencing capabilities provides great promise for genome assembly. However, the genome assemblies produced here demonstrate that short-read sequencing is sufficient for generating high-quality genomes fit for conservation purposes (Galla et al. 2019, Appendix A), and the additional costs and challenges associated with long-read sequencing should be weighed against the needs of the conservation programme. While the sequencing methods employed here were among the least expensive available, the average cost associated with producing one genome by the methods presented here was estimated at 11,375 NZD (including Illumina sequencing, computational resources, and associated person-hours; Galla et al. 2019, Appendix A), and additional long-read sequencing could substantially increase these costs. The ability to generate draft genomes using relatively cost-effective methods provides the opportunity for many avian conservation projects to make the transition to a genomic approach for management.

2.5.1 Genome-wide species comparisons

SNP discovery conducted by mapping cleaned sequences to the respective species' genomes indicated that kakī have substantially lower genomic diversity in terms of the number of SNPs discovered than the congeneric Australian pied stilt, despite similar nucleotide diversity estimates between the two species. The estimate of nucleotide diversity observed for the kakī genome of 0.5922 appears substantially different from that produced from population-level estimates for kakī by Galla et al. (2019) which produced estimates between 0.31 ± SD 0.14 and 0.35 ± SD 0.13. SNP density across the kakī genome was similar to that produced from genome-wide SNP discovery in two extinct New Zealand birds, the huia (*Heteralocha acutirostris*) and South Island kōkakō (*Callaeas cinereus*; Dussex et al. 2019). The relatively low diversity of kakī compared with Australian pied stilts may be due to the recent kakī

decline, although previous genetic analysis using microsatellite loci indicated kakī have not experienced such extensive declines in diversity as other threatened New Zealand birds (Steeves et al., 2010).

There are a number of caveats and limitations that must be taken into account when considering the results of preliminary historical population demography assessment with PSMC. Firstly, these estimates are based on only two individuals of each species, and as such may not accurately represent past effective population sizes, as indicated by the difference in patterns between the two Australian pied stilts. For genomic data sets comprising thousands of markers, accurate genetic diversity estimates can be obtained using as few as two individuals (Nazareno et al., 2017; Qu et al., 2019; Willing et al., 2012), and PSMC analysis typically relies on one or two representative individuals (Nadachowska-Brzyska et al. 2015; Mays et al. 2018, but see Liu and Hansen 2017; Vijay et al. 2018). Incorporating additional individuals, particularly for species that are widely distributed, may reveal population-specific trends, and provide more accurate estimates of temporal changes. In addition, masked regions and the haploid sex chromosome (W; as both kakī individuals and one Australian pied stilt were female) were not excluded from the analysis. There may be low coverage depth for these regions which may bias the results as indicated by Nadachowska-Brzyska et al. (2015). Generation time was estimated at six years based on the age of first breeding for kakī at 2–3 years, but is only representative of the modern kakī population. Species-specific mutation rates are not yet known, and may differ from the estimated mutation rate used here. Testing with both the high and low mutation rates estimated by Nam et al. (2010) produced no difference in the observed pattern. However, the timing of the pattern was slightly shifted (results not shown), with the decline occurring more recently. Furthermore, additional testing using different atomic time parameters may improve resolution. Despite these

caveats, these results provide a preliminary comparison of demographic patterns for these stilt species over time suggesting that kakī may have had a relatively low effective population size for the past 100,000 years compared with the estimates of much larger effective population size for Australian pied stilts.

2.5.2 Utility for conservation

Two published studies have used the genomes produced here. In the first proof-of-concept study (Galla et al. 2019, Appendix A), both genomes were used to demonstrate that for birds, genomes from closely related species—up to the level of family—can be used as a reference for SNP discovery to produce robust estimates of relevant metrics for conservation, including estimates of nucleotide diversity, relatedness, and inbreeding. Having a reference genome can substantially reduce the computational challenges of SNP discovery from reduced-representation approaches, producing a greater number of more reliable SNPs for downstream analyses, and enabling SNP annotation and robust comparisons when incorporating additional data in the future (Torkamaneh et al 2016, Shafer et al. 2016, Brandies et al 2019). Despite these benefits and the expanding resources available, most conservation genomics projects to date that include reduced-representation data (e.g., RADseq, GBS) have generated SNPs using a *de novo* approach (e.g., Rexer-Huber et al. 2019, but see Ruegg et al. 2018), perhaps due to concerns of producing genomes of insufficient quality (Leggett & MacLean, 2014) or a previous absence of empirical data to support the use of closely related genomes as reference. The use of relatively distantly related reference genomes can represent an alternative to species-specific genome assembly for resource-limited avian conservation projects, as draft genomes are currently available for at least one member of every avian order (Zhang et al., 2014b), and resources for one member of every avian family are

coming online (Zhang, 2015). This approach may be transferable to other taxa, and proof-of-concept studies in this space would be useful in further expanding the pool of resources available for other threatened species, although challenges associated with large genome size and complexity in some taxa (e.g., mammals, amphibians) may limit the available resources.

In the second published study, the kakī genome produced here was used as a reference for SNP discovery for a comparison of pedigree-, genetic-, and genomic-based estimates of pairwise relatedness between individuals (Galla et al., 2020). Accurate estimates of relatedness are essential for captive pairing decisions to avoid the negative impacts of inbreeding depression in threatened species, while maximising diversity and thus evolutionary potential (Galla et al., 2020; Giglio et al., 2016; Ivy & Lacy, 2012). Relatedness estimates based on SNP data proved more precise and accurate than those produced with microsatellites. These results indicate that future best-practice will incorporate genomic estimates of relatedness, coupled with pedigree information, and this approach has already been implemented for informing captive pairing of kākāriki karaka/orange-fronted parakeet (*Cyanoramphus malherbi*, S.J. Galla, pers. comm.). Thus the genomes produced here have contributed to two published studies (Galla et al., 2019, 2020), with the initial proof-of-concept study (Galla et al. 2019, Appendix A) directly informing conservation genomic research across a range of critically endangered New Zealand birds including kākāriki karaka, kākāpō (*Strigops habroptilus*), tara-iti /New Zealand fairy tern (*Sternula nereis davisae*), and tuturuatu/shore plover (*Thinornis novaeseelandiae*).

The genomes produced here expand the avian genomic resources available via NCBI, bringing the total reference genomes for Charadriiformes to fifteen, with additional resources in development (e.g., genome assemblies for tuturuatu and

tara-iti, T.E. Steeves, pers. comm.). The kakī genome also has the potential to further understanding of the evolution of New Zealand's endemic avifauna, alongside the assembled genomes of kākāpō, kea (*Nestor notabilis*), hihi (*Notiomystis cincta*), kiwi (brown kiwi *Apteryx australis mantelli*, great spotted kiwi *Apteryx haastii*, Okarito brown kiwi *Apteryx rowi*, and little spotted kiwi *Apteryx owenii*), and titipounamu/rifleman (*Acanthisitta chloris*), along with the extinct little bush moa (*Anomalopteryx didiformis*), huia (*Heteralocha acutirostris*), and South Island kōkakō (*Callaeas cinereus*). The sequence data produced for the five individuals here could also be leveraged in comparative genomic analyses of Charadriiformes, providing insights into the evolution and functional adaptations of these wading birds.

2.5.3 Transitioning to a conservation genomics approach with *de novo* genome assembly

For researchers transitioning to using genomics tools including *de novo* genome assembly, the array of sequencing platforms and assembly pipelines may be daunting. Although it may be tempting to test a wide variety of tools, the effort involved may be disproportionate to improvements in the quality of the assembly outcomes, and researchers should aim to produce outputs sufficient to answer the immediate questions of conservation relevance. Where genomes are anticipated to include high heterozygosity or extensive repetitive elements, or where research questions require well-annotated genomes or analysis of structural variation, additional investment in long-read sequencing and chromosomal-assembly may be required. The genome assemblies presented here indicate that for relatively small, low-complexity avian genomes, short-read sequencing is sufficient for high-quality genome assembly. SOAPdenovo2 represents a relatively straight-forward genome assembler, that produced the highest quality assemblies among the four assemblers

tested here. SOAPdenovo2 has a concise user manual available, provides the ability to adjust many parameters to improve assembly, and is widely used for genome assemblies (e.g., (de Villemereuil et al., 2019; Mays et al., 2018; Tollis et al., 2017). Furthermore, results demonstrating the utility of confamilial reference genomes for robust estimates of nucleotide diversity, individual heterozygosity, and relatedness may negate the need for species-specific genome assembly, presenting threatened species management programmes with existing genomic resources (Galla et al. 2019, Appendix A). Moreover, cross-disciplinary collaboration with experienced bioinformaticians can support upskilling and streamline the transition to a conservation genomics approach (Galla et al., 2016). The genomes presented here represent key additions to the conservation genomic resources available for kakī, and their utility has been validated (Galla et al., 2019, Appendix A). The genomes presented here represent the first step in assessing impacts of past hybridisation between kakī and poaka. The raw sequence reads produced for kakī are used in Chapter Three for assembly of a kakī mitochondrial genome for stilt phylogenetic analysis, while in Chapter Four, reduced-representation sequencing reads are mapped to the kakī reference genome for accurate SNP discovery and genome-wide analysis of introgression between kakī and poaka.

2.6 Data availability

The Australian pied stilt Whole Genome Shotgun sequences for B60480 have been deposited at DDBJ/ENA/GenBank under Accession No.: RSEF00000000, with the draft genome version described here available under NCBI Accession No.: GCA_003993805.1. Because kakī—and genomic data generated from kakī—are taonga, these data will be made available on the recommendation of the associated

iwi and hapū (see Galla et al. 2019, Appendix A). Further, a local genome browser for the kakī genome is available on request at www.ucconsert.org/data/. Scripts associated with the bioinformatic workflows included in this chapter are detailed in my GitHub 'Himantopus' repository at https://github.com/natforsdick/Himantopus.

Chapter 3: Comparison of historic and contemporary mitochondrial genomes to assess the impacts of population decline and conservation management in the critically endangered kakī (*Himantopus novaezelandiae*)

3.1 Abstract

Conservation management relies on genetic principles to reduce the extinction risk associated with the small size of threatened populations. Among the genomic resources available to inform species conservation, mitochondrial genomes represent a relatively cost-effective tool enabling comparisons between contemporary and historic populations, and thus can be used to assess impacts of past population processes, or the success of conservation management aiming to maximise genetic diversity. The Aotearoa New Zealand endemic kakī (*Himantopus novaezelandiae*) and Australian pied stilts (*H. himantopus leucocephalus*) are congeneric members of the order Charadriiformes. While Australian pied stilts self-introduced to Aotearoa and expanded to around 30,000 individuals today, habitat modification and loss and the introduction of mammalian predators contributed to kakī decline to a low of ~23 individuals in 1981. This decline resulted in anthropogenic interspecific hybridisation between these congeners. Conservation management of kakī based initially on the declining population paradigm, and later informed by genetic principles relating to inbreeding and small population size and genetic data has seen the population increase to 169 wild adults today. Here I assemble a kakī mitochondrial genome, and leverage the small size and high copy-

number of mitochondrial genomes to compare mitochondrial diversity of historic stilts with contemporary stilts prior to the initiation of conservation management. Results indicate kakī mitochondrial diversity has largely been maintained despite species decline and long-term small population size. Furthermore, mitochondrial haplotypes clearly differentiate kakī from Australian pied stilts and poaka, and thus contribute to the behavioural, morphometric, and genetic evidence that kakī conservation has successfully maintained the genetic integrity of this critically endangered taonga species.

3.2 Introduction

Extinction risk is exacerbated in small populations by the genetic processes of genetic drift, the random loss of genetic diversity due to stochasticity, and inbreeding depression, where unavoidable inbreeding results in reduced reproductive success and survival (Frankham, 2005; Keller & Waller, 2002; Nei et al., 1975; Wright et al., 2007). Conservation management programmes incorporate genetic principles through the key objectives of maximising genetic diversity, to provide the potential for populations to adapt to change over time, and minimising inbreeding, to minimise the negative effects of inbreeding depression (Frankham et al., 2017). For many species, genetic or genomic data may not be available at the initiation of conservation management, but best-practice decision-making is guided by genetic principles. Once genetic tools become available, assessment of the impacts of conservation management can be made, enabling fine-tuning of management practices to ensure the best outcomes.

3.2.1 Mitochondrial genomes for temporal comparisons of diversity

The field of paleogenomics has expanded rapidly alongside the high-throughput sequencing (HTS) revolution, with ancient DNA (aDNA) techniques improving to allow sequencing of complete mitogenomes (Krause et al., 2006) and even nuclear DNA (Miller et al., 2008; Noonan et al., 2005) from historic or ancient samples. Nevertheless, aDNA studies are complicated by DNA degradation, resulting in lower DNA quantities, increased fragmentation, and DNA damage (Pääbo et al., 2004). These factors all contribute to increase the susceptibility of aDNA to sequence contamination, and the bioinformatic challenges associated with the short sequences obtained and the increased likelihood of false mutations within these data (Leonard, 2008). Population paleogenomic studies remain limited for many species due to the small numbers of samples that are sufficiently well-preserved to produce informative nuclear genomic data. The high copy-number and relative small size (~16–20 kb) of mitochondrial genomes (mitogenomes) can be leveraged for generating genetic data from historic or ancient samples (Pakendorf & Stoneking, 2005), and so mitogenomes represent the best tool to incorporate ancient or historic data when assessing population diversity and past population processes for many species. While HTS advances have expanded the field of paleogenomics, aDNA efforts have been largely focussed on hominin evolution and human migration (e.g., Krings et al. 1997; Green et al. 2010; Knapp et al. 2012), domesticated species (e.g., Orlando et al. 2013; Botigué et al. 2017), extinct megafauna (e.g., Miller et al. 2008; Stiller et al. 2010), and microorganisms (e.g., Bos et al. 2016). As paleontological studies continue, natural history collections present a diversity of specimens that can facilitate historic and contemporary comparisons for a wider range of species (Brunson & Reich, 2019; Wandeler et al., 2007; Yeates et al., 2016), allowing investigation of temporal changes in genetic diversity in relation to geological,

climatic, or anthropogenic events (Chan et al., 2006b). Obtaining complete mitogenomes for non-model species through HTS has become relatively efficient and cost-effective, adding to the genomic resources in the conservation toolbox. Temporal comparisons of mitochondrial diversity can inform predictions of the impacts of current anthropogenic climate change on species distributions and potential future threats, assessing the success of past conservation management while informing future management (e.g., Wilmshurst et al. 2014).

3.2.2 Conservation management of kakī (*Himantopus novaezelandiae*)

The geographic isolation of New Zealand has given rise to a taxonomically distinct avifauna with high rates of endemicity (Worthy & Holdaway, 2002). Among those species present at the time of human arrival were 14 species of shorebirds (Order Charadriiformes; Trewick and Gibb 2010) including the endemic kakī (black stilt, *Himantopus novaezelandiae*; Holdaway et al. 2001). The congeneric pied stilt (*H. himantopus leucocephalus*) is widespread throughout Asia and the Pacific, occurring in sympatry with kakī in Aotearoa, where they are known as poaka. Kakī and pied stilts are hypothesised to have diverged from a common ancestor approximately one million years ago (Wallis, 1999), estimated based on observed sequence divergence in the mitochondrial control region (Chambers & Macavoy, 1999). Kakī numbers declined during the 1900s due to anthropogenic habitat modification and loss and the introduction of mammalian predators, while poaka numbers increased across the mainland to around 30,000 today (Pierce, 1984b). During kakī decline, hybridisation with poaka has occurred, resulting in the production of fertile hybrid offspring with intermediate plumage colouration between the pure black (plumage node J) kakī and the black and white poaka (Pierce 1984b; see Figure 1.5).

Conservation management began in 1981 after kakī had declined to approximately 23 wild adults (Pierce, 1984b). This management integrated genetic principles in a programme of captive breeding for translocation designed to minimise inbreeding and maximise genetic diversity, and later included the goal of maintaining species integrity in the face of hybridisation (Maloney & Murray, 2001). As genetic, and more recently, genomic resources have become available, these tools have been used to assess the extent and impacts of inbreeding (Hagen et al., 2011), genetic diversity and hybridisation (Steeves et al., 2010), and in routine management to inform captive pairing decisions (Galla et al., 2020). This recent species decline and subsequent small population size is expected to have resulted in reduced genetic diversity due to stochastic processes (Frankham, 2005; Lande, 1993; Nei et al., 1975), although estimates from nuclear markers show kakī to have moderate genetic diversity compared with other threatened New Zealand birds (Steeves et al., 2010). Amplification of a 291 bp fragment of the mitochondrial cytochrome *b* gene was incorporated with nuclear microsatellite markers to assess introgression resulting from hybridisation between kakī and poaka (Steeves et al., 2010). Four mitochondrial haplotypes were identified among the two species, with node J kakī having only haplotypes B and C, while all four haplotypes were found among hybrids and poaka. While that study confirmed that the genetic integrity of kakī had been maintained despite extensive hybridisation, the single-gene mitochondrial analysis may not be representative of total mitogenome diversity.

In combination with predator control and habitat restoration, this genetics-informed intensive management has resulted in kakī numbers increasing to 169 wild adults in 2020 (DOC, pers. comm.). Nevertheless, kakī remain critically endangered, with the species primarily limited to Te Manahuna/the Mackenzie Basin in the central South Island of Aotearoa (BirdLife International, 2018; Maloney & Murray, 2001). Over

time, genetic data has accumulated to better inform conservation management based on genetic principles. However, there is no data estimating the historic diversity of kakī available to explore what may have been lost due to the species decline, or maintained through conservation genetic management. Furthermore, hybridisation between kakī and poaka resulted in uncertainties over what constituted kakī, with early conservation efforts including individuals with dark plumage nodes G–J, as compared with management of only node J individuals today (Maloney & Murray, 2001; Reed et al., 1993a). HTS developments now enable the inclusion of mitochondrial data to assess diversity of historic samples, thus permitting temporal comparisons to resolve these uncertainties. Complete mitogenomes can provide greater information on mitochondrial diversity, including gene positions and structural features that may illuminate evolutionary relationships (Masta & Boore, 2008) and improve resolution of phylogenetic analyses compared with shorter single-gene mitochondrial sequences (Fu et al., 2013). In addition, evolutionary rates for mutation of mitochondrial genes are better calibrated than those of nuclear genes. However, downsides of the use of mitochondrial data for estimating demographic and evolutionary processes include complete mitochondrial introgression, maternal inheritance, and the single-gene nature of the mitochondrial genome (Ballard & Whitlock, 2004).

3.2.3 Aims

In this study I aim to use mitochondrial genomes of modern and historic kakī, Australian pied stilts, poaka, and kakī-poaka hybrids to evaluate the impacts of kakī decline and subsequent conservation management on mitochondrial diversity and population structure. This will require the assembly of a reliable high-quality kakī mitogenome that can be used as a reference for mapping historic and modern

mitogenome data. Mitogenomes of all available historic kakī samples will be sequenced, along with historic poaka and contemporary kakī, poaka and Australian pied stilts, and kakī-poaka hybrids. I predict that complete mitogenome analysis will identify a greater number of haplotypes among kakī than observed through single-gene analysis (Steeves et al., 2010). Comparisons of mitogenome diversity can then be made between modern and historic samples, and between kakī and pied stilts, thus allowing insights into historic kakī diversity, the impacts of the severe population bottleneck, and the success of management aimed at maintaining genetic diversity.

3.3 Methods

3.3.1 Modern stilt sample collection

Samples from two Australian pied stilts, DNA IDs B60406 and B60480, were provided by Adelaide Zoo under the Royal Zoological Society of South Australia Specimen Licence Agreement (Import Permit: 2016061954; Figure 3.1, Table 3.1). Samples from two poaka (DNA IDs B40279 and B50004, both plumage node B, see Figure 1.5) from Auckland Zoo were supplied under Auckland Zoo Animal Ethics Committee approval. A third poaka sample (DNA ID Poaka1) was received from Hawke's Bay, North Island, New Zealand. All other modern samples used in this chapter were collected under approval of the Department of Conservation Animal Ethics Committee (AEC #283) as part of routine handling practice at the kakī captive breeding facilities in Twizel (DOC) and Christchurch (Isaac Conservation and Wildlife Trust), New Zealand.

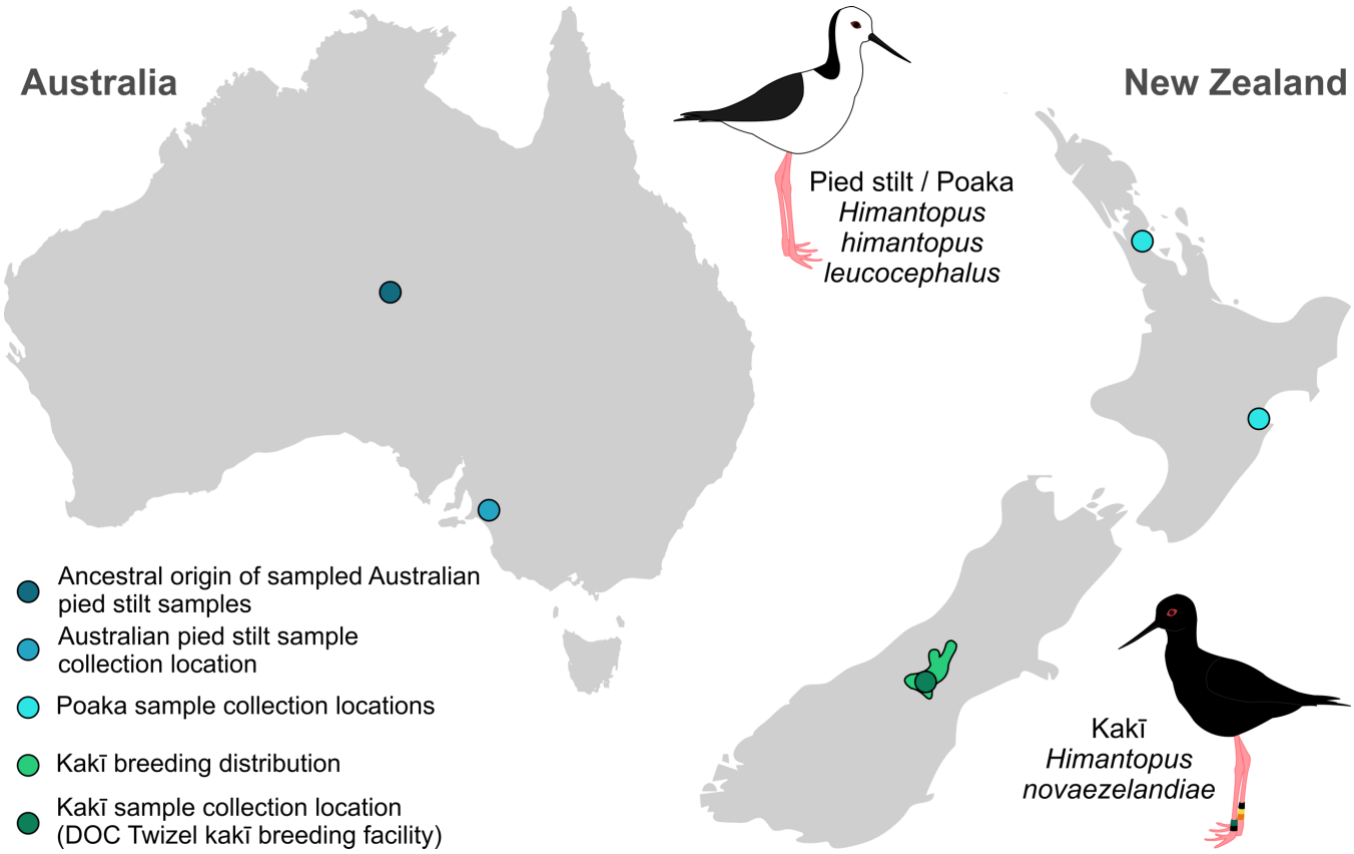


Figure 3.1: Collection locations of modern stilt samples in Australia and New Zealand.

Table 3.1: Sample information for modern stilt samples.

DNA ID	Species	Sample location
DNA1044	*Himantopus novaezelandiae*	Twizel, NZ
DNA2094	*H. h. leucocephalus* x *H. novaezelandiae*	Twizel, NZ
DNA2113	*H. h. leucocephalus* x *H. novaezelandiae*	Twizel, NZ
B40279	*H. h. leucocephalus*	Auckland Zoo, NZ
B50004	*H. h. leucocephalus*	Auckland Zoo, NZ
B60406	*H. h. leucocephalus*	Adelaide Zoo, Australia
B60480	*H. h. leucocephalus*	Adelaide Zoo, Australia
Poaka1	*H. h. leucocephalus*	Hawke's Bay, NZ

3.3.2 Historic sample collection

I collected a total of 27 samples from stilt specimens in collections at New Zealand museums Te Papa Tongarewa Museum of New Zealand (Wellington), Tāmaki Paenga Hira Auckland War Memorial Museum (Auckland), and Canterbury Museum (Christchurch; Table 3.2). Catalogue information recorded the specimens as

including twelve kakī, twelve poaka, and three hybrids, although there were some discrepancies between catalogued species identification, and morphological species identification (Table 3.2). Plumage morphology was used to confirm species identity, with verification by Liz Brown (kakī aviculturalist, DOC's Kakī Recovery Programme; see Appendix C for further discussion). This morphological classification was used unless specified in all downstream analyses, based on the relationship between plumage node and proportion of introgression described by Steeves et al. (2010). Specimens from both the North and South Island were included, along with one individual from the Chatham Islands (Figure 3.2). Five specimens dated from the 1800s, with the earliest collected in 1843. Nine specimens had no recorded collection date, and nine had no recorded collection location. Of the nine specimens from Te Papa with no recorded collection dates, catalogue numbers in the range OR000001-OR008000 were used to classify specimens as having been collected and catalogued prior to the early 1950s based on museum information (T. Schultz, Te Papa Science Collection manager, pers. comm.). All specimens were skins except for mounted specimens MS10987, MS10990, MS11003 and MS11004, and skeleton MS11001. Tissue was collected from toe-pads, except for MS11001, where dried tissue attached to the skeleton was collected. The quantity of tissue collected ranged from < 1–37.5 mg.

Table 3.2: Details of museum stilt samples. Recorded collection dates are noted where known. Grey boxes indicate no recorded information. Dates listed as 'Pre-1950' are based on specimen museum ID, with IDs lower than OR008000 being registered in the early 1950s (and specimens collected prior to this). Weights refer to the total amount of material sampled and used in DNA extractions. Individuals highlighted in green have discordant catalogue and morphological species identifications (See Appendix C). F = female, M = male, mg = milligrams, NI = North Island, SI = South Island.

Source museum	Specimen museum ID	aDNA ID	Catalogued species	Collection date	Collection location	Specimen type	Sex	Morphological identification	Weight (total collected; mg)	Weight (total used; mg)
Auckland	LB3413	MS10986	*Himantopus novaezelandiae*	01 Jul 1879	Mangere, NI	Skin	F	*Himantopus novaezelandiae*	2.3	2.3
Auckland	LB3946	MS10987	*Himantopus novaezelandiae*	Pre-1940	South Head, Kaipara Harbour, NI	Mounted		*Himantopus novaezelandiae*	2.9	2.9
Auckland	LB3414	MS10988	*Himantopus novaezelandiae*	Sep 1882	Manukau, NI	Skin	M	*Himantopus novaezelandiae*	1.8	1.8
Auckland	LB8517	MS10990	*Himantopus novaezelandiae*	Pre-1887	Waikato, NI	Mounted	M	*Himantopus novaezelandiae*	3.1	3.1
Canterbury	AV686	MS10991	*Himantopus himantopus leucocephalus*	Nov 1914	Lake Ellesmere, SI	Skin	F	*Himantopus novaezelandiae (sub-adult)*	20.4	9.6
Canterbury	AV683	MS10992	*Himantopus himantopus leucocephalus*	20 Jun 1910s	Waitaki River, SI	Skin	M	*Himantopus novaezelandiae*	30.3	
Canterbury	AV682	MS10993	*Himantopus himantopus leucocephalus*	20 Jun 1910s	Waitaki River, SI	Skin	F	*Himantopus himantopus leucocephalus*	27.1	19.1
Canterbury	AV685	MS10994	*Himantopus himantopus leucocephalus*	Nov 1914	Lake Ellesmere, SI	Skin	M	*Himantopus novaezelandiae (sub-adult)*	37.5	26.6
Canterbury	AV1938	MS10995	*Himantopus himantopus leucocephalus*	1872	Selwyn, SI	Skin	F	*Himantopus novaezelandiae*	34.2	16.7
Te Papa	OR002286	MS10996	*Himantopus novaezelandiae*	Pre-1950s		Skin		*Himantopus novaezelandiae*	28.6	15.7
Te Papa	OR002287	MS10997	*Himantopus novaezelandiae*	Pre-1950s		Skin		*Himantopus novaezelandiae*	12	6.3
Te Papa	OR002289	MS10998	*Himantopus novaezelandiae*	Pre-1950s		Skin		*Himantopus novaezelandiae*	6	4.3
Te Papa	OR004737	MS10999	*Himantopus novaezelandiae*	Pre-1950s		Skin		*Himantopus novaezelandiae*	36.1	11.5
Te Papa	OR002291	MS11000	*Himantopus novaezelandiae*	Pre-1950s		Skin		*Himantopus novaezelandiae (sub-adult)*	13.4	13.4
Te Papa	DM622-S	MS11001	*Himantopus novaezelandiae*	Jan 1959	Manukau Harbour, NI	Skeleton			10.6	10.6
Te Papa	OR22726	MS11002	*Himantopus himantopus leucocephalus* x *Himantopus novaezelandiae*	28 Oct 1981	Pauahatanui, NI	Skin	F	*Himantopus himantopus leucocephalus* x *Himantopus novaezelandiae*	21.9	21.9
Te Papa	OR014187	MS11003	*Himantopus novaezelandiae*	Post-1950s		Mounted		*Himantopus novaezelandiae*	8.3	8.3
Te Papa	OR010938	MS11004	*Himantopus novaezelandiae*	Post-1950s		Mounted		*Himantopus himantopus leucocephalus* x *Himantopus novaezelandiae*	9.2	9.2
Te Papa	OR002296	MS11005	*Himantopus himantopus leucocephalus*	May 1843		Skin		*Himantopus himantopus leucocephalus*	22.1	7.7
Te Papa	OR002292	MS11006	*Himantopus himantopus leucocephalus*	Pre-1950s		Skin		*Himantopus himantopus leucocephalus*	13.2	13.2
Te Papa	OR004736	MS11007	*Himantopus himantopus leucocephalus*	Pre-1950s	Nelson, SI	Skin		*Himantopus himantopus leucocephalus*	14.1	14.1
Te Papa	OR021957	MS11008	*Himantopus himantopus leucocephalus*	28 Sep 1977	Napier, NI	Skin	F	*Himantopus himantopus leucocephalus*	17.7	17.7
Te Papa	OR022693	MS11009	*Himantopus himantopus leucocephalus*	4 Sep 1978	Greymouth, SI	Skin	F	*Himantopus himantopus leucocephalus*	24.1	9.7
Te Papa	OR027440	MS11010	*Himantopus leucocephalus* x *novaezelandiae*	13 Jun 1995	Woodlands, Invercargill, SI	Skin	F	*Himantopus himantopus leucocephalus* or very light hybrid	29	8.3
Te Papa	OR022656	MS11011	*Himantopus himantopus leucocephalus*	14 Jul 1928	Paraparaumu, Waikanae River estuary, NI	Skin	F	*Himantopus himantopus leucocephalus*	13.2	13.2
Te Papa	OR029290	MS11012	*Himantopus himantopus leucocephalus*	21 Jan 2011	Te Whanga Lagoon, Chatham Island	Skin	M	*Himantopus himantopus leucocephalus*	17.1	17.1
Te Papa	OR026807	MS11013	*Himantopus novaezelandiae* x *leucocephalus*	19 Dec 2000	Twizel, SI	Skin			< 1	< 1

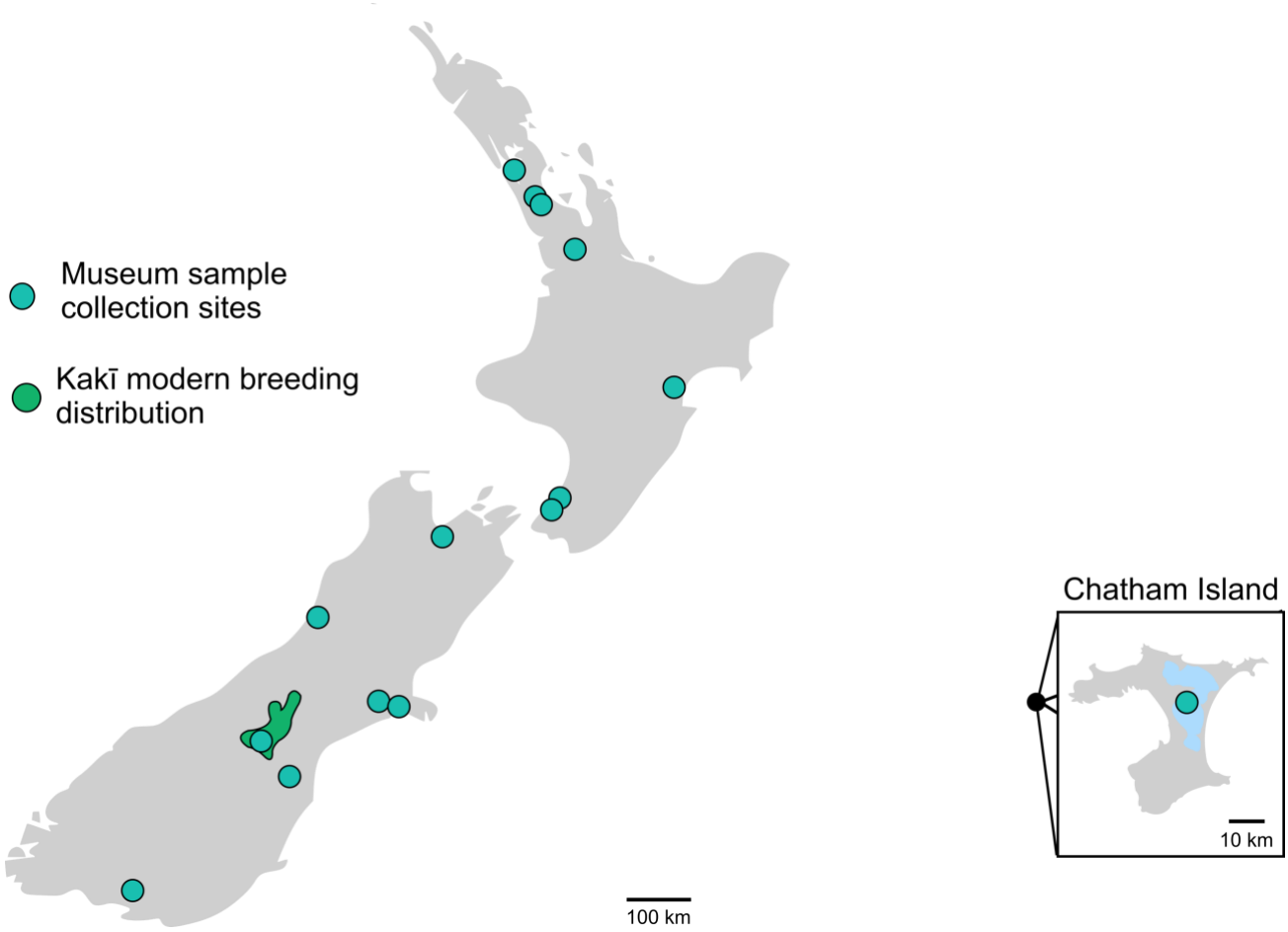


Figure 3.2: Location of origin of museum stilt specimens for specimens with recorded collection data (n = 18).

3.3.3 Mitochondrial primer design

To identify conserved regions that could be used as mitochondrial primer sites for amplification of stilt mitogenomes, in the absence of a reference mitogenome for kakī or Australian pied stilts, I constructed a conordinal proxy mitogenome from the consensus alignment of complete mitogenomes for six species within the Order Charadriiformes produced with MUSCLE v3.8.425 (Edgar, 2004) implemented in Geneious® v11.1.5. This consensus included two mitogenomes from pied avocet (*Recurvirostra avosetta*; NCBI Accession No.: KY623657.1 and NC_027420.1), one black-winged stilt (*Himantopus himantopus*; NC_035423.1), one blackish oystercatcher (*Haematopus ater*; AY074886.2), one Eurasian oystercatcher

(*Haematopus ostralegus*; NC_034237.1), and one long-billed plover (*Charadrius placidus*; KY419888.1). I identified potential primer sites in regions conserved between species using Invitrogen™ Primer3-based OligoPerfect™ in the Thermo Fisher Cloud with default parameters, with target regions customised to produce four sets of primer pairs spread across the mitogenome (Figure 3.3). Splitting the mitogenome into four sections for amplification reduces the likelihood of amplifying nuclear pseudogenes and maintains the amplification efficiency across these long fragments. These long fragments were amplified from modern samples for direct sequencing following library preparation (Section 3.3.4), and were also used as baits for mitogenome capture from historic samples (Section 3.3.6). To identify a set of four primer pairs that would span the complete mitogenome with sufficient overlap between fragments (minimum 50 bp overlap), twenty potential primer pairs were selected and manually assessed. Selected primer pairs were then assessed with the Thermo Fisher Multiple Primer Analyzer tool to estimate melting temperatures and potential for self- or cross-primer dimerisation. Primer locations were later manually confirmed against the *de novo* assembled kakī mitogenome.

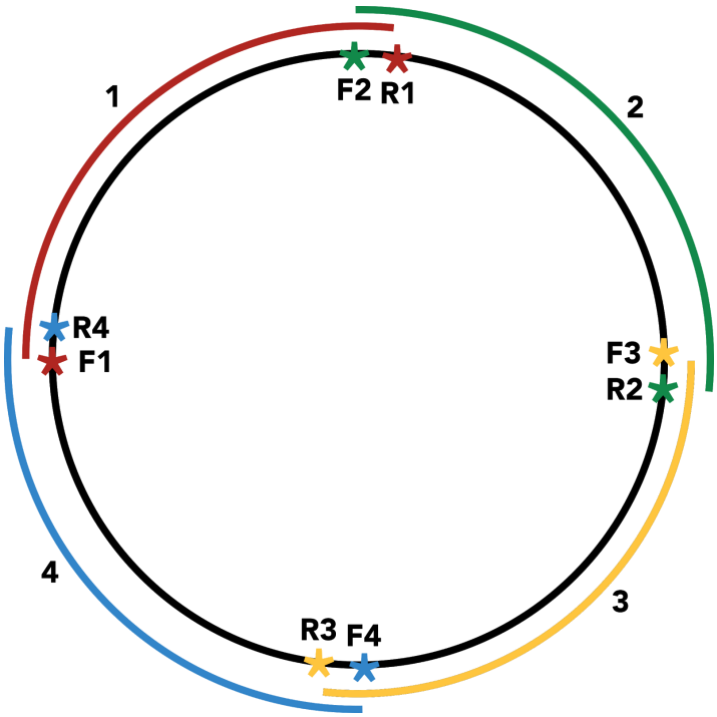


Figure 3.3: Example of placement of mitochondrial primers to amplify the four regions comprising the complete stilt mitogenome. The inner black circle represents the stilt mitogenome. Stars represent approximate positions of primers. F = forward primer, R = reverse primer.

3.3.4 Modern sample DNA preparation and sequencing

I used genomic DNA isolated for associated whole-genome sequencing and reduced-representation sequencing (see Chapters Two and Four) for initial primer testing and optimisation. Seven modern kakī samples (DNA1010, DNA1044, DNA1170, DNA1215, DNA1325, DNA1525, and DNA1691) were selected for primer optimisation based on high gDNA quantity and quality as assessed via a NanoDrop™ 8000 Spectrophotometer and via a Qubit dsDNA High Sensitivity Assay kit on a Qubit® 2.0 Fluorometer (Invitrogen). In initial primer optimisation, each 50 µL PCR consisted of 10 µL 5× KAPA magnesium-free Long Range buffer (KAPA Biosystems), 1.75–2.25 mM $MgCl_2$, 0.3 mM dNTPs, 0.5 µM of each forward and reverse primer, 0.5 U KAPA Long Range Hotstart Polymerase, and 2 µL 1:10 diluted template DNA. Thermocycling conditions incorporated a gradient to test annealing

temperatures (T_A) of 48–66°C, with initial denaturation of 94°C for 180 s, followed by ten cycles of 94°C for 25 s, T_A°C for 15 s, and 68°C for 60 s per expected kilobase (either 4 kb or 6 kb) of fragment length. This was followed by 25 cycles of 94°C for 25 s, T_A°C for 15 s, and 68°C for 60 s per expected kilobase plus 20 s per cycle, followed by a final extension step of 72°C for 60 s per expected kilobase. The final reaction mix used 1.75 mM $MgCl_2$, and T_A = 54°C for all reactions.

Following primer optimisation, modern stilt individuals were selected for targeted mitochondrial amplification and sequencing. Individuals included one kakī (DNA1044), two Australian pied stilts (B60406 and B60480), three poaka (B40279, B50004, and the individual from Hawke's Bay, DNA ID Poaka1), and two hybrid individuals (DNA 2094 and DNA 2113) with node A plumage that had been identified as non-kakī via microsatellite genotyping as part of day-to-day protocols by the Kakī Recovery Programme. The four mitochondrial regions were amplified for each modern sample using the optimised reaction mix and thermocycling protocol, before being purified using a QiaQuick® PCR Purification Kit (QIAGEN), with binding buffer (PB) equal to 5✕ the reaction volume, eluted into 30 µL elution buffer. Purified products were visualised on a 1% agarose gel at 80 V for 45 min, and quantified via Qubit.

I produced double-stranded barcoded libraries for all samples following Kircher et al. (2012). The amplified products were pooled for each sample and sonicated (nine cycles of 15 s on, 45 s off). Successful shearing of products was confirmed via gel visualisation. T4 Polynucleotide Kinase and T4 Polymerase (New England Labs Inc.) were used for blunt end repair, followed by purification using a MinElute kit, with products eluted in 30 µL elution buffer. A-tailing was performed with 1.7 µL Klenow (New England BioLabs Inc.) using 25 µL of the purified product, 3 µL NEB#2 Buffer, and 0.3 µL dATP, incubated for 30 min at 37°C, and then purified. Illumina-

compatible double-stranded adapters (Adapter_P5 and Adapter_P7) were ligated to purified products (Knapp et al., 2012b), using a common P5 adapter and a unique P7 adapter for each sample, via incubation for 1 hour at 22°C, and then purified and eluted. High-fidelity amplifications used the KAPA HiFi PCR Kit (KAPA Biosystems), with the reaction consisting of 1× KAPA HiFi Buffer (with Mg^{2+}), 0.3 mM dNTPs, 0.3 µM each of the P5 and P7 extension primers, 0.6 U KAPA HiFi DNA Polymerase, and 19 µL PCR product. Thermocycling consisted of 94°C for 3 min, twenty cycles of 94°C for 25 s, 60°C for 15 s, and 72°C for 15 s, followed by an extension step at 72°C for 15 min. Amplified products were purified as previously, and quantified via Qubit. Library fragment sizes were visualised using an Agilent 2100 BioAnalyzer. Libraries were pooled with equimolarity. An additional clean-up step was performed to improve fragment size distribution, adding 2 µL AMPure XP beads (Agencourt) to 20 µL pooled library, and incubated for 10 min at room temperature. The suspension was placed on a magnetic rack, and the supernatant was transferred to a new tube, before adding 1.8 times the volume of beads, and incubating again for 5–10 min. The supernatant was removed, and the beads washed with 200 µL 90% EtOH, before pooled DNA being eluted into 20 µL 0.1× Tris-EDTA (TE). Final DNA quantity was confirmed via Qubit. The pooled library was sequenced at the Otago Genomics and Bioinformatics Facility (OGBF) on one lane of an Illumina MiSeq Nano run, with 2 x 250 bp paired-end sequencing.

3.3.5 Mitogenome assembly from whole-genome sequencing data

I assembled a high quality kakī mitogenome from the whole-genome sequencing data of one kakī individual (DNA1914, the same individual used to generate the kakī reference genome in Chapter Two) to be used as a reference for sequence mapping and downstream analyses. I trialled four mitochondrial genome assembly pipelines:

MITObim v1.9 (Hahn et al., 2013), BioBloom Tools v2.1.0 (Chu et al., 2014), NORGAL v1.0 (Al-Nakeeb et al., 2017), and NOVOPlasty v3.4 (Dierckxsens et al., 2016). All mitogenome assembly tests used the congeneric black-winged stilt mitochondrial genome (*Himantopus himantopus*, GenBank Accession No.: KY623656.1) as a reference for mapping or baiting steps. I assessed the outputs of all assemblers for similarity to the black-winged stilt mitogenome using BLAST+ (Altschul et al., 1990).

MITObim performs mitochondrial baiting and iterative mapping to produce a mitogenome from Illumina short reads (Hahn et al., 2013). Preliminary testing using whole-genome sequence data with the MITObim pipeline produced intermediate outputs in excess of 1 TB, the maximum space available on the local computing cluster. Thus, I subsampled and interleaved 20% of the total raw forward and reverse whole-genome sequences as input for MITObim. Illumina input data was specified, with a template size of 100–500 bp estimated, set for automatic refinement. Following the initial baiting, I ran ten iterations of the subsequent baited mapping.

BioBloom categorises sequences against a reference, producing a sequence set that can then be assembled (Chu et al., 2014). Bloom filters were generated from the black-winged stilt and pied avocet (*Recurvirostra avosetta*, GenBank KY623657.1) mitogenomes. Raw sequence reads were passed to the categoriser in paired-end mode. I manually assessed summary outputs, and assessed the longest sequences for avian mitochondrial origin with BLAST+.

The NORGAL pipeline acts as a wrapper for several programmes, and does not require a reference. It first produces a whole genome assembly, looks for contigs with high sequence coverage that are more likely to be mitochondrial in origin, and performs BLAST+ searches to assess mitochondrial origin (Al-Nakeeb et al., 2017).

It also uses annotation information to identify a cytochrome oxidase I (COI) sequence that would further support mitochondrial origin. Raw sequence reads were used as input, with no further parameters required.

The circular organelle assembler NOVOPlasty v2.7.1 (Dierckxsens et al., 2016), uses a reference sequence as seed for initial mapping of reads, and then extends this to produce the complete circular organelle. Following sequence processing as described in Chapter Two, whole-genome sequences from kakī DNA1914 were passed to NOVOPlasty, with the black-winged stilt mitochondrial genome used as seed and reference to initiate assembly. Read length of 150 bp and an average insert size of 350 bp were specified, with the 'use quality scores' flag. I estimated the mitogenome size based on published mitogenomes of confamilial species at 16–20 kb, and an estimated *k*-mer of 39 was used based on that used in whole-genome assembly (Chapter Two). Output contigs were manually assessed, and overlapping regions between contigs were merged together to produce a single contig. A BLAST+ search of the NOVOPlasty mitogenome assembly was conducted to confirm mitochondrial origin. As NOVOPlasty assembly was the most successful method, cleaned whole-genome sequence data from DNA1914 was mapped against this draft mitogenome assembly to generate a consensus mitogenome sequence and correct errors. I annotated the resulting mitogenome annotation with the MITOS pipeline (Bernt et al., 2013). Scripts associated with the NOVOPlasty assembly and all following workflows are detailed in my GitHub 'Himantopus' repository at https://github.com/natforsdick/Himantopus.

3.3.6 DNA preparation and sequencing from museum samples

I conducted all gDNA extractions and library preparations of museum samples in a dedicated aDNA facility at the University of Otago using strict protocols to minimise

contamination (Knapp et al., 2012a). I used a QIAGEN DNeasy® Blood and Tissue kit (QIAGEN) spin-column protocol for tissue for gDNA extractions, with an overnight digestion step. I extracted gDNA in batches of 6–7 samples. To assess potential contamination, I processed one negative control in parallel with each batch. I made two elutions for each sample, eluting gDNA into 50 µL elution buffer for each elution. I prepared double-stranded libraries for all samples and negative controls following the methods in Greig et al. (2015) for in-solution hybridisation capture and paired-end sequencing on an Illumina platform. An additional DNA-free negative control was processed alongside each batch of samples. Between each step, libraries were purified over a MinElute PCR Purification Kit (QIAGEN) according to the manufacturer's specifications, with two PE wash steps. First, blunt-end repair using T4 Polynucleotide Kinase and T4 Polymerase was conducted. Purified products were eluted in 20 µL 0.1× TE + 0.05% Tween. Illumina-compatible adapters, Sol_adap_P5 and Sol_adap_P7-BIO (biotinylated at the 5' end) were ligated to purified products, and purified products were eluted in 25 µL 1× TE. Streptavidin-coated magnetic beads were prepared, adapter-ligated libraries were fixed to the beads and the beads were thoroughly washed to remove unincorporated adapters. Adapter-ligated libraries were denatured from the beads and eluted in 20 µL 0.1× TE. Libraries were then prepared for quantitative PCR to determine the appropriate number of cycles to minimise the production of chimeras in the subsequent PCR amplification step (Judo et al., 1998; Meyerhans et al., 1990; Odelberg et al., 1995; Thompson et al., 2002). Each qPCR consisted of 1× SYBR Green Master Mix (Thermo Fisher Scientific, Inc.), 0.25 µM of each primer (Sol_quant_P5 and Sol_quant_P7), 1µL template DNA, and ultrapure water to make up a total volume of 25 µL. qPCR amplification was conducted using a QuantStudio 3 instrument (Applied Biosystems), and consisted of 95°C for 10 min, and forty cycles of 94°C for

30 s, 58°C for 30 s, 72°C for 60 s, followed by an extension step of 72°C for 10 min. I recorded the cycle number at which amplification plateaued for each sample, and visualised 10 µL of each qPCR product on a 2% agarose gel run at 100 V for 1 hour.

Libraries were double-indexed with one of two distinct P5 barcoding primers (Sol_prim_ext_P5) and a unique P7 (Sol_prim_ext_p7) barcoding primer for each sample. Indexing reactions consisted of 1× Taq Buffer, 0.5 mM $MgCl_2$, 1 mM dNTPs, 0.2 µM of each identifying P5 and P7 primer, 3.75 U AmpliTaq Gold polymerase (Thermo Fisher Scientific, Inc.), with 19 µL prepared library, and ultrapure water added to a total reaction volume of 50 µL. The prepared reaction mix was transported from the ancient DNA lab to the modern lab facility for amplification. All PCR amplifications were conducted in a modern lab facility using a SensoQuest Labcycler (Dnature). The thermocycling protocol consisted of 95°C for 12 min, 15–20 cycles (determined by previous qPCR) of 94°C for 30 s, 58°C for 30 s, and 72°C for 60 s, with a final extension step of 72°C for 10 min. Following purification, I eluted libraries in 20 µL 0.1× TE with 0.05% Tween. Libraries were further amplified prior to target enrichment via hybridisation capture. High fidelity amplification used 1× KAPA HiFi Buffer, 0.3 mM dNTPs, 0.3 µM of each Sol_amp_p5 and Sol_amp_p7 primer, 1 U KAPA HiFi DNA Polymerase, 1 µL prepared library, and ultrapure water up to a total volume of 50 µL. Amplification consisted of 94°C for 5 min, ten cycles of 94°C for 20 s, 55°C for 55 s, 72°C for 15 s, and an extension step of 72°C for 5 min. Amplified products were purified and eluted in 20 µL 0.1× TE. I quantified libraries on a NanoDrop™ 8000 Spectrophotometer to ensure sufficient DNA quantities (minimum 2 µg) for hybridisation capture.

To ensure successful capture for all historic samples comprising kakī, poaka, and interspecific hybrids, I prepared a combined stilt mitochondrial bait for hybridisation

capture from one modern kakī (DNA1044) and one modern poaka (B40279) following a modified protocol based on Maricic et al. (2010). To produce sufficient bait for targeted capture of all prepared museum stilt libraries, three long-range amplifications were performed for each of the four mitochondrial primer sets for both samples, using the optimised reaction mix with 1 µL template DNA and standard thermocycling conditions. Amplified products were purified using a QiaQuick PCR Purification kit (QIAGEN), quantified using a NanoDrop™ 8000 Spectrophotometer and visualised on a 1% agarose gel run at 80 V for 45 minutes. Long-range products were sonicated to ~500 bp using a Bioruptor® Pico (Diagenode, Inc.), with volumes standardised to 100 µL, run for 9 cycles at 4°C with 15 s on, 45 s off. Successful shearing was confirmed by running sheared and non-sheared products on a gel for comparison. I quantified sheared products with a NanoDrop™ 8000 Spectrophotometer before pooling all products from both individuals in equimolar amounts with a minimum total requirement of 1.3 µg per prepared library. Blunt-end repair and phosphorylation of the long-range pooled products used T4 Polynucleotide Kinase and T4 Polymerase (both New England BioLabs Inc.) with an incubation of 12°C for 15 min and 25°C for 15 min. I purified products through MinElute silica spin columns and eluted them in 15 µL 0.1× TE. I ligated biotinylated (Bio-T/B) adapters (Maricic et al., 2010) to blunt-end repaired fragments by incubation at 22°C for 1 hour, before purification and elution in 15 µL 0.1× TE. To confirm a minimum DNA quantity of 500 ng per prepared historic library, I quantified the ligated products via a NanoDrop™ 8000 Spectrophotometer.

I conducted hybridisation of libraries to the modern DNA bait following Maricic et al. (2010) with modification. Briefly, I aliquoted 500 ng of the prepared bait with an equal volume of BWT (Bind and Wash and Tween Buffer; 2 M NaCl, 10 mM Tris (pH 8.0), 1 mM EDTA (pH 8.0), 0.1% Tween) and denatured the mixture for 1 min at 98°C. I

then added single-stranded bait to prepared streptavidin-coated beads, and rotated this at room temperature for 20 min to allow for binding. The solution was placed on a magnetic rack and beads were washed twice with 1× BWT at 50°C to remove unincorporated bait. Beads were then resuspended in 50 µL 1× TE with 0.05% Tween, and stored at 4°C until hybridisation capture. To prepare libraries for hybridisation capture, I added approximately 2 mg DNA of each prepared library to 0.9× Agilent hybridisation buffer, 0.9× Agilent blocking agent, and 1.8 µM of each of eight blocking oligonucleotides. Libraries were made single-stranded by denaturing at 95°C for 3 min, followed by 37°C for 30 min. I removed the buffer from the baited beads, and resuspended the beads in the hybridisation mix. To hybridise the libraries to the bait, the bead mix was rotated at 12 rpm for 48 hours in a 65°C hybridisation oven. Following hybridisation, I performed a series of three bead washes with 1× 200 µL BWT, two washes with 200 µL HWT (2.5 mM $MgCl_2$, 50 mM KCl, 15 mM Tris-HCl (pH 8.0), 0.1% Tween) at 60°C, one wash with 1× BWT, and a final wash with 100 µL 1× TE with 0.05% Tween. I resuspended libraries in 15 µL 0.1× TE, and detached the libraries from the beads at 95°C for 3 min. Captured libraries were transferred to fresh siliconised tubes and stored at -20°C.

I conducted a final post-capture amplification of historic libraries prior to sequencing. The reaction mix consisted of 1× KAPA HiFi Buffer (with Mg^{2+}), 0.3 mM dNTPs, 0.3 µM of each of the amplification primers Sol_amp_p5 and Sol_amp_p7, 1 U KAPA HiFi DNA Polymerase, and 15 µL captured library. Thermocycling conditions consisted of 94°C for 5 min, twenty cycles of 94°C for 20 s, 55°C for 55 s, 72°C for 15 s, and a final extension of 72°C for 5 min. I purified the captured libraries and eluted them in 20 µL 0.1× TE. Final prepared libraries were quantified via Qubit, and pooled with equimolarity. I concentrated the pooled library through a MinElute spin

column with a single PE wash step, and eluted this into 20 µL 0.1× TE before re-quantification. I assessed the fragment length distribution of the pooled library with a QIAxcel Advanced (QIAGEN) using a DNA High Resolution kit and QIAxcel ScreenGel® software. The final pooled library was diluted to 10 nM and sequenced on one lane of MiSeq v2 with 2 × 75 bp paired-end sequencing. Negative controls, having been processed in the same manner alongside individual samples, were pooled for sequencing with negative controls from three other aDNA projects within the lab, comprising a total of 29 negative controls. The fragment length of the pool was assessed with the QIAxcel Advanced. To avoid the failure of the negative sequencing run due to potential lack of cluster formation associated with the low quantities and short fragments of DNA present, the negative pool was sequenced on one lane of a MiSeq Nano 2 × 75 bp paired-end sequencing run with a PhiX spike of 5–10%.

3.3.7 Modern sample bioinformatic processing

Modern samples were demultiplexed as part of the OGBF in-house pipeline, and sequence quality was confirmed via FastQC v0.11.5 (Andrews, 2010). To increase the sample size for kakī, I incorporated whole-genome resequencing data for 24 additional modern kakī generated as part of an aligned project (Galla et al., 2019) in downstream analyses. These data included sequencing from twelve kakī expected to be particularly diverse based on pedigree data (Galla et al., 2020).

I processed the modern stilt targeted mitochondrial data and modern kakī resequencing data independently through an in-house pipeline to produce complete mitochondrial genomes for each individual. The kakī mitogenome assembly was first indexed with BWA v0.7.17 (Li & Durbin, 2009), and sequence read-group information was collected from raw sequence reads. AdapterRemoval v2.1.7 (Schubert et al.,

2016) was used to remove adapters, collapse paired-end reads with a maximum mismatch of ⅓, trim low quality bases (Q < 20) and Ns from the end of reads, and remove trimmed reads shorter than 25 bp. Paired-end reads were aligned with BWA-MEM v0.7.17 with read-group information attached and shorter read hits marked as secondary, and sorted by reference coordinate with Picard v2.18.0 *SortSam* (Picard Toolkit, 2019). SAMtools v1.7 (Li et al., 2009) *merge* was used to merge the aligned BAM files produced from the collapsed and uncollapsed reads. The alignment was indexed, and the total numbers of mapped and unmapped reads were collected with *idxstats*. Picard *CleanSam* was used to fix mapping quality scores to 0 for unmapped reads and to soft-clip overhanging reads. Unmapped reads were removed with SAMtools *view*, and duplicates were removed with Picard *MarkDuplicates*. SAMtools *depth* was used to generate mapping depths across the reference mitogenome for each sample, which were visualised using a custom R script. Histograms of insert sizes were generated from the merged sorted BAM files for each sample using Picard *CollectInsertSizeMetrics*.

3.3.8 Historic sample bioinformatic processing

Museum samples and negative controls were demultiplexed as part of the OGBF in-house pipeline, and sequence quality was confirmed via FastQC v 0.11.5 prior to processing these data independently with an in-house pipeline. This pipeline began in the same manner as the modern sample processing, with read-group information collected and quality trimming performed with AdapterRemoval v2.2.2. As shorter reads were produced from the historic samples compared to the modern samples, BWA-ALN was used to find the suffix array coordinates of the reads, with a maximum edit distance of 0.03, a maximum of 2 gap opens for each of the collapsed and uncollapsed outputs from AdapterRemoval. Paired-end reads were aligned to

the reference using BWA-SAMPE, while single-end reads were aligned with BWA-SAMSE. SAMtools v1.7 *view* was used to collect alignment counts for both the paired-end and collapsed reads, filtering alignments with mapping quality < 20, and producing BAM files. Alignments were then sorted and indexed with SAMtools *sort* and *index*. Removal of unmapped reads was confirmed with SAMtools *view*, and then duplicates were removed from both single- and paired-end reads with SAMtools *rmdup*, before indexing and collection of the total numbers of mapped and unmapped reads with SAMtools *index* and *idxstats*. Picard *MarkDuplicates* was used to confirm removal of PCR duplicates specified previously, and the PALEOMIX (Schubert et al., 2014) *rmdup_collapsed.py* script was used to ensure removal of duplicates from the collapsed reads. MapDamage v2.0.2 (Jónsson et al., 2013) was used to quantify patterns of DNA damage for historic samples (length of overhangs, nick frequency, and cytosine deamination) and rescale the quality scores appropriately. Picard *AddOrReplaceReadGroups* was used to add read-group information to BAM files, which were then indexed with SAMtools. Depth of coverage generated with SAMtools depth was visualised with an in-house R script. Read lengths were also visualised with an in-house R script. Samples were included in downstream analyses if they resulted in > 90% coverage across the mitogenome.

3.3.9 Variant calling

Following sequence mapping, I conducted variant-calling for each data set independently. The reference mitogenome was indexed with SAMtools v1.9 *faidx*, and a sequence dictionary created with Picard *CreateSequenceDictionary*. GATK v3.8.0 (McKenna et al., 2010) *HaplotypeCaller* was used to call variants for haploid mitochondrial data from individual sample BAM files (sorted, indexed BAMs with duplicates removed for modern samples, or the merged, sorted, rescaled BAM files

with duplicates removed for historic samples). Joint genotyping based on the group variant files was then conducted with GATK's *GenotypeGVCFs*. SNPs and indels were independently identified using *SelectVariants*, and independently filtered using *VariantFiltration*. Variants were filtered based on the VCF INFO fields, to exclude SNPs with quality by depth (QD) < 2.0, Fisher Strand (FS; Phred-scale probability of strand bias) > 60, MQ (root mean-square mapping quality) < 40, and to exclude indels with QD < 2.0, FS > 200, or ReadPosRankSum < -20.0 (indels found near ends of reads more often than expected). All SNPs and indels were then recombined with GATK *RecombineVariants*, with priority given to SNPs, and variants that had failed the filtering were removed with *SelectVariants*. The group VCF generated was then used to produce individual VCFs for each sample with GATK *SelectVariants*, excluding non-variant and filtered variants. GATK *DepthOfCoverage* was used to add coverage information for sample VCFs, and an in-house R script was used to exclude sites with coverage < 10× for all samples. GATK *FastaAlternateReferenceMaker* was then used to produce FASTA files from the filtered individual VCFs.

3.3.10 Bioinformatic processing of negative controls

I processed the eight negative controls prepared alongside historic samples independently of all other sequence data using the same pipeline as the historic samples. Instead of using the kakī mitogenome assembly as a reference, negative control sequencing was mapped against a composite FASTA containing mitochondrial genomes of species previously processed in the aDNA lab and other common reagent contaminants including cattle (*Bos taurus*, NC_006853.1), domestic chicken (*Gallus gallus*, NC_001323.1), domestic dog (*Canis lupus familiaris*, NC_002008.4), guinea pig (*Cavia porcellus*, NCBI Accession No.:

NC_000884.1), human (*Homo sapiens*, NC_012920.1), kakī (this chapter), kākā (*Nestor meridionalis*, Martini 2020), Pacific rat (*Rattus exulans*, NC_012389.1), and pig (*Sus scrofa*, NC_000845.1). An additional set of negative control sequences that did not match the list of individual barcodes provided at the time of sequencing was also processed. Where negative controls produced contaminating sequences mapping to kakī, the samples processed in the associated batch were excluded from all downstream analyses.

3.3.11 Bayesian phylogenetic analysis

To estimate the time to the most recent common ancestor (TMRCA) for kakī and Australian pied stilts and confirm their position within the Order Charadriiformes, I used mitogenomes with BEAST v2.6.0 (Bouckaert et al., 2019) to conduct Bayesian phylogenetic analyses in a two-phase approach (as in Morin et al. 2015; Morin et al. 2018). In Phase I, I estimated TMRCA for stilts by implementing a time-calibrated phylogenetic analysis for all Charadriiformes. I randomly selected one kakī and one Australian pied stilt individual from among the modern samples, and aligned these mitogenomes with forty other Charadriiformes mitogenomes available on GenBank, along with the chicken (*Gallus gallus*; NCBI Accession No.: MH732978) and Australasian grebe (*Tachybaptus novaehollandiae*, Order Podicipediformes; NCBI Accession No.: EF532936) as outgroups using MUSCLE implemented in Geneious® (Table 3.3). MUSCLE was selected as an appropriate alignment tool based on the number of samples and sequence lengths, and was utilised with default settings. I identified a potential duplication in the mitogenome of kakī and pied stilts from 14,242 bp onwards based on coverage plots of resequencing data (see Results, Figure 3.6), so alignments were truncated to 1–14,239 bp to exclude this anomalous region. The only conordinal mitogenome excluded from the available mitogenomes

was that of the Vega gull (*Larus vegae*), due to known taxonomic issues (Yang et al., 2017) and preliminary maximum likelihood analysis ascribing it as sister to Charadriidae rather than within the Laridae clade (results not shown).

*Table 3.3: List of species included in Bayesian and maximum likelihood analyses. All species are in the Order Charadriiformes except the outgroups of Australasian grebe (Podicipediformes) and the domestic chicken (Galliformes). * = representative kakī and Australian pied stilts samples, with the DNA ID as the identifier. The mitogenome of Australian pied stilt has not been independently assembled, but mitogenome size is estimated at 17,300–17,600 bp based on the mitogenomes of black-winged stilts and kakī (but see Appendix D).*

Family	Species	Common name	GenBank ID	Mitogenome size (bp)
Alcidae	*Pinguinus impennis*	Great auk	KU158188.1	16,784
	Synthliboramphus antiquus	Ancient murrelet	AP009042.1	16,730
	Synthliboramphus wumizusume	Japanese murrelet	NC_029328.1	16,714
Charadriidae	*Charadrius placidus*	Long-billed plover	KY419888.1	16,895
	Pluvialis fulva	Pacific golden plover	NC_033966.1	16,854
	Vanellus cinereus	Grey-headed lapwing	NC_025514.1	17,074
	Vanellus vanellus	Northern lapwing	NC_025637.1	16,795
	Charadrius alexandrinus	Kentish plover	NC_041118.1	16,905
Haematopodidae	*Haematopus ostralegus*	Eurasian oystercatcher	NC_034237/1	16,798
	Haematopus ater	Blackish oystercatcher	AY074886.2	16,791
Jacanidae	*Hydrophasianus chirurgus*	Pheasant-tailed jacana	NC_039096.1	16,855
	Jacana jacana	Wattled jacana	KJ631049.1	16,975
	Jacana spinosa	Northern jacana	KJ631048.1	17,079
Laridae	*Chroicocephalus ridibundus*	Black-headed gull	KM577662.1	16,807
	Saundersilarus saundersii	Saunder's gull	KJ631624.1	16,739
	Chroicocephalus brunnicephalus	Brown-headed gull	JX155863.1	16,769
	Ichthyaetus relictus	Relict gull	KC760146.1	16,586
	Larus crassirostris	Black-tailed gull	KM507782.1	16,746
	Larus dominicanus	Kelp gull	AY293619.1	16,701
	Sternula albifrons	Little tern	KT350612.1	16,357
	Gelochelidon nilotica	Gull-billed tern	NC_036344.1	16,748
	Sterna hirundo	Common tern	NC_046345.1	16,707

Table 3.3 cont.: List of species included in Bayesian and maximum likelihood analyses.

Family	Species	Common name	GenBank ID	Mitogenome size (bp)
Recurvirostridae	*Recurvirostra avosetta*	Pied avocet	KY623657.1	16,856
	Himantopus himantopus	Black-winged stilt	NC_035423.1	17,378
	*Himantopus himantopus leucocephalus**	Australian pied stilt	B60480	
	*Himantopus novaezelandiae**	Kakī	DNA451	17,566
Scolopacidae	*Arenaria interpres*	Ruddy turnstone	AY074885.2	16,725
	Calidris ruficollis	Rufous-necked stint	NC_040990.1	16,860
	Phalaropus lobatus	Red-necked phalarope	KY765409.1	16,714
	Limosa lapponica baueri	Bar-tailed godwit	KX371106.1	16,732
	Tringa glareola	Wood sandpiper	NC_039096.1	16,804
	Tringa guttifer	Nordmann's greenshank	NC_044665.1	16,835
	Tringa nebularia	Common greenshank	NC_044651.1	16,682
	Tringa totanus	Common redshank	NC_044648.1	16,818
	Eurynorhynchus pygmeus	Spoon-billed sandpiper	KY434065.1	16,709
	Xenus cinerus	Terek sandpiper	KX644890.1	16,817
	Tringa ochropus	Green sandpiper	NC_033974.1	16,904
	Tringa semipalmata	Willet	MF036175.1	16,603
	Gallinago stenura	Pin-tailed snipe	KY888681.1	18,153
	Numenius phaeopus	Whimbrel	KP308149.1	17,091
	Scolopax rusticola	Eurasian woodcock	KM434134.1	16,984
Stercorariidae	*Stercorarius maccormicki*	South polar skua	KM401546.1	16,669
Order Podicipediformes, Family Podicipedidae	*Tachybaptus novaehollandiae*	Australasian grebe	EF532936.1	18,002
Order Galliformes, Family Phasianidae	*Gallus gallus*	Domestic chicken	MH732978.1	16,785

To determine the appropriate site partitioning scheme, the alignment was passed to PartitionFinder2 (Lanfear et al., 2017) with four models assessed (GTR, GTR+G, GTR+I+G, and JC+I+G) using corrected AIC for model selection and implementing the Greedy algorithm (Guindon & Gascuel, 2003; Lanfear et al., 2012). Based on PartitionFinder2's results (scheme AICc = 296810.201, scheme lnL = -148244.294), coding genes were partitioned into first-, second-, and third-codon positions; tRNAs into a combined partition of the first- and second-"codon" positions and a partition with the third-"codon" positions, and sRNAs were also partitioned into a combined partition of the first- and second-"codon" positions and a partition with the third-"codon" positions of the combined sRNAs. In reference to the tRNAs and sRNAs, it should be noted that the "codon" partitioning is a limitation of PartitionFinder2, as unlike protein-coding genes, there is not an *a priori* reason to expect similar site substitution patterns by "codon". Non-coding positions were not included, as these had been removed due to the putative duplicated region identified earlier. I used BEAUti v2.6.0 to generate input files for BEAST, linking tree and clock models across all partitions. The Gamma site model (with five categories) was used across all partitions. Initially, all partitions identified by PartitionFinder2 were allocated a separate site model, with GTR (Generalised Time Reversible) defined as the substitution model. However, the failure of initial runs to reach stationarity due to over-parameterization led to all tRNAs positions being included in a single partition and assigned the TN93 (Tamura & Nei, 1993) substitution model. GTR was used for all other partitions. The analysis was implemented with a relaxed log normal clock model to allow for among-lineage rate variation (Drummond et al., 2006), and using a calibrated Yule model so I could extract divergence time information (Heled & Drummond, 2012). To calibrate the analysis, I used log normal priors (mean in real space) based on fossil evidence (Smith, 2015): crown Charadriiformes (divergence

of Charadrii from the other Charadriiformes) was given a mean and minimum (using offsets) age of 41.3 Mya, and the divergence of skuas (Stercorariidae) and auks (Pan-Alcidae) a mean and minimum age of 34.2 Mya. The standard deviation of the log-transformed distribution for both priors was 1.25. Log normal priors were given for both calibration points (see Supplementary File 1 for final parameters). Two chains of 100 million states, logged every 1000 states were run. The first 50% of each run was discarded as burn-in, and then Tracer v1.7.1 (Rambaut et al., 2018) was used to confirm both runs had reached stationarity and convergence. To further assess convergence, maximum clade credibility trees were constructed for the two chains independently with TreeAnnotator v2.5.1, with a burn-in of 50%. Once convergence was confirmed, the two log and tree files were combined with LogCombiner v2.6.1 after removing the first 50% as burn-in. The combined annotated tree was visualised with FigTree v1.4.3.

In Phase II, a subsequent Bayesian analysis was conducted to assess congeneric relationships among kakī, poaka, hybrid individuals, and Australian pied stilts using parameter estimates obtained in Phase I to inform the analysis (see Supplementary File 1 and Supplementary File 2 for the final parameters). Complete mitogenomes for 49 stilts comprising 34 kakī, 8 poaka, 5 kakī-poaka hybrids, and 2 Australian pied stilts were aligned with MUSCLE using default settings in Geneious®. All individuals were classified by morphological identification rather than catalogue records, with the exception of two individuals from the Kakī Recovery Programme (DNA2094 and DNA2113, node A individuals, classified as hybrids). The stilt skeleton MS11001 catalogued as kakī was classified here as kakī. Partitions were implemented as in the conordinal Bayesian analysis, with exclusion of non-coding sites (i.e., trimming the total alignment to a length of 14,239 bp) due to the anomalous region observed in kakī. As for the conordinal analysis, the tRNA partition used the TN93 substitution

model (Tamura & Nei, 1993), and the GTR substitution model used for all other partitions. All site model parameters were derived from the Phase I conordinal analysis (see Supplementary File 1). I implemented a relaxed log-normal clock with free rates (Drummond et al., 2006) as the conordinal analysis indicated rate variation between the kakī and pied stilt lineages (see Results, Figure 3.7). As for the conordinal analysis, clock and tree models were linked across all partitions. However, as multiple individuals were sampled per species, the Coalescent Bayesian Skyline tree model (Drummond et al., 2005) was implemented instead of the Calibrated Yule tree model. The tree height was given a uniform prior with lower and upper bounds (0.4157–1.1633 Mya) based on the credibility interval for the divergence of kakī and Australian pied stilts in the conordinal analysis (see Results, Figure 3.8 and Supplementary File 3). Two chains of 10 million states logging every 1,000 states were run. Outputs were assessed for stationarity and convergence, combined using a burn-in of 10%, annotated, and visualised as for the conordinal analysis.

3.3.12 Maximum likelihood phylogenetic analysis

For comparison with the results of Bayesian analysis, I used maximum likelihood analysis to construct a phylogenetic tree for the Order Charadriiformes. First, I selected the optimal nucleotide-substitution model for the aligned mitogenomes using jModelTest v2.1 (Darriba et al., 2012; Guindon & Gascuel, 2003), with 88 candidate models and 11 substitution schemes, using the Bayesian Information Criterion (BIC) to select the most appropriate model, which was GTR+I+G (General Time Reversible model with I = proportion of invariable sites and G = Gamma distribution). I then used IQ-Tree v1.6.6 (Nguyen et al., 2015) to generate a maximum likelihood consensus tree with the nucleotide-substitution model selected,

and ultrafast bootstrapping for 10,000 bootstraps (Hoang et al., 2018). I visualised the consensus tree with FigTree v.1.4.3. To assess whether using the truncated whole genomes resulted in any substantial differences in taxonomic relationships within the order, I compared the output trees with previously published Charadriiformes phylogenies.

3.3.13 Haplotype network analysis

To assess haplotype diversity and differentiation among the sampled stilts, I produced Median Joining Networks of haplotypes in PopART (Bandelt et al., 1999; Leigh & Bryant, 2015). Traits blocks were created based on morphological species identification information (kakī, poaka, Australian pied stilts, kakī-poaka hybrids). I used TempNet (Prost & Anderson, 2011) to visualise the temporal separation of haplotype networks based on statistical parsimony. I classified samples as 'Historic' (catalogued prior to the early 1950s) or 'Modern' (catalogued after the 1950s). Samples without recorded collection dates were designated 'Historic' or 'Modern' based on museum ID number, where samples with IDs between OR000001 and OR008000 were known to be recorded in the database prior to the early 1950s, and those with collection numbers beyond OR8000 were catalogued more recently. Thus, samples MS11000, MS11001, MS11006, and MS11007 were designated 'Historic', while MS11003 and MS11004 were designated 'Modern'. I selected the 1950s as the cut-off between historic and modern samples because no individuals were known to have been collected in the period 1960–1975, so everything with collection dates post-1950s stems from the peak of the kakī decline towards the end of the 1970s. Individuals collected after 1975 are assumed to be representative of the diversity remaining among contemporary kakī.

3.3.14 Mitochondrial diversity and differentiation between species

I used DnaSP v6 (Rozas et al., 2017) to assess nucleotide diversity (π) across the truncated mitogenome. The alignment was specified as a haploid mitochondrial sequence and the 14,239 bp region was regarded as a single locus for diversity estimates. Overall statistics for the number of haplotypes, haplotype diversity, nucleotide diversity, and related metrics were produced, and then each species was assessed independently. All pied stilts were grouped for comparison with kakī, as sample size limited comparisons between Australian pied stilts and poaka. Following the visualisation of haplotype networks, I excluded individuals with haplotypes that did not correspond to their classified group (e.g., kakī skeleton MS11001) and those with discordant catalogue-morphology classifications from the species-specific estimates to avoid overestimating diversity metrics due to inclusion of interspecific individuals. The small number of samples for each morphological group over time was insufficient to consider comparative temporal diversity metrics.

3.4 Results

3.4.1 Historic sample morphological identification

Species designation of museum specimens based on plumage morphology produced several discrepancies between catalogued species identification and morphological species identification. Four individuals recorded as *H. h. leucocephalus* were identified as *H. novaezelandiae* based on plumage (MS10991, MS10992, MS10994, MS10995; Table 3.2, see Appendix C for photographs and further discussion relating to the discrepancies described here). Two of these individuals with some white breast feathers were likely sub-adult kakī, which may have complicated initial identification. One individual recorded as *H. novaezelandiae*

(MS11004) was identified by plumage as a kakī-poaka hybrid, and one individual recorded as a kakī-poaka hybrid (MS11010) was noted to be either *H. h. leucocephalus* or a very light (node A–C) hybrid (see Appendix C). Five of these samples entered the museum collections before the early 1950s, prior to the categorisation of stilts by plumage node described by Pierce (1984b). The remaining 19 specimens (excluding the skeleton MS11001) had concordant morphological and catalogued species identifications.

3.4.2 Mitogenome assembly from whole genome sequence data

All mitochondrial genome assemblers except NOVOPlasty produced either very short (< 2,000 bp) contigs, or sequences with low BLAST sequence similarity to the target mitochondrial genomes of pied avocet or black-winged stilt (e.g., NORGAL output mitochondrial genome contig was 1,655 bp, with 91.04% match to the pied avocet reference mitogenome). The NOVOPlasty mitogenome assembly from kakī DNA1914 WGS sub-sampled 32.01% of the input sequence reads, totalling 134 million reads. Of these, 9,106 reads were aligned to the reference black-winged stilt mitogenome, and 8,062 were subsequently assembled, producing two contigs of length 15,754 and 1,849 bp, with an average sequence coverage depth of 78×. Visual assessment of these contigs detected a region of overlap of 37 bp, and the two contigs were thus merged to produce a single circular mitochondrial genome of 17,566 bp (Figure 3.4), within the expected length for Charadriiformes mitogenomes (~16–18 kb; Table 3.3). Base composition was 31.91% A, 30.97% C, 13.22% G, and 23.82% T. BLAST results against the nucleotide database produced the top match to the black-winged stilt mitogenome, with 99% query cover and 99% identity, followed by 94% query cover and 93% identity to the confamilial pied avocet. Mitogenome

annotation with MITOS identified 22 tRNAs all with a regular clover-leaf shape, two rRNAs, and 13 protein-coding genes in the typical avian order.

3.4.3 Sequencing outputs

Primers were designed for long-read amplification of four regions comprising the complete kakī mitogenome (Table 3.4). Sequencing of targeted long-read amplified mitochondrial genomes from modern stilts was less successful than expected due to over-clustering at sequencing. An unexpectedly high proportion of short fragments libraries resulted in an underestimation of the concentration of the library pool, and subsequent loss of sequencing intensity beyond 150 cycles.

Thus, reads were demultiplexed to use only the first 150 cycles of sequencing (150 bp). Despite this, sufficient high-quality sequence data was produced for downstream analyses. An average of 254,904.5 ± SD 56,661.93 sequence reads were produced per sample (Table 3.5). All 27 museum samples produced sequence reads, with an average 353,738.22 ± SD 240,534.17 total sequence reads per sample, with read length 76 bp (Table 3.6). Sequence yield for samples MS10993, MS11008, MS11012 and MS11013 was very low (< 1000 reads per sample). In subsequent bioinformatic processing, mapDamage reports showed the expected damage patterns for DNA extracted from historic samples (see Figure 3.5 for an example of the mapDamage outputs, Appendix C for those of all samples). C-to-T misincorporations are increased at the 5’ sequence end, and correspondingly G-to-A misincorporations are more frequent at the 3’ end. The relatively recent ages of these samples results in less pronounced misincorporation frequencies than would be observed from ancient samples as these misincorporations increase over time (Ginolhac et al. 2011).

Sequencing of the lab group pool of 29 aDNA negative controls produced 1.3 Gb sequence data, including 68 Mb of sequence data with barcodes that had not been specified. Following demultiplexing, negative controls included in this study produced 4–35,254 sequence reads (Table 3.7). Only one negative control produced any substantial reads mapping to the kakī mitogenome (> 50 reads), with 16.23% of reads produced from DNA-ve2 (sequence ID OG5003-13-0-1) mapping to kakī, resulting in 97.7% coverage of the mitogenome, with 27× depth. Thus, samples prepared in the batch associated with this negative (n = 7) were excluded from downstream analysis. Among the sequence reads with barcodes not included in the sequence set, no barcodes had been used in sample preparation for this study, and only 133 (0.014%) of the resulting sequence reads mapped to the kakī genome.

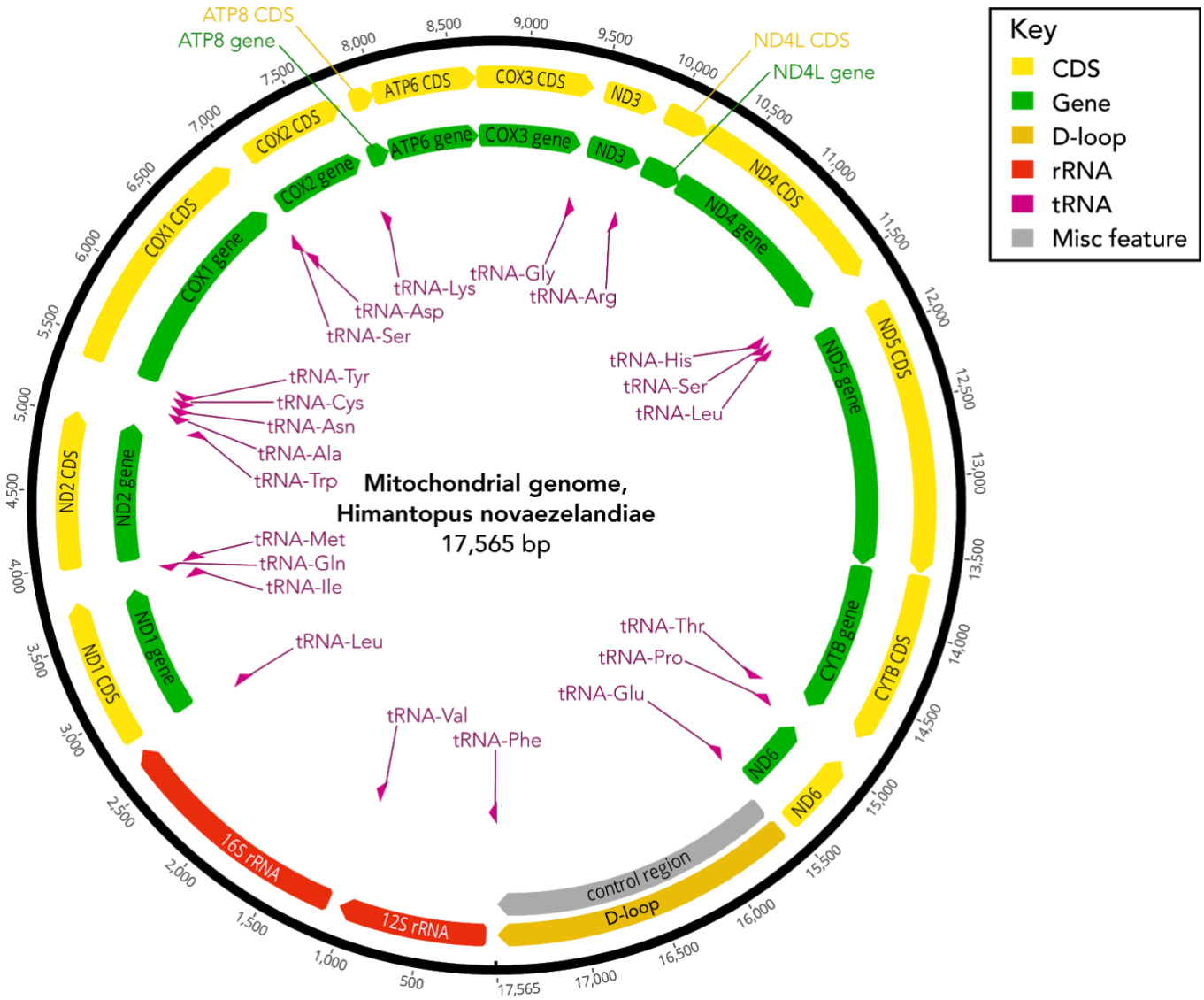


Figure 3.4: Visualisation of the draft mitochondrial genome for kakī Himantopus novaezelandiae, assembled from Illumina whole-genome sequencing of a single individual with NOVOPlasty v2.7.1. Gene features were annotated with MITOS and confirmed against the annotated black-winged stilt H. himantopus mitogenome. Green corresponds to gene regions, yellow represents coding regions, red shows rRNA regions, tRNA regions are in pink, the D-loop is in gold, with miscellaneous features in grey. The numbers on the outer ring correspond to the base position on the mitogenome. Arrows indicate the direction of transcription of features.

Table 3.4: Charadriiformes consensus-derived mitochondrial primer characteristics. Start positions and estimated region length are relative to the assembled kakī mitogenome. bp = base pairs, F = forward primer, R = reverse primer.

Primer	Forward primer sequence	Reverse primer sequence	Start position F (bp)	Start position R (bp)	Estimated region length (bp)
KakiM1	CAAACCCACCTAGAGGAGCC	GAGTGGTTTGATGCGGTTGG	630	6,981	6,300
KakiM2	TTTCAAGCCAACCGCATCAA	CTAGTTGGCTGGATGTGGAGAA	6,955	12,747	5,800
KakiM3	AAACCCCAACACTCCCCCTA	GGCCCTGACATAGGAACCAG	12,436	16,102	3,800
KakiM4	CTCCAACTCCCAAAGCTGGT	ACAGGCAACCAGCTATCACC	14,898	1,504	4,200

Table 3.5: Results of modern stilt targeted sequencing and read mapping to the kakī mitogenome assembly. bp = base pairs, Mb = megabase pairs.

DNA ID	Sequence ID	Barcodes	Average sequence length (bp)	Total sequence reads produced	Yield (Mb)	Mitogenome coverage (%)	Coverage depth (×)
DNA1044	OG4109-02	P5-1-P7-89	150	216,990	32.549	99.8	478.2
DNA2094	OG4109-03	P5-1-P7-90	150	188,528	28.279	99.7	367.6
DNA2113	OG4109-04	P5-1-P7-91	150	207,018	31.053	99.6	368.0
B40279	OG4109-05	P5-1-P7-92	150	204,766	30.715	99.7	401.7
B50004	OG4109-06	P5-1-P7-93	150	324,726	48.709	99.8	342.7
B60406	OG4109-07	P5-1-P7-94	150	289,896	43.484	99.6	470.4
B60480	OG4109-08	P5-1-P7-95	150	325,688	48.853	99.7	440.8
Poaka1	OG4109-09	P5-1-P7-96	150	281,624	42.244	99.7	360.9

Table 3.6: Sequence and mapping results for historic museum samples. Barcodes describe the sequencing barcode combination used for identification. Mb = megabase pairs, bp = base pairs.

Specimen ID	aDNA ID	Sequencing ID	Barcodes	Yield (Mb)	Average sequence length (bp)	% duplication	% GC content	Total sequences	Mitogenome coverage (%)	Coverage depth (×)
LB3413	MS10986	OG4983-01-0-1	P5-9-P7-4	48	76	19.93	48	628846	97.9	283.9
LB3946	MS10987	OG4983-02-0-1	P5-7-P7-13	44	76	25.56	47	579314	97.7	370.2
LB3414	MS10988	OG4983-03-0-1	P5-7-P7-14	27	76	32.01	46	361204	98.1	286.5
LB8517	MS10990	OG4983-04-0-1	P5-9-P7-5	20	76	12.61	48	261510	97.3	70.1
AV686	MS10991	OG4983-05-0-1	P5-9-P7-6	21	76	22.87	45	272700	98.3	196
AV683	MS10992	OG4983-06-0-1	P5-9-P7-13	15	76	17.58	46	203066	97.6	196
AV682	MS10993	OG4983-07-0-1	P5-7-P7-1	0	76	1.41	49	284	16.3	0.2
AV685	MS10994	OG4983-08-0-1	P5-7-P7-2	15	76	16.32	48	192450	97.1	112.5
AV1938	MS10995	OG4983-09-0-1	P5-9-P7-1	29	76	23.19	48	386008	97.7	159.8
OR002286	MS10996	OG4983-10-0-1	P5-7-P7-3	10	76	17.31	47	135734	97.6	84.9
OR002287	MS10997	OG4983-11-0-1	P5-7-P7-4	26	76	35.81	45	344110	98	335.4
OR002289	MS10998	OG4983-12-0-1	P5-7-P7-5	48	76	27.08	48	634050	97.9	331.8
OR004737	MS10999	OG4983-13-0-1	P5-7-P7-6	28	76	29.91	46	373666	98.1	261.9
OR002291	MS11000	OG4983-14-0-1	P5-9-P7-16	71	76	29.48	44	932626	98.7	501.3
DM622-S	MS11001	OG4983-15-0-1	P5-9-P7-17	27	76	24.11	50	355744	98	156.1
OR22726	MS11002	OG4983-16-0-1	P5-9-P7-18	16	76	16.81	45	214380	98.1	150
OR014187	MS11003	OG4983-17-0-1	P5-7-P7-15	45	76	24.44	47	586348	98	356.4
OR010938	MS11004	OG4983-18-0-1	P5-7-P7-16	51	76	24.32	46	673632	97.9	391.5
OR002292	MS11006	OG4983-20-0-1	P5-7-P7-17	22	76	26.86	45	287992	98.6	208.3
OR004736	MS11007	OG4983-21-0-1	P5-9-P7-22	25	76	18.39	48	325952	97.6	119.2
OR021957	MS11008	OG4983-22-0-1	P5-7-P7-18	0	76	1.48	47	270	13.5	0.2
OR022693	MS11009	OG4983-23-0-1	P5-7-P7-19	31	76	19.56	49	406058	97.9	147.8

Table 3.6 cont.: Sequence and mapping results for historic museum samples. Barcodes describe the sequencing barcode combination used for identification. Mb = megabase pairs, bp = base pairs.

Specimen ID	aDNA ID	Sequencing ID	Barcodes	Yield (Mb)	Average sequence length (bp)	% duplication	% GC content	Total sequences	Mitogenome coverage (%)	Coverage depth (×)
OR027440	MS11010	OG4983-24-0-1	P5-7-P7-20	31	76	26.29	47	404246	98	148.3
OR022656	MS11011	OG4983-25-0-1	P5-9-P7-23	19	76	19.49	46	244346	97.6	156.9
OR029290	MS11012	OG4983-26-0-1	P5-9-P7-24	0	76	2.77	48	794	30.8	0.4
OR026807	MS11013	OG4983-27-0-1	P5-9-P7-25	0	76	1.33	54	150	0	0

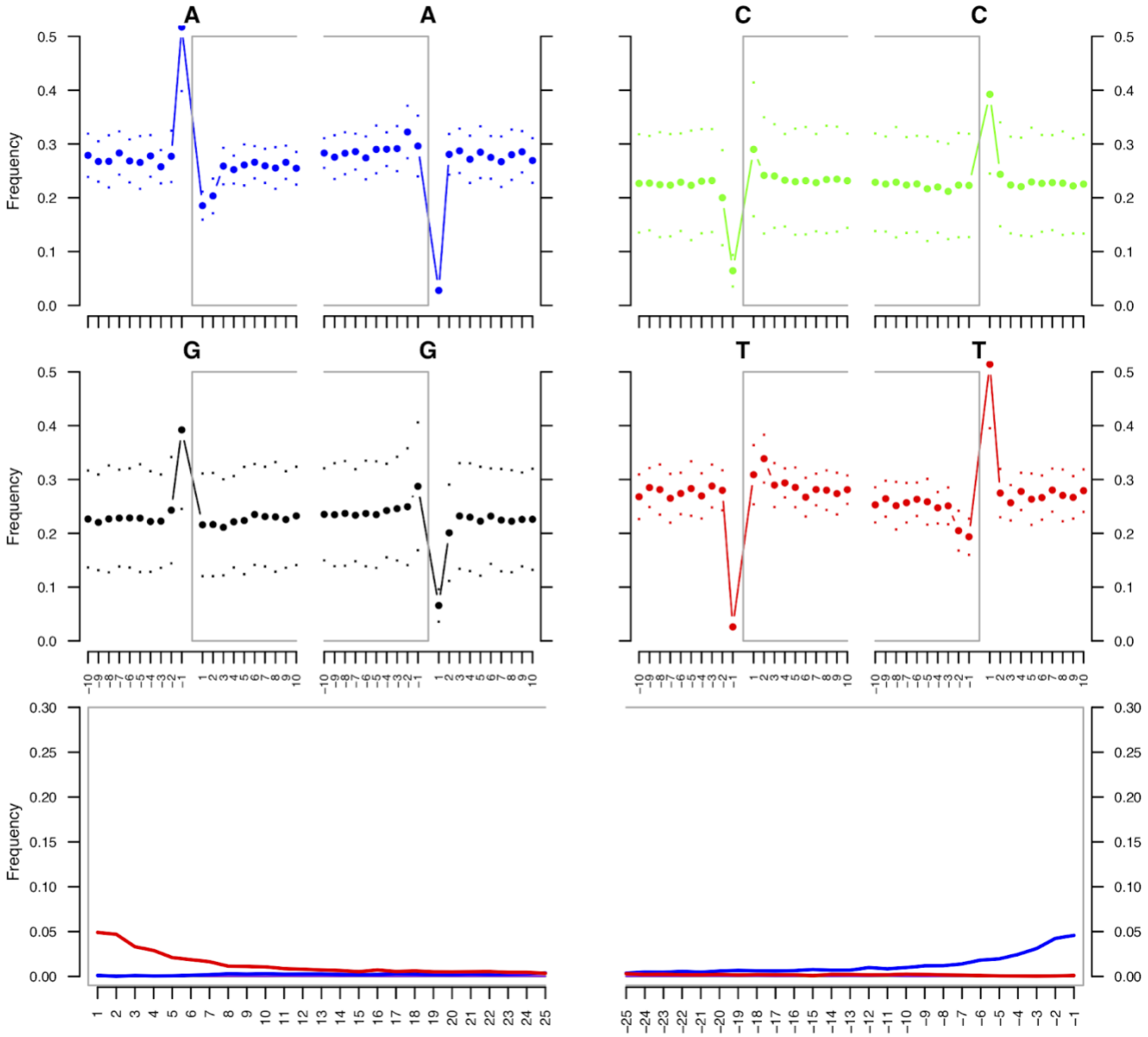


Figure 3.5: Example mapDamage base misincorporation results for individual aDNA ID MS10986. In the bottom plots, the red line represents the frequency of C-to-T substitutions while the blue line represents the frequency of G-to-A substitutions in the first and last 25 bp of sequence reads.

Table 3.7: Sequencing results of negative controls prepared alongside sample batches during initial DNA extraction (DNA-ve) and double-stranded library preparation (Neg). None of the barcodes identified among the sequence pool corresponded to any of the barcodes used here for blanks or museum samples. Undetermined refers to those sequenced barcodes that did not match the barcodes recorded among the pooled blanks.

Blank ID	Sequence ID	Barcodes	Samples prepared alongside	Yield (Mb)	Average sequence length	% duplication	% GC content	Total sequence reads	Collapsed reads mapped	% reads mapped	Mitogenome coverage (%)	Coverage depth (×)
DNA-ve1	OG5003-12-0-1	P5-9-P7-14	MS10986–MS10992	0	76	26.90	39	1,160	2	0.172	0	0
DNA-ve2	OG5003-13-0-1	P5-9-P7-2	MS10993–MS10999	3	76	6.43	46	35,254	5720	16.225	97.7	26.9
DNA-ve3	OG5003-14-0-1	P5-9-P7-20	MS11000–MS11006	0	76	16.81	42	904	50	5.531	17.5	0.2
DNA-ve4	OG5003-15-0-1	P5-9-P7-26	MS11007–MS11013	0	76	20.06	47	668	15	2.246	0	0
Neg1	OG5003-16-0-1	P5-9-P7-3	MS10993–MS10999	0	76	20.00	51	10	0	0	0	0
Neg2	OG5003-17-0-1	P5-9-P7-15	MS10986–MS10992	0	76	50.00	46	4	0	0	0	0
Neg3	OG5003-18-0-1	P5-9-P7-21	MS11000–MS11006	0	76	50.00	51	8	0	0	0	0
Neg4	OG5003-19-0-1	PF-9-P7-28	MS11007–MS11013	0	76	33.96	52	106	1	0.943	0	0
Undetermined				68	76		49	899,556	133	0.015	27.9	0.4

3.4.4 Mapping sequences to the assembled kakī mitochondrial genome

Sequencing of targeted long-read amplification of the mitogenome for modern stilts mapped successfully to the mitogenome assembly, with an average of 99.7% of the mitogenome covered at a depth of 403.79× (Table 3.5, see Appendix C for coverage plots for all samples). Mapping of museum samples produced mitogenomes with > 97% of bases covered for all samples except MS10993, MS11008, MS11012, and MS11013 due to low sequence yield (Table 3.6). Excluding these four samples, an average of 97.89% of the mitogenome was covered with an average depth of 233.89× across all samples. The resequencing data for 24 kakī produced an average of 154,998,902 reads per individual. All data were processed with the mapping pipeline, and resulted in 100% sequence coverage for all individuals, with an average depth of 250.93× (Table 3.8). Mapping of this resequencing data identified a region from 14,239 bp onwards (including the NAD6 gene and the control region) with twice the mapping depth compared with the rest of the mitochondrial genome (Figure 3.6). This was hypothesised to represent a region of duplication in the kakī mitogenome, or a nuclear pseudogene (see Appendix D). The observed increased coverage depth across this region suggests that mean coverage depth of resequenced individuals is likely inflated.

Table 3.8: Results of modern kakī whole-genome resequencing data mapping to the kakī mitogenome.

Sequencing ID	Stilt DNA ID	Sequencing strategy	Total raw sequence reads	Mitogenome coverage (%)	Coverage depth (×)
H01383	DNA240	High coverage	108,247,282	100	307.1
H01384	DNA451	High coverage	323,915,460	100	257.4
H01385	DNA452	High coverage	141,757,794	100	313.6
H01386	DNA453	High coverage	192,931,784	100	345.4
H01387	DNA639	High coverage	170,267,148	100	456.1
H01388	DNA1376	High coverage	239,378,718	100	628.7
H01389	DNA1377	High coverage	197,094,168	100	276.2
H01390	DNA1429	High coverage	201,384,664	100	469.8
H01391	DNA1469	High coverage	228,489,236	100	365.3
H01392	DNA1565	High coverage	230,785,686	100	398.1
H01393	DNA1659	High coverage	178,588,964	100	116.9
H01394	DNA1661	High coverage	246,479,770	100	303.8
H01407	DNA1738	Low coverage	77,264,826	100	42.0
H01408	DNA1872	Low coverage	108,325,914	100	190.2
H01409	DNA1892	Low coverage	93,887,144	100	171.7
H01410	DNA1934	Low coverage	95,889,628	100	271.1
H01411	DNA1936	Low coverage	79,991,160	100	59.6
H01412	DNA1980	Low coverage	113,637,902	100	228.6
H01413	DNA2012	Low coverage	149,576,634	100	200.9
H01414	DNA2023	Low coverage	172,893,122	100	198.2
H01415	DNA2032	Low coverage	187,531,438	100	204.0
H01416	DNA2035	Low coverage	54,170,050	100	74.1
H01417	DNA2074	Low coverage	61,238,394	100	85.4
H01418	DNA2091	Low coverage	66,246,754	100	58.2

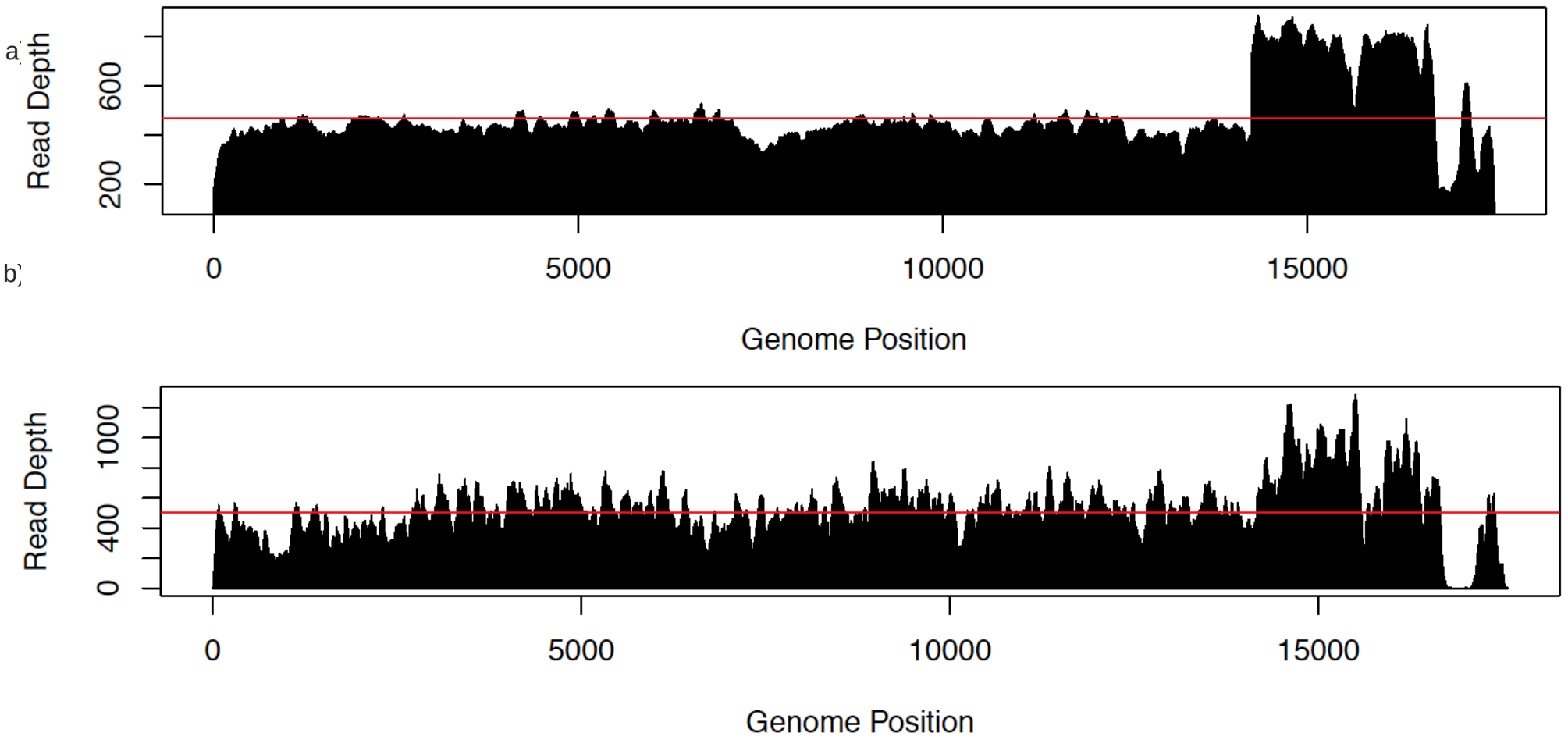


Figure 3.6: Example of the mapping coverage distribution for a) resequenced kakī H01390 (100% coverage, 469.8× average depth), showing a region of duplication with approximately double the coverage depth from ~14.2–16.5 kb, and b) a historic stilt sample, kakī MS11000 (98.7% coverage with 501.3× average depth). While less clear than for that of modern resequencing individuals, the increased coverage depth for the historic sample in the region ~14.2–16.5 kb lends support to the hypothesis that the kakī mitochondrial genome contains a region of gene duplication.

3.4.5 Bayesian and maximum likelihood phylogenetic analyses

Bayesian phylogenetic analysis and maximum likelihood analysis on the mitogenome alignment of forty members of the Order Charadriiformes, with the Australasian grebe (*Tachybaptus novaehollandiae*, Order Podicipediformes) and chicken (*Gallus gallus*, Order Galliformes) as outgroups, and one representative kakī and Australian pied stilt (total n = 44; Table 3.3) produced trees with identical topologies (Figure 3.7), consistent with those of previous phylogenetic studies of the Order Charadriiformes using mitochondrial data (e.g., Baker et al. 2007; Yang et al. 2017). The black-winged stilt (*Himantopus himantopus*) was sister to the combined clade of kakī and Australian pied stilts. Based on the Bayesian analysis, TMRCA for kakī and Australian pied stilts was approximately 0.750 Mya (95% highest posterior density (HPD) = 0.416–1.163 Mya), and TMRCA for the *Himantopus* clade was estimated at approximately 1.480 Mya (95% HPD = 0.883–2.214 Mya; Figure 3.8). The divergence estimate for kakī and Australian pied stilts was more recent than all other congeneric species pairs within Charadriiformes, except those of the plovers *Charadrius placidus* and *C. alexandrinus* (Charadriidae, TMRCA estimated at 0.118 Mya (95% HPD = 0.041–0.209 Mya)) and the gulls *Chroicocephalus brunnicephalus* and *C. ridibundus* (Laridae, TMRCA estimated at 0.473 Mya (95% HPD = 0.240–0.754 Mya)). Only limited impacts of hybridisation were observed between the well-supported kakī and pied stilt clades in the congeneric Bayesian analysis (posterior probability = 1, estimated divergence 0.663 Mya (95% HPD = 0.416–1.062 Mya; Figure 3.9), with a single kakī sample found in the pied stilt clade, and a single poaka (and no Australian pied stilts) found in the kakī clade. One kakī-hybrid individual occurred in the kakī clade, with two occurring in the pied stilt clade, along with the skeleton sample originally recorded as kakī (MS11001, Figure 3.9). Diversification of lineages within kakī and pied stilts appear to have occurred during a similar time

period approximately 0.010–0.193 Mya, with the more recent diversification potentially associated with the end of the Ōtira glacial period.

3.4.6 Haplotype networks

Median-joining networks produced with PopART further confirmed differentiation between species. All haplotypes identified through network analysis represented strongly supported nodes in congeneric Bayesian analysis (Figure 3.9). Four kakī-type haplotypes were observed (A–D), and eight pied-type haplotypes (E–L; Figure 3.10). One poaka (MS11006) and one hybrid (MS11002) were observed to have kakī-type haplotypes A and C respectively. The skeleton specimen (MS11001) was the only individual catalogued as kakī that was observed to have a pied stilt haplotype. The remaining three hybrid individuals all had pied stilt haplotypes. Among the two Australian pied stilts, one had a haplotype shared by a hybrid and a poaka (haplotype G), while the other possessed a unique haplotype (haplotype E). Comparison of these haplotypes with those specific to mitochondrial haplotypes used by Steeves et al. (2010) identified the cytochrome *b* (Cyt*b*) region sequenced corresponds to the assembled Cyt*b* region at 13,760–14,050 bp. Among the haplotypes described here, haplotype A corresponds to Cyt*b* haplotype B, haplotype C corresponds to Cyt*b* haplotype C, and haplotype E corresponds to Cyt*b* haplotype A. No haplotype representing the Cyt*b* haplotype D was detected among stilt mitogenomes in this study. When individuals were categorised as ‘modern’ (1978 to the present) or ‘historic’ (pre-1960), among kakī, one haplotype was present in historic samples that is not represented among the modern samples, while one haplotype is observed among modern samples but not captured among the historic samples (Figure 3.11).

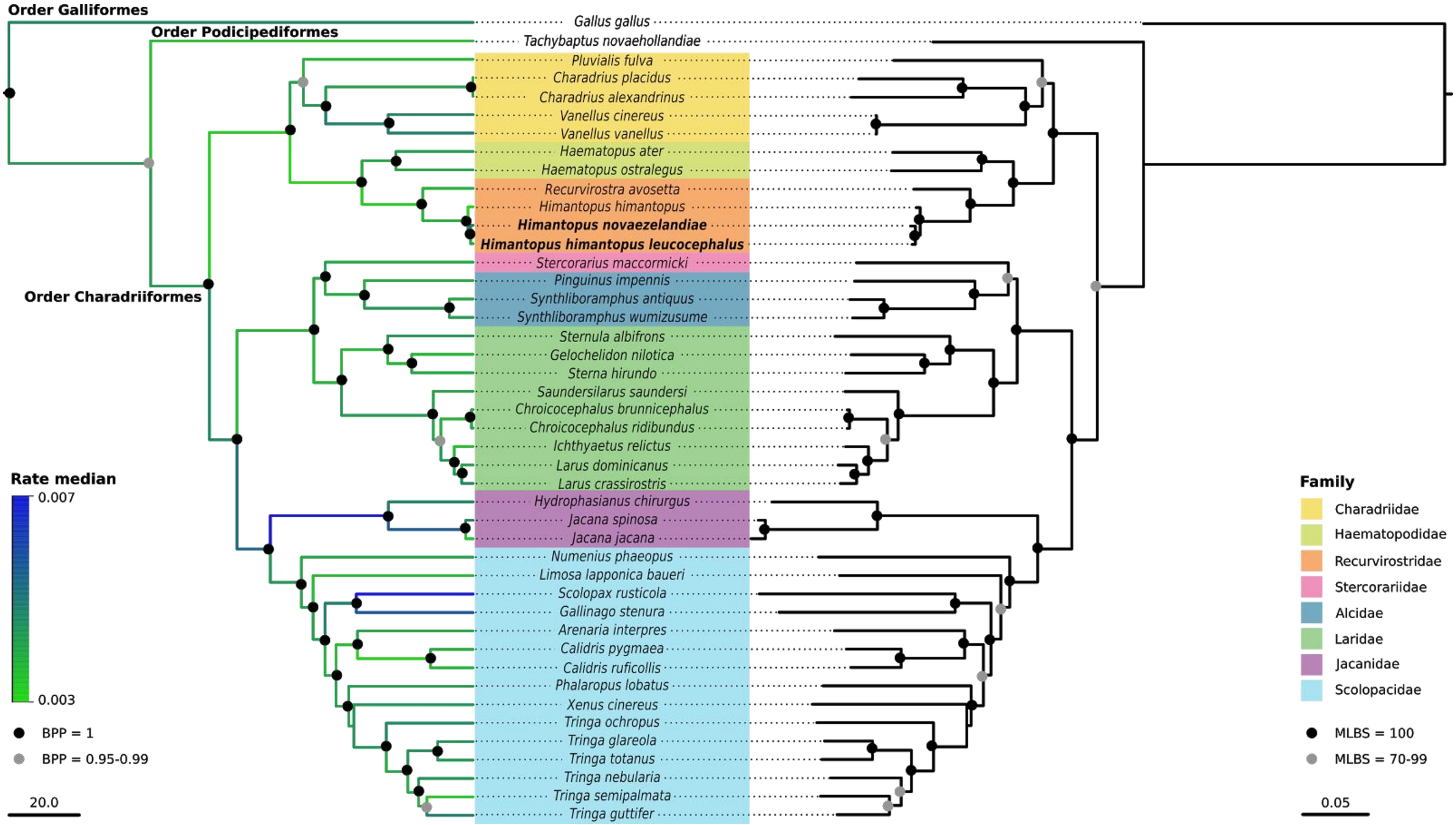


Figure 3.7: BEAST maximum clade credibility tree (MCC tree, left) and maximum likelihood tree (ML tree, right) constructed with IQ-Tree v1.6.6 for the Order Charadriiformes based on a region of 14,239 bp of the aligned mitochondrial genomes. Bayesian posterior probability (BPP) values > 0.95 indicate strong node support in the MCC tree, while maximum likelihood bootstrap node support (MLBS) values > 70% indicate strong node support in the ML tree. Trees are visualised with FigTree v1.4.3. The domestic chicken (Order Galliformes, Gallus gallus) and Australasian grebe (Order Podicipediformes, Tachybaptus novaehollandiae) are included as outgroups. All other Charadriiformes species with complete mitogenomes available at the time of analysis are included, except for the Vega gull (Larus vegae). The representative kakī included here is individual H01384, and the representative Australian pied stilt is individual B60480. Species are coloured by family within the Order Charadriiformes. Branches on the MCC tree are coloured according to the rate median. The scale bar on the left represents time since the present in million years for the MCC tree, and the scale bar on the right represents the number of substitutions per site for the ML tree.

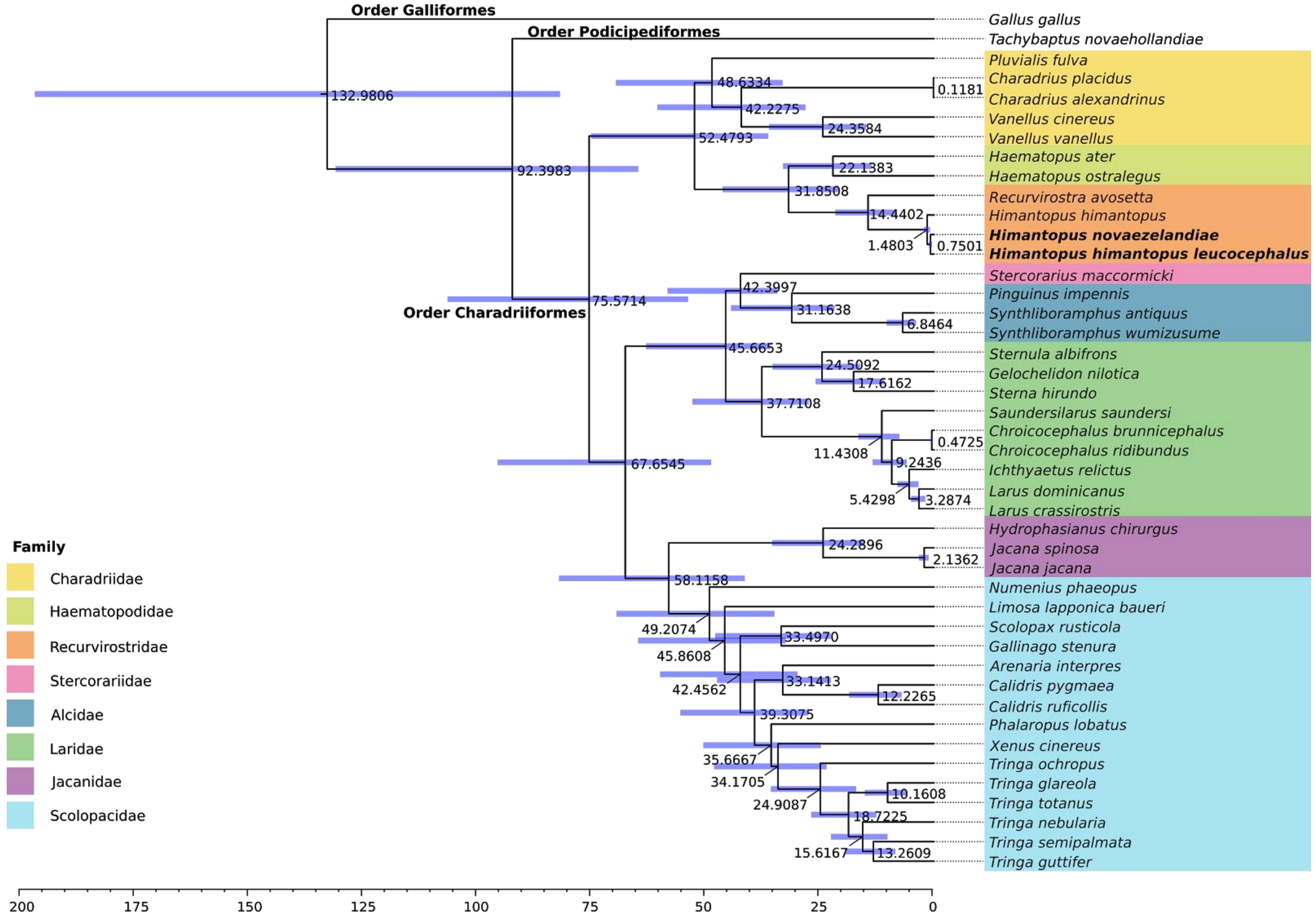


Figure 3.8: BEAST maximum clade credibility tree produced from the truncated mitogenome alignment for the Order Charadriiformes, visualised with FigTree v1.4.3. Estimated divergence times (million years ago, Mya) are displayed at branch nodes, with 95% highest posterior density intervals around these estimated times visualised as horizontal blue bars. Species are coloured by family within the Order Charadriiformes. The scale bar represents time since the present (Mya). The calibrated nodes were divergence of Stercorariidae from Alcidae, and the divergence of the clade comprising Charadriidae, Haematopodidae, and Recurvirostridae from all other Charadriiformes.

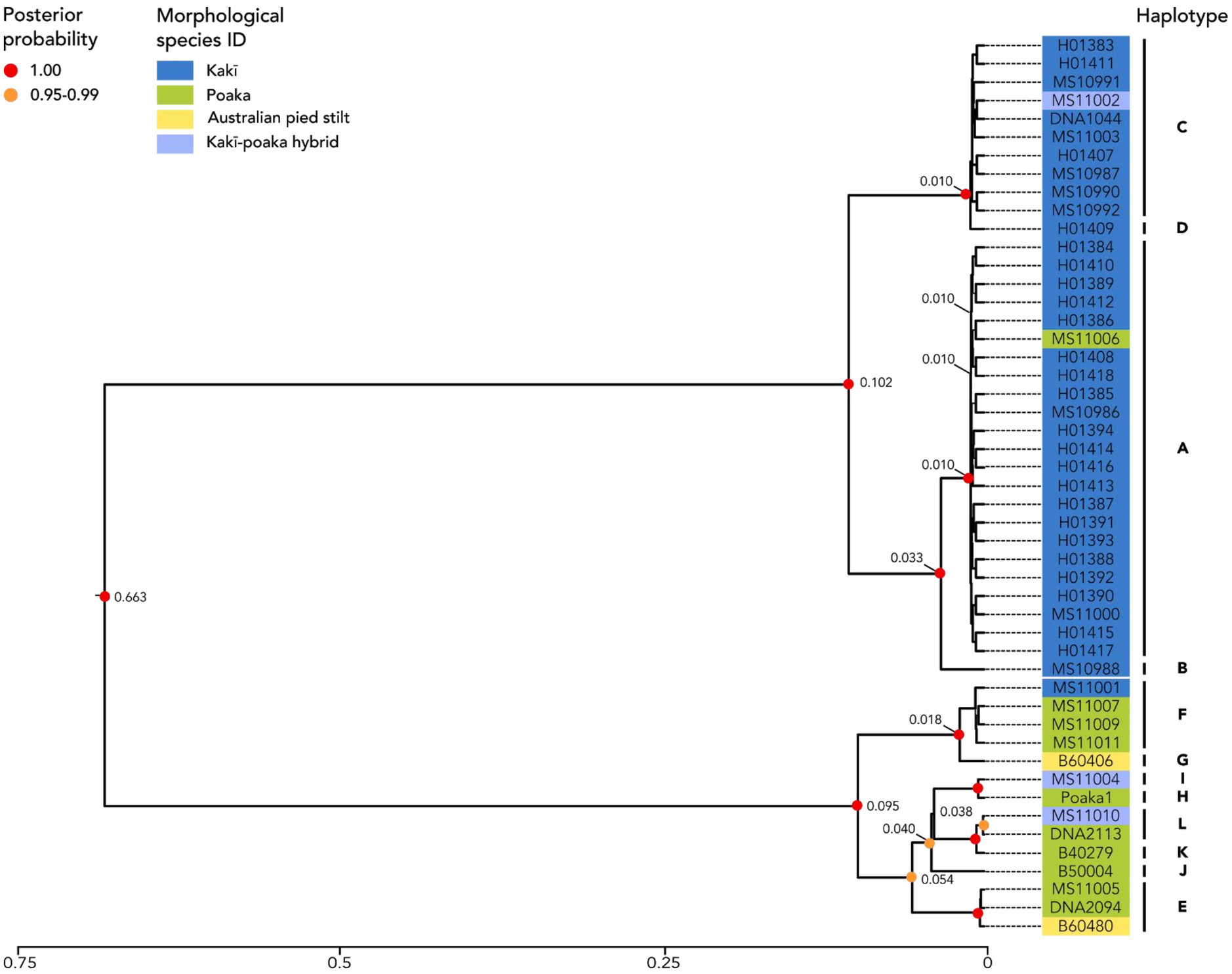


Figure 3.9: BEAST maximum clade credibility tree produced from congeneric analysis the mitogenome region 1–14,239 bp of pre-defined kakī, Australian pied stilts, poaka, and interspecific hybrids, visualised with FigTree v1.4.3. Circle colour represents the posterior probability associated with that node for those well-supported nodes. Numbers at nodes and branch lengths represent estimated time since divergence (TMRCA; Mya). Haplotype labels correspond to those identified in network analysis (see Figure 3.10). Individuals are coloured according to morphological species identity. Scale bar represents time (Mya).

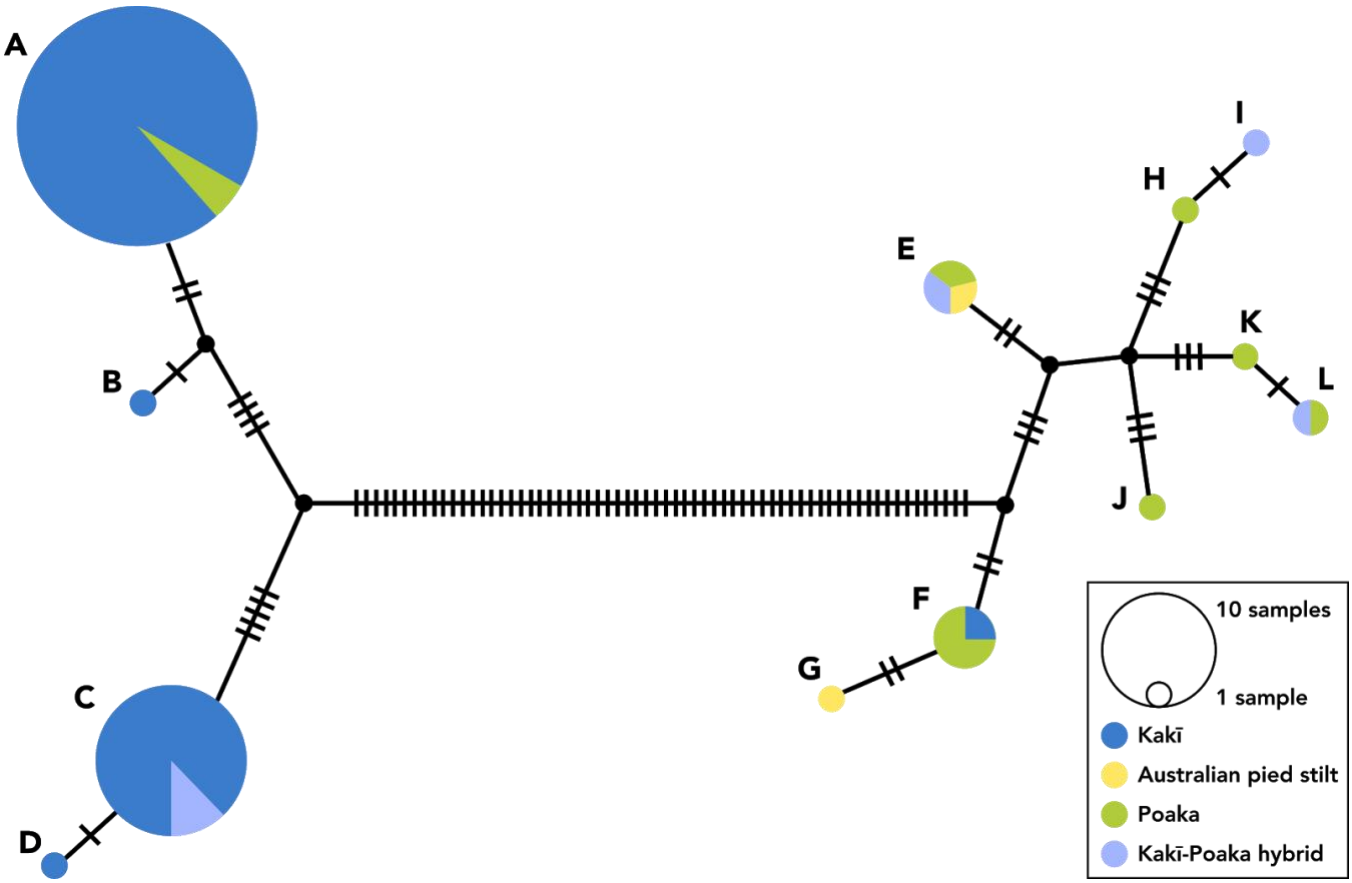


Figure 3.10: Median-joining network of historic and modern kakī (n = 34), poaka (n = 8), Australian pied stilts (n = 2), and kakī-poaka hybrids (n = 5) for the mitogenome region 14,239 bp, showing two distinct haplogroups: 1) kakī, haplotypes A–D, and 2) pied stilts, haplotypes E–L. Haplotype A corresponds to the cytochrome-b (Cytb) haplotype B in Steeves et al. (2010), haplotype C corresponds to Cytb haplotype C, and haplotype E corresponds to Cytb haplotype A. Haplotypes are coloured according to morphological species identification of individuals within that haplotype. Cross-hatched lines represent the number of variant sites differentiating haplotypes. Circle size represents the number of individuals sharing haplotypes.

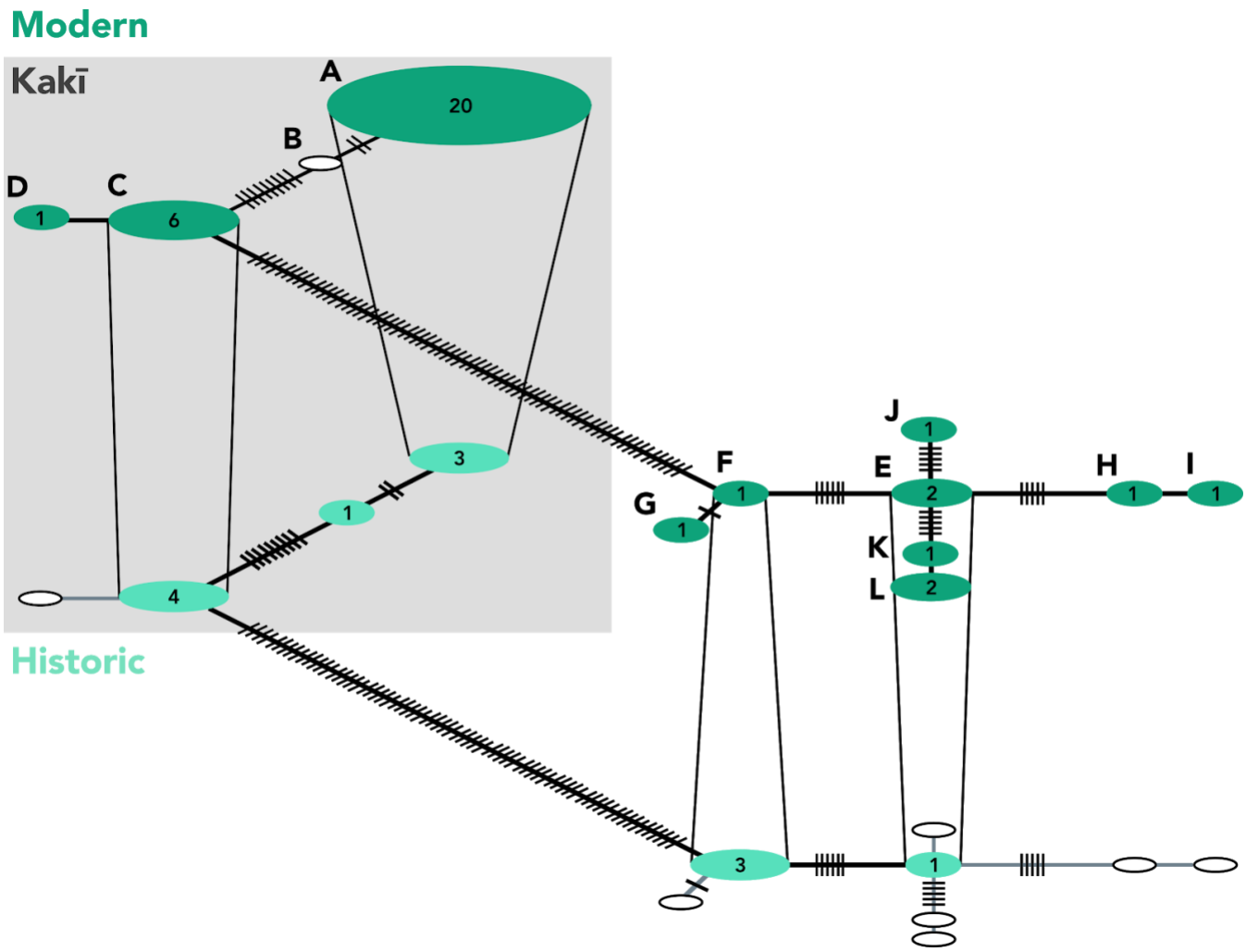


Figure 3.11: Temporal haplotype network derived from the 14,239 bp mitochondrial alignment of 49 modern and historic stilt samples, visualised with TempNet to compare diversity of modern and historic populations. Kakī haplotypes are highlighted in the grey box. Samples collected prior to 1960 were categorised as 'Historic', while those collected from the 1960s onwards were classed as 'Modern'. Colour distinguishes presence of haplotypes between the modern and historic groups. White circles represent haplotypes that were not present among the samples included in that time period. Numbers within circles indicate the number of individuals with that haplotype, and letters alongside haplotypes correspond to the haplotypes described in Figure 3.10. Hatch-marks represent the number of variant sites differentiating haplotypes.

3.4.7 Mitochondrial diversity metrics

Diversity and differentiation metrics were estimated from 49 kakī, Australian pied stilts, poaka, and kakī-poaka hybrids across the extracted 14,239 bp of the mitogenome. In this region, 45 sites contained gaps or missing data, with 98

biallelic sites. Eight of these polymorphic sites were singletons, each occurring in only a single individual, leaving ninety parsimony-informative sites. Twelve unique haplotypes were observed among all individuals (Figure 3.10), with haplotype diversity (*Hd*) 0.730 ± SD 0.056, and nucleotide diversity (π) 0.00238 ± SD 0.00029. There was an average of 33.838 nucleotide differences between individuals. Following classification of individuals by morphology and assessing haplotypes for any discordance between catalogued species identity and haplotype identity, known hybrids and individuals with unknown/uncertain plumage were excluded, leaving 34 kakī and nine pied stilts (comprising Australian pied stilts and poaka) for interspecific comparison. Among kakī, there were 13 polymorphic sites, representing four unique haplotypes (*Hd* = 0.4849, π = 0.00034 ± SD 0.00005), while among pied stilts, there were 19 polymorphic sites and six unique haplotypes (*Hd* = 0.8889, π = 0.00046 ± SD 0.00006). There were no shared polymorphisms between the two species. There was an average of 4.870 nucleotide differences among kakī individuals, and 6.500 among pied stilts, and an average of 74.261 nucleotide differences between the two species.

3.5 Discussion

Here I produced the first annotated mitogenome assembly for kakī, and used this to investigate mitochondrial diversity and divergence between kakī and the congeneric Australian pied stilt. All analyses supported differentiation between these congeners, with divergence of kakī and Australian pied stilts from a common ancestor estimated at 750,000 ya (95% HPD = 0.415–1.163 Mya) based on Bayesian inference. Modern and historic sequence mapping and alignment to the mitogenome revealed greater mitochondrial diversity among kakī than previously

detected through single-gene mitochondrial analyses (Steeves et al., 2010), with mitogenome diversity largely maintained among kakī despite the strong population bottleneck.

3.5.1 Mitogenome assembly and identification of a potential mitochondrial gene duplication

The abundance of mitochondrial reads present among the short-read whole-genome sequencing data produced in Chapter Two enabled kakī mitogenome assembly. NOVOPlasty was identified as the superior mitogenome assembly tool among those tested, producing an assembly of the expected length (16–18 kb), with close similarity to the published mitogenome of the congeneric black-winged stilt, and the full complement of avian mitochondrial genes as detected through MITOS annotation. Sequence mapping to the assembled mitogenome was successful for all samples that produced sufficient sequence data, producing overall high depth and consistent coverage across the mitogenome. However, mapping high-coverage whole-genome resequencing data to the assembled kakī mitogenome revealed a region of the mitogenome covered at twice the depth of the remaining assembly. Not only did this inflate the mean coverage depth for the modern resequenced kakī samples, it indicated an underlying issue with the mitogenome assembly. I hypothesise that this anomalous region of high coverage represents a mitochondrial gene duplication that has been collapsed during assembly due to the close similarity between the duplicated sequences. Such duplications have been observed in a wide range of avian species (including those in the orders Bucerotiformes (Sammler et al., 2011), Passeriformes (Gibb et al., 2015; Singh et al., 2008), Pelecaniformes (Cho et al., 2009; Gibb et al., 2013), Procellariiformes (Abbott et al., 2005; Eda et al., 2010; Torres et al., 2019),

Psittaciformes (Eberhard et al., 2001; Eberhard & Wright, 2016; Schirtzinger et al., 2012), and Sulliformes (Morris-Pocock et al., 2010)) including the conordinal ruff (*Calidris pugnax*; Verkuil et al. 2010). This may result in incorrect estimation of diversity and inference of phylogeny when apparent nucleotide differences are the result of differences between duplicated regions within an individual, and I excluded this region to avoid any such complications. This hypothesis warrants further investigation (as discussed in Future Directions, this chapter, and Appendix D). Despite the exclusion of this region from downstream analyses, I contend that the results of these analyses are robust and representative of the complete mitogenome.

3.5.2 Stilt divergence dating

Mitogenomes are particularly valuable for phylogenetic analysis and estimation of divergence times, as many more mitogenomes than nuclear genomes are available in this order (Table 3.3 *cf*. the 14 Charadriiformes nuclear genome assemblies currently available, Table 2.9). Here these mitogenome resources were used to estimate the timing of mitochondrial divergence for kakī and Australian pied stilts using Bayesian phylogenetic analysis. These results indicated a TMRCA for kakī and Australian pied stilts of approximately 0.750 Mya, one of the most recent estimates of divergence for a congeneric species pair within the order Charadriiformes (Figure 3.7). This is the first calibrated divergence estimate for these stilts, and is similar to the divergence date of 1 Mya estimated from mitochondrial control region divergence (Wallis, 1999). One limitation of these Bayesian analyses arises from the limited divergence date calibrations available within the order with complete mitogenomes available for analysis (Smith, 2015). Those calibrations available represent very deep divergence within

Charadriiformes, and may result in overestimation of the TMRCA for shallow splits, such as that seen between kakī and Australian pied stilts.

Although there are many known instances of hybridisation within the Order Charadriiformes (McCarthy, 2006), the northern and wattled jacanas (*Jacana spinosa* and *J. jacana*) are the only other species known to hybridise with one another among the species represented in the conordinal comparative phylogenetic analyses here (Miller et al., 2014) , and so the relationship between these congeners can be compared with that between kakī and poaka. These jacanas have a more distant estimated TMRCA (2.1362 Mya, 95% HPD = 1.905–4.969 Mya) than kakī and Australian pied stilts (Figure 3.7), and hybridisation between these congeners across the range overlap in Western Panama has resulted in mitochondrial introgression, as assessed using a 651 bp region of the mitochondrial COI gene (Miller et al., 2014). In comparison, potential mitochondrial introgression was observed in only a single poaka individual (MS11006), with no evidence of introgression from poaka into kakī.

Congeneric Bayesian analysis indicated a recent divergence of haplotypes within both kakī and pied stilts. This similarity in intraspecific haplotype divergence despite differences in range and estimated population sizes through time (see Figure 2.10) may be an artefact of the limited number of pied stilt specimens available for analysis, which are unlikely to represent species-wide diversity.

3.5.3 Comparison of kakī and poaka mitochondrial diversity and identification of mitochondrial introgression

When assessing mitochondrial diversity in stilts, four distinct haplotypes were identified among kakī (Figure 3.10), twice the number previously identified with single-gene mitochondrial analysis (Steeves et al., 2010). Three of these

haplotypes were detected among historic samples and three haplotypes were detected among modern samples (Figure 3.11). This supports the prediction that mitogenome analysis would reveal greater diversity than previously detected in single-gene mitochondrial studies, with two additional haplotypes identified among kakī here. Nevertheless, this represents relatively low haplotype diversity. Such low diversity is not unexpected given the demographic history of kakī including a severe recent bottleneck. Furthermore, the results of PSMC analysis in Chapter Two (Figure 2.10) suggested that kakī have existed in low numbers over a long period. If the initial kakī founding in New Zealand began with only a few individuals, and remained limited, potentially due to restricted habitat availability during interglacial periods, the species may have never had the opportunity to expand and diversify.

Haplotype A represented across the greatest number of kakī may be over-represented here, due to the individuals selected for resequencing conducted as part of an aligned project (Galla et al., 2019). While twelve of these individuals were selected to represent diverse lineages among the kakī pedigree, there are three clusters of related individuals among the subset included here, with members of two of these groups comprising ten of the individuals sharing this common haplotype. Conversely, this may indeed represent the true frequency of the haplotype in the species, given the high levels of relatedness among kakī due to their unavoidable inbreeding in the small population.

Haplotype diversity among modern pied stilts was more than twice that of kakī, although mitochondrial nucleotide diversity was similar for both species. Pied stilt diversity is likely underestimated here due to the small sample size and limited sampling locations, but allow comparison between this more widespread species and threatened kakī. The greater number of haplotypes detected among pied stilts

compared with the previous study (Steeves et al., 2010) is due to the wider sampling distribution, including individuals from Australia and the North Island of New Zealand, as well as South Island sites beyond Te Manahuna. Haplotype E is shared by an Australian pied stilt, a poaka, and a morphological kakī-poaka hybrid, which indicates that the Australian pied stilts and poaka have not yet become genetically differentiated at the mitogenome level. This may be due to a large, genetically diverse founding population arriving in New Zealand with little loss of diversity following arrival, or due to ongoing migration maintaining gene flow. The observation of one poaka with a kakī haplotype (individual MS11006 with Haplotype A) is most likely due to mitochondrial introgression resulting from hybridisation. This individual was a museum specimen collected prior to the early 1950s, demonstrating genetic evidence of hybridisation throughout the 1900s in addition to the morphological evidence described by Pierce (1984b). Similar mitochondrial introgression is observed in individual MS11002, a specimen with hybrid plumage collected in 1981, which has kakī haplotype C. All other individuals identified as hybrids had pied stilt haplotypes, and no individuals morphologically identified as kakī had non-kakī haplotypes (except the skeleton).

3.5.4 Temporal changes in haplotype diversity

Sequencing and downstream analyses incorporating historic museum specimens was informative for investigating temporal changes in patterns of diversity, despite the small number of historic kakī samples included (n = 8). These samples comprised all known historic kakī samples held in national museum collections, with the exception of two kakī skins collected by early naturalist Sir Walter Buller. Evidence of stilts in the New Zealand fossil record is limited (Holdaway & Worthy, 1997; Worthy, 1998a, 1998b), the most significant of which consists of fossil

deposits of five stilts in Pyramid Valley, North Canterbury, including the partial skeleton of one individual identified by morphological data as kakī (Holdaway, 1995).There are estimated to only be 10–15 stilt bones held in national museum collections (R. P. Scofield, Canterbury Museum, pers. comm.), but the rarity and value of these specimens (and the Buller skins), and concerns regarding the required destructive sampling precluded their use in genomic analyses here. Identification of both kakī haplotypes corresponding to those previously identified in kakī by Steeves et al. (2010) confirms that conservation management of the species aiming to maintain diversity and species integrity has been successful at the mitogenome level. As nuclear data from the past is difficult to access, this mitochondrial data, while limited, represents the best available measure of temporal changes in genetic diversity.

All except one kakī haplotype present in the modern population was detected in the historic population (Haplotype D), with one haplotype observed among historic samples that was not present among the modern individuals assessed (Haplotype B). As the sample of modern kakī represents ~20% of the modern wild adult population, Haplotype B either occurs at very low frequency in the population, and thus is likely to be lost due to stochastic processes associated with small population size, or has already been lost.

3.5.5 Conclusions

In this study I used mitogenomes to compare genetic diversity through time for kakī and pied stilts. I resolved the timing of divergence for kakī and Australian pied stilts at approximately 0.750 Mya, and detected additional mitochondrial diversity among kakī undetected in previous single-gene mitochondrial analysis. These results indicate that mitochondrial diversity may have already been limited prior to

kakī decline. These results provide evidence that kakī management based on genetic data (Hagen et al., 2011; Maloney & Murray, 2001; Steeves et al., 2010) has successfully maintained mitochondrial diversity. Mitochondrial DNA may not necessarily be a good proxy for nuclear diversity, but, acknowledging these limitations, these mitochondrial results are consistent with estimates of nuclear diversity from microsatellite data (Steeves et al. 2010, see Forsdick et al. 2017 for a comparison with other threatened New Zealand birds), and, in combination, indicate that genetic diversity has been maintained despite the bottleneck and subsequent long-term small population size. Comparison of modern and historic pied stilt mitochondrial diversity is limited due to the small sample sizes which may not have captured the true extent of historic haplotype diversity. Despite these limitations, these results also confirm clear differentiation between kakī and poaka, and no evidence of mitochondrial introgression among kakī. This suggests that while reproductive barriers to hybridisation remain semi-permeable, the two species have independent evolutionary trajectories, and conservation management of kakī has achieved the goal of maintaining genetic integrity. In Chapter Four I will use nuclear genomic markers to investigate whether these findings are representative of genome-wide species differentiation.

HTS techniques successfully enabled the inclusion of data from historic stilt samples that was previously unobtainable via genetic methods. The relatively recent origins of historic samples included here facilitated the sequencing of high-quality data for most samples. Recovering mitogenomes from historic museum samples provides a glimpse into the past, revealing that although there was one mitochondrial haplotype present among historic kakī that is not present in the modern population, much of this mitochondrial diversity has been maintained due in part to genetics-informed conservation management. Although the single-copy

nature of mitochondrial genomes should produce accurate inferences based on single gene regions, when a larger region (or the complete mitogenome) is incorporated, a greater number of variable sites included in analyses will produce more robust estimates of diversity, differentiation, and increase confidence in phylogenetic inferences. When access to nuclear genomic data remains beyond the scope of for many non-model threatened species due to limited resources, mitogenome sequencing can provide representative estimates of diversity and population structure in an efficient manner, with utility for questions involving taxonomic uncertainty, population differentiation, and interspecific hybridisation. In addition, while the availability of complete nuclear genomes is increasing (see Appendix B), such resources remain outnumbered by the available mitogenomes that can be used as references for non-model species and may provide a more efficient pathway for analysis than *de novo* mitogenome assembly. Therefore, while nuclear genomic data may provide additional data for robust phylogenetic analysis and estimation of conservation-relevant nuclear genomic metrics, the lower costs associated with mitogenome sequencing, combined with the ability to include temporal assessments of diversity, ensure that mitogenomes remain a valuable tool when informing conservation management for many non-model species.

3.6 Data availability

Kakī are a taonga species to Māori, and as such, genomic data derived from kakī are also recognised as taonga in their own right. Due to the tapu nature of these data, the data produced here are hosted on a password-protected database at www.ucconsert.org/data/, and will be made available at the discretion of the

kaitiaki of the iwi and hapū associated with kakī. These data include the preliminary and corrected kakī mitochondrial genome assemblies presented here, raw demultiplexed mitochondrial sequence reads for each individual, the group VCF files produced from sequence mapping for each of the resequencing, modern targeted, and historic museum sequencing sets, and raw kakī long-read sequencing. Scripts associated with the bioinformatic pipelines included in this chapter are detailed in my GitHub 'Himantopus' repository at https://github.com/natforsdick/Himantopus.

Chapter 4: Genotyping-by-sequencing confirms no introgression resulting from past hybridisation between kakī (*Himantopus novaezelandiae*) and poaka (*H. h. leucocephalus*)

4.1 Abstract

Genetic swamping due to interspecific hybridisation poses a risk to species persistence with disproportionate impacts on small threatened populations. The development of high-throughput sequencing to generate large numbers of genomic markers enables robust analysis of the impacts of interspecific hybridisation on the genomes of threatened populations and can reveal introgression previously undetected using small numbers of genetic markers. Anthropogenic impacts have resulted in hybridisation and subsequent backcrossing of the critically endangered New Zealand endemic kakī (*Himantopus novaezelandiae*) with the non-threatened self-introduced congeneric poaka (the New Zealand population of pied stilts, *Himantopus himantopus leucocephalus*), yet genetic analyses with a small set of microsatellite markers revealed no introgression of poaka genetic material in kakī, excluding one individual. Here I conduct genotyping-by-sequencing (GBS) and develop a workflow for reference-guided marker discovery and analysis to reassess the genomic extent of introgression resulting from hybridisation between kakī and poaka. No introgression from poaka to kakī was detected (all morphological kakī were assigned as kakī with > 95% probability). These results support the previous findings that a combination of strong positive assortative mating by kakī, outbreeding

depression, and conservation management actively excluding hybrids are maintaining the genetic integrity of this critically endangered wading bird. These results indicate that the existing genetic microsatellite markers provide a robust, cost-effective approach to analysis of introgression in kakī, and thus the continued use of this approach is recommended for kakī conservation management associated with hybridisation. Further, conservation practitioners should continue to implement management strategies aimed at minimising the potential for hybridisation between these two species to avoid hybridisation threatening the continued recovery of kakī.

4.2 Introduction

4.2.1 Hybridisation and conservation

The process of interspecific hybridisation (herein, hybridisation), the interbreeding between closely related species, is common among birds (Grant & Grant, 1992; McCarthy, 2006; Ottenburghs et al., 2015). Although hybridisation is a naturally occurring process that can facilitate rapid adaptation and speciation (Abbott et al., 2013; Anderson & Stebbins, 1954; Arnold, 1992; Dowling & Secor, 1997; Hedrick, 2013; Seehausen, 2004), anthropogenic impacts that alter species distributions, population demographics, and habitat availability can result in hybridisation that adversely affects threatened species (Allendorf et al., 2001; Rhymer & Simberloff, 1996; Todesco et al., 2016), and such impacts are likely to increase with the effects of anthropogenic climate change (Chunco, 2014). Hybridisation negatively impacts threatened species through the waste of reproductive and energetic resources resulting from interspecific breeding that can reduce reproductive output (i.e., demographic swamping; Allendorf et al., 2001; Wolf et al., 2001). In addition,

hybridisation may result in introgression, where hybridisation and subsequent backcrossing to the parental species mediates the incorporation of genetic material from one species into the genome of another, which at its most extreme, may result in extinction-by-hybridisation (Allendorf et al., 2001; Fitzpatrick et al., 2010; Rhymer & Simberloff, 1996; Riley et al., 2003; Taylor et al., 2006; Todesco et al., 2016). This introgression may lead to outbreeding depression, through the breakdown of coadapted gene complexes or the introduction of maladaptive traits resulting in the decreased fitness of hybrid offspring (Arnold, 1997; Edmands, 2007; Lynch, 1991). Thus, hybridisation resulting from anthropogenic impacts that facilitate contact between previously isolated species, particularly when the two species include a threatened endemic hybridising with a more prolific species, are of concern for conservation.

Genetic tools are employed to assist conservation management programmes in assessing the extent and impacts of hybridisation, and for identification of cryptic hybrid offspring morphologically indistinguishable from parental types (Chan et al., 2006a; Ma & Lambert, 1997; Milián-García et al., 2015; Pierpaoli et al., 2003). However, to date, most conservation studies have used a small number of genetic markers (e.g., microsatellites) that may not be representative of genome-wide diversity, particularly among threatened species where population bottlenecks have left populations genetically depauperate (Taylor, 2015; Taylor et al., 2015; Väli et al., 2008). Over the past decade, advances in genomic sequencing technologies and rapidly declining costs have enabled the sequencing and assembly of complete genomes for threatened non-model organisms (e.g., Li et al., 2010; Sutton et al., 2018; Chapter Two). While genomes are integral resources for investigation of evolutionary processes such as functional adaptation (e.g., Le Duc et al., 2015; Marra et al., 2019; Yu et al., 2016), thousands of genomic markers distributed

throughout the genome (i.e., single-nucleotide polymorphisms (SNPs)) are sufficient to facilitate population-level estimation of metrics including diversity, relatedness, population structure, and introgression in an efficient, cost-effective manner (Ba et al., 2017; Chen et al., 2016; Peek et al., 2019; Rexer-Huber et al., 2019; Rick et al., 2019). Population-level reduced-representation sequencing (including restriction-enzyme associated DNA sequencing (RADseq), double-digest RADseq (ddRADseq), and genotyping-by-sequencing (GBS)) is an approach that can produce thousands of variant sites for high-resolution population genomic analyses (Davey et al., 2011; Davey & Blaxter, 2010; Elshire et al., 2011; Narum et al., 2013) and as such has wide applicability for conservation (Andrews et al., 2016; Seabury et al., 2011; Wright et al., 2019).

Although a growing number of studies are using genomic tools to explore avian hybridisation, studies with a conservation focus are thus far limited. For example, extensive genetic research has been carried out among duck species (Family Anatidae) where hybridisation is common (e.g., Fowler et al., 2008; Lavretsky et al., 2014; McCracken et al., 2013; McCracken & Wilson, 2011; Muñoz-Fuentes et al., 2013; Williams et al., 2005) and genomic tools have been found to outperform genetic marker sets to differentiate between species and detect introgression (e.g., Lavretsky et al., 2019b, 2019a). Nevertheless, few genomic studies to date have explored the impacts of hybridisation on threatened ducks (but see Brown et al., 2020; Peters et al., 2016; Wells et al., 2019). By leveraging the greater power provided by genomic tools to assess the extent of hybridisation and subsequent introgression, conservation practitioners can prioritise conservation efforts and implement informed strategies to maintain genetic integrity of hybridising species.

4.2.2 Study system, kakī (black stilt, Himantopus novaezelandiae)

The kakī (black stilt, *Himantopus novaezelandiae*) is a critically endangered New Zealand endemic wading bird (BirdLife International, 2018; Robertson et al., 2016). Anthropogenic impacts resulted in population decline, with numbers falling to approximately 23 individuals comprising a single population in Te Manahuna/the Mackenzie Basin in 1981 (Pierce, 1984b; Steeves et al., 2010). Adaptive conservation management for kakī, including predator control throughout Te Manahuna and a programme of captive breeding and rearing for translocation have been integral to kakī recovery to 169 wild adults in 2020 (Hagen et al., 2011; Heezik et al., 2005; Keedwell et al., 2002; Maloney & Murray, 2001; Reed et al., 1993; Steeves et al., 2010, DOC, pers. comm.).

Along with predation and changes in habitat availability, interspecific hybridisation has been an additional threat to kakī. The New Zealand population of congeneric pied stilts (poaka, *Himantopus himantopus leucocephalus*) self-introduced from Australia at least 200 years ago (Pierce, 1984b). While anthropogenic impacts reduced suitable habitat for kakī, these same impacts made new habitat available to the more generalist poaka (Pierce, 1984b). Unlike kakī, poaka have evolved alongside mammalian predators, further enhancing their successful expansion across New Zealand (Pierce, 1986b). At the peak of kakī decline, an estimated 1,500 poaka were present in Te Manahuna (DOC, pers. comm.). Limited mate choice and a male sex-bias among kakī at this time contributed to increasing hybridisation between kakī and poaka, producing viable, fertile hybrid offspring (Pierce, 1984a; Steeves et al., 2010). These hybrid offspring exhibit reduced fitness in terms of hatching and fledging success and survival, which combined with management to exclude hybrids, appears to have prevented the formation of a hybrid swarm as assessed through a genetic study by Steeves et al. (2010).

Both genetic and pedigree data have been incorporated during over forty years of conservation management of kakī, with the goals of maximising genetic diversity, minimising inbreeding, and maintaining the genetic integrity of the species (Maloney & Murray, 2001). Molecular sexing of individuals (Millar et al., 1997), analysis of genetic diversity and effects of inbreeding (Hagen et al., 2011), brood parasitism (Overbeek et al., 2017), pedigree validation (Overbeek et al., In Review), and assessment of genetic integrity and hybrid identification (Chambers & Macavoy, 1999; Greene, 1999; Steeves et al., 2010) have used allozyme, mitochondrial, and microsatellite markers as these tools became available. More recently, a combined pedigree and genomics approach has been implemented to assess relatedness for captive pairing decisions (Galla et al., 2020). Previous assessment of hybridisation found that not only were kakī genetically distinct from poaka, thus warranting conservation, but that introgression had been minimal despite extensive hybridisation (Steeves et al., 2010). For some threatened species, hybridisation has been considered as a potential conservation management strategy akin to genetic rescue (Whiteley et al., 2015), with the goal of increasing genetic diversity and thus long-term evolutionary potential for genetically depauperate species (Chan et al., 2019). However, once the genetic integrity of kakī was established, such a strategy was deemed inappropriate given the moderate level of genetic diversity in kakī compared with other threatened New Zealand endemic birds, along with the reduced fitness of hybrid offspring (Steeves et al., 2010). Thus hybrids have been actively excluded from kakī conservation management, despite previous inclusion of dark hybrids (Maloney & Murray, 2001; Reed et al., 1993a; Steeves et al., 2010; Wallis, 1999).

These previous genetic studies have been invaluable to the conservation management of kakī, allowing management decisions to be made with the best

information available at the time to ensure the persistence of this critically endangered taonga species. The primary limitation of these studies is the small number of markers that are assumed to be representative of the genome, and thus may lack the power to robustly detect introgression resulting from past hybridisation (Brumfield, 2010; Steeves et al., 2010). The improved accuracy of genomic approaches has already been demonstrated for kakī with estimates of relatedness derived from whole-genome resequencing (WGS) data providing more accurate estimates of relatedness than those produced with microsatellites when informing captive breeding (Galla et al., 2020), and may have similar benefits for analysis of introgression. Despite intensive observational monitoring of wild kakī pairs during the breeding season, non-kakī chicks are frequently detected from clutches of eggs collected in the wild and reared in captivity (Overbeek et al., 2017; Overbeek et al., In Review). Any individuals with anomalous plumage are genetically assessed with a microsatellite panel to confirm species identity (Overbeek et al., 2017; Overbeek et al., In Review; Steeves et al., 2008). However, unintentional incubation and rearing of non-kakī eggs and chicks detracts from the resources available to kakī management (Overbeek et al., 2017; Overbeek et al., In Review), and any increase in prevalence of hybridisation may compromise kakī recovery. While it is too early to determine whether a genomic approach will outperform the established microsatellite panel for assessing introgressive hybridisation between these stilts, verifying the impacts of this hybridisation is essential to ensure the implementation of appropriate management strategies to maintain the genetic integrity and support recovery of this critically endangered taonga species.

4.2.3 Aims

The primary aim of this chapter is to use genomic markers generated through population-level GBS for kakī, along with known hybrids, poaka, and Australian pied stilts, to reassess the extent of introgression due to hybridisation between kakī and poaka in the contemporary kakī population. One overall goal of this thesis is to produce genome assemblies for stilts that can be used as reference for variant discovery and downstream conservation genomics analyses, and thus the genomes produced in Chapter Two are used as references for a multi-pipeline approach to variant discovery from stilt GBS. I will produce a set of biallelic SNPs for use in introgression analysis with the objective of determining the extent and pattern of any introgression from poaka into kakī. There are two potential introgression scenarios that would be of particular concern for conservation practitioners and thus require further investigation if detected. In the first scenario, a substantial amount of introgression would be identified among a small number of kakī due to recent undetected hybridisation in the wild. In this first scenario, accurate identification of any such individuals will allow practitioners to make decisions regarding potential exclusion of these individuals from breeding with non-introgressed kakī, with the goal of maintaining the genetic integrity of the species. In addition, this may indicate that current observational efforts are insufficient to detect these rare hybridisation events in the wild. In the second scenario, a moderate amount of introgression previously undetected with genetic markers may be observed across a greater number of kakī, resulting from the presence of a small number of dark hybrids among the founding individuals in the captive breeding programme (Galla et al., 2020), which may have been exacerbated by subsequent selection maintaining introgressed genes. In this second scenario, management may be concerned with taking the proportion of introgression into account for captive breeding decision-making aimed at maintaining

kakī genetic integrity. Further investigation of the genes associated with any widespread introgression would be beneficial to determine the functional impacts of this introgression. In both scenarios, accurate detection of introgression is required to determine future research and management needs. Thus I define a stringent threshold of > 95% individual probability to assign individuals as kakī, informed by the clear genetic differentiation observed between these two species (Steeves et al., 2010; Chapter Three). These data will provide essential information supporting the ongoing recovery of kakī. An additional objective of this chapter is to investigate whether any introgression from kakī is detected among poaka through comparison with assignment probabilities of Australian pied stilts that are geographically isolated from kakī, and thus should demonstrate no signal of introgression. Further, after assessing a range of SNP discovery methods, filtering strategies, and analysis pipelines, I will produce a streamlined bioinformatic workflow for assessing introgression in any additional individuals sequenced in the future. This workflow can be readily adapted for use in other threatened birds where analysis of introgression may provide valuable information to determine appropriate conservation management practices.

4.3 Methods

4.3.1 Sample collection and DNA extraction

Following Steeves et al. (2010), individuals sampled herein were grouped by plumage morphology. Poaka and pied stilts (plumage nodes A–C2; see Figure 1.5) were labelled ‘pied’, completely black node J individuals were labelled ‘kakī’, and individuals of intermediate plumage (nodes D1–I/J) or known hybrid parentage were labelled ‘hybrid’ (Supplementary Table 1). Extracted genomic DNA (gDNA) was

available for eighty stilt samples including kakī, Australian pied stilts and poaka, and kakī hybrids that had been extracted using a DNeasy® Blood and Tissue Kit (QIAGEN) in the previous genetic study (Steeves et al., 2010), or for assessments of relatedness to inform captive pairing decisions by the Kakī Recovery Programme. I quantified extracted gDNA using a NanoDrop™ 8000 Spectrophotometer to assess the viability for GBS. GBS guidelines recommended a total of 1 µg of DNA at a concentration of 80–150 ng/µL, and 260/280 and 260/230 ratios of approximately 1.80 and 1.0 respectively.

I extracted gDNA from an additional 155 feather samples (Figure 4.1) collected as part of regular handling practices for kakī under Department of Conservation ethics approvals (AEC #283) at the DOC's Kakī Recovery Programme, Twizel, and the Isaac Conservation and Wildlife Trust kakī captive rearing facility, Christchurch, New Zealand. Pedigree information is recorded for all kakī individuals as part of routine Kakī Recovery Programme management, extending up to seven generations at the initiation of management. Blood samples were collected as part of routine health checks from two Australian pied stilts at Adelaide Zoo (provided under a Royal Zoological Society of South Australia Specimen Licence Agreement; Import Permit #2016061954), two node B poaka from Auckland Zoo (under Auckland Zoo animal ethics approval), and tissue was provided from one deceased poaka found in Hawke's Bay. Poaka from the North Island were preferentially sampled due to a low likelihood of recent contact with kakī, minimising the chance of these individuals having recent hybrid ancestry. I extracted gDNA from all samples using a Thermo Scientific™ MagJET™ Genomic DNA kit, following Protocol E (manual genomic DNA purification from up to 20 mg tissue). To maximise extraction efficiency of feather samples, I finely cut up the feather tips and used liquid nitrogen to facilitate grinding of the cut feather tips to increase surface area during digestion and

maximise DNA yield. I adapted the protocol such that feather samples were incubated for eight hours during the digestion step, and two 50 µL elutions of DNA in elution buffer were produced for each sample. I used a NanoDrop™ 8000 Spectrophotometer to quantify extracted gDNA. Twenty extractions including those with the highest and lowest DNA concentrations were visualised via gel electrophoresis on a 1% agarose gel at 100 V for 30 min to ensure the production of a single, strong band at ~20 kb with minimal evidence of degradation.

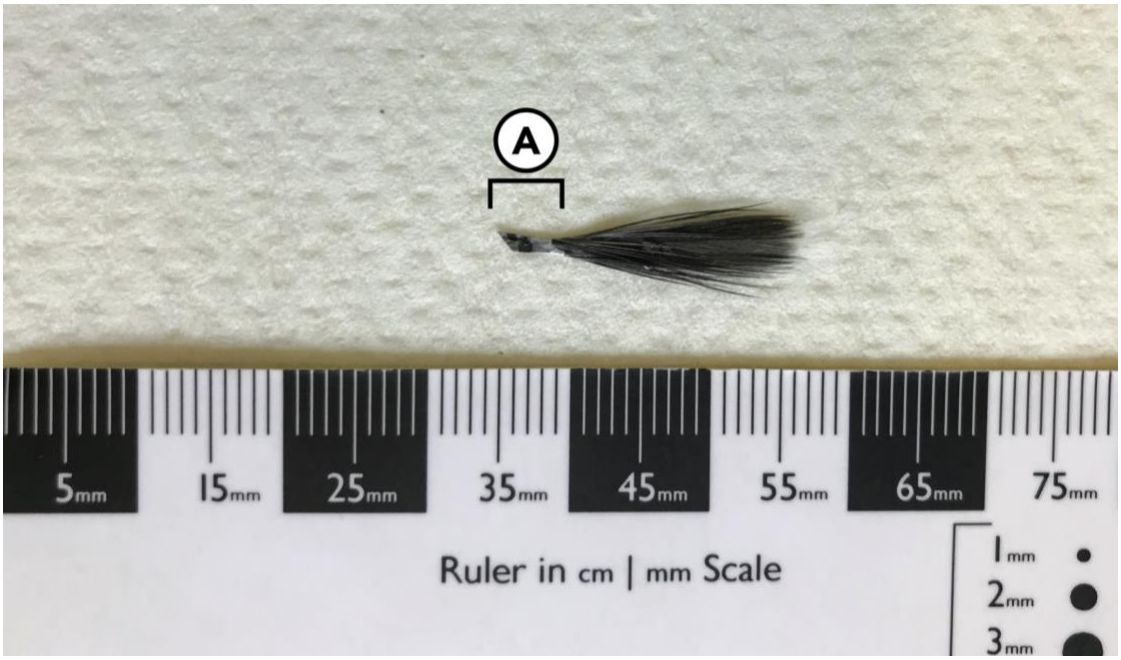


Figure 4.1: Example kakī, where region A is the feather tip used for DNA extraction.

4.3.2 Genotyping-by-sequencing

Among all available DNA extracts, 145 samples (130 kakī, six pied stilts, and nine hybrids) had DNA of sufficient quality and quantity for GBS. This included 74 males and 62 females, with nine individuals of unknown sex. The list of all sequenced individuals, DNA concentrations, sequence outputs and mapping success can be found in Supplementary Tables 1 and 2. I prepared two 96-well plates of samples for GBS, containing 145 samples, two negative controls (DNA-free controls) per plate,

and three positive controls (replicate samples) across plates. I diluted samples with gDNA concentration > 150 ng/µL to 100 ng/µL, and supplied ~1 µg of DNA for all samples.

GBS optimisation and sequencing was conducted by AgResearch Ltd. (Mosgiel, New Zealand). Initial optimisation of restriction enzymes used 8.6 µL of DNA at 100 ng/µL. Enzyme digestion was assessed using three enzymes: ApeKI, PstI, ApeKI/MspI, and PstI/MspI, and confirmed through gel electrophoresis. In double-digest samples (i.e., ApeKI/MspI and PstI/MspI), ApeKI was digested first and MspI was digested second, while PstI and MspI were digested together at the same time. The ApeKI digestion was performed at 75°C for two hours. For PstI/MspI digestion was performed at 37°C for 2 hrs followed by 65°C for 30 min. Following digestion, barcodes were ligated to each sample in duplicate using a ligation protocol consisting of 22°C for 60 min and 65°C for 30 min. Samples were purified through an Omega Bio-Tek column, and eluted in 30 µL elution buffer. PCR amplification was then conducted, with a thermocycling protocol of 72°C for 5 min, 98°C for 30 s, 18 cycles of 98°C for 30 s, 65°C for 10 s, and 72°C for 30 s, followed by a final extension step of 72°C for 5 min. Duplicate samples were then pooled and purified through a QIAGEN column and eluted in 30 µL elution buffer. Amplified products were assessed on an Agilent 2100 BioAnalyzer. The optimised GBS protocol used a double-digest with enzymes PstI-MspI. A single library was generated for all 145 samples and controls with a fragment length filter of 193–500 bp including adapter sequences. This library was sequenced on one lane of an Illumina HiSeq 2500 v4 sequencing run for 101 cycles.

4.3.3 Variant discovery

Scripts associated with the following workflows are detailed in my GitHub 'Himantopus' repository at https://github.com/natforsdick/Himantopus. I assessed sequence quality with FastQC v0.11.5 (Andrews, 2010), and assessed negative controls for the presence of contamination through BLAST searches against the nucleotide database (Altschul et al., 1990). As part of the AgResearch GBS quality control pipeline, preliminary *de novo* SNP discovery was performed with the UNEAK (Lu et al., 2013), and SNPs with minor allele frequency of zero or sample depth < 0.01 were removed. Limited results of this preliminary analysis are reported here as this was primarily used to confirm success of the GBS pipeline. Using a species-specific reference genome for variant discovery provides greater accuracy in variant-calling and subsequent analyses (Galla et al., 2019), and so I mapped GBS data to the stilt reference genomes produced in Chapter Two.

I conducted all sequence cleaning, mapping, and variant discovery and filtering using a local (University of Otago) computing cluster with 32 cores and 1 TB RAM. I demultiplexed and filtered raw sequences with Sabre v1.0 (Joshi, 2013) and adapter trimmed with Cutadapt v1.17 (Martin, 2011). I assessed the relationship between DNA concentration and raw sequence output using a Kendall rank correlation test for non-parametric data implemented with the *ggpubr* v0.2.3 package in R v3.5.1 (R Core Team, 2018). I compared mapping success of GBS data to the kakī and pied stilt reference genomes produced in Chapter Two (Galla et al., 2019). I indexed genomes with BWA v0.7.17 (Li & Durbin, 2009), and mapped the cleaned GBS reads to the genomes with BWA-MEM. To compare the proportion of successfully mapped reads to each reference genome, I used the SAMtools v1.2 (Li et al., 2009) *view* tool. The observation of higher overall mapping rates of GBS data to the kakī genome led to the decision to proceed with downstream analyses using these kakī-

mapped reads for all samples. The use of the kakī genome as reference could introduce biases when mapping data from pied stilts and kakī-poaka hybrids, such that species-specific SNPs may be identified at a lower rate for non-kakī individuals than for kakī, and pied stilt variability may also be underestimated using this approach. Nevertheless, the higher mapping success produced when mapping data from all individuals to the kakī reference genome results in a greater number of mapped reads available for variant discovery, which is likely to produce a greater number of variants overall, and thus greater confidence in downstream analyses. Furthermore, mapping to a single reference allows comparisons between SNP sets, and any additional future sequencing for investigation of hybridisation should use this same reference to maintain comparability between data sets. I pre-processed the mapped reads for variant discovery by adding read-group information with SAMtools, marking duplicates with Picard v2.18.0 (Picard Toolkit, 2019), and realigning indels with the Genome Analysis Toolkit (GATK) v3.5 (McKenna et al., 2010).

As this work was conducted at a time when the differences in outputs between the various variant discovery pipelines had not yet been evaluated for reduced-representation sequencing approaches for non-model species (but see now Wright et al., 2019), I compared five independent pipelines for reference-guided variant discovery (Figure 4.2). The 'GATK' pipeline used GATK's *HaplotypeCaller* and *GenotypeGVCFs* to call variants. The 'Samtools' pipeline used SAMtools v1.7 *mpileup* and BCFtools v1.6 variant caller (Li, 2011). The 'Platypus' pipeline used the *callVariants* tool in Platypus v0.8.1 (Rimmer et al., 2014) with minimum mapping quality of 20, minimum base quality of 20, minimum depth to call a variant of 2, and flag to generate indels set. The 'Stacks' pipeline implemented Stacks v2.2 (Catchen et al., 2013) reference-guided pipeline with default parameters. The mapped

sequence reads were passed as input to GATK, Samtools, Platypus, and Stacks.

The fifth pipeline, 'Tassel', was run

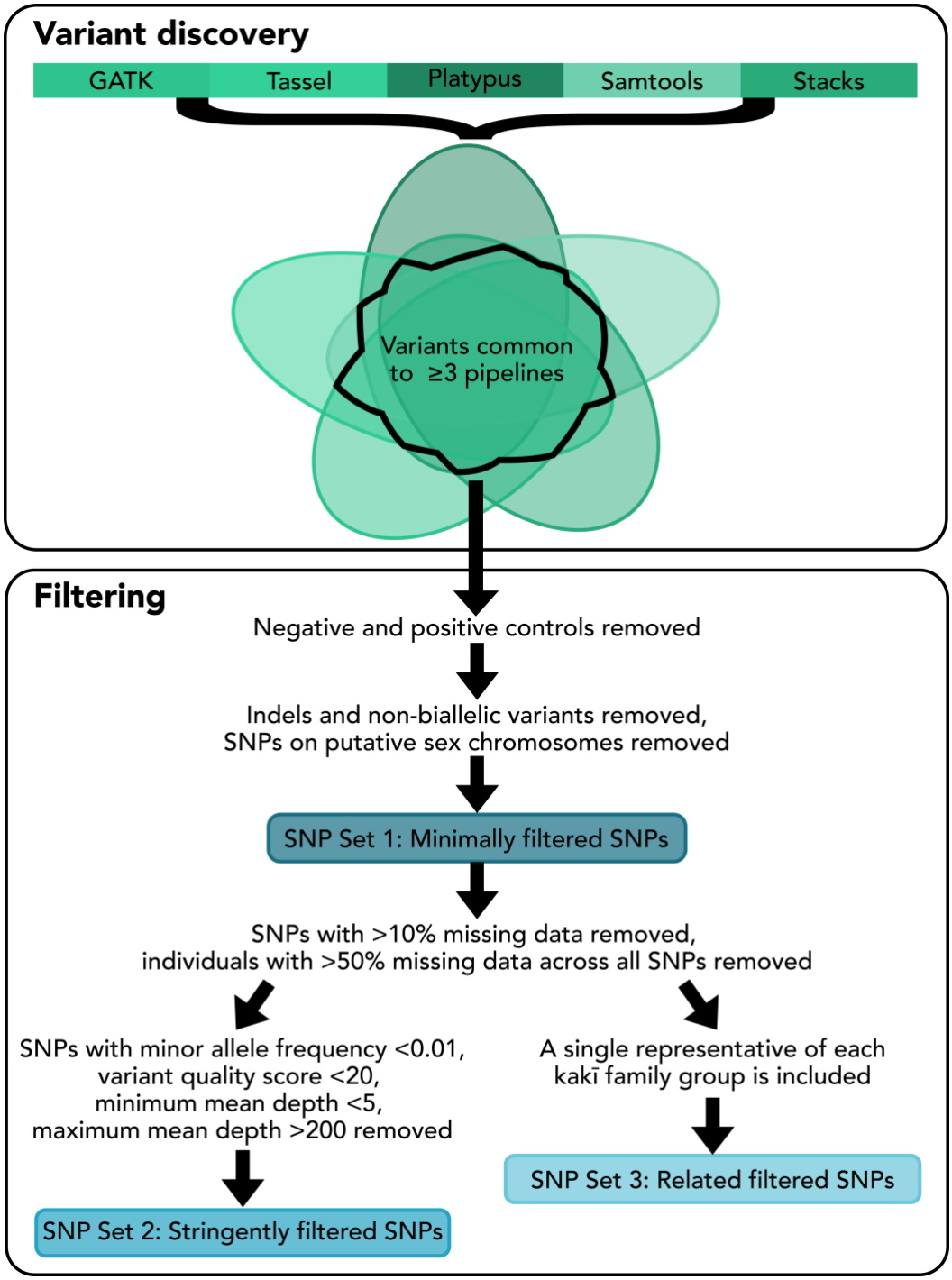


Figure 4.2: Variant discovery and SNP filtering pipelines implemented for GBS data from kakī, Australian pied stilts and poaka, and kakī-poaka hybrids to produce three SNP sets for downstream analyses of hybridisation.

independently with the raw multiplexed GBS data passed as input to TASSEL5-GBS2 v5.2.39 (Glaubitz et al., 2014). Tags were extracted from the data set with a

minimum quality score of 10, and then passed to BWA v0.7.12 for alignment against the reference genome. The resulting SAM file was passed back to TASSEL5-GBS2 for variant discovery with default settings, and all SNPs with quality ≥ 10 were retained.

4.3.4 Variant processing

By using a reference-guided approach, genomic location data was available for all variants, and so I could compare variants produced across the five pipelines using VCFtools v0.1.15 (Danecek et al., 2011) *vcf-compare* following standardisation of variant call format files with *vcf-convert*. I visualised the intersections of common variants among pipelines with the R package *UpSetR* (Conway et al., 2017; Lex et al., 2014). To improve confidence that the SNPs discovered were true SNPs rather than the result of sequencing or mapping error, I produced a single variant call file (VCF) comprising all variants detected via at least three pipelines from the intersections of variants common to multiple pipelines generated using VCFtools *vcf-isec*, *vcf-merge* and *vcf-sort*. This method provided confidence that true variants, rather than those resulting from sequencing or mapping error, were incorporated in the concordant variant set for downstream analysis. To produce a set of biallelic SNPs to investigate admixture between kakī and poaka, I conducted preliminary variant filtering of the composite variant set involving the removal of indels and multiallelic SNPs using VCFtools. To confirm absence of contamination and replicability of lab processes, preliminary filtering tests and downstream analyses retained negative and positive controls, and following these quality checks, these controls were removed for final filtering and analyses.

To compare the impacts of differing filtering strategies on results of downstream analyses, I produced three SNP sets: a minimally filtered, stringently filtered, and

related filtered sets (herein SNP Set 1, SNP Set 2, and SNP Set 3, respectively; Figure 4.2). To produce SNP Set 1, I first excluded the negative and control samples from the combined SNP set, and then removed indels and multiallelic SNPs (SNP Set 1 n = 145). Due to incomplete data regarding individual sex, I excluded SNPs located on putative sex chromosomes (Z and W, identified through BLAST alignments of the kakī genome assembly against the chicken genome in Chapter Two). To produce SNP Set 2, I further filtered SNP Set 1 to exclude sites with > 10% missing data, a minor allele frequency of < 0.01, and a minimum quality score < 20. SNPs with mean depth over all individuals between 5× and 200× were retained, while individuals with > 50% missing data across all sites were excluded (SNP Set 2 n = 140). I generated SNP Set 3 to disentangle the impacts of relatedness on population structure by again filtering SNP Set 1 to exclude loci with > 10% missing data, and retained a single representative from groups of related individuals, with the final set consisting of forty individuals. As the inclusion of congeneric individuals means this sample set does not meet the assumptions of Hardy-Weinberg Equilibrium (HWE), and the population history of kakī means that many individuals will be related to one another, I chose not to filter SNPs based on deviation from HWE and linkage disequilibrium (LD). Preliminary filtering investigation incorporated filtering of SNPs deviating from HWE and two methods of LD filtering (one by statistical correlation between SNPs, and one implementing a thinning approach to remove co-located SNPs) and found no observable difference in results of analyses (preliminary data not shown), but the reduced number of SNPs available is likely to limit the power of these sets to detect introgression.

Following filtering, I used VCFtools to generate summary statistics for each SNP set including SNP density and distribution across the kakī genome, per-individual rates of missing data, coverage depth per site, heterozygosity (measured in terms of

nucleotide diversity; π), deviation from HWE ($P < 0.001$), and population differentiation between kakī and pied stilts using the method of Weir & Cockerham (1984). I visualised these data with the R package *ggplot2* v3.2.1 (Wickham, 2016). Further identification of loci deviating from HWE used the False Discovery Rate correction for multiple tests implemented in R with $\alpha = 0.05$. I then converted SNP sets from VCF with STACKS *populations --plink* to produce MAP and PED files, and *--structure* to STRUCTURE input files, before conversion with PLINK v1.9 (Purcell et al., 2007) *--make-bed* to binary files in BED, BIM, and FAM formats, or with *--recode A* to produce raw additive component files (both using the *--allow-extra-chr* flag) for downstream analyses. During this conversion process, SNPs with malformed VCF fields due to version format issues, or that did not meet sample or population constraints were removed, and additional summary statistics were produced, including genotyping rates and population-specific metrics for the mean number of samples per locus, nucleotide diversity, and numbers of polymorphic sites and private alleles.

4.3.5 Discriminant Analysis of Principal Components

In an exploratory multivariate approach to population clustering, I conducted Discriminant Analysis of Principal Components (DAPC) with the R package *adegenet* v2.1.1 (Jombart, 2008; Jombart et al., 2010; Jombart & Ahmed, 2011; Jombart & Collins, 2015b), taking raw additive component files (RAW) and the associated MAP files containing variant information generated from each SNP set as input. DAPC attempts to partition variance in a between-group and within-group manner to maximise the discrimination between groups. Using a multivariate approach allows for fine-scale assessment of population structure, without relying on population genetic models, and so is independent of the assumptions of HWE or

linkage equilibrium associated with ADMIXTURE and fastSTRUCTURE analyses (Jombart et al., 2010).

DAPC uses *a priori* information of the number of clusters present in the data set, and then assesses the discriminants that best explain those clusters. To prevent overfitting of the data, I optimised DAPC parameters using the Bayesian Information Criterion (BIC) and *a*-scores, and performed cross-validation following Jombart & Collins (2015). The *a*-score measures the trade-off between the power of discrimination and potential to overfitting the data, using a randomisation of the data to determine when cluster assignment is successful due to the analysis or due to random discrimination, and penalises the reassignment score by the number of retained principal components (PCs). Cross-validation confirms the appropriate numbers of PCs, using a random seed to produce 1,000 replicate runs with a training set of 80% of the data across up to sixty PCs (reduced to thirty for the smaller sample size in SNP Set 3). The accuracy of the retained PCs was then tested with the remaining 20% of the data, and the PCs retained for the final DAPC were based on that which produced the lowest mean squared error and highest mean success. The optimised DAPC analyses were visualised to infer species differentiation and individual clustering.

4.3.6 Analysis of introgression with ADMIXTURE

To estimate individual assignment to population clusters and detect introgression, I analysed each SNP set with a maximum likelihood method implemented in ADMIXTURE v1.3.0 (Alexander et al., 2009). Although ADMIXTURE can incorporate sex-linked markers (Shringarpure et al., 2016), putative sex-linked markers located on the kakī Z and W pseudochromosomes had been excluded in the filtering process due to their putative nature and incomplete individual sex data. To minimise

stochasticity across multiple runs, I conducted 100 iterations of ADMIXTURE analysis with each SNP set for K = 1–6, where K represents the hypothesised number of population clusters, in multithreaded mode via an array on the NeSI (New Zealand eScience Infrastructure) high performance computing (HPC) 'Mahuika' platform, using a random seed, ten-fold CV, and with point estimation terminating when the change in log-likelihood increased by < 0.0001. The range of K-values was selected to allow for differentiation between the two species, kakī and pied stilts, along with potential population structuring among kakī, or differentiation between Australian pied stilts and poaka. To determine the most appropriate value of K for each SNP set, I averaged CV errors across the 100 iterations and visualised the results, with the lowest CV error representing the most likely K. I visualised mean assignment probabilities (Q-values)[3] across all iterations with *pophelper* v2.3.0 (Francis, 2017) in R. I used *pophelper* for file conversion for input to CLUMPP to handle label switching. Consensus Q-values for each individual were calculated with the Greedy algorithm over 100 iterations in CLUMPP vMacOSX 1.1.2 (Jakobsson & Rosenberg, 2007), and I visualised the results with *pophelper* in R. I manually assessed the final Q-matrices for all individuals using the predefined assignment threshold to assign individuals as kakī. To test the correlation between mean assignment probabilities across each SNP set I used a Kendall rank correlation test for non-parametric data implemented with *ggpubr* v0.2.3 in R.

[3] ADMIXTURE refers to the resulting Q-values as the admixture or ancestry coefficient of individuals for the respective population clusters, while STRUCTURE and associated programs refer to these as admixture proportions. While I acknowledge these differences, they have the same essential meaning, and so here I refer to Q-values as assignment probabilities regardless of the analysis method from which they were derived.

4.3.7 Analysis of introgression with fastSTRUCTURE

To compare the results of ADMIXTURE analyses using a different statistical approach based on the same underlying population genetic model, I used fastSTRUCTURE (Raj et al., 2014) , a version of STRUCTURE (Falush et al., 2003; Pritchard et al., 2000) that implements a variational Bayesian inference algorithm optimised for large genomic data-sets. Here, I passed the same input files from ADMIXTURE analyses for each SNP set to fastSTRUCTURE implemented via NeSI's Mahuika HPC platform, with K = 1–6, a simple prior, a random seed, a convergence criterion of 1×10^{-5}, and ten-fold CV. Replicate runs of fastSTRUCTURE were unnecessary, as each run consists of multiple iterations. fastSTRUCTURE uses both CV error and maximum likelihood values to determine the most likely *K*-value, and these were visualised with *ggplot2* in R for comparison with the *K*-values indicated by ADMIXTURE. I visualised results of individual posterior mean admixture proportions with *pophelper* in R. I assessed correlations between assignment probabilities produced from fastSTRUCTURE analyses of each SNP set, and between the outputs of ADMIXTURE and fastSTRUCTURE with a Kendall rank correlation test for non-parametric data implemented with the *ggpubr* v0.2.3 package in R. To investigate whether there was a similar relationship to that observed by Steeves et al. (2010) where individuals with darker plumage had a higher probability of assignment to the kakī cluster, I visualised the relationship between plumage node and assignment probability with *ggplot2* in R. No statistical test was performed here due to the small number of individuals representing each of the light plumage nodes.

4.3.8 Combining pedigree data with genomic population assignment data

Following all admixture and population clustering analyses, only a small number of node J kakī individuals were identified with < 100% kakī assignment across any of the analyses conducted. A small number of hybrids individuals are included among the founders in the kakī pedigree (Galla, 2019), and so to determine if these assignment probabilities could be attributed to recorded hybrid ancestry (< 8 generations deep), I investigated the ancestry of these individuals in the kakī pedigree (Galla et al., 2020).

4.4 Results

Genotyping-by-sequencing for 145 stilts and a reference-guided approach to multi-pipeline variant discovery produced 140,948 SNPs for use in downstream analyses that detected no evidence of introgression from poaka among the kakī assessed, verifying the findings of the previous genetic analysis by Steeves et al. (2010). These results lend support to the finding that despite interspecific hybridisation, the combined effects of genetic-informed conservation management and natural processes including outbreeding depression as assessed by Steeves et al. (2010) have resulted in the maintenance of the genetic integrity of kakī.

4.4.1 Genotyping-by-sequencing outputs and quality control

Of the 250 gDNA samples available, 145 extractions contained DNA of the required quantity and quality for GBS, including 66 of the 106 (63.2%) adults alive in the wild kakī population when this work began in 2017 (Supplementary Table 1). GBS of the pooled set of kakī, Australian pied stilts and poaka, and kakī-poaka hybrids produced

a total of 303,639,199 raw sequences with length 35–101 bp and high sequence quality (Appendix E). Demultiplexing produced an average of 2,024,530 ± SD 1,031,208.21 reads per sample (Supplementary Table 2), and no samples failed to sequence. Negative controls produced a low number of reads (mean = 2,585 ± SD 1554.76 reads per negative). Contamination checks of negative controls produced no matches to the BLAST nucleotide database, while the top matches for sample sequences were to other avian species: North Island brown kiwi (*Apteryx australis mantelli)*, American golden eagle (*Aquila chrysaetos canadensis)*, ruff (*Calidris pugnax),* killdeer (*Charadrius vociferus)*, and crested ibis (*Nipponia nippon)*. A total of 38,272 SNPs were discovered with the *de novo* UNEAK pipeline used by AgResearch for quality control, with an average depth of 4.15 reads/SNP. Following removal of negatives, and filtering removing SNPs with a minor allele frequency of zero or sample depth < 0.01, 38,250 SNPs remained.

4.4.2 Variant discovery and filtering

Mapping of trimmed, filtered reads for all samples to the reference kakī genome produced an average 1,138,306.05 ± SD 597,406.95 mapped reads per individual (Supplementary Table 2). This represents an average of 85.4% of reads per sample successfully mapped to the reference genome, compared with an average of 77.4% of reads successfully mapped to the Australian pied stilt genome (Table 4.1). Mapping success to the kakī genome was higher than to the Australian pied stilt genome for all individuals regardless of species (Supplementary Table 2).

Table 4.1: Sequencing outputs and mapping success of GBS data from kakī, Australian pied stilts and poaka, and kakī-poaka hybrids averaged by species. APS = Australian pied stilt, n = total individuals. Overall includes all samples along with negative and positive controls.

	Mean reads	Mean cleaned reads	Mean reads mapped to kakī genome	Mean reads mapped to APS genome
Kakī (n = 130)	2,094,197.47	1,378,947.26	1,211,024.32	1,090,390.84
Pied stilts (n = 6)	1,722,470.33	966,686.33	879,153.17	799,878.33
Hybrids (n = 9)	1,181,117.00	812,732.00	732,418.11	668,224.33
Overall	1,971,321.13	1,293,796.12	1,138,306.05	1,025,813.82

Variant calling with GATK produced 35,441 variants, even fewer than the *de novo* UNEAK pipeline, while SAMtools produced the most at 488,940 variants (Figure 4.3). There were 177,437 variants common to ≥ 3 pipelines (Figure 4.3). Despite GATK producing the fewest variants among the five pipelines, the majority (92.68%) were retained in the common variant set. SAMtools had the lowest proportion (35.14%) of discovered variants retained in the common set. Total SNPs produced following the three filtering strategies ranged from 15,851 to 140,948 (Figure 4.4). The five individuals that had produced the fewest raw sequences (26,725–313,884 reads) were subsequently excluded from the stringently filtered SNP Set 2 due to high (> 50%) levels of missing data. Total genotyping rate for the minimally filtered set SNP Set 1 was 50.88%, compared with 96.91% for the SNP Set 2 and 93.74% for the related filtered SNP Set 3. Conversion from VCF for downstream analysis identified a small number of fixed SNPs among SNP Set 1 and SNP Set 3 (7,916 and 308 respectively).

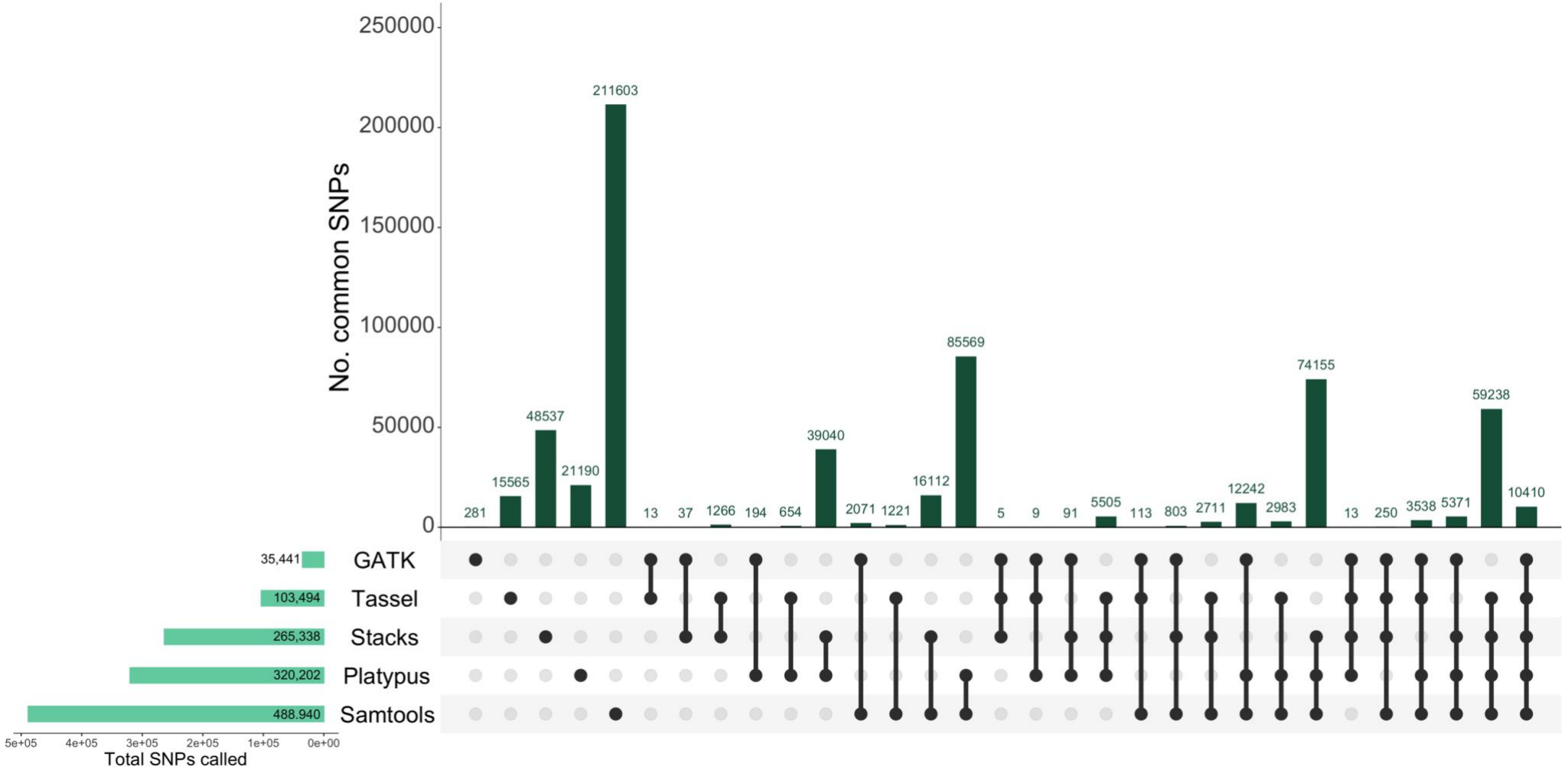


Figure 4.3: UpSetR plot of the intersections of the total variants discovered from GBS data for kakī, Australian pied stilts and poaka, and interspecific hybrids across five variant discovery pipelines: GATK, Platypus, Samtools, Stacks, and Tassel,. Bottom left bars represent the total number of variants discovered with each pipeline, while the main bar plot represents the number of variants common to multiple pipelines as indicated by the points below.

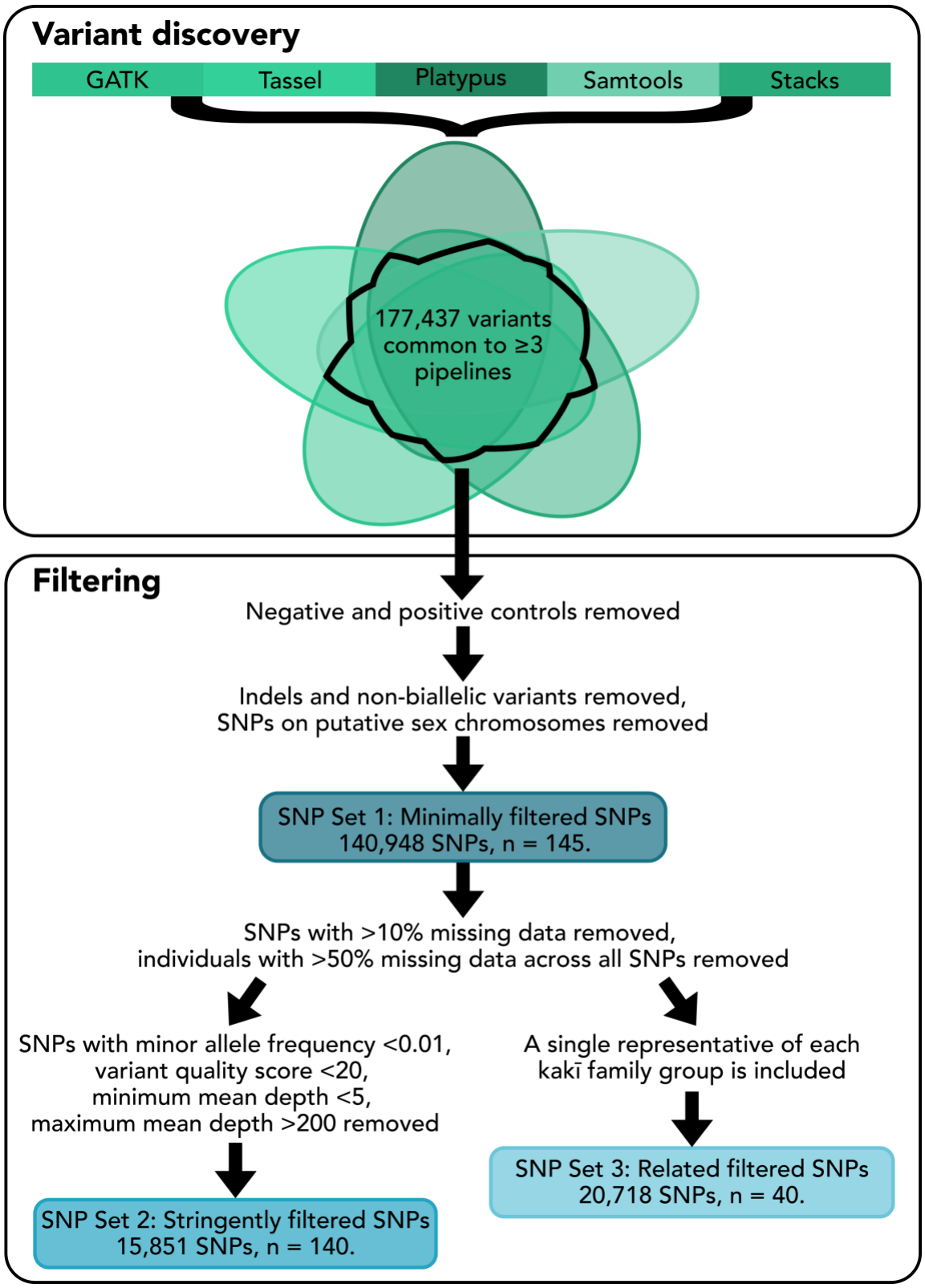


Figure 4.4: Results of variant discovery from GBS data generated for kakī, Australian pied stilt and poaka, and kakī-poaka hybrids, and subsequent application of three filtering strategies to produce SNP Set 1, SNP Set 2, and SNP Set 3. n = final number of individuals included in data set, SNP = single-nucleotide polymorphism.

Exploratory statistics for the three SNP sets indicated an even distribution of SNPs throughout the genome, with a mean density ranging from 0.014–0.128 SNPs/kb across sets (Table 4.2, Appendix E). Mean depth of coverage per individual ranged from 3.007–11.663×. Mean nucleotide diversity (π) was highest for SNP Set 3 (π = 0.139 ± SD 0.0977), and consistently lowest among kakī (Table 4.2, Table 4.3). Kakī were well-differentiated from pied stilts with mean F_{ST} ranging from 0.423 (SNP Set 3) to 0.622 (SNP Set 2; Table 4.2).

Table 4.2: Summary statistics for three SNP sets produced from GBS data for stilts. Mean ± standard deviation (SD). SNPs = single-nucleotide polymorphisms, KB = kilobase, HWE = Hardy-Weinberg Equilibrium, FDR = False Discovery Rate, FST = measure of population differentiation, SNP = single-nucleotide polymorphism Ts/Tv = transitions to transversions.

	SNP Set 1	**SNP Set 2**	**SNP Set 3**
Total SNPs	140,948	15,851	20,718
Total samples	145	140	40
Total Kakī / Pied / Hybrid	130 / 6 / 9	125 / 6 / 9	27 / 6 / 7
Mean depth per SNP per individual	3.007 ± 1.507	11.663 ± 6.107	8.245 ± 4.489
Mean per SNP depth	435.995 ± 957.198	1,632.77 ± 668.176	329.819 ± 492.634
Mean SNP quality	565.67 ± 1479.86	1,248.19 ± 2447.43	1,010.6 ± 2624.91
Mean frequency of missing data per individual	0.520 ± 0.130	0.031 ± 0.073	0.063 ± 0.087
Mean SNPs/KB	0.128 ± 0.557	0.014 ± 0.158	0.019 ± 0.186
Total singletons/private doubletons	39,778	0	2,762
Mean nucleotide diversity; π	0.134 ± 0.163	0.0912 ± 0.100	0.139 ± 0.098
SNPs deviating from HWE (FDR-corrected P ≤ 0.05)	17,132	758	361
Ts/Tv ratio	3.341	3.894	3.996
Weir & Cockerham mean F_{ST} (Kakī v Pied)	0.487	0.622	0.423
Weir & Cockerham weighted mean F_{ST} (Kakī v Pied)	0.585	0.637	0.525

Table 4.3: Population summary statistics from the three SNP sets produced from GBS data for kakī, Australian pied stilts and poaka, and kakī-poaka hybrids, as calculated during format conversion from VCF to PLINK.

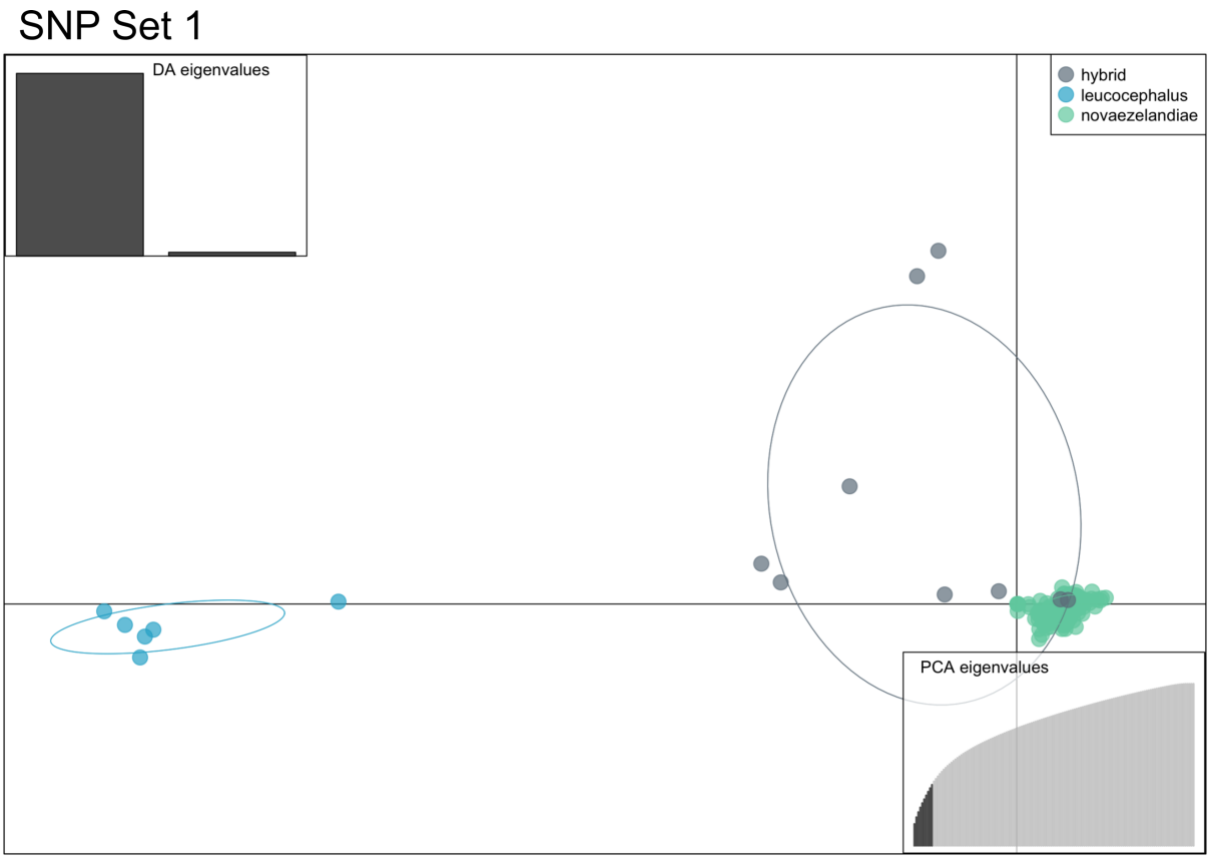
	SNP set 1			SNP set 2			SNP set 3		
	Kakī	**Pied**	**Hybrid**	**Kakī**	**Pied**	**Hybrid**	**Kakī**	**Pied**	**Hybrid**
Mean samples per locus	68.785	3.241	4.376	121.990	5.542	8.141	25.894	5.354	6.250
Polymorphic sites	64,279	85,258	67,241	6,729	14,699	13,621	5,917	18,422	14,765
Private alleles	20,481	24,959	8,050	389	1,100	122	521	4,419	819
Mean nucleotide diversity (π)	0.090	0.386	0.242	0.057	0.396	0.235	0.045	0.370	0.210

While a high number of singletons (alleles occurring only once across all individuals) or private doubletons (alleles homozygous in a single individual) were observed in SNP Set 1 (Table 4.2), the stringent filtering strategy including filtering on a minor allele frequency of 0.01 employed for SNP Set 2 removed all of these. Following False Discovery Rate correction for multiple tests, 17,132 loci from SNP Set 1 were found to significantly deviate from HWE (corrected *P*-value ≤ 0.05), compared with 758 and 361 loci from SNP sets 2 and 3 respectively.

Format conversion of SNP Set 1 from VCF to PLINK for all downstream analyses identified 132,202 variant SNPs among the set of 140,948 SNPs included in the VCF (Figure 4.4), with most of those excluded due to presence as singletons or with very low minor allele frequencies. No SNPs were excluded in the conversions of SNP sets 2 and 3, which produced 15,851 and 20,410 variant SNPs, respectively. Per-population summary statistics consistently identified pied stilts as having higher diversity in terms of nucleotide divergence, a greater number of variant sites, and more private alleles than either kakī or hybrids (Table 4.3). Kakī always displayed the lowest nucleotide diversity and fewest polymorphic sites.

4.4.3 Discriminant Analysis of Principal Components

DAPC analyses implemented for each SNP set with *adegenet* produced concordant clustering patterns across all SNP sets, with Australian pied stilts and poaka assigned to one distinct cluster and kakī assigned to a second cluster, with hybrids intermediate to the two species, though grouping more closely with kakī than poaka (Figure 4.5). Two node I/J hybrid individuals (DNA777 and DNA779) consistently clustered among kakī in all DAPC analyses.



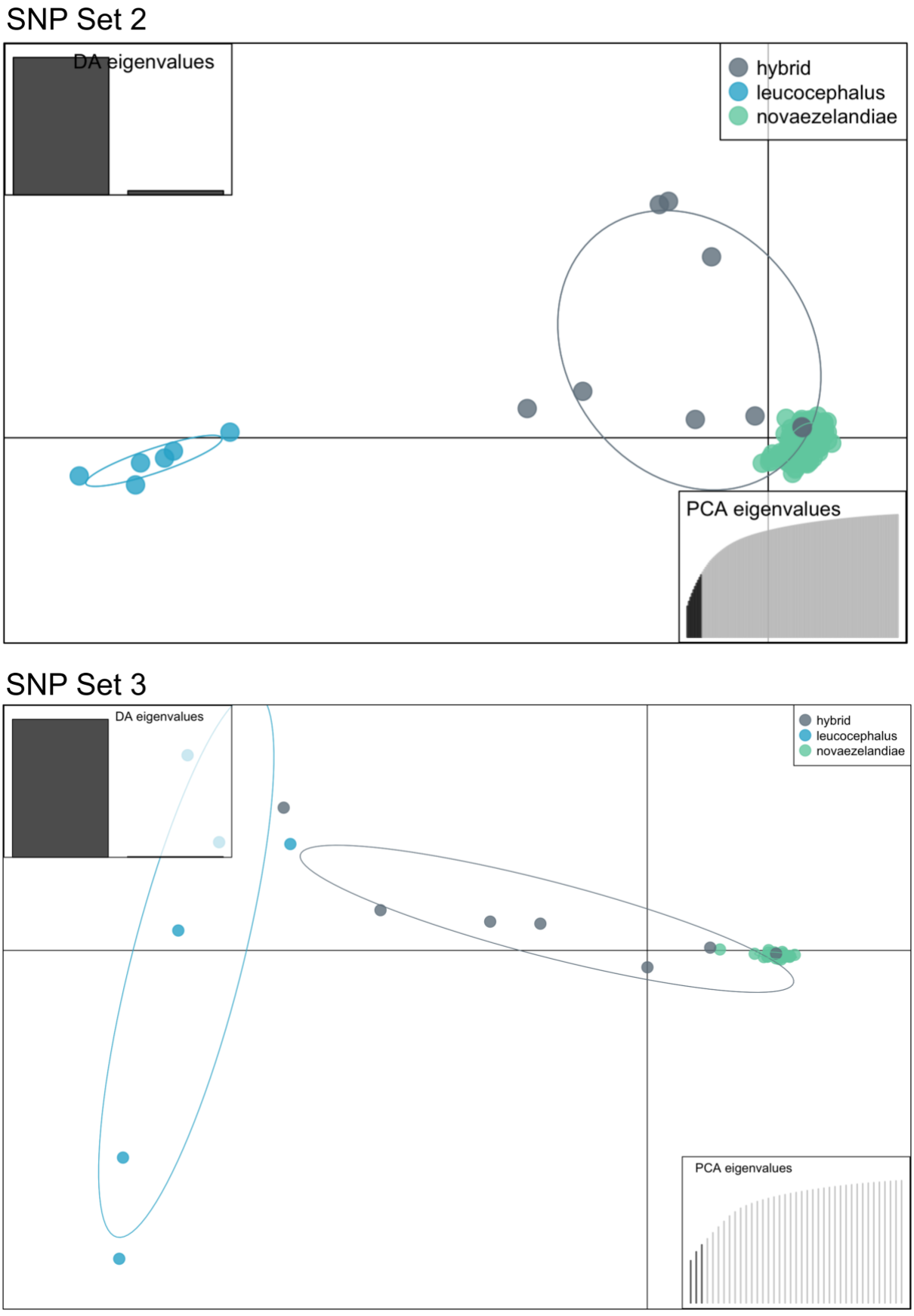


Figure 4.5: Scatterplots of a Discriminant Analysis of Principal Components (DAPC) produced from each of SNP Set 1, SNP Set 2, and SNP Set 3 from GBS data for kakī, Australian pied stilts and poaka, and kakī-poaka hybrids conducted in adegenet. Individuals

are coloured according to predefined population information as one of: novaezelandiae = kakī, leucocephalus = pied stilt, hybrid = kakī-pied stilt hybrid. The closer individuals are to one another, the more likely they are to have shared genetic ancestry. DAPC analysis was optimised using a-score, cross-validation, and BIC to derive the appropriate number of principal components (SNP sets 1 and 2: 2 discriminants, 10 PCs, SNP Set 3: 2 discriminants, 3 PCs). The 67% inertial ellipses around each cluster represent the variance of the clusters depicted. The insert of PCA eigenvalues represents the variation explained by the PCs, and the insert of DA eigenvalues represents the magnitude of variation explained by the two discriminants.

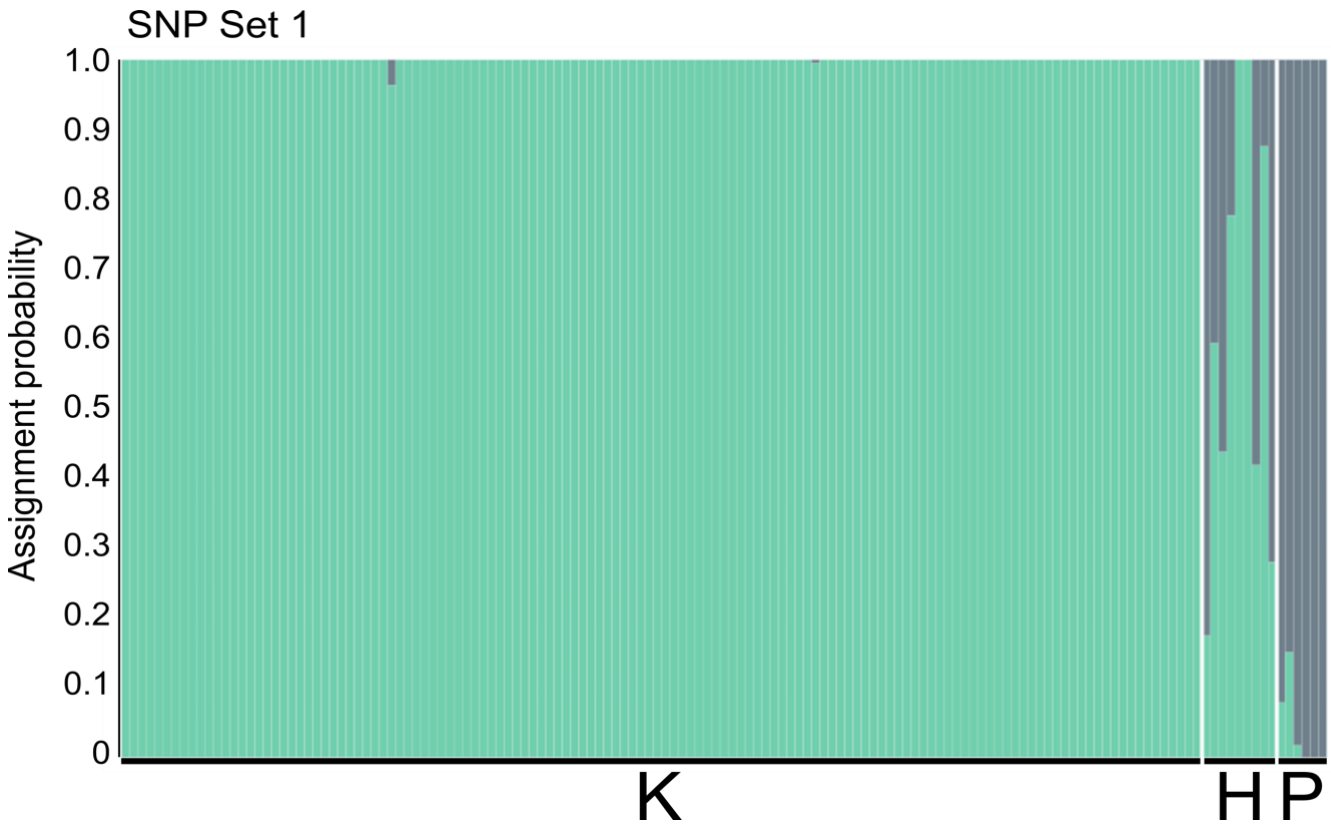
4.4.4 Estimation of the most appropriate *K*-value

K = 2 was indicated as the most likely number of clusters for both ADMIXTURE and fastSTRUCTURE analyses of all SNP sets (Appendix E) based on CV error and maximum likelihood values, concordant with the results of DAPC identifying two distinct clusters. These two clusters differentiated kakī from Australian pied stilts and poaka. Thus, only results for *K* = 2 for ADMIXTURE and fastSTRUCTURE analyses are reproduced below.

4.4.5 ADMIXTURE analysis of introgression

In the results of ADMIXTURE analyses, all individuals categorised as kakī (node J individuals) had assignment probabilities to the kakī cluster above the pre-defined 95% threshold, regardless of SNP set, and only nine individuals were assigned as kakī with a probability below 100% (Figure 4.6, Table 4.4, Supplementary Table 3). Only individuals DNA1252 and DNA1429 were assigned as kakī with less than 100% probability in analysis of more than one SNP set (Table 4.4). Standard errors around the assignment probabilities produced from bootstrapping were low across all three SNP sets (maximum error about the individual *Q*-values averaged across 100 ADMIXTURE runs ranged from 0.0242 to 0.0258; Supplementary Table 3b). The Australian pied stilts and poaka individual Poaka1 were consistently assigned with

100% probability to the pied stilt cluster (mean assignment probability of all poaka and Australian pied stilts to the pied cluster over all SNP sets = 0.9329 ± SD 0.0828; Table 4.5). In a preliminary ADMIXTURE analysis of SNP Set 1, these three individuals were removed from the data to determine whether this was creating some skew in the results. This resulted in the remaining two poaka individuals attaining 100% pied stilt assignment, but did not alter kakī assignment probabilities, thus all Australian pied stilts were retained in all downstream analyses. Assignment probabilities for hybrids ranged from 17.43% (DNA2113) to 100% (DNA777 and DNA779) to the kakī cluster (mean *Q* = 0.6337 ± SD 0.2790; Table 4.5, Supplementary Table 3).



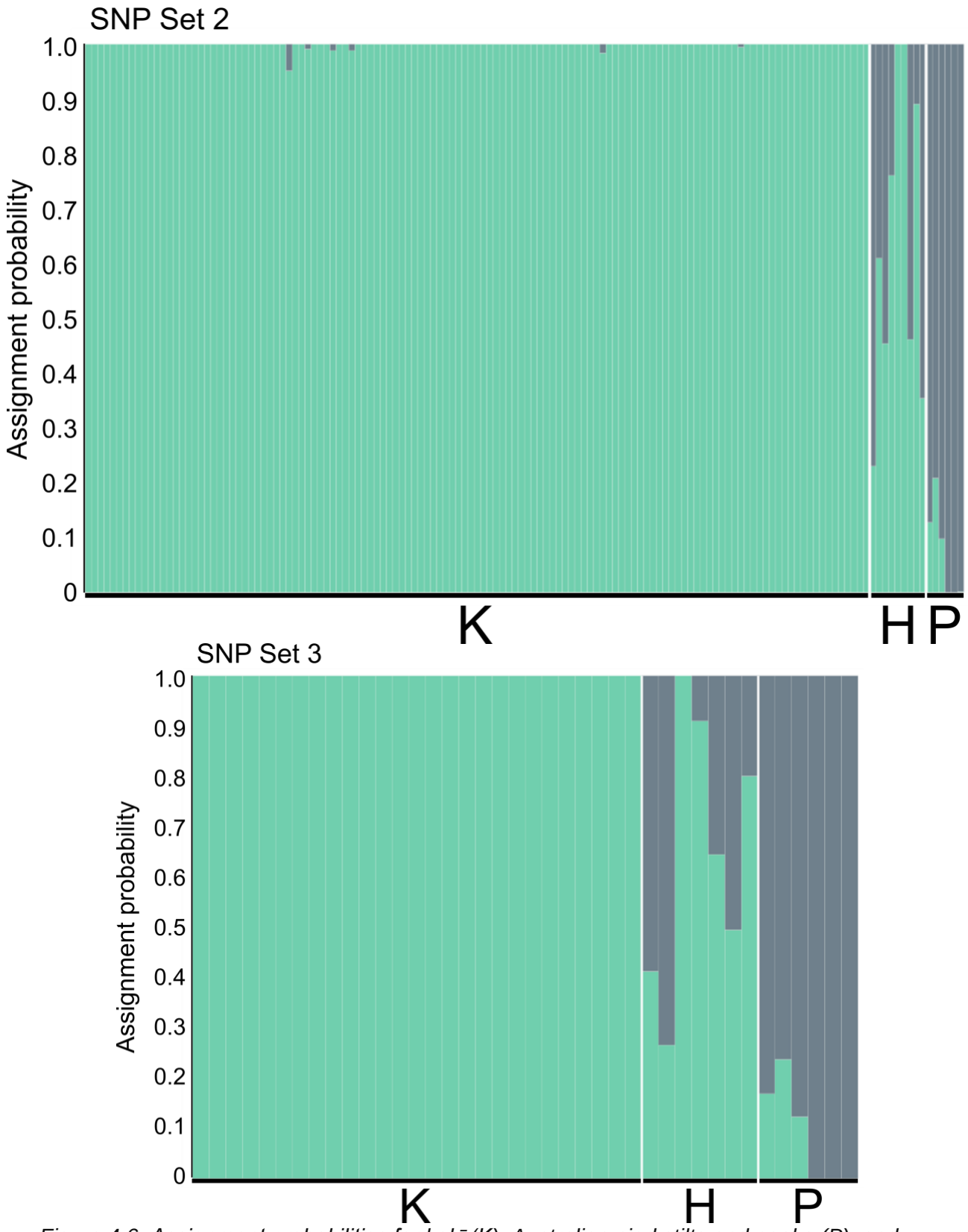


Figure 4.6: Assignment probabilities for kakī (K), Australian pied stilts and poaka (P), and kakī -poaka hybrids (H) produced via pophelper visualisation of CLUMPP-permuted ADMIXTURE results for SNP sets 1, 2, and 3 with K = 2. Each individual is represented by a vertical bar, with colours indicating the assignment probability to the kakī (green) or pied stilt (grey) cluster.

Table 4.4: Individual assignment probabilities (Q-values) for kakī that were assigned to the kakī cluster with probabilities above the 95% threshold, but below 100% in at least one of the population assignment analyses. Admix = ADMIXTURE analysis, FS = fastSTRUCTURE analysis. Green indicates the assignment probabilities below 100%.

	SNP Set 1		SNP Set 2		SNP Set 3	
DNA ID	**Admix Q**	**FS Q**	**Admix Q**	**FS Q**	**Admix Q**	**FS Q**
DNA1252	0.9633	0.9888	0.9514	0.9728		
DNA1429	0.9954	1.0000	0.9838	1.0000	1.0000	1.0000
DNA1483	1.0000	0.9998				
DNA1540	0.9999	1.0000	1.0000	1.0000		
DNA1620	1.0000	1.0000	0.9882	1.0000		
DNA1694	1.0000	1.0000	0.9945	1.0000		
DNA1699	0.9999	1.0000	1.0000	1.0000		
DNA1707	1.0000	0.9999				
DNA1717	1.0000	0.9997				
DNA240	1.0000	0.9999				
DNA451	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000
DNA897	1.0000	1.0000	0.9882	1.0000	1.0000	1.0000
DNA932	1.0000	1.0000	0.9910	1.0000	1.0000	1.0000

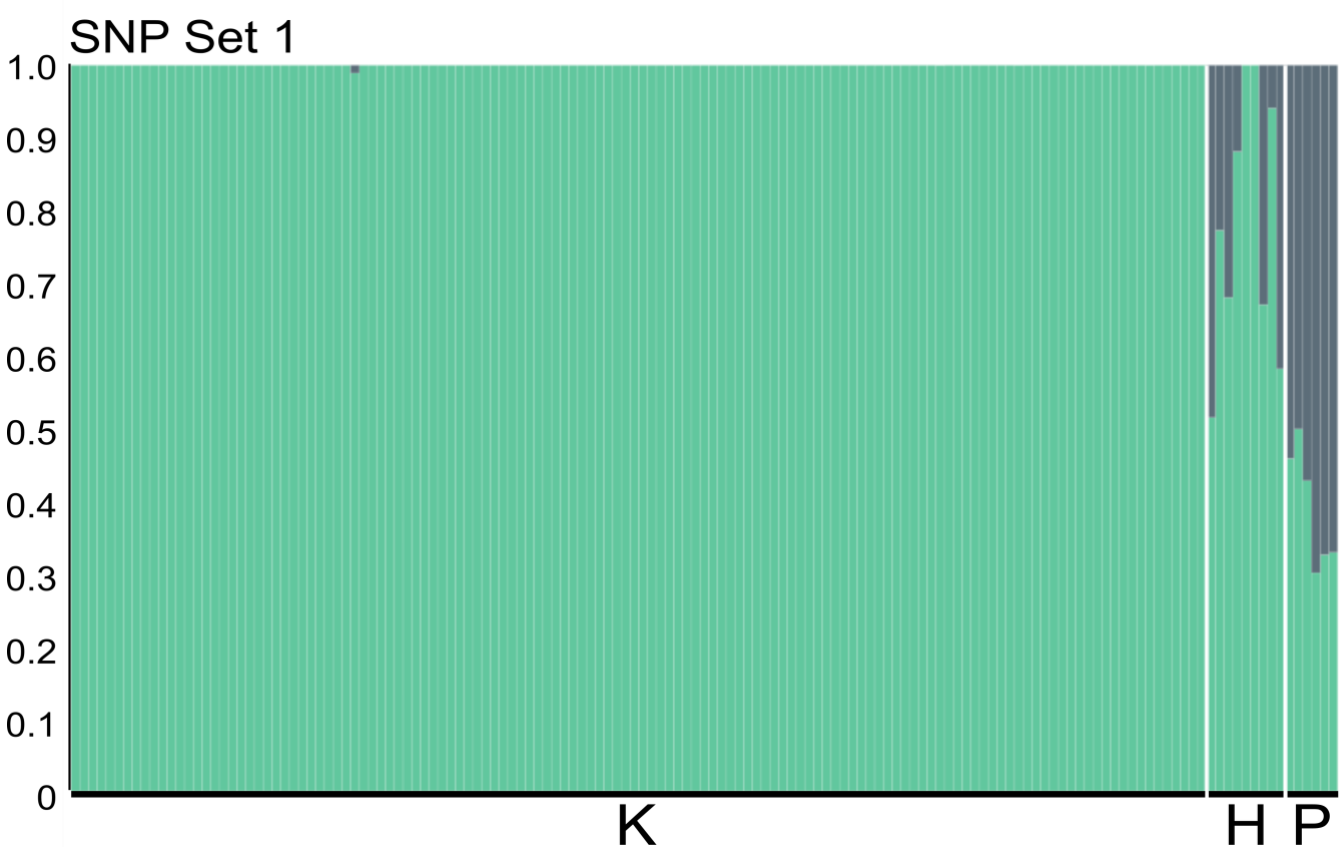
Table 4.5: Average assignment probability ± SD of pre-defined kakī, Australian pied stilts and poaka, and kakī-poaka hybrids to the respective cluster as calculated from ADMIXTURE analysis of each of SNP sets 1, 2, and 3, derived from stilt GBS data.

	SNP Set 1	SNP Set 2	SNP Set 3
Kakī to the kakī cluster	0.9997 ± 0.0032	0.9992 ± 0.0049	1.0000 ± 0.0000
Pied to the pied cluster	0.9590 ± 0.0617	0.9276 ± 0.0868	0.9121 ± 0.1030
Hybrid to the kakī cluster	0.6175 ± 0.3098	0.6399 ± 0.2859	0.6466 ± 0.2713

4.4.6 fastSTRUCTURE analyses of introgression

fastSTRUCTURE analysis with *K* = 2 assigned all pre-defined kakī to the kakī cluster with assignment probabilities above the 95% threshold (Figure 4.7), and only five individuals were assigned with probabilities below 100% (Table 4.4). Analysis of SNP Set 1 found three kakī had assignment probabilities above the 95% threshold but below 100%, including two individuals that had not been identified by any

previous analyses (DNA1483 and DNA1717, Table 4.4). These two individuals had been excluded from SNP Set 2 due to high levels of missing data (> 50%). Node I/J hybrid individuals DNA777 and DNA779 were always assigned to the kakī cluster with 100% probability. No Australian pied stilts or poaka were assigned with 100% probability to the pied stilt cluster. Calculation of mean *Q*-values across all SNP sets found Australian pied stilts and poaka were assigned to the pied stilt cluster with probabilities ranging from 50.12% to 75.62% across all SNP sets (Table 4.6). Hybrid assignment probabilities to the pied stilt cluster ranged from 1.40% to 54.31%.



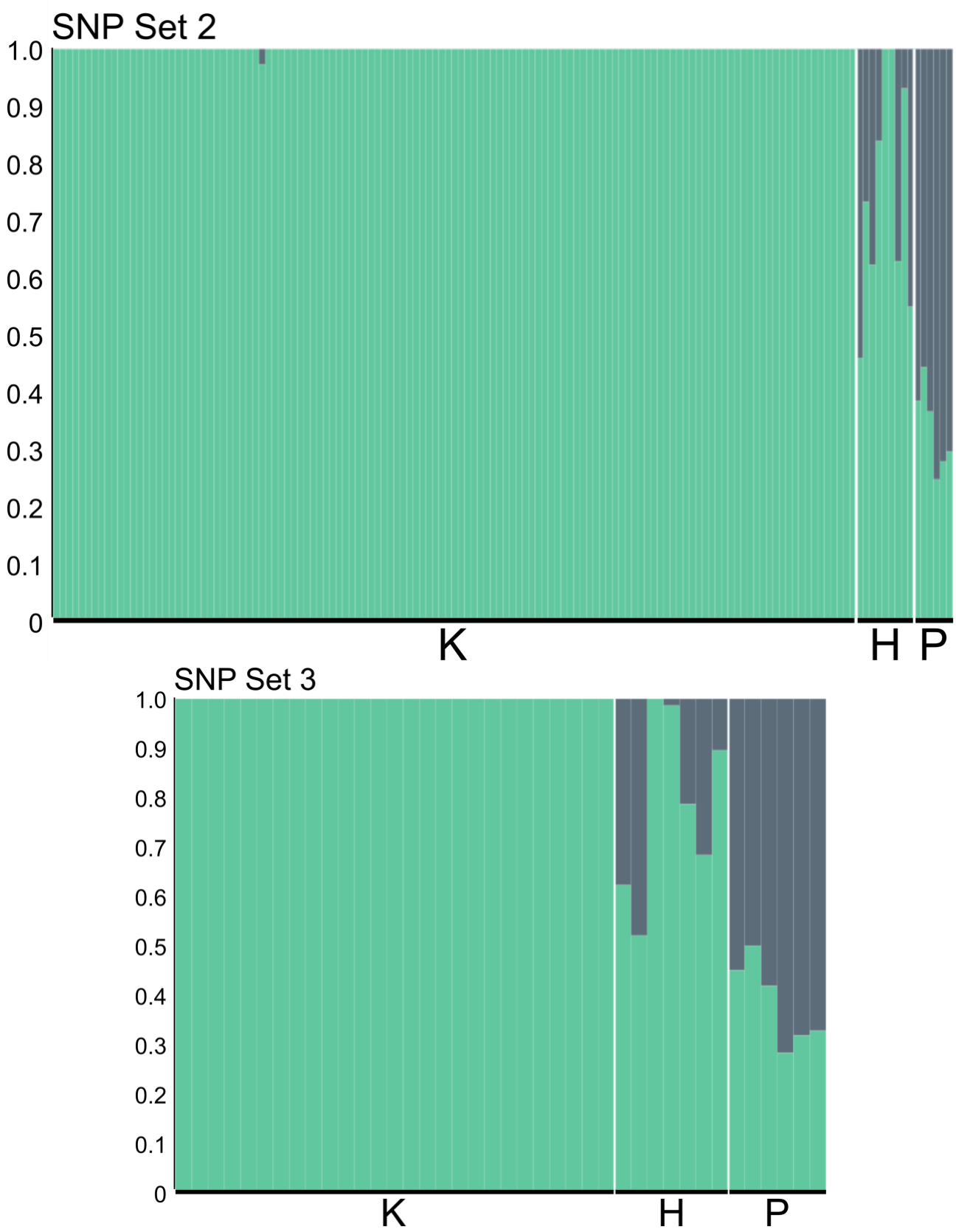


Figure 4.7: Assignment probabilities produced from fastSTRUCTURE analysis (K = 2) for kakī (K), Australian pied stilts and poaka (P), and kakī-poaka hybrids (H) visualised with pophelper from stilt GBS data across SNP sets 1, 2, and 3. Each vertical bar represents one individual, with the colour representing the assignment probability to the kakī (green) or pied stilt (grey) cluster.

Table 4.6: Average assignment probability ± SD of pre-defined kakī, Australian pied stilts and poaka, and kakī-poaka hybrids to the respective cluster as calculated from fastSTRUCTURE analysis of each of SNP sets 1, 2, and 3, derived from stilt GBS data.

	SNP Set 1	SNP Set 2	SNP Set 3
Kakī to the kakī cluster	0.9999 ± 0.0010	0.9998 ± 0.0024	1.0000 ± 0.0000
Pied to the pied cluster	0.6100 ± 0.0821	0.6671 ± 0.0746	0.6202 ± 0.0860
Hybrid to the kakī cluster	0.7823 ± 0.1819	0.7501 ± 0.2015	0.7838 ± 0.1861

4.4.7 Correlations between individual assignment probabilities derived from different SNP sets and analyses methods

A highly-significant correlation was observed between assignment probabilities (mean *Q*-values) derived from ADMIXTURE and fastSTRUCTURE analyses of the three SNP sets (Figure 4.8). Similarly, there was a strong significant correlation between individual assignment probabilities derived from each SNP set for both ADMIXTURE and fastSTRUCTURE analyses (Figure 4.9). When individual assignment probabilities to the kakī cluster (mean *Q*-values) were plotted against plumage node, there appeared to be a trend of individuals with darker plumage having higher likelihood of belonging to the kakī cluster (Figure 4.10).

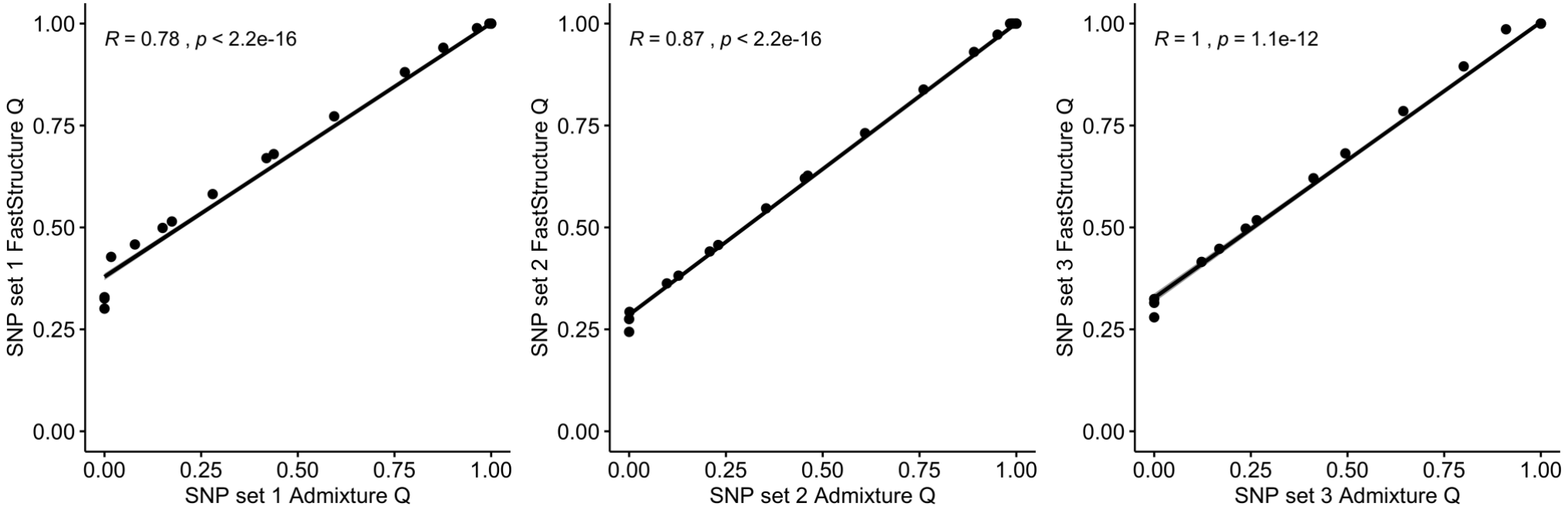


Figure 4.8: Scatterplots comparing mean individual assignment probability (Q) to the kakī cluster from ADMIXTURE analysis with those produced from fastSTRUCTURE analysis from each of SNP Set 1, SNP Set 2, and SNP Set 3 (K = 2). Trend lines are in black, with 95% confidence intervals about the trend line in grey. Kendall rank correlation coefficient R and associated significance value P included in the top left corner.

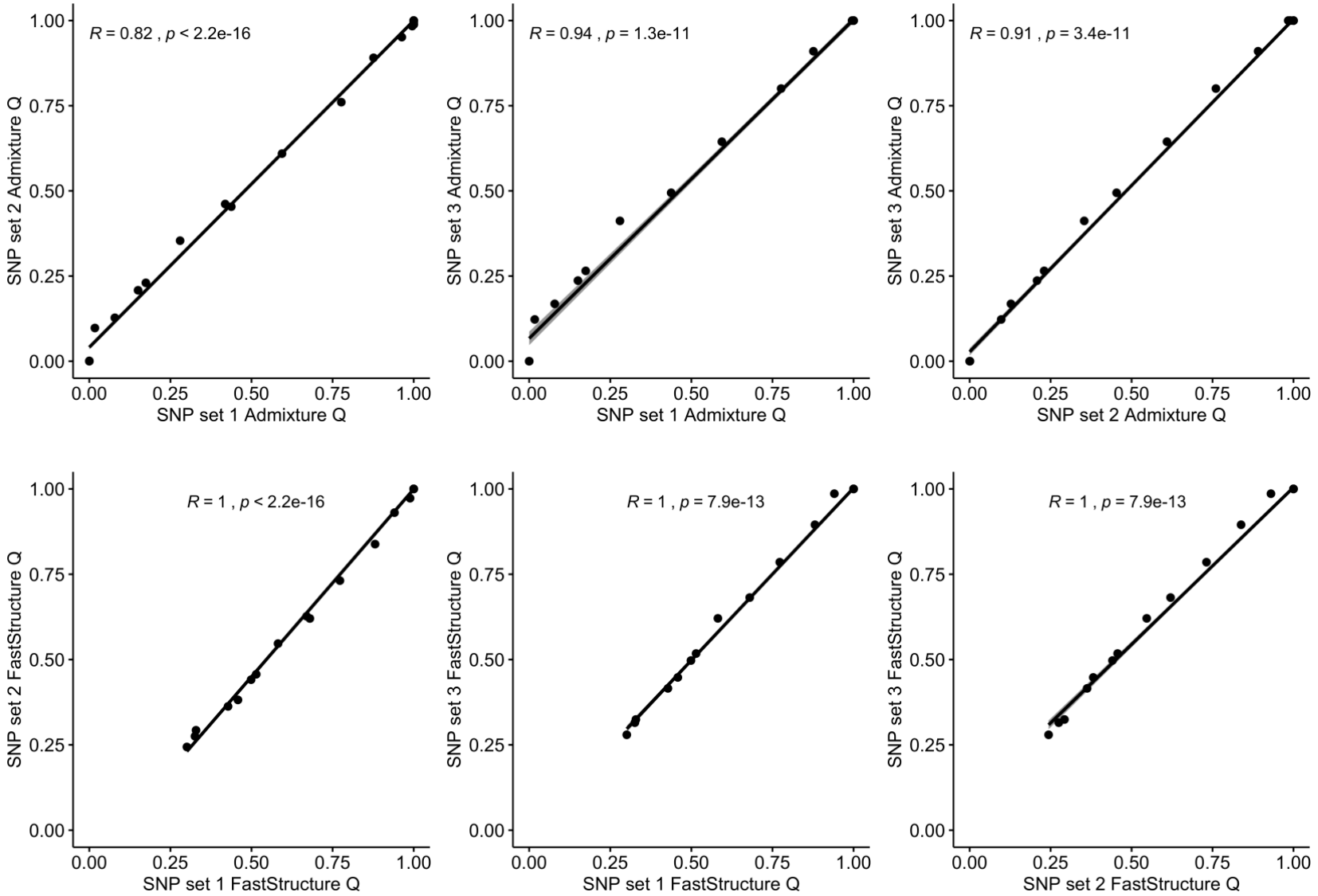


Figure 4.9: Scatterplots comparing mean individual assignment probability (Q) to the kakī cluster generated by ADMIXTURE and fastSTRUCTURE analyses (K = 2) between each of SNP sets 1, 2, and 3. Trend lines are in black, with 95% confidence intervals about the trend line in grey. Kendall rank correlation coefficient R and associated significance value P are included in the top left corners.

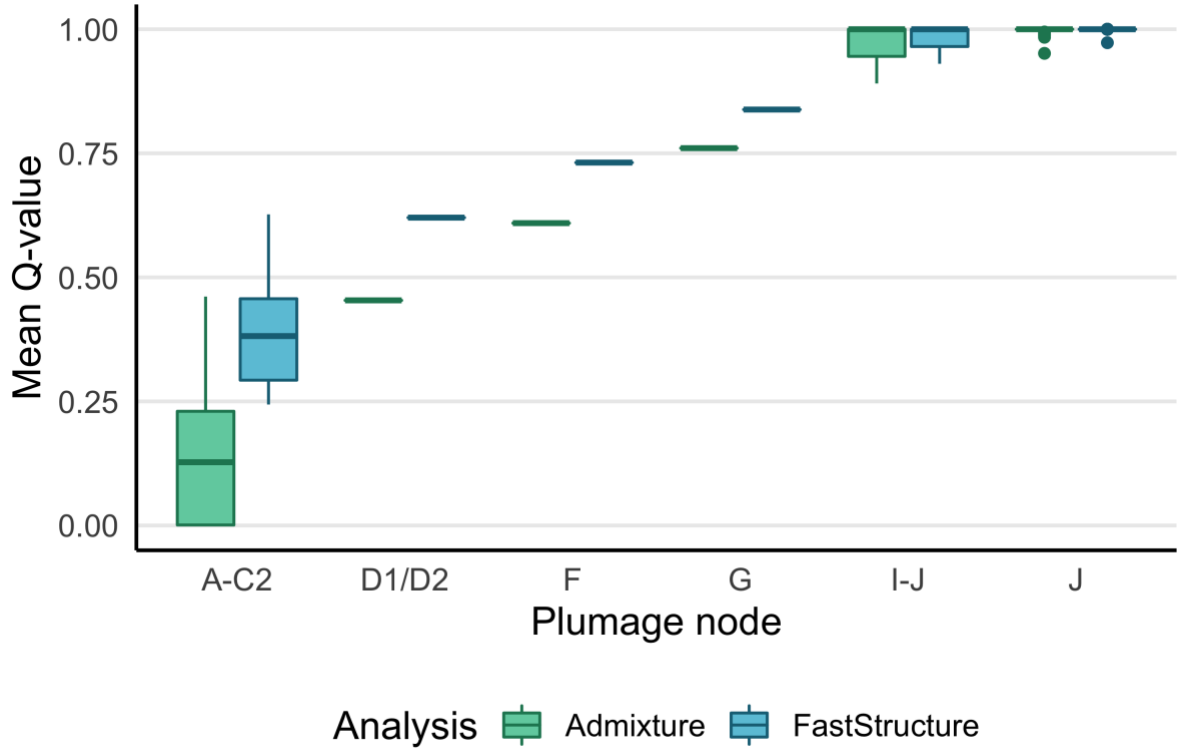


Figure 4.10: Boxplot displaying the median and interquartile range of mean individual assignment probability to the kakī cluster as determined through ADMIXTURE and fastSTRUCTURE analyses produced from GBS data including kakī, Australian pied stilts and poaka, and kakī-poaka hybrids across a range of stilt plumage nodes. Mean assignment probabilities for each individual were calculated across SNP sets 1, 2, and 3. Nodes A - C2 are predominantly white-bodied individuals representing Australian pied stilts and poaka, while node J are completely black kakī. Single points represent outliers, and whiskers describe the minimum and maximum values across the range.

4.4.8 Incorporating pedigree data with results of genomic analysis of introgression for individuals of interest

All node J individuals were assigned as kakī with probabilities above the 95% threshold, but a small number of these individuals had assignment probabilities in the range 95.00–99.99%, which may be due to recent hybrid ancestry (as recorded in the kakī pedigree; Galla et al., 2020). Pedigree assessment revealed only 17 kakī individuals included in this study had no recorded hybrid ancestry (i.e., all of their recorded ancestors were individuals with plumage node J representing completely

black birds). Among the node J individuals identified as having 95.00–99.99% probability of assignment to the kakī cluster by any of the analyses conducted here, all individuals had at least one node I or I/J individual in their recorded ancestry. Individual DNA897 had the most recent hybrid ancestry, with an I/J ancestor three generations deep. Individual DNA1252 was most frequently identified as deviating from 99.99% kaki assignment, and had the lowest probability of assignment to the kakī cluster of all kakī individuals with assignment probabilities in the 95.00–99.99% range (Table 4.4, Table 4.6). While individual DNA1252 did not have more frequent or more recent recorded hybrid ancestry than other kakī here, there is no recorded ancestry for the paternal lineage. The mothers of individuals DNA1252 and DNA397, and full sibling of individual DNA1659 (all falling in the 95.00–99.99% kakī membership range from at least one analysis method/SNP set) were included in sequencing and analyses, and were not identified as having any hybrid ancestry despite DNA397 having a node I/J ancestor four generations deep, suggesting a paternal lineage origin for the relatively higher proportion of admixture identified in DNA1252. Individual DNA1694 has a relatively deep pedigree among kaki, spanning seven generations. This individual also has the most frequent incidence of recorded hybrid ancestry among these individuals, with a node I/J ancestor four generations deep, three further I/J ancestors six generations deep, and a node I ancestor seven generations deep. The two hybrid individuals that were consistently assigned to the kakī cluster were siblings DNA777 and DNA779, with a node I/J mother, and an I/J ancestor three generations deep.

4.5 Discussion

With the advent of HTS techniques, conservation management of threatened species can now incorporate genomic tools to explore threats to species survival, such as anthropogenic hybridisation, with improved accuracy (Allendorf et al., 2010; Avise, 2010; Ekblom & Galindo, 2011; Gompert, 2012; Primmer, 2009). As discussed by Allendorf et al. (2001), hybridisation of a threatened species with a more common species is likely to result in the formation of a hybrid swarm, and potential extinction of the threatened species. Studies to date indicate that while using a small number of genetic markers can provide important baseline data for conservation management (e.g., genetic sexing, Steeves et al., 2010; brood parasitism, (Overbeek et al., In Review, 2017), such marker sets have proved less robust than genomic markers for analysis of parentage (Tokarska et al., 2009), relatedness (Galla et al., 2020), intraspecific population structure (McCartney-Melstad et al., 2018), and introgression (Parejo et al., 2018). Thus, re-examining the extent of introgression between critically endangered kakī and non-threatened congeneric poaka is essential to ascertain the efficacy of conservation management aimed at minimising gene flow following anthropogenic hybridisation between these birds.

4.5.1 Impacts of hybridisation in stilts

Hybridisation has occurred between kakī and poaka throughout the period of kakī decline through the 1900s (Pierce, 1984a; Steeves et al., 2010). All 130 kakī—representing 63% of the contemporary adult population at the time of sampling—were assigned as kakī with > 95% probabilities. These data confirm an absence of gene flow from poaka into kakī despite hybridisation, and are concordant with previous analysis of a small number of microsatellite markers (Steeves et al., 2010).

These results come in marked contrast to those of a study by Wells et al. (2019) which detected widespread introgression resulting from hybridisation between the endangered koloa maoli (Hawaiian duck, *Anas wyvilliana*) and the invasive mallard (*A. platyrhynchos*). Koloa persistence is directly threatened by hybridisation with mallards, and so it was important to determine the extent of introgression between species across the koloa range to inform conservation actions. Using 3,308 ddRADseq loci, sequencing across islands revealed that populations on all except one island (Kaua'i) represent a hybrid swarm. With this genomic information, conservation can be prioritised on Kaua'i where non-admixed koloa still persist, and removal of feral mallards should be undertaken to prevent koloa extinction through hybridisation. In comparison, no such management is required to support kakī conservation.

The decline of kakī throughout the 1900s and subsequent small population size has had minimal impact on the retained genetic diversity of the species, as assessed with mitochondrial data (Chapter 3). Previous genetic studies show kakī to have moderate levels of diversity compared with other threatened New Zealand birds (Steeves et al., 2010; see comparison in Forsdick et al., 2017), yet genomic diversity is much lower than non-threatened pied stilts in terms of numbers of polymorphic sites and private alleles, and mean nucleotide diversity (Table 4.3). This corroborates the estimates from whole-genome sequencing data in Chapter Two indicating kakī diversity was reduced compared with Australian pied stilts. This reduced diversity among kakī has also impacted the diversity present in hybrids, with hybrids genotyped here displaying levels of diversity intermediate to the two species. No node J kakī were identified as cryptic hybrids, with all individuals assigned to the kakī cluster with probabilities above the predefined threshold of 95%. Among the sampled individuals, 29 kakī had at least one hybrid ancestor three generations in

the past, and these ancestors were all dark hybrids (node I or darker). No kakī individuals included in this study had recorded hybrid ancestors with plumage lighter than node I. The results here corroborate the finding by Steeves et al. (2010) that dark hybrids have a lower proportion of introgressed material from poaka, and so we can assume that these individuals are contributing a very limited proportion of introgressed material to any offspring. With every generation following hybridisation, the proportion of introgressed genes is increasingly likely to be replaced by kakī genetic material. Indeed, two individuals identified as hybrids (DNA777 and DNA779) with a node I/J mother were assigned as kakī with 100% probability across all analyses. This may indicate that even a small number of generations of backcrossing with kakī can rapidly overwhelm introgression from poaka. However, the nature of GBS as a reduced-representation approach means that despite this large increase in data compared with the eight microsatellite loci used previously, this still only represents < 1% of the 1.1 Gb kakī genome (Chapter Two). Despite no signal of introgression detected in these two individuals, the current kakī recovery strategy excludes these individuals from management (i.e., inclusion in the captive breeding programme) based on the previous genetic results (Steeves et al., 2010). The unexpected finding of no introgression from poaka to kakī likely results from a combination of factors. First, the management strategy enacted by the DOC's Kakī Recovery Programme to maintain kakī genetic integrity has successfully limited hybridisation. This management has been responsive to the results of genetic analysis leading to exclusion of all non-node J individuals from management (Steeves et al., 2010; DOC, pers. comm.). Intensive population monitoring of breeding pairs and assessment of putative hybrids using the microsatellite panel has enabled practitioners to break up mixed pairs and exclude hybrids from the breeding programme (Maloney & Murray, 2001; DOC, pers. comm.). Ongoing kakī recovery

has produced a relatively balanced sex ratio in the wild, and combined with the strong positive assortative mating of kakī has minimised the likelihood of kakī breeding with non-kakī (Steeves et al., 2010). Moderate outbreeding depression and stochastic processes have also contributed to reduce the reproductive success of hybrids (Steeves et al., 2010). From a genetic perspective, any introgression resulting from hybridisation and backcrossing in the past is unlikely to be maintained during subsequent generations of backcrossing with kakī.

4.5.2 Implications for kakī conservation

The results presented here provide evidence that active conservation management designed to minimise hybridisation can be effective in maintaining species integrity, and support the ongoing management strategy of poaka and hybrid exclusion from the captive management programme based on the results of genetic analysis (Steeves et al., 2010). The goals of current kakī management to maintain the genetic integrity of the species are appropriate due to the strong differentiation between species observed here and the previous observation of moderate levels of genetic diversity among kakī compared with other threatened endemics (Steeves et al., 2010). Despite the increased resolution of genomic data, when individuals with anomalous plumage are observed in the captive breeding and rearing facility, microsatellite genotyping remains the most cost- and time-efficient, low-complexity method for confirming species status (Table 4.7; Overbeek et al., 2017; Overbeek et al., In Review).

Table 4.7: Comparison of the costs and benefits associated with a genetic approach (i.e., microsatellite panel) for assessing hybridisation in kakī with a genomics approach (i.e., genotyping-by-sequencing and SNP discovery). With both platforms already established for kakī, future cost per sample is the primary deciding factor. SNPs = single-nucleotide polymorphisms. All cost estimates are in New Zealand dollars (NZD).

	Microsatellite panel	**Genotyping-by-sequencing**
Platform development cost	< 10,000 NZD. Includes development and screening of ~20 polymorphic loci, and genotyping of up to 94 individuals (based on a known decrease in costs since the estimate of Galla et al. (2016)).	~5,500 NZD. Includes testing of restriction enzymes and adapter barcoding, plus DNA extraction and sequencing of up to 94 samples, with expected output of 30–60K SNPs discovered via a de novo approach.
Platform development time	3–4 months	3–4 months
Cost per sample for additional samples	15 NZD. Includes DNA extraction and quantification, and microsatellite genotyping at eight optimised loci. Excludes associated person-hours.	55 NZD. Includes DNA extraction and quantification (but not the associated person-hours), and GBS.
Lab time for additional samples	< 1 week	1–3 months
Analysis time	< 1 week	4 weeks
Analysis requirements	Access to a standard desktop computer. Access to allele-calling software (e.g., GeneMarker v2.2). Access to and experience with population clustering tools optional.	Access to a high-capacity computing system (e.g., NeSI) or capable computing cluster, access to and experience with a variety of bioinformatic tools.
Additional benefits	Previous uncertainty regarding how representative of genome-wide patterns of introgression these data are is now resolved for kakī by comparison with GBS data, indicating the eight loci are sufficient for robust identification of non-kakī individuals. All wet-lab work and analysis for additional samples can be run in-house.	Increased confidence in detection of introgression with large genome-wide SNP set compared with microsatellites. No ascertainment bias associated with marker generation. Once these data are generated, they can be implemented for a variety of downstream uses, including genetic sexing. Potential for these data to have additional future applications with genomic analysis developments.
Limitations	Caveats associated with how the microsatellite is generated, and what it was developed for (e.g., active avoidance of regions under selection, ascertainment bias from selecting the most heterozygous markers). This limits the type of data produced, and thus the types of analyses that can be performed. Moderate level of human error associated with manual allele-calling, but mitigated with experience.	Additional sampling requires the SNP discovery pipeline to be run with all previous samples included every time, and so analysis does not become more efficient over time. Potential for batch effects between multiple sequencing rounds (Leigh et al., 2018). Analyses limited by software and models available in a developing field.

Under optimal circumstances, kakī recovery will continue, leading to increased numbers of kakī within Te Manahuna in the short-term, and the potential for natural expansion beyond the basin in the long-term. In the short-term recovery scenario, active management of the species may be scaled back (DOC, pers. comm.). This may see a reduction in the number of adult kakī maintained in captivity for breeding, although management of wild nests, including egg-pulling and captive rearing are likely to be continued to maximise population growth while kakī remain critically endangered. As kakī are one of the few threatened New Zealand birds to have maintained a population on the mainland despite the presence of invasive predators, and are capable of travelling long distances, active translocations are unlikely to be necessary to support natural expansion beyond Te Manahuna. In addition, it is unlikely that active management of any such expansion would be feasible (DOC, pers. comm.). As such, management to minimise the likelihood of hybridisation within Te Manahuna will continue, but with the wide distribution of poaka across the country, future expansion into areas with high poaka densities may result in the increased prevalence of hybridisation that could once again compromise genetic integrity. Therefore, maintaining the integrity of the source population within Te Manahuna should be a high priority for conservation.

4.5.3 Impacts of hybridisation on poaka

The identification of poaka with pied stilt assignment probabilities < 95% may be a result of initial small population size on arrival to New Zealand, and subsequent hybridisation with kakī prior to species decline. Kakī only occur as vagrants in the North Island of New Zealand, observed in very low numbers since at least the 1950s (Pierce, 1984b). Given the limited recent contact between kakī and poaka in the North Island, I expected the poaka samples sourced from the North Island to

produce assignment probabilities similar to those of the Australian pied stilts. However, both individuals sourced from Auckland Zoo had assignment probabilities to the pied stilt cluster < 95%. The only North Island individual with population assignment probability comparable to those of the Australian pied stilts was the individual from Hawke's Bay (Poaka1). This suggests that hybridisation early in the establishment of poaka may have resulted in introgression of kakī genetic material into an initially small poaka population that was not frequently supplemented by any substantial number of new immigrants, with introgressed material maintained in the expanding population despite subsequent backcrossing. Kakī introgression into poaka is supported by the observation of node A poaka having tarsal lengths outside the range observed among Australian pied stilts with no history of hybridisation, and poaka presenting a greater proportion of black plumage than is typical for Australian pied stilts (Pierce, 1984a).

4.5.4 Potential limitations associated with sample size skew

A potential limitation of this study arises from the unequal sample sizes obtained for kakī, Australian pied stilts and poaka, and kakī-poaka hybrids, resulting from the limited number of captive poaka and Australian pied stilts available, and the degradation of some DNA extracts among those samples included in the 2010 study. Furthermore, Known difficulties in catching birds in the wild precluded additional sampling of poaka and hybrids, and so only samples previously collected through the Kakī Recovery Programme were available. Nevertheless, these samples are sufficient to detect introgression in kakī. DNA degradation during long-term storage resulted in exclusion of many of the hybrid and poaka samples used in previous genotyping (Steeves et al., 2010) from GBS for this study, although comparison of gDNA concentrations with sequencing outputs indicated that even samples with very

low concentrations (at least as low as 13.6 ng/µL) may have produced sufficient sequences for inclusion. Genetic studies typically require a minimum of 25 individuals to produce accurate estimates of population genetic diversity (Hale et al., 2012). However, as genomic studies capture a greater proportion of population-level diversity within the much larger marker sets, they require fewer individuals to produce representative diversity estimates. Simulation studies to determine appropriate sample sizes when using genomic markers indicate that as few as four individuals may be sufficient for accurately estimating population differentiation with sets of at least 3,000 markers (Willing et al., 2012), while studies in plants (Nazareno et al., 2017) and insects (Qu et al., 2019) indicate that as few as two individuals can provide accurate estimates of population genetic diversity and differentiation when using at least 1,000 markers. Kakī-specific estimates produced from kakī whole-genome resequencing and GBS data indicate that sampling of eight individuals is sufficient to resolve population heterozygosity for this critically endangered species (Collier-Robinson, 2019). Comparison with data from Buller's albatross (*Thalassarche bulleri*), a species with approximately 33,000 breeding pairs across two populations, suggests that 10–12 individuals are required to accurately sample larger populations with inherently greater diversity such as poaka (Collier-Robinson, 2019). Indeed, comparisons of the numbers of private alleles present among kakī, Australian pied stilts and poaka, and kakī-poaka hybrids indicate that the limited number of pied stilts sampled may not have captured the true level of diversity of pied stilts. As hybrids should represent admixtures of kakī and poaka genetic material, all alleles present in hybrids would be assumed to be found within either the kakī or poaka, thus we would expect numbers of private alleles among hybrids to be close to zero. This was not the case, indicating sampling has not captured complete diversity across poaka or Australian pied stilts, although there were fewer private

alleles among hybrids than among pied stilts or kakī (except in SNP Set 3). However, as I am interested in characterising differentiation between kakī and pied stilts rather than diversity within these species, I assert that the relatively low number of pied stilts included here is nevertheless sufficient. Similar sampling sizes to these presented here have been used in other genomic studies of introgression (e.g., Oswald et al., 2019) with no obvious negative impacts. Further, as one of the primary goals of this study was to identify any signal of introgression among modern kakī, the inclusion of a much greater number of kakī than non-kakī was appropriate. Investigation of the effects of unequal sample sizes on measures of population differentiation by Willing et al. (2012) suggests that the F_{ST} calculation of Weir and Cockerham (1984) may result in underestimates of differentiation, but remains relatively robust. The exclusion of close relatives in SNP Set 3 was primarily used to confirm the appropriate number of population clusters, and to reduce the effects of inbreeding (although as a threatened species with a limited number of founders, inbreeding is unavoidable among kakī). Using this set also reduced the sample size skew, with more similar numbers of kakī and non-kakī. Thus the F_{ST} value of 0.423 produced from this SNP set is likely the most accurate estimate of differentiation between kakī and pied stilts produced here. This value represents a substantial level of differentiation, confirming that the two species are genetically distinct and that introgression can be accurately detected.

4.5.5 The challenges of defining species assignment thresholds in the genomics era

Determining what constitutes substantial introgression, and at what level introgression is considered to be of concern for threatened species management is not straightforward, requiring a balance between the possibility of misassignments

attributed to low efficiency (where true hybrids are misassigned to the parental species) or low accuracy (where individuals of the parental species are misassigned as hybrids) of the marker set (as reviewed by McFarlane & Pemberton, 2019). Simulation studies can be implemented to determine appropriate thresholds (e.g., using the program HYBRIDLAB (Nielsen et al., 2006), as in van Heugten et al., 2017), but such simulations are best designed for detection of recent hybridisation and backcrossing (up to F2 hybrids or backcrosses). Although most of the individuals included here have at least one individual with hybrid ancestry (dark hybrids, plumage nodes H–I/J), these are typically recorded as ≥ 4 generations deep in the pedigree. I argue that it is more useful for the Kakī Recovery Programme to identify individuals with substantial evidence of introgression that has been maintained despite multiple generations of backcrossing, or results from recent undetected hybridisation, as outlined in the aims of this chapter. Capturing the pattern associated with this deep hybridisation through simulation would require a large number of replicates with no way of verifying the accuracy, with the potential for exponential increases in the error around such simulations with every additional generation, and as such, I argue that conducting such simulations would be inappropriate here. Further, because of the rapid advances in the field of genomics, when programs have not been maintained in line with these developments (i.e., with very large SNP sets), the results of these programs should be considered with scepticism until they have been critically evaluated. HYBRIDLAB, released in 2006, was developed for small marker sets (Nielsen et al., 2006), and has not been maintained or critically evaluated for large genomic marker sets. As such, the uncertainty associated with implementing a simulation approach to define a species assignment threshold makes this process unreliable, and so the 95% threshold defined here is most appropriate until the reliability of this approach can be

ascertained with genomic data. The use of multiple methods of analysis, and in particular, the use of bootstrapping to produce standard errors to examine variation around assignments as implemented through ADMIXTURE improves confidence in these assignments, in lieu of simulation-derived thresholds.

4.5.6 Use of a reference genome assists variant discovery for introgression analysis

Comparison of the *de novo* approach to SNP discovery used as part of the AgResearch GBS quality control pipeline with that of the reference-guided approach implemented to produce the final SNP sets identified the reference-guided approach to be superior in terms of total marker output, with over four times more variants discovered prior to filtering. Similar results were obtained when comparing *de novo* and reference-guided approaches to SNP discovery from ddRADseq for Mexican grey wolves (*Canis lupus baileyi*) and bighorn sheep (*Ovis canadensis*; Andrews et al., 2018). While a *de novo* approach is valid in the absence of a reference genome, a reference-guided approach is recommended where possible, providing a greater number of high-quality SNPs that can then be stringently filtered to reduce potential biases. As demonstrated by (Galla et al., 2019), reference genomes of closely related species can assist variant discovery for estimating conservation-relevant metrics.

Selection of the most appropriate reference genome for sequence mapping and variant discovery maximised the power of the data produced across both species included here. Higher mapping success from all samples to the kakī genome further confirmed the higher quality of the kakī genome as compared to that of the Australian pied stilt (Chapter Two), and further validated the conservation utility of these genomes. Not only has the use of a species-specific reference genome

increased the total variant output, it has also enabled comparisons between reference-guided SNP discovery approaches, and will allow robust and comparable SNP discovery when incorporating additional sequencing in the future.

4.5.7 Implementation of multiple SNP filtering strategies

Putative sex-linked SNPs were excluded primarily due to the challenges of including sex chromosomes in analyses of introgression relating to the haploid nature of the W chromosome in avian females. Sex information for sampled individuals was incomplete, potentially increasing the amount of missing data for these individuals if sex-linked SNPs were included. Furthermore, there may be incongruities associated with the identification of sex chromosomes in the kakī reference genome based on comparison with the chicken genome assembly. Comparisons via BLAST searches indicated that the putative avian sex chromosomes Z and W had regions of similarity, indicating potential misassembly of these chromosomes. While no other contigs were definitively identified as originating from sex chromosomes, the chicken represents a relatively distantly-related species from kakī. Confirmation of correct assembly of the kakī sex chromosomes will require additional sequencing and reassembly beyond the scope of this thesis. Despite the possibility for inclusion of haploid and/or sex-linked regions with some (e.g., ADMIXTURE; Alexander et al., 2009) of the analysis methods used, exclusion of SNPs present in putative sex chromosomes was the most appropriate strategy given these limitations.

Comparison of three SNP filtering strategies identified minimal differences in the results of population clustering and individual assignment between SNP sets. This strong concordance between SNP sets is likely a result of the initial merging of variants across multiple variant discovery pipelines producing a common set of high quality variants from which the three filtered SNP sets were derived. The exploratory

analysis of SNP filtering strategies and stringencies, and the preliminary tests of downstream analyses produced concordant results, providing high confidence in the final filtering strategies implemented. Despite the stringent filtering strategy of SNP Set 2 producing the fewest SNPs, with 85.45% of sites excluded from SNP Set 1 due to levels of missing data > 10%, the 20,718 autosomal SNPs remaining were sufficient to robustly assess genome-wide introgression. While a larger marker set increases statistical power of analysis, particularly useful for detecting advanced backcrosses (Boecklen & Howard, 1997; McFarlane & Pemberton, 2019; Vähä & Primmer, 2006), the increased depth and quality resulting from the filtering strategy for SNP Set 2 provides greater confidence while maintaining statistical power.

The sampling regime and subsequent analysis methods implemented here meant that some commonly used SNP filtering strategies were unnecessary (e.g., filtering for HWE, more stringent minor allele frequency cut-offs). The population history of kakī and the inclusion of individuals from two different species mean that some loci would be expected to deviate from HWE. Although SNP sets were not specifically filtered to exclude those deviating from HWE, the stringent and related filtered SNP sets had relatively low numbers of SNPs deviating from HWE (4% and 2%, respectively, of the total number of SNPs deviating from HWE in SNP Set 1; Table 4.2), indicating that other filtering steps eliminated many of these SNPs. Preliminary filtering tests had included SNP thinning (reducing the set to one SNP per 100 bp) to reduce potential impacts of non-independence of co-located SNPs, with no impacts on downstream results.

4.5.8 Implementation of a multi-pipeline analysis method

The implementation of three distinct analysis methods (DAPC, ADMIXTURE, fastSTRUCTURE) to investigate the effects of hybridisation produced concordant

results for kakī, providing high confidence in the result of no introgression in kakī. Implementation of DAPC analysis was particularly useful in assessing population differentiation within and among the two species, independent of any population genetic assumptions. Careful optimisation of the retained number of principal components was implemented to avoid overfitting the data, which could result in overestimates of divergence between clusters. These results showed that the kakī and pied stilts were well-differentiated, with greater differentiation within pied stilts than kakī likely due to the inclusion of pied stilts from larger populations across a wide geographic range. These results suggest that along with morphological, behavioural and ecological divergence from Australian pied stilts (Pierce, 1984a), poaka have also begun to diverge genetically since arrival in New Zealand, perhaps in part due to introgressive hybridisation early in their arrival. Analyses with fastSTRUCTURE and ADMIXTURE produced individual assignment probabilities to the two species clusters that can be incorporated in management decisions aiming to maintain kakī genetic integrity. While the assumptions underlying both methods should be identical, bootstrapping to produce standard error enabled assessment of variation about the assignment probabilities produced with ADMIXTURE, finding similarly low variation across the SNP sets. Thus I contend that an exploratory DAPC combined with ADMIXTURE analysis is the most informative approach to assessing introgression in this system. Indeed, in ADMIXTURE analyses, pied stilts attained higher assignment probabilities to the pied stilt cluster than demonstrated from fastSTRUCTURE analyses. Thus, with species differentiation confirmed here, additional future analysis could implement an approach using ADMIXTURE without the need for validation via multiple methods.

4.5.9 Comparison of genetic and genomic approaches to introgression analysis

Reduced-representation sequencing approaches have proven to be efficient, robust, and cost-effective for variant discovery (Andrews et al., 2016; Davey et al., 2011; Elshire et al., 2011; Peterson et al., 2012). Here GBS was used as a cost-effective approach to population-level genomic sequencing of non-model species, producing species-discriminating SNP sets. Initial development of a GBS system is markedly less expensive than development of a microsatellite panel (~5500 NZD for GBS development using 94 samples in this study in 2018 compared with ~10,000 NZD development and testing of a microsatellite panel of approximately ten loci using 94 samples based on the estimate of Galla et al. (2016); Table 4.7). However, ongoing costs of the microsatellite panel per sample remain considerably lower than that of GBS (~15 NZD/sample for a microsatellite panel compared with ~50 NZD/sample for GBS; Table 4.7), and the time required from individual sampling to completion of analysis is substantially faster. There are also fewer barriers to analysing microsatellite data (e.g., microsatellite genotyping and analyses can be conducted on a standard desktop computing system compared with the requirement of a high-performance computing system with a wide range of bioinformatic tools for analysis of GBS data, Table 4.7). While the substantial increase in data produced and subsequent robustness of results makes a genomics approach desirable for conservation projects, the associated costs may limit uptake, especially when providing data for time-dependent decisions. Despite the increasing uptake of genomics approaches to answer questions pertinent to conservation management (Galla et al., 2016), the current greater costs and other transitional challenges (e.g., bioinformatic knowledge) will likely maintain the conservation genomics gap for some time yet. Declining costs associated with whole-genome resequencing may overtake

reduced-representation sequencing in the near future, and if concerns once again rise regarding hybridisation among these stilts, reassessment with a whole-genome resequencing approach may be beneficial. While captive breeding continues, a combined pedigree and genomic approach to pairing recommendations to minimise inbreeding is being implemented (T.E. Steeves, pers. comm.), based on the finding of the improved accuracy of this approach over a genetic approach by (Galla et al., 2020). Suspected hybrid individuals could be incorporated in whole-genome resequencing alongside those individuals of interest for captive breeding, and assessed to determine the utility of this approach for hybrid detection. By using a reference genome, sequence data produced from additional individuals should produce SNPs comparable with those produced here, and thus can be incorporated into analyses to confirm hybrid status with relative ease. However, any further genomic reassessment is of low priority while hybridisation is infrequent, and at this point in time, genomic resequencing is not yet cost-effective for such reassessment. I anticipate that such an approach would corroborate the findings here, but that additional benefits could be provided as the improved resolution of the data may allow assessment of fitness impacts of introgression. Implementation of genomics approaches for investigating hybridisation in other threatened species will require careful evaluation of the associated costs and benefits. Conservation genomics researchers would do well to ensure that practitioners are fully apprised of the potential costs, benefits, and limitations associated with these approaches, to ensure the research conducted is aligned with the best interests of species recovery.

4.5.10 Streamlined bioinformatics pipeline for introgression analysis from GBS data

An additional objective of this study was to produce a streamlined bioinformatics pipeline for analysis of introgression from GBS data. The availability of such pipelines can assist in reducing the conservation genomics gap (Holderegger et al., 2019), and can be implemented for additional GBS data generated for these stilts. After investigating a variety of methods for sequence mapping, SNP discovery, and analysis methods, I have developed a streamlined workflow to process and analyse GBS data to investigate the impacts of interspecific hybridisation in kakī and other threatened birds (Appendix F). No such direct workflow for introgression analysis of birds existed when this project was initiated, and this workflow will support any future investigation of introgression in kakī, with the potential to adapt the workflow as required for other taxa. Development of this workflow drew on similar comparisons of variant discovery pipelines for a model plant (Torkamaneh et al., 2016) and the SNP filtering methods for analysis of differing reduced-representation sequencing for non-model animals (Tasmanian devil, *Sarcophilus harrisii*; pink-footed goose, *Anser brachyrhynchus*; Wright et al., 2019). The resulting workflow reproduced in Appendix F (see Figure 4.11 for an overview) is annotated to assist users entering the conservation genomics space (although experience with command-line processes is recommended), with the aim of reducing the barriers to entry for conservation geneticists making this transition. As with any such workflow, users should consult the documentation associated with the various tools included here to ensure the processes described are appropriate for their data. Users should also note that there are a range of alternative tools available that could be substituted at various stages of this workflow.

The final pipeline processes raw GBS data using existing well-documented bioinformatic tools, and guides the user through all the steps from receipt of sequence data through to visualisation of the results of population assignment analysis. Raw sequences are processed with Sabre v1.0 and Cutadapt v1.17, mapped against a reference genome with SAMtools and BWA, and SNP discovery is then performed with the reference-guided Stacks v2.2 pipeline. Stacks is becoming the preferred option for SNP discovery from reduced-representation data for non-model organisms (Wright et al., 2019) and proved robust in pipeline comparisons presented here (Figure 4.3). Although I implemented a multi-pipeline approach to SNP discovery that allowed pipeline comparison and improved confidence in the resulting SNPs, the use of a single variant discovery pipeline should produce sufficient numbers of SNPs that can then be filtered to ensure robustness in downstream analyses, producing concordant results to that of a multi-pipeline approach. Furthermore, such an approach will allow comparisons with other such studies using the well-documented, well-supported Stacks pipeline.

With a variety of SNP filtering strategies possible, a strategy based on that used to produce the stringently-filtered SNP Set 2 is presented here to generate a robust marker set for downstream analyses. Filtering strategies must be optimised for the data and intended downstream analyses, and guidance is given in suggesting metrics to examine as part of this optimisation process (Appendix F). Such considerations may include appropriate levels of missing data, minimum allele frequencies, and deviations from HWE (see O'Leary et al., 2018; Wright et al., 2019 for further discussion, along with additional filtering options described in the VCFtools manual, http://vcftools.sourceforge.net/man_latest.html, and in a SNP filtering tutorial at http://ddocent.com/filtering/, Puritz et al., 2014). Assignment probabilities were highly correlated between ADMIXTURE and fastSTRUCTURE

analyses, and so the two programs are largely interchangeable. In this workflow I implement an array method using a high-performance computing system to minimise ADMIXTURE run-time such that a panel of up to ~140K SNPs can be analysed with 100× replication in less than two hours real-time. ADMIXTURE can also be implemented on lower-capacity systems, although associated run-times will be extended when multiple iterations of the analysis are run in series rather than in parallel. Finally, an Rmarkdown workflow incorporating *pophelper* is used for data wrangling and visualisation of the results of multiple ADMIXTURE runs. Interpretation of results relies on user knowledge of the relevant evolutionary patterns and processes, and the implementation of appropriate predetermined thresholds are recommended to support robust population assignments and identification of introgression.

This workflow was produced for efficient analysis of any future GBS data generated from putative cryptic hybrid stilts if required. This workflow can also be modified for whole-genome resequencing data, to enable direct comparison between these sequencing approaches on results of introgression among stilts in the future as such sequencing becomes increasingly cost-effective. While this workflow is designed based on utility for this study, it has applicability for other threatened species impacted by hybridisation, or for assessment of intraspecific population structuring.

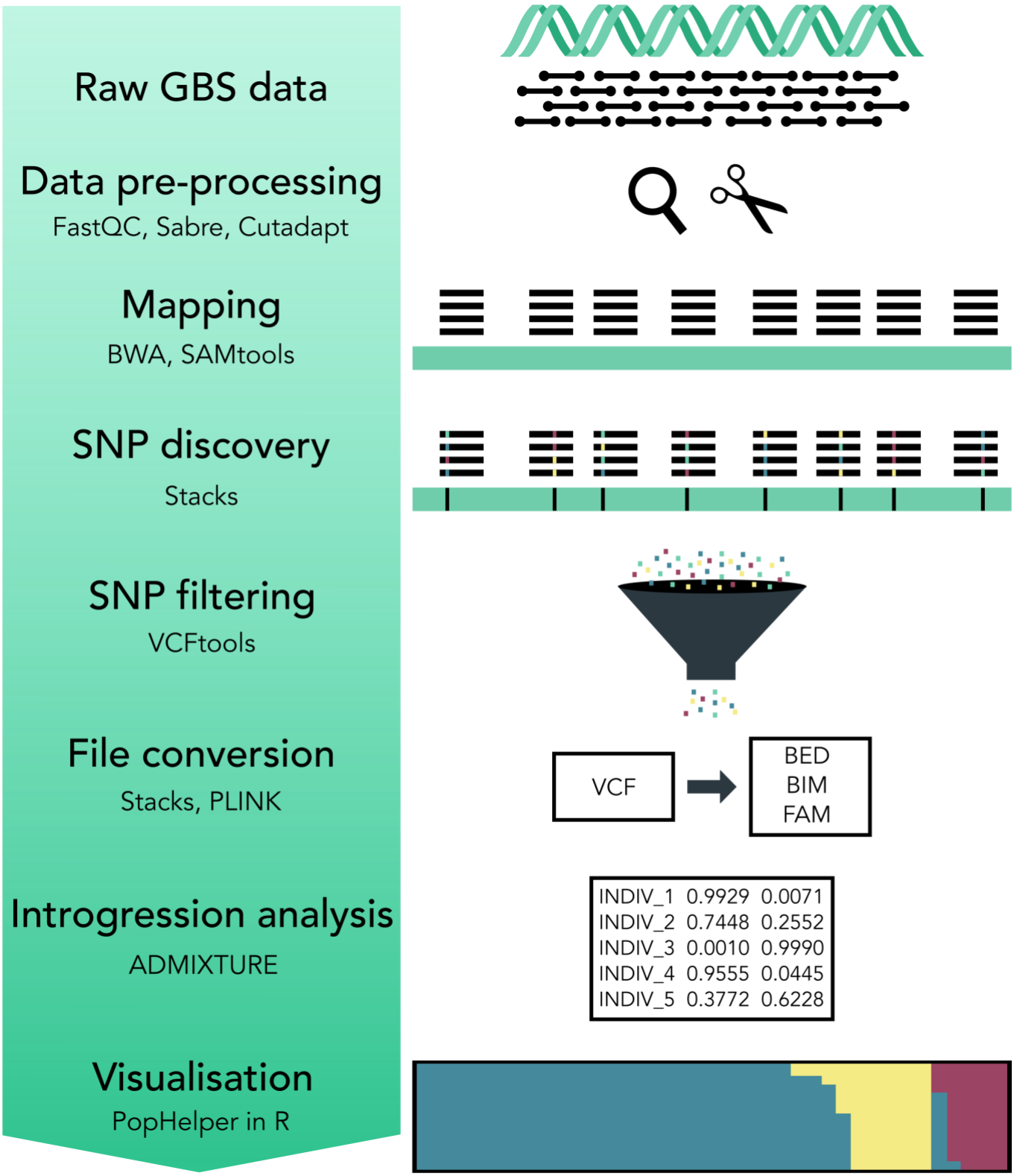


Figure 4.11: A visual representation of the streamlined workflow for analysis of introgression using genotyping-by-sequencing data and a reference genome for mapping and SNP discovery with relevant tools listed at each stage, from receipt of raw sequence data through to visualisation of results. Scripts associated with thisworkflows are detailed in my GitHub 'Himantopus' repository at https://github.com/natforsdick/Himantopus.

4.5.11 Conclusions

This study has relevance to other bird species threatened by hybridisation, the number of which is likely to rise due to anthropogenic effects including climate change, habitat modification, and increased international trade altering species ranges (Chunco, 2014). Indeed, this study is only the second to use a genomic approach to investigate hybridisation in a threatened New Zealand bird, along with a study that identified mitochondrial but no nuclear introgression from tarāpunga (red-billed gulls, *Larus novaehollandiae scopulinus*) across the range of the endangered tarāpuka (black-billed gulls, *L. bulleri*; Mischler et al., 2018). However, a direct comparison with this study is not appropriate, as the focus therein was to investigate population structure among tarāpuka, rather than hybridisation between the two species, and so tarāpunga were not deliberately included for GBS except for two individuals erroneously included due to species misidentification. After excluding these individuals from analysis of GBS data after preliminary assessment revealed the misidentification, hybridisation was assessed using mitochondrial markers in combination with phenotypic data. Due to the identification of these misidentified tarāpunga as clear outliers in preliminary clustering analysis from GBS, Mischler et al. concluded that hybridisation between these gulls had resulted in mitochondrial but no nuclear introgression. A similar approach using reduced-representation sequencing is being used to investigate the genetic integrity of pārera (New Zealand grey duck, *Anas superciliosa*), where preliminary data indicate that extensive hybridisation with introduced mallards (*A. platyrhynchos*) has resulted in extensive introgression, with genetic integrity only appearing to have been maintained in isolated populations on the West Coast of the South Island (Brown et al., 2020). The large marker sets assessed here produced confidence in the results confirming no introgression from poaka to kakī. Studies comparing the utility of microsatellites

with genomic sequencing for generating estimates of population genetic diversity and differentiation indicate that genomics approaches outperform the limited marker sets typically used in genetic studies (Hauser et al., 2011; Hohenlohe et al., 2013; Santure et al., 2010; Weinman et al., 2015). While results are similar between the two approaches here, this may not be the case for other species, particularly when hybridisation may be widespread. Should hybridisation become a significant threat to kakī recovery in the future, then a comparison of GBS with whole-genome resequencing data for introgression analysis may be useful to explore the potential greater accuracy attainable from whole-genome resequencing data. Whole-genome resequencing has lower stringencies on input DNA concentrations, and so gDNA from individuals included in previous genetic analysis that had been excluded from GBS due to DNA degradation could be incorporated in the future. For now, this research is directly informing conservation management of a critically endangered New Zealand endemic bird.

4.6 Data availability

Kakī are a taonga species, and as such, genomic data derived from kakī are also recognised as taonga in their own right. Due to its tapu nature, the data presented here are hosted on a password-protected database at www.ucconsert.org/data/, and will be made available at the discretion of the kaitiaki of the iwi and hapū associated with kakī. These data include raw genotyping-by-sequencing, along with the VCF that includes the common set of unfiltered variants derived from multi-pipeline SNP discovery. Scripts associated with the bioinformatic workflows included in this chapter are detailed in my GitHub ‘Himantopus’ repository at https://github.com/natforsdick/Himantopus.

Chapter 5: General discussion and future directions

5.1 Synopsis

In this thesis, I produced a range of genomic resources to support conservation management of the critically endangered Aotearoa New Zealand endemic kakī (*Himantopus novaezelandiae)*. Included among these resources are the assembled genomes for kakī and the non-threatened congeneric Australian pied stilt (*H. h. himantopus*), representing the first reference genomes available for the genus Himantopus (Chapter Two). The conservation utility of these genomes was demonstrated in a Proof-of-Concept study that found that for birds, genomes up to the confamilial level can be used as references for SNP discovery to generate accurate estimates of metrics relevant to conservation (Galla et al., 2019, Appendix A). With more than 180 avian genomes available for use as proxy reference genomes for threatened birds (see Chapter Two), these findings reduce the conservation genomics gap, enabling more rapid uptake of genomic tools for threatened species management. The kakī genome has also been used in a comparison of pedigree-, genetic-, and genomic-based approaches for estimating relatedness (Galla et al., 2020), with direct application for conservation management, where captive pairing recommendations will be made using pedigree data in tandem with genomic resequencing data moving forwards. In addition, these genomes were used as references for SNP discovery to investigate the impacts of interspecific hybridisation between kakī and poaka in Chapter Four.

In Chapter Three, I produced a complete mitogenome for kakī (but see Appendix D), and with mitogenome resequencing data from contemporary and historic stilt

samples, indicated that kakī conservation management has largely maintained mitochondrial diversity despite kakī decline during the 1900s. Furthermore, the divergence date estimated in Chapter Three combined with the strong differentiation observed between kakī and poaka in Chapter Four mean that conservation management has maintained the genetic integrity of kakī. Further, through close collaboration with the DOC's Kakī Recovery Programme, I assessed the veracity of the recorded catalogue data of stilts in New Zealand natural history collections with morphological (plumage-based) and mitochondrial population identification (see Appendix C). This identified a number of specimens with incorrect catalogue species identifications, highlighting the need for any future research using stilt samples obtained from natural history collections in New Zealand and internationally to include similar assessments.

Finally, I used genotyping-by-sequencing (GBS) for kakī, Australian pied stilts, poaka, and kakī-poaka hybrids with the assembled reference genomes for SNP discovery and analysis of introgression (Chapter Four). These data further support the strong differentiation between kakī and poaka as observed from mitochondrial data in Chapter Three and previous genetic analysis using microsatellite markers (Steeves et al., 2010), and identified no introgression into kakī, concordant with the previous genetic analysis. To support the uptake of genomics to elucidate the impacts of hybridisation in other threatened species, and to streamline the process if additional samples are generated for these stilts in the future, I developed a bioinformatic workflow that guides the user from receipt of GBS data through to visualisation of the results of admixture analysis (Appendix F).

As detailed below, each of the genomic resources produced here (i.e., reference genomes, mitochondrial genomes, and the genome-wide SNP set) have directly contributed to conservation decision-making for kakī, by confirming that current

conservation is achieving the goal of maintaining species integrity, as well as contributing resources (i.e., the kakī reference genome) to inform other aspects of management such as captive pairing decision-making, and demonstrating the utility of a genomics approach moving forward. Further, in addition to providing invaluable information regarding the evolution of this endemic New Zealand species and its close congener, the approaches used to develop these tools (i.e., short-read whole-genome sequencing and GBS) demonstrated that relatively cost-effective sequencing techniques are sufficient to generate high-quality genomic resources (i.e., reference genomes, mitogenomes, and genomic markers) that meet the needs of conservation management. In combination, this thesis provides valuable tools to understand the evolution and impacts of interspecific hybridisation in this critically endangered taonga species, and confirms that the actions implemented by the DOC's Kakī Recovery Programme aiming to maintain genetic integrity of this critically endangered endemic species have been successful.

5.2 Kakī conservation in the genomics era

The results presented in this thesis provide assurance to the Kakī Recovery Programme that the management decisions made in lieu of genetic data, and later refined using genetic data as it became available, have achieved the management objective of maintaining the genetic integrity of this critically endangered wading bird. Genetic diversity (as assessed using temporal comparisons from mitogenome data, Chapter 3) has been largely maintained despite the species decline during the 1900s, with only a single mitochondrial haplotype known to have been lost. Kakī and poaka were genetically differentiated prior to the initiation of conservation management of kakī (Chapter 3), and the genetic integrity of this taonga species has

been maintained (Chapter 4), confirming the results of previous genetic analysis (Steeves et al., 2010). The information is assisting conservation decision-making by the DOC's Kakī Recovery Programme to determine appropriate management strategies for current and future management. In addition to the results presented here, observational data indicates that while there are increasing numbers of individuals with intermediate plumage nodes observed in Te Manahuna in recent years, few of these appear to be the result of inappropriate pairings (DOC, pers. comm.), and thus hybridisation is currently of low concern to the Kakī Recovery Programme. Continuing with the established strategies to maintain kakī genetic integrity should be sufficient for continued maintenance of genetic integrity in the short term. These strategies include breaking up inappropriate pairs in the wild, using the species-discriminating microsatellite panel to assess any individuals with anomalous plumage in the captive rearing facility, and removal of such individuals from the facility when non-kakī status is confirmed (DOC, pers. comm.). The identification of two plumage node I/J hybrid individuals assigned to the kakī cluster based on genomic analyses raises the question of the potential value of such dark hybrids for kakī recovery, and this is on the agenda for discussion with practitioners at the next Kakī Scientific Advisory Group meeting (scheduled for May 2020). I anticipate that such individuals will continue to be excluded from active management until further evidence can resolve this.

These findings have wider long-term implications for kakī recovery. It is not yet possible to disentangle the impacts of outbreeding depression that may have limited the extent of poaka introgression into kakī (Steeves et al., 2010) from the effects of conservation management aimed at maintaining genetic integrity of the species. As kakī recovery continues, we anticipate two stages of recovery over different temporal scales. In the short-term, the number of wild kakī present in Te Manahuna will

increase, and in turn, the numbers of wild pairs will increase. These increases may result in changes in the active management of the species in Te Manahuna, such as a reduction in captive breeding, and an increased focus on captive brooding and rearing of chicks produced from wild pairs (DOC, pers. comm.). Monitoring of the 27 wild breeding pairs spread throughout Te Manahuna is already challenging, as demonstrated by a wild pair that was undetected at the Tasman River delta site until after having fledged four chicks during the 2017/2018 breeding season (DOC, pers. comm.), the increasing number of wild pairs may result in some inappropriate pairs being undetected. However, I anticipate that the increased numbers of kakī will reduce the likelihood of hybridisation in Te Manahuna, due to the strong positive assortative mating exhibited by kakī (Steeves et al., 2010).

With species recovery we also anticipate natural range expansion beyond Te Manahuna in the long-term, and the likelihood of hybridisation becomes difficult to predict. As kakī expanding into new habitats will be outnumbered by poaka initially, limited breeding opportunities may increase the risks of hybridisation. Although the impacts of outbreeding depression contribute to minimising subsequent introgression (Steeves et al., 2010), there is a risk that hybridisation may become more widespread, with negative impacts on kakī due to the associated reproductive effort directed away from species recovery. In this case, outbreeding depression is most likely the result of intrinsic genomic incompatibilities (Steeves et al., 2010). Thus, differences in extrinsic selection pressures between Te Manahuna and potential future expansion sites will likely have negligible impacts on the intrinsic fitness impacts on hybrid offspring. However, active management is unlikely to be enacted beyond Te Manahuna due to operational challenges (DOC, pers. comm.). Therefore, continued prioritisation of the genetic integrity of kakī in Te Manahuna remains important to mitigate any future hybridisation beyond the basin.

5.3 Consideration of appropriate tools to support kakī conservation

Alongside the data produced here supporting conservation management for kakī, recommendations regarding the use of genetic and genomic tools are also beneficial. The genomes assembled here have already had wide utility (but see Future Directions), and for estimates of relatedness, a genomic approach has proven to be more robust than the genetic approach used previously, with relatedness metrics more accurately estimated with genomic SNPs than with genetic microsatellite markers (Galla et al., 2020; Hagen et al., 2011). However, the findings presented here indicate that a genetic approach (as in Steeves et al., 2010) is sufficient to confirm the species status of putative hybrid individuals as part of routine kakī management (also see Overbeek et al., 2017; Overbeek et al., In Review). Moreover, as discussed in Chapter Four, this approach remains more cost- and time-efficient than implementing GBS, due to reduced cost-efficiency of GBS for small sample sizes. As only small numbers of individuals (typically < 5) are identified as putative hybrids each breeding season (DOC, pers. comm.), employing GBS to assess species status of these individuals would be inefficient. In addition, the time required for sequencing and subsequent analysis would slow down the decision-making process of whether to invest resources in these putative hybrids, potentially at the expense of node J kakī. If hybridisation does become more prevalent following kakī expansion in the future, implementing genomic sequencing may prove more efficient if assessment of a large number of individuals is required, and may yield additional information than can be obtained via a genetic approach (e.g., if a high-quality genome annotation was available to investigate functional data associated

with introgression). With captive pairing decisions likely to be based on pedigree-informed genomic data using a whole-genome resequencing approach (Galla et al., 2020), it is possible that a similar approach could be implemented for hybrid assessment, as whole-genome resequencing may support this additional research. While we are not at the stage to require this approach, particularly while resequencing costs remain high relative to microsatellite genotyping and GBS, the bioinformatic workflow presented in Chapter Four can be readily adapted to incorporate resequencing data for analysis of introgression. As sequencing costs and efficiency continue to improve, I anticipate the method of hybrid assessment may change in the future, but the current efficiency achieved through microsatellite genotyping means this existing approach remains the most appropriate, particularly when management actions are dependent on receiving results in a timely manner.

5.4 Bridging the conservation genomics gap beyond kakī

The results presented in this thesis contribute to the knowledge regarding the impacts of interspecific hybridisation in threatened species, and the ways in which genomic tools can inform conservation management in these situations. Here I demonstrated that a genomic approach employing thousands of genome-wide SNPs produced results concordant with those of a genetic approach using a small microsatellite panel. This indicates that when there is strong differentiation between congeners, microsatellite genotyping is able to accurately detect this differentiation, and thus genetic tools will continue to offer robust and efficient identification of hybrid individuals for routine management. This is in contrast to studies that have identified conflicting results of genetic and genomic approaches, with genomic data proving more reliable for fine-scale assessments. RADseq methods were used for SNP

discovery to assess introgression in westslope cutthroat trout (*Oncorhynchus clarkii lewisi*) resulting from hybridisation with invasive rainbow trout (*O. mykiss*), and found that all individuals assigned as westslope cutthroat trout in microsatellite analysis (Boyer et al., 2008) presented a signal of rainbow trout introgression (Hohenlohe et al., 2013), and similar findings have been produced from analysis of introgression between native red deer (*Cervus elaphus*) and Japanese sika (*C. nippon*) in Scotland (McFarlane et al., 2020). Broader comparisons would be useful to determine whether such discordance between genetic and genomic analyses can be attributed to the reduced power of genetic marker sets to resolve interspecific differentiation resulting from more recent species divergence.

When a genomic approach is chosen to support conservation, additional consideration is required to determine which types of genomic tools are required. Among genomic sequencing methods, short-read sequencing is demonstrated here to represent a suitable approach to avian genome assembly to inform conservation of non-model species (Chapter Two). Where *de novo* assembly remains inaccessible to conservation programmes due to costs and assembly challenges, there are a wealth of existing genomic resources (e.g., the ~180 avian genomes available to date) that can be used to support reference-guided genome assembly or variant discovery using reduced-representation sequencing approaches. Reduced-representation sequencing is an accessible and accurate (McLennan et al., 2019) tool available to resource-limited conservation programmes transitioning to conservation genomics. However, there are limitations to this sequencing, namely that it is 'reduced-representation' and thus still only represents a fraction of the genome. In addition, for species with large or complex genomes (e.g., amphibians (Rogers et al., 2018; Smith et al., 2019), cartilaginous fishes (Hara et al., 2018; Marra et al., 2019), reduced-representation sequencing may indeed be the most

appropriate method for the foreseeable future, especially for those threatened species lacking genomic resources. The whole-genome resequencing data produced for kakī (Galla et al., 2019) has had a wide range of uses thus far, including assessing the optimal relatedness of reference genomes for marker discovery and robust estimates of conservation-relevant metrics (Galla et al., 2019, Appendix A), comparison of pedigree-, genetic- and genomic-based estimates of relatedness to inform captive breeding for species recovery (Galla et al., 2020), estimation of required sample sizes to obtain accurate genomic diversity metrics (Collier-Robinson, 2019), and to produce mitogenomes for modern kakī to assess haplotype diversity (Chapter Three). Such data sets present wider opportunities for use than those produced with genetic methods or through reduced-representation sequencing approaches where inferences beyond the focal research questions may be limited by the caveats associated with those data types (e.g., marker neutrality). Despite the greater costs of whole-genome resequencing compared with current reduced-representation sequencing, whole-genome resequencing methods may represent a better long-term investment, yielding more information (including the ability to investigate functional or structural variation) for wider applicability. As yet, the full extent of applications of genomic data have not been revealed, but as bioinformatic capabilities continue to evolve and collaborations between researchers and practitioners identify additional research questions, the benefits of whole-genome resequencing approaches are likely to become even more apparent.

5.5 Conservation and hybridisation in the genomics era

In addition to these findings confirming the success of conservation management for kakī, this thesis has implications for threatened species recovery more broadly. In

lieu of genetic or genomic data, conservation actions based on genetic principles can achieve the objectives, but where possible should incorporate genetic or genomic assessment to confirm that objectives are being met, and to implement adaptive management strategies where required. For conservation programmes affected by interspecific hybridisation, I recommend that when genetic assessment has not been conducted or there is remaining uncertainty as to whether genetic analysis has accurately captured the true extent of hybridisation, where resources allow, a genomic approach will enable robust assessment of the extent of introgression and validate management strategies aimed at maintaining genetic integrity. For example, while genetic tools proved adequate to detect hybridisation between the critically endangered Chatham Island black robin (*Petroica traversi*) and Chatham Island tomtit (*P. macrocephala chathamensis*; Cubrinovska et al., 2016), discriminating between hybridising New Zealand parakeets (*Cyanoramphus* spp.) has proved challenging and reassessment with genomic data would be beneficial (DOC, pers. comm.). For those programmes that have implemented a genetic approach to assessment of introgression, validation of the results via a genomics approach should be conducted where resources permit, due to the variation in concordance between approaches (McFarlane et al., 2020 *cf*. this thesis). When the results of genetic and genomic analyses agree, conservation managers can confidently continue to use genetic tools if these remain more efficient and cost effective. Although the field of conservation is in transition to a conservation genomics approach, when genetic tools remain robust and cost-effective, there is no reason to discontinue their use.

These results also have wider implications in relation to the IUCN Species Survival Commission Guidelines for Reintroductions and Other Conservation Translocations (IUCN Species Survival Commission, 2013). Implementation of these guidelines by

conservation practitioners can be hindered due to uncertainties regarding genetic concerns including interspecific hybridisation, particularly when conservation translocations move threatened species to areas where they may come into contact with close relatives. Based on the results herein, I hypothesise that when substantial differentiation between species is detected using genetic markers, introgressive hybridisation is unlikely to be detected with genomic markers, and the probability of detecting introgression with genomic markers will be low. Combined with additional studies of hybridisation in other threatened taxa, these findings will provide greater certainty for conservation practitioners when assessing the risks of interspecific hybridisation prior to translocations.

5.6 Shifting perspectives regarding hybridisation in light of genomic knowledge

I anticipate that as genomic approaches further reveal the impacts of hybridisation in both contemporary and evolutionary history contexts, the ways in which researchers consider hybridisation will change, especially for threatened species. There is already early evidence of this in the discussions regarding deliberate intraspecific hybridisation as a form of genetic rescue for genetically depauperate species (e.g., between subspecies of yellow-tufted honeyeaters, *Lichenostomus melanops*; Harrisson et al., 2016), or as a method akin to de-extinction to support ecosystem recovery (e.g., Galápagos tortoises, (Quinzin et al., 2019). In addition to these discussions, I also anticipate changes in the usage of terminology associated with hybridisation. As growing genomic evidence suggests that hybridisation is an integral part of the evolutionary history for many species (e.g., North American wolves, vonHoldt et al., 2011, 2016, but also see Hohenlohe et al., 2017; Rutledge et al.,

2015, and subsequent discussions), I anticipate that there will be a decline in the usage of terms with strong negative connotations (e.g., genetic 'pollution', 'contamination', and 'purity') as they are no longer regarded as appropriate descriptors for the observed evolutionary processes. In Web Of Sciences searches conducted April 3rd 2020 to assess the usage of terminology associated with hybridisation[4], I found 21 articles using the term 'genetic purity', 19 using 'genetic contamination', 48 using 'genetic pollution', and 181 using 'genetic integrity', from a total of 18,397 articles related to genetic or genomic research associated with hybridisation. Terminology usage by conservation practitioners outside of published research cannot be easily ascertained, but I anticipate that the use of 'genetic integrity' will continue to be widely used, with the use of terminology with strong negative connotations likely to decline.

5.7 Future directions for research

In this thesis, the reference genomes assembled for kakī and Australian pied stilts have been a key tool in the transition to a genomics approach for the conservation management of kakī. These genomes supported the successful completion of the objectives of this thesis. However, kakī recovery would benefit from targeted research to elucidate the underlying genetic basis of inbreeding depression in kakī (as revealed by Hagen et al., 2011) through investigation of structural genomic variants (see Deakin et al., 2019) or other studies of functional adaptation (e.g.,

4 Search terms: ALL FIELDS: (genetic OR genomic) AND (hybrid* OR introgress*) AND (species). Results were then refined by: Year published: 1970–2020; Document types excluded: 'news item', 'editorial material', 'meeting abstract', 'correction', 'reprint', 'retraction', and 'biographical item'; Web Of Science categories excluded: a wide range of apparently unrelated categories were excluded including categories related to veterinary science, geology, geography, medical science, microbiology, art, sociology, reproductive and developmental biology, and psychology. Following refinement, searches were conducted within the results for the terms 'genetic integrity', 'genetic pollution', 'genetic contamination', and 'genetic integrity'. This search is likely to underrepresent the use of these terms as it does not include the main body text of articles.

immunocompetence genes; Grueber et al., 2014; Wang et al., 2012). Such studies investigating structural variation and functional adaptation require an accurately annotated genome with improved contiguity (i.e., a gene-annotated chromosomally-assembled genome). The rapid development of technologies in the genomics space (in terms of both sequencing and analyses) now enables such research and thus kakī genome reassembly has been initiated as part of the High Quality Genomes project of Genomics Aotearoa, a national collaborative platform developing genomic capacity (https://www.genomics-aotearoa.org.nz/projects/high-quality-genomes). The development of a high-quality chromosomally-assembled genome will incorporate long-read PacBio sequencing alongside the existing short-read data to close gaps (Rhoads & Au, 2015), transcriptomic data for genome annotation (He et al., 2016), and Hi-C analyses (van Berkum et al., 2010). As such, the kakī genome produced here remains useful in the short-term for supporting captive pairing recommendations, but will eventually be superseded. Although such high-quality chromosomally-assembled genomes may be produced for other relatively well-funded threatened species, a short-read genome assembly such as these developed here will be useful for the vast majority of resource-limited threatened species. Discrepancies between morphological and recorded species identities and the data supporting morphological classification of specimens from natural history collections (Chapter 3, Appendix C) have been communicated to the relevant curators at Te Papa Tongarewa Museum of New Zealand and Canterbury Museum, and the discrepancies identified here may be validated through additional morphological assessment (see Holdaway, 1995; Pierce, 1984). Such discrepancies should encourage researchers to verify catalogue data when incorporating specimens from natural history collections in analyses (also see Verry et al., 2019), with particular

care taken for species where hybridisation may hinder species identification due to the production of cryptic hybrids.

Additional work is required to complete the kakī mitogenome assembly as described in Chapter Three, (see preliminary reassembly conducted in Appendix D), to ensure robust resolution of the observed region of gene duplication. Further work can also be conducted to ascertain the presence of a similar region of mitochondrial gene duplication in pied stilts. I recommend that a wide reassessment of 'complete' mitogenomes be conducted within Charadriiformes to resolve the evolution of this duplication, and improve phylogenetic inferences within this order. Nevertheless, the exclusion of the hypothesised region of gene duplication in the analyses conducted in Chapter Three allowed robust estimation of mitochondrial diversity and differentiation among stilts, and were sufficient to address the needs of conservation management.

DNA degradation over time following organism death makes the extraction of intact nuclear data from historic and ancient samples challenging (Hofreiter et al., 2001). The relatively recent origin of the majority of the historic samples incorporated in Chapter Three, along with excellent specimen preservation in natural history collections has enabled the extraction of high-quality mitochondrial sequence data. With extracted DNA available for many of the historic samples used here, there is the potential to generate nuclear genomic data via high-throughput sequencing, and in combination with the data from modern samples (Chapter Four) could enable temporal comparisons of nuclear introgression (reviewed in Schaefer et al., 2016). With conservation management of kakī shifting towards the inclusion of low-coverage whole-genome resequencing alongside pedigree data to inform captive pairing decisions, reassessment of introgression with resequencing data could be considered in the future if sequencing costs continue to decline and hybridisation

becomes of renewed concern for kakī conservation. The workflow developed for assessing introgression from reduced-representation data in Chapter Four can be readily adapted to incorporate whole-genome resequencing data as input, although will require additional consideration of the SNP filtering methods employed, and how these may need to be adjusted for resequencing data. I anticipate that a larger SNP set will be produced from whole-genome resequencing data than from reduced-representation data, and so more stringent filtering could be implemented with less reduction in resolution than was observed from preliminary filtering assessments of the GBS data used here. Nevertheless, I predict that the findings here indicating no introgression from poaka are robust and would be corroborated by analysis of resequencing data, but in combination with the improved kakī genome assembly may allow elucidation of the underlying genomic basis of outbreeding depression observed in hybrid offspring (Steeves et al., 2010). In addition, preliminary assessment of SNP sets including putative sex chromosomes did not identify any difference in introgression from that determined by the final SNP sets that excluded sex chromosomes. The observed outbreeding depression resulting from hybridisation between these species manifests as reduced fitness of hybrid females, consistent with Haldane's rule that hybridisation will negatively impact the heterogametic sex (in birds, females with ZW sex chromosomes; Orr, 1997), indicating a likely association to the sex chromosomes. If resequencing was conducted for analysis of introgression, there is the potential to examine any differences in levels of introgression between autosomal and sex-linked loci as implemented in the study of hybridisation between koloa and mallards in Hawaii (Wells et al., 2019), or to investigate the underlying processes driving the observed pattern dark plumage correlating with high kakī assignment probability using the chromosomally assembled annotated genome.

Following our results demonstrating that confamilial reference genomes produce robust estimates of metrics relevant to conservation in birds (Galla et al., 2019, Appendix A), similar Proof-of-Concept studies to identify appropriate reference genomes across a wider range of taxa would be beneficial to broaden the genomic resources available for other threatened species, further reducing the conservation genomics gap. With genomic data sets available for an increasing breadth of non-model taxa, we also now have the potential to empirically evaluate the impacts of different variant filtering strategies on the accuracy of estimation of a range of metrics, including relatedness, genome-wide diversity, inbreeding, and introgression across a diversity of taxa. There is the potential for these considerations to be overlooked in this space as priority is placed on application, but consideration of the underlying processes is recommended to ensure accuracy of results and to minimise biases in the data. Differences in results observed across variant-calling pipelines have been identified previously (e.g., in model species, O'Rawe et al., 2013; Torkamaneh et al., 2016). In addition, varying rates of missing data have been observed to have significant effects on the results of downstream analyses (e.g., in analysis of effective population size, Marandel et al., 2020). Broader comparisons of the impacts of various filtering strategies on diversity and differentiation estimates are currently underway, with a number of research groups developing bioinformatic pipelines and best practices for variant discovery and filtering to achieve robust, accurate estimates and minimise biases (e.g., Andrews et al., 2018; McLennan et al., 2019; O'Leary et al., 2018; Puritz et al., 2014; Wright et al., 2019), and in some instances have recognised that variations in bioinformatic processing of data can have differential impacts on the results of downstream analyses (Shafer et al., 2018). Therefore, empirical studies comparing filtering methods across a range of threatened taxa are recommended. Such studies may include the development of

pipelines with broad applicability for variant filtering regardless of the population dynamics of the focal species, thus further supporting the transition to conservation genomics for threatened species.

5.8 Contributing to bridging the conservation genomics gap

In addition to improving outcomes for the critically endangered kakī, this thesis presents resources and methods to accelerate the uptake of genomics for threatened species conservation. The thesis objectives were designed with the needs of the DOC's Kakī Recovery Programme at the forefront, with a collaborative approach to research development, and co-production of research outputs with practitioners (i.e., through co-authorship on research papers). In addition to sharing and discussing the results of this research with practitioners, I have also engaged with the day-to-day practicalities of kakī conservation alongside practitioners, including assisting with translocations, gaining a better understanding of the biological, ecological, and practical challenges associated with kakī conservation. Through this, a collaboration based on open and honest communication to develop mutual respect and understanding has strengthened this collaboration, resulting in the findings presented here having already been implemented in the decision-making processes to support kakī recovery. Through collaborations and knowledge-sharing (with the DOC's Kakī Recovery Programme, and genomics researchers and conservation practitioners more widely within the New Zealand conservation landscape, e.g., at the annual MapNet meeting for genomic researchers across the primary production and conservation research spaces), the production of open access research outputs (e.g., Galla et al., 2019) and public science communication

associated with this research (e.g., through involvement in Art + Genetics, see Forsdick & Kelly, 2017; Wilkinson & Forsdick, 2017, https://issuu.com/dunedinschoolofart/docs/art_and_genetics_2017_catalogue, Figure 5.1; New Zealand Bird of the Year, see https://sciblogs.co.nz/guestwork/2018/10/10/novel-science-communication-sees-bird-of-the-year-take-on-tinder/ co-authored by myself and Stephanie Galla, https://www.stuff.co.nz/environment/107549959/bird-of-the-year-how-an-endangered-bird-ended-up-on-tinder), and working within a Kindness In Science framework supporting diversity and inclusivity to improve research outcomes (see http://www.kindnessinscience.org/, https://bioheritage.nz/kindness-in-science/), the resources and outcomes produced herein contribute to reducing the conservation genomics gap, and this research is an exemplar for bridging the research-implementation gap (see recommendations in Jarvis et al., in review).

Figure 5.1: 'Kakī - on cutting edge.' Woodcut, dry point colour etching, coloured pencil, inspired by the research presented in this thesis, by Brigitte Kammlein, 2017. This was one of three art pieces produced by three New Zealand artists inspired by this research as part of the 2017 Art + Genetics Project. All three pieces were presented at the Genetics Society of AustralAsia/New Zealand Society for Biochemistry and Molecular Biology joint 2017 conference and were subsequently on public display. A print in this series was gifted to the DOC's Kakī Recovery Programme by the artist.

Bibliography

Abbott, C. L., Double, M. C., Trueman, J. W. H., Robinson, A., & Cockburn, A. (2005). An unusual source of apparent mitochondrial heteroplasmy: Duplicate mitochondrial control regions in Thalassarche albatrosses. *Molecular Ecology*, *14*(11), 3605–3613. https://doi.org/10.1111/j.1365-294X.2005.02672.x

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Åkesson, M., Liberg, O., Sand, H., Wabakken, P., Bensch, S., & Flagstad, Ø. (2016). Genetic rescue in a severely inbred wolf population. *Molecular Ecology*, *25*(19), 4745–4756. https://doi.org/10.1111/mec.13797

Alexander, A. M., Su, Y.-C., Oliveros, C. H., Olson, K. V., Travers, S. L., & Brown, R. M. (2017). Genomic data reveals potential for hybridization, introgression, and incomplete lineage sorting to confound phylogenetic relationships in an adaptive radiation of narrow-mouth frogs. *Evolution*, *71*(2), 475–488. https://doi.org/10.1111/evo.13133

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

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

Allendorf, F. W., Leary, R. F., Spruell, P., & Wenburg, J. K. (2001). The problems with hybrids: Setting conservation guidelines. *Trends in Ecology & Evolution*, *16*(11), 613–622. https://doi.org/10.1016/S0169-5347(01)02290-X

Allendorf, F. W., Luikart, G., & Aitken, S. N. (2013). *Conservation and the genetics of populations* (2nd ed.). Wiley-Blackwell,.

Al-Nakeeb, K., Petersen, T. N., & Sicheritz-Pontén, T. (2017). Norgal: Extraction and de novo assembly of mitochondrial DNA from whole-genome sequencing data. *BMC Bioinformatics*, *18*(1), 510. https://doi.org/10.1186/s12859-017-1927-y

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anderson, E., & Stebbins, G. L. (1954). Hybridization as an evolutionary stimulus. *Evolution*, *8*(4), 378–378. https://doi.org/10.2307/2405784

Andrews, K. R., Adams, J. R., Cassirer, E. F., Plowright, R. K., Gardner, C., Dwire, M., … Waits, L. P. (2018). A bioinformatic pipeline for identifying informative SNP panels for parentage assignment from RADseq data. *Molecular Ecology Resources*, *18*(6), 1263–1281. https://doi.org/10.1111/1755-0998.12910

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*(2), 81–92. https://doi.org/10.1038/nrg.2015.28

Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data.* http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Ardui, S., Ameur, A., Vermeesch, J. R., & Hestand, M. S. (2018). Single molecule real-time (SMRT) sequencing comes of age: Applications and utilities for medical diagnostics. *Nucleic Acids Research*, *46*(5), 2159–2168. https://doi.org/10.1093/nar/gky066

Arlettaz, R., Schaub, M., Fournier, J., Reichlin, T. S., Sierro, A., Watson, J. E. M., & Braunisch, V. (2010). From publications to public actions: When conservation biologists bridge the gap between research and implementation. *BioScience*, *60*(10), 835–842. https://doi.org/10.1525/bio.2010.60.10.10

Arnold, M. L. (1992). Natural hybridization as an evolutionary process. *Annual Review of Ecology and Systematics*, *23*(1), 237–261. https://doi.org/10.1146/annurev.es.23.110192.001321

Arnold, M. L. (1997). *Natural hybridization and species concepts*. Oxford University Press. https://books.google.co.nz/books?hl=en&lr=&id=Jh8jVjEuDfUC&oi=fnd&pg=PR11&ots=PbObh3Qp32&sig=fligYJ6xTvI37WapjEhBMlzlPVs&redir_esc=y#v=onepage&q&f=false

Arnold, M. L. (2016). *Divergence with genetic exchange*. Oxford University Press.

Avise, J. C. (2010). Perspective: Conservation genetics enters the genomics era. *Conservation Genetics*, *11*(2), 665–669. https://doi.org/10.1007/s10592-009-0006-y

Ba, H., Jia, B., Wang, G., Yang, Y., Kedem, G., & Li, C. (2017). Genome-wide SNP discovery and analysis of genetic diversity in farmed sika deer (*Cervus nippon*) in northeast China using double-digest restriction site-associated DNA

sequencing. *G3: Genes, Genomes, Genetics*, *7*(9), 3169–3176. https://doi.org/10.1534/g3.117.300082

Baker, A. J., Pereira, S. L., & Paton, T. A. (2007). Phylogenetic relationships and divergence times of Charadriiformes genera: Multigene evidence for the Cretaceous origin of at least 14 clades of shorebirds. *Biology Letters*, *3*(2), 205–209. https://doi.org/10.1098/rsbl.2006.0606

Ballard, J.W.O., & Whitlock, M.C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*(4), 729–744. https://doi.org/10.1046/j.1365-294x.2003.02063.x

Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*(1), 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Banker, S. E., Wade, E. J., & Simon, C. (2017). The confounding effects of hybridization on phylogenetic estimation in the New Zealand cicada genus Kikihia. *Molecular Phylogenetics and Evolution*, *116*, 172–181. https://doi.org/10.1016/j.ympev.2017.08.009

Bao, E., Jiang, T., & Girke, T. (2014). AlignGraph: Algorithm for secondary de novo genome assembly guided by closely related references. *Bioinformatics*, *30*(12), i319–i328. https://doi.org/10.1093/bioinformatics/btu291

Bayat, A., Gaëta, B., Ignjatovic, A., & Parameswaran, S. (2017). Improved VCF normalization for accurate VCF comparison. *Bioinformatics*, *33*(7), 964–970. https://doi.org/10.1093/bioinformatics/btw748

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., … Stadler, P. F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, *69*(2), 313–319. https://doi.org/10.1016/j.ympev.2012.08.023

BirdLife Australia. (2019). *Working list of Australian birds version 3*. https://www.birdlife.org.au/conservation/science/taxonomy

BirdLife International. (2017). *One in eight of all bird species is threatened with global extinction*. http://www.birdlife.org

BirdLife International. (2018). *Himantopus novaezelandiae* (amended version of 2016 assessment). *The IUCN Red List of Threatened Species*. http://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T22693690A129560535.en

Boecklen, W. J., & Howard, D. J. (1997). Genetic analysis of hybrid zones: Numbers of markers and power of resolution. *Ecology*, *78*(8), 2611–2616. https://doi.org/10.1890/0012-9658(1997)078[2611:GAOHZN]2.0.CO;2

Boessenkool, S., Taylor, S. S., Tepolt, C. K., Komdeur, J., & Jamieson, I. G. (2007). Large mainland populations of South Island robins retain greater genetic diversity than offshore island refuges. *Conservation Genetics*, *8*(3), 705–714. https://doi.org/10.1007/s10592-006-9219-5

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Boon, W. M., Kearvell, J. C., Daugherty, C. H., & Chambers, G. K. (2000). Molecular systematics of New Zealand Cyanoramphus parakeets: Conservation of Orange-fronted and Forbes' Parakeets. *Bird Conservation International*, *10*(3), 211–239. https://doi.org/10.1017/S0959270900000198

Bos, K. I., Herbig, A., Sahl, J., Waglechner, N., Fourment, M., Forrest, S. A., … Poinar, H. N. (2016). Eighteenth century *Yersinia pestis* genomes reveal the long-term persistence of an historical plague focus. *ELife*, *5*, e12994. https://doi.org/10.7554/eLife.12994

Botigué, L. R., Song, S., Scheu, A., Gopalan, S., Pendleton, A. L., Oetjens, M., … Veeramah, K. R. (2017). Ancient European dog genomes reveal continuity since the Early Neolithic. *Nature Communications*, *8*(1), 1–11. https://doi.org/10.1038/ncomms16082

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., … Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, *15*(4), e1006650. https://doi.org/10.1371/journal.pcbi.1006650

Boyer, M. C., Muhlfeld, C. C., & Allendorf, F. W. (2008). Rainbow trout (*Oncorhynchus mykiss*) invasion and the spread of hybridization with native westslope cutthroat trout (*Oncorhynchus clarkii lewisi*). *Canadian Journal of Fisheries and Aquatic Sciences*, *65*(4), 658–669. https://doi.org/10.1139/f08-001

Bradnam, K. R., Fass, J. N., Alexandrov, A., Baranay, P., Bechner, M., Birol, I., … Korf, I. F. (2013). Assemblathon 2: Evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience*, *2*(1), 10–10. https://doi.org/10.1186/2047-217X-2-10

Brandies, P., 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

Brown, C. T., Howe, A., Zhang, Q., Pyrkosz, A. B., & Brom, T. H. (2012). A reference-free algorithm for computational normalization of shotgun sequencing data. *ArXiv:1203.4802 [q-Bio]*. http://arxiv.org/abs/1203.4802

Brown, J. I., Mohl, J., Sheppard, J. L., & Lavretsky, P. (2020). *Mallard introductions to New Zealand result in extensive hybridization with endemic grey ducks*. The Plant and Animal Genome Conference XXVIII, San Diego, California, USA. https://plan.core-apps.com/pag_2020/abstract/60658e794c2a6d06f3aa20dcfb5c9172

Brumfield, R. T. (2010). Speciation genetics of biological invasions with hybridization. *Molecular Ecology*, *19*(23), 5079–5083. https://doi.org/10.1111/j.1365-294X.2010.04896.x

Brunson, K., & Reich, D. (2019). The Promise of Paleogenomics Beyond Our Own Species. *Trends in Genetics*, *35*(5), 319–329. https://doi.org/10.1016/j.tig.2019.02.006

Buschke, F. T., Botts, E. A., & Sinclair, S. P. (2019). Post-normal conservation science fills the space between research, policy, and implementation. *Conservation Science and Practice*, *1*(8), e73. https://doi.org/10.1111/csp2.73

Butchart, S. H. M., Walpole, M., Collen, B., Strien, A. van, Scharlemann, J. P. W., Almond, R. E. A., … Watson, R. (2010). Global biodiversity: Indicators of recent declines. *Science*, *328*(5982), 1164–1168. https://doi.org/10.1126/science.1187512

Butler, D., & Merton, D. (1992). *The black robin: Saving the world's most endangered bird*. Oxford University Press,.

Caballero, A., Rodríguez-Ramilo, S. T., Ávila, V., & Fernández, J. (2010). Management of genetic diversity of subdivided populations in conservation programmes. *Conservation Genetics*, *11*(2), 409–419. https://doi.org/10.1007/s10592-009-0020-0

Card, D. C., Schield, D. R., Reyes-Velasco, J., Fujita, M. K., Andrew, A. L., Oyler-McCance, S. J., … Castoe, T. A. (2014). Two low coverage bird genomes and a comparison of reference-guided versus de novo genome assemblies. *PLOS ONE*, *9*(9), e106649. https://doi.org/10.1371/journal.pone.0106649

Caruso, B. S. (2006). Project River Recovery: Restoration of braided gravel-bed river habitat in New Zealand's high country. *Environmental Management*, *37*(6), 840–861. https://doi.org/10.1007/s00267-005-3103-9

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354

Chakraborty, M., Baldwin-Brown, J. G., Long, A. D., & Emerson, J. J. (2016). Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Research*, *44*(19), e147–e147. https://doi.org/10.1093/nar/gkw654

Chambers, G. K., & Macavoy, E. S. (1999). Molecular genetic analysis of hybridisation. *Science for Conservation*, *105*, 1–29.

Chan, C. H., Ballantyne, K. N., Aikman, H., Fastier, D., Daugherty, C. H., & Chambers, G. K. (2006a). Genetic analysis of interspecific hybridisation in the world's only Forbes' parakeet (*Cyanoramphus forbesi*) natural population. *Conservation Genetics*, *7*(4), 493–506. https://doi.org/10.1007/s10592-005-9060-2

Chan, W. Y., Hoffmann, A. A., & Oppen, M. J. H. van. (2019). Hybridization as a conservation management tool. *Conservation Letters*, *e12652*, e12652. https://doi.org/10.1111/conl.12652

Chan, Y. L., Anderson, C. N. K., & Hadly, E. A. (2006b). Bayesian estimation of the timing and severity of a population bottleneck from ancient DNA. *PLoS Genetics*, *2*(4). https://doi.org/10.1371/journal.pgen.0020059

Chapman, J. A., Ho, I. Y., Goltsman, E., & Rokhsar, D. S. (2017). Meraculous2: Fast accurate short-read assembly of large polymorphic genomes. *ArXiv*. http://arxiv.org/abs/1608.01031

Chapple, D. G., Birkett, A., Miller, K. A., Daugherty, C. H., & Gleeson, D. M. (2012). Phylogeography of the endangered Otago skink, *Oligosoma otagense*: Population structure, hybridisation and genetic diversity in captive populations. *PLOS ONE*, *7*(4), e34599. https://doi.org/10.1371/journal.pone.0034599

Charlesworth, B., & Charlesworth, D. (1999). The genetic basis of inbreeding depression. *Genetics Research*, *74*(3), 329–340. https://doi.org/10.1017/S0016672399004152

Check Hayden, E. (2014). Technology: The $1,000 genome. *Nature News*, *507*(7492), 294. https://doi.org/10.1038/507294a

Chen, N., Cosgrove, E. J., Bowman, R., Fitzpatrick, J. W., & Clark, A. G. (2016). Genomic consequences of population decline in the endangered Florida scrub-jay. *Current Biology*, *26*(21), 2974–2979. https://doi.org/10.1016/j.cub.2016.08.062

Chikhi, R., & Medvedev, P. (2014). Informed and automated k-mer size selection for genome assembly. *Bioinformatics*, *30*(1), 31–37. https://doi.org/10.1093/bioinformatics/btt310

Cho, H.-J., Eda, M., Nishida, S., Yasukochi, Y., Chong, J.-R., & Koike, H. (2009). Tandem duplication of mitochondrial DNA in the black-faced spoonbill, *Platalea min*or. *Genes & Genetic Systems*, *84*(4), 297–305. https://doi.org/10.1266/ggs.84.297

Chu, J., Sadeghi, S., Raymond, A., Jackman, S. D., Nip, K. M., Mar, R., … Birol, I. (2014). BioBloom tools: Fast, accurate and memory-efficient host species sequence screening using bloom filters. *Bioinformatics*, *30*(23), 3402–3404. https://doi.org/10.1093/bioinformatics/btu558

Chunco, A. J. (2014). Hybridization in a warmer world. *Ecology and Evolution*, *4*(10), 2019–2031. https://doi.org/10.1002/ece3.1052

Collier-Robinson, L. F. M. (2019). *A genomics approach to the conservation of kōwaro (Canterbury mudfish, Neochanna burrowsius)*. University of Canterbury.

Connell, J. (2009). Birdwatching, twitching and tourism: Towards an Australian perspective. *Australian Geographer*, *40*(2), 203–217. https://doi.org/10.1080/00049180902964942

Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics*, *33*(18), 2938–2940. https://doi.org/10.1093/bioinformatics/btx364

Coyne, J. A., & Orr, H. A. (1989). Patterns of speciation in Drosophila. *Evolution*, *43*(2), 362–381. https://doi.org/10.1111/j.1558-5646.1989.tb04233.x

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer Associates.

Cremona, T., Spencer, P., Shine, R., & Webb, J. K. (2017). Avoiding the last supper: Parentage analysis indicates multi-generational survival of re-introduced ‘toad-smart’ lineage. *Conservation Genetics*, *18*(6), 1475–1480. https://doi.org/10.1007/s10592-017-0973-3

Crusoe, M. R., Alameldin, H. F., Awad, S., Boucher, E., Caldwell, A., Cartwright, R., … Brown, C. T. (2015). The khmer software package: Enabling efficient

nucleotide sequence analysis. *F1000Research*, *4*, 900. https://doi.org/10.12688/f1000research.6924.1

Cruz, J., Pech, R. P., Seddon, P. J., Cleland, S., Nelson, D., Sanders, M. D., & Maloney, R. F. (2013). Species-specific responses by ground-nesting Charadriiformes to invasive predators and river flows in the braided Tasman River of New Zealand. *Biological Conservation*, *167*, 363–370. https://doi.org/10.1016/j.biocon.2013.09.005

Cubrinovska, I., Massaro, M., & Hale, M. L. (2016). Assessment of hybridisation between the endangered Chatham Island black robin (*Petroica traversi*) and the Chatham Island tomtit (*Petroica macrocephala chathamensis*). *Conservation Genetics*, *17*(2), 259–265. https://doi.org/10.1007/s10592-015-0778-1

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and high-performance computing. *Nature Methods*, *9*(8), 772. https://doi.org/10.1038/nmeth.2109

Davey, J. W., & Blaxter, M. L. (2010). RADSeq: Next-generation population genetics. *Briefings in Functional Genomics*, *9*(5–6), 416–423. https://doi.org/10.1093/bfgp/elq031

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, *12*(7), 499–510. https://doi.org/10.1038/nrg3012

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*(5), 889-894.e3. https://doi.org/10.1016/j.cub.2019.01.072

Deakin, J. E., Potter, S., O'Neill, R., Ruiz-Herrera, A., Cioffi, M. B., Eldridge, M. D. B., … Ezaz, T. (2019). Chromosomics: Bridging the gap between genomes and chromosomes. *Genes*, *10*(8). https://doi.org/10.3390/genes10080627

Der Sarkissian, C., Ermini, L., Schubert, M., Yang, M. A., Librado, P., Fumagalli, M., … Orlando, L. (2015). Evolutionary genomics and conservation of the endangered Przewalski's horse. *Current Biology*, *25*(19), 2577–2583. https://doi.org/10.1016/j.cub.2015.08.032

Dierckxsens, N., Mardulyn, P., & Smits, G. (2016). NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research*, *45*(4), gkw955–gkw955. https://doi.org/10.1093/nar/gkw955

Dierickx, E.G., Sin, S.Y.W., van Veelen, H.P.J., Brooke, M. de L., Liu, Y., Edwards, S.V., & Martin, S.H. (2020). Genetic diversity, demographic history and neo-sex chromosomes in the Critically Endangered Raso lark. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1922). https://doi.org/10.1098/rspb.2019.2613

Dowling, T. E., & Secor, C. L. (1997). The role of hybridization and introgression in the diversification of animals. *Annual Review of Ecology and Systematics*, *28*(1), 593–619. https://doi.org/10.1146/annurev.ecolsys.28.1.593

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLOS Biology*, *4*(5), e88. https://doi.org/10.1371/journal.pbio.0040088

Drummond, A. J., Rambaut, A., Shapiro, B., & Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, *22*(5), 1185–1192. https://doi.org/10.1093/molbev/msi103

Dubois, N. S., Gomez, A., Carlson, S., & Russell, D. (2020). Bridging the research-implementation gap requires engagement from practitioners. *Conservation Science and Practice*, *2*(1), e134. https://doi.org/10.1111/csp2.134

Duchêne, D. A., Bragg, J. G., Duchêne, S., Neaves, L. E., Potter, S., Moritz, C., … Eldridge, M. D. B. (2018). Analysis of phylogenomic tree space resolves relationships among marsupial families. *Systematic Biology*, *67*(3), 400–412. https://doi.org/10.1093/sysbio/syx076

Duncan, R. P., & Blackburn, T. M. (2004). Extinction and endemism in the New Zealand avifauna. *Global Ecology and Biogeography*, *13*(6), 509–517. https://doi.org/10.1111/j.1466-822X.2004.00132.x

Duncan, R. P., & Blackburn, T. M. (2007). Causes of extinction in island birds. *Animal Conservation*, *10*(2), 149–150. https://doi.org/10.1111/j.1469-1795.2007.00110.x

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

Eaton, D. A. R. (2014). PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, *30*(13), 1844–1849. https://doi.org/10.1093/bioinformatics/btu121

Eberhard, J. R., & Wright, T. F. (2016). Rearrangement and evolution of mitochondrial genomes in parrots. *Molecular Phylogenetics and Evolution*, *94*, 34–46. https://doi.org/10.1016/j.ympev.2015.08.011

Eberhard, J. R., Wright, T. F., & Bermingham, E. (2001). Duplication and concerted evolution of the mitochondrial control region in the parrot genus Amazona. *Molecular Biology and Evolution*, *18*(7), 1330–1342. https://doi.org/10.1093/oxfordjournals.molbev.a003917

Eda, M., Kuro-o, M., Higuchi, H., Hasegawa, H., & Koike, H. (2010). Mosaic gene conversion after a tandem duplication of mtDNA sequence in Diomedeidae (albatrosses). *Genes & Genetic Systems*, *85*(2), 129–139. https://doi.org/10.1266/ggs.85.129

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Edmands, S. (2007). Between a rock and a hard place: Evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology*, *16*(3), 463–475. https://doi.org/10.1111/j.1365-294X.2006.03148.x

Ekblom, R., Brechlin, B., Persson, J., Smeds, L., Johansson, M., Magnusson, J., … Ellegren, H. (2018). Genome sequencing and conservation genomics in the Scandinavian wolverine population. *Conservation Biology*, *32*(6), 1301–1312. https://doi.org/10.1111/cobi.13157

Ekblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, *107*(1), 1–15. https://doi.org/10.1038/hdy.2010.152

Ekblom, R., & Wolf, J. B. W. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, *7*(9), 1026–1042. https://doi.org/10.1111/eva.12178

Eldridge, M. D. B., King, J. M., Loupis, A. K., Spencer, P. B. S., Taylor, A. C., Pope, L. C., & Hall, G. P. (1999). Unprecedented low levels of genetic variation and inbreeding depression in an island population of the black-footed rock-wallaby.

Conservation Biology, *13*(3), 531–541. https://doi.org/10.1046/j.1523-1739.1999.98115.x

Elliott, G. P., Merton, D. V., & Jansen, P. W. (2001). Intensive management of a critically endangered species: The kakapo. *Biological Conservation*, *99*(1), 121–133. https://doi.org/10.1016/S0006-3207(00)00191-9

Ellstrand, N. C., Biggs, D., Kaus, A., Lubinsky, P., McDade, L. A., Preston, K., … Schierenbeck, K. A. (2010). Got hybridization? A multidisciplinary approach for informing science policy. *BioScience*, *60*(5), 384–388. https://doi.org/10.1525/bio.2010.60.5.8

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, *6*(5), e19379–e19379. https://doi.org/10.1371/journal.pone.0019379

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, *164*(4), 1567–1587.

Fitzpatrick, B. M., Johnson, J. R., Kump, D. K., Smith, J. J., Voss, S. R., & Shaffer, H. B. (2010). Rapid spread of invasive genes into a threatened native species. *Proceedings of the National Academy of Sciences*, *107*(8), 3606–3610. https://doi.org/10.1073/pnas.0911802107

Forsdick, N. J., Cubrinovska, I., Massaro, M., & Hale, M. L. (2017). Genetic diversity and population differentiation within and between island populations of two sympatric Petroica robins, the Chatham Island black robin and tomtit. *Conservation Genetics*, *18*(2), 275–285. https://doi.org/10.1007/s10592-016-0899-1

Forsdick, N., & Kelly, M. (2017). Kakī reared and sampled. *Junctures*, *2017*(18), 58–64.

Fowler, A. C., Eadie, J. M., & Engilis, A. (2008). Identification of endangered Hawaiian ducks (*Anas wyvilliana*), introduced North American mallards (*A. platyrhynchos*) and their hybrids using multilocus genotypes. *Conservation Genetics*, *10*(6), 1747. https://doi.org/10.1007/s10592-008-9778-8

Francis, R. M. (2017). pophelper: An R package and web app to analyse and visualise population structure. *Molecular Ecology Resources*, *17*(1), 27–32. https://doi.org/10.1111/1755-0998.12509

Frankham, R. (1997). Do island populations have less genetic variation than mainland populations? *Heredity*, *78*(3), 311–327. https://doi.org/10.1038/hdy.1997.46

Frankham, R. (2005). Genetics and extinction. *Biological Conservation*, *126*(2), 131–140. https://doi.org/10.1016/j.biocon.2005.05.002

Frankham, R., Ballou, J. D., Eldridge, M. D. B., Lacy, R. C., Ralls, K., Dudash, M. R., & Fenster, C. B. (2011). Predicting the probability of outbreeding depression. *Conservation Biology*, *25*(3), 465–475. https://doi.org/10.1111/j.1523-1739.2011.01662.x

Frankham, R., Ballou, J. D., Ralls, K., Eldridge, M., Dudash, M. R., Fenster, C. B., … Sunnucks, P. (2017). *Genetic management of fragmented animal and plant populations*. Oxford University Press.

Fu, Q., Mittnik, A., Johnson, P. L. F., Bos, K., Lari, M., Bollongino, R., … Krause, J. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. *Current Biology*, *23*(7), 553–559. https://doi.org/10.1016/j.cub.2013.02.044

Galla, S. J. (2019). *Conservation genomic management of two critically endangered New Zealand birds*. https://ir.canterbury.ac.nz/handle/10092/17759

Galla, S. J., Buckley, T. R., Elshire, R., Hale, M. L., Knapp, M., Mccallum, J., … Steeves, T. E. (2016). Building strong relationships between conservation genetics and primary industry leads to mutually beneficial genomic advances. *Molecular Ecology*. https://doi.org/10.1111/mec.13837

Galla, S. J., Forsdick, N. J., Brown, L., Hoeppner, M., Knapp, M., Maloney, R. F., … Steeves, T. E. (2019). Reference genomes from distantly related species can be used for discovery of single nucleotide polymorphisms to inform conservation management. *Genes*, *10*(1), 9. https://doi.org/10.3390/genes10010009

Galla, S. J., Moraga, R., Brown, L., Cleland, S., Hoeppner, M. P., Maloney, R. F., … 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*. 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). https://doi.org/10.1093/gigascience/giz111

Garner, B. A., Hand, B. K., Amish, S. J., Bernatchez, L., Foster, J. T., Miller, K. M., … Luikart, G. (2016). Genomics in conservation: Case studies and bridging the gap between data and application. *Trends in Ecology and Evolution*, *31*(2), 81–82. https://doi.org/10.1016/j.tree.2015.10.009

Genome 10K Community of Scientists. (2009). Genome 10K: A proposal to obtain whole-genome sequence for 10,000 vertebrate species. *The Journal of Heredity*, *100*(6), 659–674. https://doi.org/10.1093/jhered/esp086

Gibb, G. C., England, R., Hartig, G., McLenachan, P. A. (Trish), Taylor Smith, B. L., McComish, B. J., … 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

Gibb, G. C., Kennedy, M., & Penny, D. (2013). Beyond phylogeny: Pelecaniform and ciconiiform birds, and long-term niche stability. *Molecular Phylogenetics and Evolution*, *68*(2), 229–238. https://doi.org/10.1016/j.ympev.2013.03.021

Giglio, R. M., Ivy, J. A., Jones, L. C., & Latch, E. K. (2016). Evaluation of alternative management strategies for maintenance of genetic variation in wildlife populations. *Animal Conservation*, *19*(4), 380–390. https://doi.org/10.1111/acv.12254

Gill, B. J. (Ed.). (2010). *Checklist of the birds of New Zealand, Norfolk and Macquarie Islands, and the Ross Dependency, Antarctica* (Vol. 4). Te Papa Press.

Gillespie, G. D. (1985). Hybridization, introgression, and morphometric differentiation between Mallard (*Anas platyrhynchos*) and Grey Duck (*Anas superciliosa*) in Otago, New Zealand. *The Auk*, *102*(3), 459–469. https://doi.org/10.1093/auk/102.3.459

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS ONE*, *9*(2), e90346–e90346. https://doi.org/10.1371/journal.pone.0090346

Gómez-Sánchez, D., Olalde, I., Sastre, N., Enseñat, C., Carrasco, R., Marques-Bonet, T., … Ramírez, O. (2018). On the path to extinction: Inbreeding and admixture in a declining grey wolf population. *Molecular Ecology*, *27*(18), 3599–3612. https://doi.org/10.1111/mec.14824

Gompert, Z. (2012). Population genomics as a new tool for wildlife management. *Molecular Ecology*, *21*(7), 1542–1544. https://doi.org/10.1111/j.1365-294X.2012.05471.x

Goodman, S. J., Barton, N. H., Swanson, G., Abernethy, K., & Pemberton, J. M. (1999). Introgression through rare hybridization: A genetic study of a hybrid zone between red and sika deer (genus Cervus) in Argyll, Scotland. *Genetics*, *152*(1), 355–371.

Gordon, D., Huddleston, J., Chaisson, M. J. P., Hill, C. M., Kronenberg, Z. N., Munson, K. M., … Eichler, E. E. (2016). Long-read sequence assembly of the gorilla genome. *Science*, *352*(6281). https://doi.org/10.1126/science.aae0344

Grabherr, M. G., Russell, P., Meyer, M., Mauceli, E., Alföldi, J., di Palma, F., & Lindblad-Toh, K. (2010). Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics*, *26*(9), 1145–1151. https://doi.org/10.1093/bioinformatics/btq102

Grant, P. R., & Grant, B. R. (1992). Hybridization of bird species. *Science (New York, N.Y.)*, *256*(5054), 193–197. https://doi.org/10.1126/science.256.5054.193

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., … Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science (New York, N.Y.)*, *328*(5979), 710–722. https://doi.org/10.1126/science.1188021

Greene, B. (1999). Genetic variation and hybridisation of black stilts (*Himantopus novaezelandiae*) and pied stilts (*H. h. leucocephalus*), Order Charadriiformes. *New Zealand Journal of Zoology*, *26*(4), 271–277. https://doi.org/10.1080/03014223.1999.9518196

Gregory, T. R. (2001). *Animal Genome Size Database*. http://www.genomesize.com/

Greig, K., Boocock, J., Prost, S., Horsburgh, K. A., Jacomb, C., Walter, R., & Matisoo-Smith, E. (2015). Complete mitochondrial genomes of New Zealand's first dogs. *PLOS ONE*, *10*(10), e0138536. https://doi.org/10.1371/journal.pone.0138536

Grueber, C. E. (2015). Comparative genomics for biodiversity conservation. *Computational and Structural Biotechnology Journal*, *13*, 370–375. https://doi.org/10.1016/j.csbj.2015.05.003

Grueber, C. E., Wallis, G. P., & Jamieson, I. G. (2014). Episodic positive selection in the evolution of avian toll-like receptor innate immunity genes. *PLoS ONE*, *9*(3). https://doi.org/10.1371/journal.pone.0089632

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, *52*(5), 696–704. https://doi.org/10.1080/10635150390235520

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

Gurr, L. (1967). Interbreeding of *Larus novaehollandiae scopulinus* and *Larus bulleri* in the wild in New Zealand. *Ibis*, *109*(4), 552–555. https://doi.org/10.1111/j.1474-919X.1967.tb00024.x

Hagen, E. N., Hale, M. L., Maloney, R. F., & Steeves, T. E. (2011). Conservation genetic management of a critically endangered New Zealand endemic bird: Minimizing inbreeding in the Black Stilt *Himantopus novaezelandiae*. *Ibis*, *153*(3), 556–561. https://doi.org/10.1111/j.1474-919X.2011.01137.x

Hahn, C., Bachmann, L., & Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—A baiting and iterative mapping approach. *Nucleic Acids Research*, *41*(13), e129–e129. https://doi.org/10.1093/nar/gkt371

Haig, S. M., & Allendorf, F. W. (2006). Hybrids and policy. In J. M. Scott, D. D. Goble, & F. W. Davis (Eds.), *The Endangered Species Act at Thirty, Volume 2: Conserving Biodiversity in Human-Dominated Landscapes* (pp. 150–163). Island Press. https://digitalcommons.unl.edu/usgsstaffpub/708/

Hale, M. L., Burg, T. M., & Steeves, T. E. (2012). Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. *PLOS ONE*, *7*(9), e45170. https://doi.org/10.1371/journal.pone.0045170

Hänfling, B., Bolton, P., Harley, M., & Carvalho, G. R. (2005). A molecular approach to detect hybridisation between crucian carp (*Carassius carassius*) and non-indigenous carp species (*Carassius* spp. and *Cyprinus carpio*). *Freshwater Biology*, *50*(3), 403–417. https://doi.org/10.1111/j.1365-2427.2004.01330.x

Hara, Y., Yamaguchi, K., Onimaru, K., Kadota, M., Koyanagi, M., Keeley, S. D., … Kuraku, S. (2018). Shark genomes provide insights into elasmobranch evolution and the origin of vertebrates. *Nature Ecology & Evolution*, *2*(11), 1761–1771. https://doi.org/10.1038/s41559-018-0673-5

Haridas, S., Breuill, C., Bohlmann, J., & Hsiang, T. (2011). A biologist's guide to de novo genome assembly using next-generation sequence data: a test with fungal

genomes. *Journal of Microbiological Methods*, *86*(3), 368–375. https://doi.org/10.1016/j.mimet.2011.06.019

Harrisson, K. A., Pavlova, A., Goncąlves Da Silva, A., Rose, R., Bull, J. K., Lancaster, M. L., … Sunnucks, P. (2016). Scope for genetic rescue of an endangered subspecies though re-establishing natural gene flow with another subspecies. *Molecular Ecology*, *25*(6), 1242–1258. https://doi.org/10.1111/mec.13547

Hartono, S. R., Korf, I. F., & Chédin, F. (2015). GC skew is a conserved property of unmethylated CpG island promoters across vertebrates. *Nucleic Acids Research*, *43*(20), 9729–9741. https://doi.org/10.1093/nar/gkv811

Hauser, L., Baird, M., Hilborn, R., Seeb, L. W., & Seeb, J. E. (2011). An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. *Molecular Ecology Resources*, *11*(SUPPL. 1), 150–161. https://doi.org/10.1111/j.1755-0998.2010.02961.x

He, X., Johansson, M. L., & Heath, D. D. (2016). Role of genomics and transcriptomics in selection of reintroduction source populations. *Conservation Biology*, *30*(5), 1010–1018. https://doi.org/10.1111/cobi.12674

Heber, S., Varsani, A., Kuhn, S., Girg, A., Kempenaers, B., & Briskie, J. (2013). The genetic rescue of two bottlenecked South Island robin populations using translocations of inbred donors. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1752). https://doi.org/10.1098/rspb.2012.2228

Hedrick, P. W. (2013). Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, *22*(18), 4606–4618. https://doi.org/10.1111/mec.12415

Heezik, Y. van, Lei, P., Maloney, R. F., & Sancha, E. (2005). Captive breeding for reintroduction: Influence of management practices and biological factors on survival of captive kaki (black stilt). *Zoo Biology*, *24*(5), 459–474. https://doi.org/10.1002/zoo.20065

Heezik, Y. van, Maloney, R. F., & Seddon, P. J. (2009). Movements of translocated captive-bred and released Critically Endangered kaki (black stilts) *Himantopus novaezelandiae* and the value of long-term post-release monitoring. *Oryx*, *43*(4), 639–647. https://doi.org/10.1017/S0030605309990081

Heled, J., & Drummond, A. J. (2012). Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Systematic Biology*, *61*(1), 138–149. https://doi.org/10.1093/sysbio/syr087

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Hoban, S. M., Hauffe, H. C., Pérez-Espona, S., Arntzen, J. W., Bertorelle, G., Bryja, J., … Bruford, M. W. (2013). Bringing genetic diversity to the forefront of conservation policy and management. *Conservation Genetics Resources*, *5*(2), 593–598. https://doi.org/10.1007/s12686-013-9859-y

Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M., & Pääbo, S. (2001). Ancient DNA. *Nature Reviews Genetics*, *2*(5), 353–359. https://doi.org/10.1038/35072071

Hogg, C. J., Grueber, C. E., Pemberton, D., Fox, S., Lee, A. V., Ivy, J. A., & Belov, K. (2017). "Devil Tools & Tech": A Synergy of Conservation Research and Management Practice. *Conservation Letters*, *10*(1), 133–138. https://doi.org/10.1111/conl.12221

Hohenlohe, P. A., Amish, S. J., Catchen, J. M., Allendorf, F. W., & Luikart, G. (2011). Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, *11*(SUPPL. 1), 117–122. https://doi.org/10.1111/j.1755-0998.2010.02967.x

Hohenlohe, P. A., Day, M. D., Amish, S. J., Miller, M. R., Kamps-Hughes, N., Boyer, M. C., … Luikart, G. (2013). Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Molecular Ecology*, *22*(11), 3002–3013. https://doi.org/10.1111/mec.12239

Hohenlohe, P. A., Rutledge, L. Y., Waits, L. P., Andrews, K. R., Adams, J. R., Hinton, J. W., … White, B. N. (2017). Comment on "Whole-genome sequence analysis shows two endemic species of North American wolf are admixtures of the coyote and gray wolf." *Science Advances*, *3*(6), e1602250–e1602250. https://doi.org/10.1126/sciadv.1602250

Holdaway, R. N. (1995). A fossil record of the black stilt *Himantopous novaezelandiae*, Gould, 1841. *New Zealand Natural Sciences*, *22*, 69–74.

Holdaway, R. N., & Worthy, T. H. (1997). A reappraisal of the late Quaternary fossil vertebrates of Pyramid Valley Swamp, North Canterbury, New Zealand. *New*

Zealand Journal of Zoology, *24*(1), 69–121. https://doi.org/10.1080/03014223.1997.9518107

Holdaway, R. N., Worthy, T. H., & Tennyson, A. J. D. (2001). A working list of breeding bird species of the New Zealand region at first human contact. *New Zealand Journal of Zoology*, *28*(2), 119–187. https://doi.org/10.1080/03014223.2001.9518262

Holderegger, R., Balkenhol, N., Bolliger, J., Engler, J. O., Gugerli, F., Hochkirch, A., … Zachos, F. E. (2019). Conservation genetics: Linking science with practice. *Molecular Ecology*, *28*(17), 3848–3856. https://doi.org/10.1111/mec.15202

Hubbs, C. L. (1955). Hybridization between fish species in nature. *Systematic Zoology*, *4*(1), 1–20. JSTOR. https://doi.org/10.2307/2411933

Hunt, M., Kikuchi, T., Sanders, M., Newbold, C., Berriman, M., & Otto, T. D. (2013). REAPR: a universal tool for genome assembly evaluation. *Genome Biology*, *14*(5), R47–R47. https://doi.org/10.1186/gb-2013-14-5-r47

Ingvarsson, P. K. (2001). Restoration of genetic variation lost – the genetic rescue hypothesis. *Trends in Ecology & Evolution*, *16*(2), 62–63. https://doi.org/10.1016/S0169-5347(00)02065-6

International Chicken Genome Sequencing Consortium. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, *432*(7018), 695–716.

Iqbal, M. (2008). Notes on the breeding records of the White-headed Stilt in the floodplain of Ogan Komering Lebaks, South Sumatra, Indonesia. *Stilt*, *53*, 3–5.

IUCN Species Survival Commission. (2013). *Guidelines for reintroductions and other conservation translocations*. https://portals.iucn.org/library/node/10386

Ivy, J. A., & Lacy, R. C. (2012). A comparison of strategies for selecting breeding pairs to maximize genetic diversity retention in managed populations. *The Journal of Heredity*, *103*(2), 186–196. https://doi.org/10.1093/jhered/esr129

Jackman, S. D., Vandervalk, B. P., Mohamadi, H., Chu, J., Yeo, S., Hammond, S. A., … Birol, I. (2017). ABySS 2.0: Resource-efficient assembly of large genomes using a Bloom filter. *Genome Research*, *27*(5), 768–777. https://doi.org/10.1101/gr.214346.116

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*(14), 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jamieson, I. G. (2009). Loss of genetic diversity and inbreeding in New Zealand's threatened bird species. *Science for Conservation*, *293*.

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., … Zhang, G. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, *346*(6215), 1320–1331. https://doi.org/10.1126/science.1253451

Jarvis, R. M., Borrelle, S. B., Breen, B. B., & Towns, D. R. (2015). Conservation, mismatch and the research-implementation gap. *Pacific Conservation Biology*, *21*(2), 105–107. https://doi.org/10.1071/PC14912

Jarvis, R. M., Borrelle, S. B., Forsdick, N. J., Pérez-Hammerle, K.-V., Dubois, N. S., Griffin, S., … Prohaska, B. K. (in review). *Navigating spaces between conservation research and practice: Are we making progress?*

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jombart, T., & Collins, C. (2015a). *A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 2.0.0*. http://adegenet.r-forge.r-project.org/files/tutorial-dapc.pdf

Jombart, T., & Collins, C. (2015b). *Analysing genome-wide SNP data using adegenet 2.0.0*. http://adegenet.r-forge.r-project.org/files/tutorial-genomics.pdf

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 94–94. https://doi.org/10.1186/1471-2156-11-94

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Joshi, N. (2013). *Sabre—A barcode demultiplexing and trimming tool for FastQ files.* https://github.com/najoshi/sabre

Judo, M. S. B., Wedel, A. B., & Wilson, C. (1998). Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Research*, *26*(7), 1819–1825. https://doi.org/10.1093/nar/26.7.1819

Kadykalo, A. N., Cooke, S. J., & Young, N. (2020). Conservation genomics from a practitioner lens: Evaluating the research-implementation gap in a managed freshwater fishery. *Biological Conservation*, *241*, 108350. https://doi.org/10.1016/j.biocon.2019.108350

Kearns, A. M., Joseph, L., White, L. C., Austin, J. J., Baker, C., Driskell, A. C., … Omland, K. E. (2016). Norfolk Island robins are a distinct endangered species: Ancient DNA unlocks surprising relationships and phenotypic discordance within the Australo-Pacific robins. *Conservation Genetics*, *17*(2), 321–335. https://doi.org/10.1007/s10592-015-0783-4

Keedwell, R. J., Maloney, R. F., & Murray, D. P. (2002). Predator control for protecting kaki (*Himantopus novaezelandiae*)—Lessons from 20 years of management. *Biological Conservation*, *105*(3), 369–374. https://doi.org/10.1016/S0006-3207(01)00220-8

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

Kim, J., Larkin, D. M., Cai, Q., Asan, Zhang, Y., Ge, R.-L., … Ma, J. (2013). Reference-assisted chromosome assembly. *Proceedings of the National Academy of Sciences*, *110*(5), 1785–1790. https://doi.org/10.1073/pnas.1220349110

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, *40*(1), e3. https://doi.org/10.1093/nar/gkr771

Knapp, M., Clarke, A. C., Horsburgh, K. A., & Matisoo-Smith, E. A. (2012a). Setting the stage—Building and working in an ancient DNA laboratory. *Annals of Anatomy*, *194*(1), 3–6. https://doi.org/10.1016/j.aanat.2011.03.008

Knapp, M., Horsburgh, K. A., Prost, S., Stanton, J.-A., Buckley, H. R., Walter, R. K., & Matisoo-Smith, E. A. (2012b). Complete mitochondrial DNA genome sequences from the first New Zealanders. *Proceedings of the National Academy of Sciences*, *109*(45), 18350–18354. https://doi.org/10.1073/pnas.1209896109

Knight, A. T., Cowling, R. M., Rouget, M., Balmford, A., Lombard, A. T., & Campbell, B. M. (2008). Knowing but not doing: Selecting priority conservation areas and the research–implementation gap. *Conservation Biology*, *22*(3), 610–617. https://doi.org/10.1111/j.1523-1739.2008.00914.x

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, *27*(5), 722–736. https://doi.org/10.1101/gr.215087.116

Krause, J., Dear, P. H., Pollack, J. L., Slatkin, M., Spriggs, H., Barnes, I., … Hofreiter, M. (2006). Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. *Nature*, *439*(7077), 724–727. https://doi.org/10.1038/nature04432

Krings, M., Stone, A., Schmitz, R. W., Krainitzki, H., Stoneking, M., & Pääbo, S. (1997). Neandertal DNA sequences and the origin of modern humans. *Cell*, *90*(1), 19–30. https://doi.org/10.1016/S0092-8674(00)80310-4

Küpper, C., Stocks, M., Risse, J. E., dos Remedios, N., Farrell, L. L., McRae, S. B., … Burke, T. (2016). A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics*, *48*(1), 79–83. https://doi.org/10.1038/ng.3443

Lamichhaney, S., Fan, G., Widemo, F., Gunnarsson, U., Thalmann, D. S., Hoeppner, M. P., … Andersson, L. (2016). Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nature Genetics*, *48*(1), 84–88. https://doi.org/10.1038/ng.3430

Lande, R. (1993). Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *The American Naturalist*, *142*(6), 911–927. https://doi.org/10.1086/285580

Lanfear, R., Calcott, B., Ho, S. Y. W., & Guindon, S. (2012). PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, *29*(6), 1695–1701. https://doi.org/10.1093/molbev/mss020

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017). PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, *34*(3), 772–773. https://doi.org/10.1093/molbev/msw260

Larson, W. A., Seeb, L. W., Everett, M. V., Waples, R. K., Templin, W. D., & Seeb, J. E. (2014). Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolutionary Applications*, *7*(3), 355–369. https://doi.org/10.1111/eva.12128

Laurance, W. F., Koster, H., Grooten, M., Anderson, A. B., Zuidema, P. A., Zwick, S., … Anten, N. P. R. (2012). Making conservation research more relevant for conservation practitioners. *Biological Conservation*, *153*, 164–168. https://doi.org/10.1016/j.biocon.2012.05.012

Laver, T., Harrison, J., O'Neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., & Studholme, D. J. (2015). Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomolecular Detection and Quantification*, *3*, 1–8. https://doi.org/10.1016/j.bdq.2015.02.001

Lavretsky, P., DaCosta, J. M., Sorenson, M. D., McCracken, K. G., & Peters, J. L. (2019a). DdRAD-seq data reveal significant genome-wide population structure and divergent genomic regions that distinguish the mallard and close relatives in North America. *Molecular Ecology*, *28*(10), mec.15091-mec.15091. https://doi.org/10.1111/mec.15091

Lavretsky, P., Hernández-Baños, B. E., & Peters, J. L. (2014). Rapid radiation and hybridization contribute to weak differentiation and hinder phylogenetic inferences in the New World Mallard complex (*Anas* spp.). *The Auk*, *131*(4), 524–538. https://doi.org/10.1642/AUK-13-164.1

Lavretsky, P., Janzen, T., & McCracken, K. G. (2019b). Identifying hybrids & the genomics of hybridization: Mallards & American black ducks of eastern North America. *Ecology and Evolution*, *9*(6), 3470–3490. https://doi.org/10.1002/ece3.4981

Le Duc, D., Renaud, G., Krishnan, A., Almén, M. S., Huynen, L., Prohaska, S. J., … Schöneberg, T. (2015). Kiwi genome provides insights into evolution of a nocturnal lifestyle. *Genome Biology*, *16*(1), 1–15. https://doi.org/10.1186/s13059-015-0711-4

Leggett, R. M., & MacLean, D. (2014). Reference-free SNP detection: Dealing with the data deluge. *BMC Genomics*, *15*(4), 1–7. https://doi.org/10.1186/1471-2164-15-S4-S10

Leigh, D. M., Lischer, H. E. L., Grossen, C., & Keller, L. F. (2018). Batch effects in a multiyear sequencing study: False biological trends due to changes in read lengths. *Molecular Ecology Resources*, *18*(4), 778–788. https://doi.org/10.1111/1755-0998.12779

Leigh, J. W., & Bryant, D. (2015). popart: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110–1116. https://doi.org/10.1111/2041-210X.12410

Lemopoulos, A., Prokkola, J. M., Uusi-Heikkilä, S., Vasemägi, A., Huusko, A., Hyvärinen, P., … Vainikka, A. (2019). Comparing RADseq and microsatellites for estimating genetic diversity and relatedness—Implications for brown trout conservation. *Ecology and Evolution*, *9*(4), 2106–2120. https://doi.org/10.1002/ece3.4905

Leonard, J. A. (2008). Ancient DNA applications for wildlife conservation. *Molecular Ecology*, *17*(19), 4186–4196. https://doi.org/10.1111/j.1365-294X.2008.03891.x

Leonardi, M., Librado, P., Der Sarkissian, C., Schubert, M., Alfarhan, A. H., Alquraishi, S. A., … Orlando, L. (2017). Evolutionary patterns and processes: Lessons from ancient DNA. *Systematic Biology*, *66*(1), e1–e29. https://doi.org/10.1093/sysbio/syw059

Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., … Zhang, G. (2018). Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences*, *115*(17), 4325–4333. https://doi.org/10.1073/pnas.1720115115

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., & Pfister, H. (2014). UpSet: Visualization of intersecting sets. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 1983–1992. https://doi.org/10.1109/TVCG.2014.2346248

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

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., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493–496. https://doi.org/10.1038/nature10231

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., … Wang, J. (2010). The sequence and de novo assembly of the giant panda genome. *Nature*, *463*(7279), 311–317. https://doi.org/10.1038/nature08696

Lischer, H. E. L., & Shimizu, K. K. (2017). Reference-guided de novo assembly approach improves genome reconstruction for related species. *BMC Bioinformatics*, *18*(1), 474. https://doi.org/10.1186/s12859-017-1911-6

Liu, S., & Hansen, M. M. (2017). PSMC (pairwise sequentially Markovian coalescent) analysis of RAD (restriction site associated DNA) sequencing data. *Molecular Ecology Resources*, *17*(4), 631–641. https://doi.org/10.1111/1755-0998.12606

Lu, F., Lipka, A. E., Glaubitz, J., Elshire, R., Cherney, J. H., Casler, M. D., … Costich, D. E. (2013). Switchgrass genomic diversity, ploidy, and evolution: Novel insights from a network-based SNP discovery protocol. *PLoS Genetics*, *9*(1), e1003215. https://doi.org/10.1371/journal.pgen.1003215

Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore MinION sequencing and genome assembly. *Genomics, Proteomics & Bioinformatics*, *14*(5), 265–279. https://doi.org/10.1016/j.gpb.2016.05.004

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., … Wang, J. (2012). SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *GigaScience*, *1*(1), 18–18. https://doi.org/10.1186/2047-217X-1-18

Lynch, M. (1991). The genetic interpretation of inbreeding depression and outbreeding depression. *Evolution; International Journal of Organic Evolution*, *45*(3), 622–629. https://doi.org/10.1111/j.1558-5646.1991.tb04333.x

Lynch, M., Conery, J., & Burger, R. (1995). Mutational meltdowns in sexual populations. *Evolution*, *49*(6), 1067–1080. JSTOR. https://doi.org/10.2307/2410432

Ma, W., & Lambert, D. (1997). Minisatellite DNA markers reveal hybridisation between the endangered black robin and tomtit. *ELECTROPHORESIS*, *18*(9), 1682–1687. https://doi.org/10.1002/elps.1150180936

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*(5), 229–237. https://doi.org/10.1016/j.tree.2005.02.010

Mallet, J. (2007). Hybrid speciation. *Nature*, *446*(7133), 279–283. https://doi.org/10.1038/nature05706

Maloney, R. F., & Murray, D. P. (2001). *Kaki (black stilt) Recovery Plan 2001-2011*. Department of Conservation, Wellington, New Zealand.

Maloney, R. F., Rebergen, A. L., Nilsson, R. J., & Wells, N. J. (1997). Bird density and diversity in braided river beds in the Upper Waitaki Basin, South Island, New Zealand. *Notornis*, *44*, 219–232.

Marandel, F., Charrier, G., Lamy, J.-B., Cam, S. L., Lorance, P., & Trenkel, V. M. (2020). Estimating effective population size using RADseq: Effects of SNP selection and sample size. *Ecology and Evolution*, *10*(4), 1929–1937. https://doi.org/10.1002/ece3.6016

Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, *27*(6), 764–770. https://doi.org/10.1093/bioinformatics/btr011

Maricic, T., Whitten, M., & Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLOS ONE*, *5*(11), e14004. https://doi.org/10.1371/journal.pone.0014004

Marra, N. J., Stanhope, M. J., Jue, N. K., Wang, M., Sun, Q., Bitar, P. P., … Shivji, M. S. (2019). White shark genome reveals ancient elasmobranch adaptations associated with wound healing and the maintenance of genome stability. *Proceedings of the National Academy of Sciences*, *116*(10), 4446–4455. https://doi.org/10.1073/pnas.1819778116

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*, 10–12. http://dx.doi.org/10.14806/ej.17.1.200

Masta, S. E., & Boore, J. L. (2008). Parallel evolution of truncated transfer RNA genes in arachnid mitochondrial genomes. *Molecular Biology and Evolution*, *25*(5), 949–959. https://doi.org/10.1093/molbev/msn051

Matasci, N., Hung, L.-H., Yan, Z., Carpenter, E. J., Wickett, N. J., Mirarab, S., … Wong, G. K.-S. (2014). Data access for the 1,000 Plants (1KP) project. *GigaScience*, *3*(1), 17. https://doi.org/10.1186/2047-217X-3-17

Matz, M. V. (2018). Fantastic beasts and how to sequence them: Ecological genomics for obscure model organisms. *Trends in Genetics*, *34*(2), 121–132. https://doi.org/10.1016/j.tig.2017.11.002

Mays, H. L., Hung, C.-M., Shaner, P.-J., Denvir, J., Justice, M., Yang, S.-F., … Primerano, D. A. (2018). Genomic analysis of demographic history and ecological niche modeling in the endangered Sumatran rhinoceros *Dicerorhinus sumatrensis*. *Current Biology*, *28*(1), 70-76.e4. https://doi.org/10.1016/j.cub.2017.11.021

McCarthy, E. M. (2006). *Handbook of avian hybrids of the world*. Oxford University Press.

McCartney-Melstad, E., Vu, J. K., & Shaffer, H. B. (2018). Genomic data recover previously undetectable fragmentation effects in an endangered amphibian. *Molecular Ecology*, *27*(22), 4430–4443. https://doi.org/10.1111/mec.14892

McCracken, K. G., & Wilson, R. E. (2011). Gene flow and hybridization between numerically imbalanced populations of two duck species in the Falkland Islands. *PLoS ONE*, *6*(8). https://doi.org/10.1371/journal.pone.0023173

McCracken, K. G., Wilson, R. E., & Martin, A. R. (2013). Gene flow and hybridization between numerically imbalanced populations of two duck species on the subantarctic island of South Georgia. *PLOS ONE*, *8*(12), e82664. https://doi.org/10.1371/journal.pone.0082664

McFarlane, S. E., Hunter, D. C., Senn, H. V., Smith, S. L., Holland, R., Huisman, J., & Pemberton, J. M. (2020). Increased genetic marker density reveals high levels of admixture between red deer and introduced Japanese sika in Kintyre, Scotland. *Evolutionary Applications*, *13*(2), 432–441. https://doi.org/10.1111/eva.12880

McFarlane, S. E., & Pemberton, J. M. (2019). Detecting the true extent of introgression during anthropogenic hybridization. *Trends in Ecology & Evolution*, *34*(4), 315–326. https://doi.org/10.1016/j.tree.2018.12.013

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

McLennan, E. A., Wright, B. R., Belov, K., Hogg, C. J., & Grueber, C. E. (2019). Too much of a good thing? Finding the most informative genetic data set to answer conservation questions. *Molecular Ecology Resources*, *19*(3), 659–671. https://doi.org/10.1111/1755-0998.12997

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

Meyerhans, A., Vartanian, J.-P., & Wain-Hobson, S. (1990). DNA recombination during PCR. *Nucleic Acids Research*, *18*(7), 1687–1691. https://doi.org/10.1093/nar/18.7.1687

Milián-García, Y., Ramos-Targarona, R., Pérez-Fleitas, E., Sosa-Rodríguez, G., Guerra-Manchena, L., Alonso-Tabet, M., … Russello, M. A. (2015). Genetic evidence of hybridization between the critically endangered Cuban crocodile and

the American crocodile: Implications for population history and in situ/ex situ conservation. *Heredity*, *114*(3), 272–280. https://doi.org/10.1038/hdy.2014.96

Millar, C. D., Reed, C. E. M., Halverson, J. L., & Lambert, D. M. (1997). Captive management and molecular sexing of endangered avian species: An application to the black stilt *Himantopus novaezelandiae* and hybrids. *Biological Conservation*, *82*(1), 81–86. https://doi.org/10.1016/S0006-3207(96)00163-2

Miller, J. M., Poissant, J., Hogg, J. T., & Coltman, D. W. (2012). Genomic consequences of genetic rescue in an insular population of bighorn sheep (*Ovis canadensis*). *Molecular Ecology*, *21*(7), 1583–1596. https://doi.org/10.1111/j.1365-294X.2011.05427.x

Miller, M. J., Lipshutz, S. E., Smith, N. G., & Bermingham, E. (2014). Genetic and phenotypic characterization of a hybrid zone between polyandrous Northern and Wattled Jacanas in Western Panama. *BMC Evolutionary Biology*, *14*(1), 227. https://doi.org/10.1186/s12862-014-0227-7

Miller, W., Drautz, D. I., Ratan, A., Pusey, B., Qi, J., Lesk, A. M., … Schuster, S. C. (2008). Sequencing the nuclear genome of the extinct woolly mammoth. *Nature*, *456*(7220), 387–390. https://doi.org/10.1038/nature07446

Minton, C., Hassell, C., Clarke, R. H., & Jackett, N. (2017). White-headed stilt (*Himantopus leucocephalus*) now an international migratory wader? *Stilt*, *71*, 51–53.

Mischler, C., Veale, A., Van Stijn, T., Brauning, R., McEwan, J. C., Maloney, R. F., & Robertson, B. C. (2018). Population connectivity and traces of mitochondrial introgression in New Zealand black-billed gulls (*Larus bulleri*). *Genes*, *9*(11), 544. https://doi.org/10.3390/genes9110544

Morin, P. A., Foote, A. D., Baker, C. S., Hancock-Hanser, B. L., Kaschner, K., Mate, B. R., … Alexander, A. (2018). Demography or selection on linked cultural traits or genes? Investigating the driver of low mtDNA diversity in the sperm whale using complementary mitochondrial and nuclear genome analyses. *Molecular Ecology*, *27*(11), 2604–2619. https://doi.org/10.1111/mec.14698

Morin, P. A., Martien, K. K., & Taylor, B. L. (2009). Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources*, *9*(1), 66–73. https://doi.org/10.1111/j.1755-0998.2008.02392.x

Morin, P. A., Parsons, K. M., Archer, F. I., Ávila-Arcos, M. C., Barrett-Lennard, L. G., Rosa, L. D., … Foote, A. D. (2015). Geographic and temporal dynamics of a

global radiation and diversification in the killer whale. *Molecular Ecology*, *24*(15), 3964–3979. https://doi.org/10.1111/mec.13284

Morris-Pocock, J. A., Taylor, S. A., Birt, T. P., & Friesen, V. L. (2010). Concerted evolution of duplicated mitochondrial control regions in three related seabird species. *BMC Evolutionary Biology*, *10*(1), 14. https://doi.org/10.1186/1471-2148-10-14

Muñoz-Fuentes, V., Green, A. J., & Negro, J. J. (2013). Genetic studies facilitated management decisions on the invasion of the ruddy duck in Europe. *Biological Invasions*, *15*(4), 723–728. https://doi.org/10.1007/s10530-012-0331-9

Nabholz, B., Kunstner, A., Wang, R., Jarvis, E. D., & Ellegren, H. (2011). Dynamic evolution of base composition: Causes and consequences in avian phylogenomics. *Molecular Biology and Evolution*, *28*(8), 2197–2210. https://doi.org/10.1093/molbev/msr047

Nadachowska-Brzyska, K., Li, C., Smeds, L., Zhang, G., & Ellegren, H. (2015). Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. *Current Biology*, *25*(10), 1375–1380. https://doi.org/10.1016/j.cub.2015.03.047

Nam, K., Mugal, C., Nabholz, B., Schielzeth, H., Wolf, J. B., Backström, N., … Ellegren, H. (2010). Molecular evolution of genes in avian genomes. *Genome Biology*, *11*(6), R68. https://doi.org/10.1186/gb-2010-11-6-r68

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, *22*(11), 2841–2847. https://doi.org/10.1111/mec.12350

Nazareno, A. G., Bemmels, J. B., Dick, C. W., & Lohmann, L. G. (2017). Minimum sample sizes for population genomics: An empirical study from an Amazonian plant species. *Molecular Ecology Resources*, *17*(6), 1136–1147. https://doi.org/10.1111/1755-0998.12654

Nei, M., Maruyama, T., & Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution*, *29*(1), 1–10. JSTOR. https://doi.org/10.2307/2407137

Ngāi Tahu Claims Settlement Act. (1998). http://www.legislation.govt.nz/act/public/1998/0097/latest/DLM429090.html

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood

phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Nielsen, E. E. G., Bach, L., & Kotlicki, P. (2006). HYBRIDLAB (version 1.0): A program for generating simulated hybrids from population samples. *Molecular Ecology Notes*, *6*(4), 971–973. https://doi.org/10.1111/j.1471-8286.2006.01433.x

Noonan, J. P., Hofreiter, M., Smith, D., Priest, J. R., Rohland, N., Rabeder, G., … Rubin, E. M. (2005). Genomic sequencing of Pleistocene Cave Bears. *Science*, *309*(5734), 597–599. https://doi.org/10.1126/science.1113485

Nuijten, R. J. M., Bosse, M., Crooijmans, R. P. M. A., Madsen, O., Schaftenaar, W., Ryder, O. A., … Megens, H. J. (2016). The use of genomics in conservation management of the endangered Visayan warty pig (*Sus cebifrons*). *International Journal of Genomics*, *2016*. https://doi.org/10.1155/2016/5613862

Odelberg, S. J., Weiss, R. B., Hata, A., & White, R. (1995). Template-switching during DNA synthesis by *Thermus aquaticus* DNA polymerase I. *Nucleic Acids Research*, *23*(11), 2049–2057. https://doi.org/10.1093/nar/23.11.2049

Ogden, R., Gharbi, K., Mugue, N., Martinsohn, J., Senn, H., Davey, J. W., … Congiu, L. (2013). Sturgeon conservation genomics: SNP discovery and validation using RAD sequencing. *Molecular Ecology*, *22*(11), 3112–3123. https://doi.org/10.1111/mec.12234

O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., & Portnoy, D. S. (2018). These aren't the loci you're looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular Ecology*, *27*(16), 3193–3206. https://doi.org/10.1111/mec.14792

Oliehoek, P. A., Windig, J. J., van Arendonk, J. A. M., & Bijma, P. (2006). Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. *Genetics*, *173*(1), 483–496. https://doi.org/10.1534/genetics.105.049940

O'Rawe, J., Jiang, T., Sun, G., Wu, Y., Wang, W., Hu, J., … Lyon, G. J. (2013). Low concordance of multiple variant-calling pipelines: Practical implications for exome and genome sequencing. *Genome Medicine*, *5*(3). https://doi.org/10.1186/gm432

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., … Willerslev, E. (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, *499*(7456), 74–78. https://doi.org/10.1038/nature12323

Orr, H. A. (1997). Haldane's Rule. *Annual Review of Ecology and Systematics*, *28*(1), 195–218. https://doi.org/10.1146/annurev.ecolsys.28.1.195

Oswald, J. A., Harvey, M. G., Remsen, R. C., Foxworth, D. U., Dittmann, D. L., Cardiff, S. W., & Brumfield, R. T. (2019). Evolutionary dynamics of hybridization and introgression following the recent colonization of Glossy Ibis (Aves: *Plegadis falcinellus*) into the New World. *Molecular Ecology*, *28*(7), 1675–1691. https://doi.org/10.1111/mec.15008

Ottenburghs, J., Ydenberg, R. C., Hooft, P. V., Wieren, S. E. V., & Prins, H. H. T. (2015). The Avian Hybrids Project: Gathering the scientific literature on avian hybridization. *Ibis*, *157*(4), 892–894. https://doi.org/10.1111/ibi.12285

Ouborg, N. J., Pertoldi, C., Loeschcke, V., Bijlsma, R. K., & Hedrick, P. W. (2010). Conservation genetics in transition to conservation genomics. *Trends in Genetics*, *26*(4), 177–187. https://doi.org/10.1016/j.tig.2010.01.001

Overbeek, A. L., Galla, S. J., Cleland, S., Thyne, C., Maloney, R. F., & Steeves, T. E. (In Review). *Pedigree validation using genetic markers in an intensively-managed taonga species, the critically endangered kakī (Himantopus novaezelandiae)*.

Overbeek, A. L., Hauber, M. E., Brown, E., Cleland, S., Maloney, R. F., & Steeves, T. E. (2017). Evidence for brood parasitism in a critically endangered Charadriiform with implications for conservation. *Journal of Ornithology*, *158*(1), 333–337. https://doi.org/10.1007/s10336-016-1375-x

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., … Hofreiter, M. (2004). Genetic analyses from ancient DNA. *Annual Review of Genetics*, *38*, 645–679. https://doi.org/10.1146/annurev.genet.37.110801.143214

Pakendorf, B., & Stoneking, M. (2005). Mitochondrial DNA and human evolution. *Annual Review of Genomics and Human Genetics*, *6*(1), 165–183. https://doi.org/10.1146/annurev.genom.6.080604.162249

Parejo, M., Henriques, D., Pinto, M. A., Soland-Reckeweg, G., & Neuditschko, M. (2018). Empirical comparison of microsatellite and SNP markers to estimate introgression in *Apis mellifera mellifera*. *Journal of Apicultural Research*, *57*(4), 504–506. https://doi.org/10.1080/00218839.2018.1494894

Parks, M., Subramanian, S., Baroni, C., Salvatore, M. C., Zhang, G., Millar, C. D., & Lambert, D. M. (2015). Ancient population genomics and the study of evolution.

Philosophical Transactions of the Royal Society B: Biological Sciences, *370*(1660), 1–10. https://doi.org/10.1098/rstb.2013.0381

Parody Merino, Á. M. (2018). *Genetics of migration timing in bar-tailed godwits*. Massey University.

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, *25*(11), 2337–2360. https://doi.org/10.1111/mec.13557

Peek, R. A., Bedwell, M., O'Rourke, S. M., Goldberg, C., Wengert, G. M., & Miller, M. R. (2019). Hybridization between two parapatric ranid frog species in the northern Sierra Nevada, California, USA. *Molecular Ecology*, *28*(20), 4636–4647. https://doi.org/10.1111/mec.15236

Peona, V., Weissensteiner, M. H., & Suh, A. (2018). How complete are "complete" genome assemblies? - An avian perspective. *Molecular Ecology Resources*, *18*(6), 1188–1195. https://doi.org/10.1111/1755-0998.12933

Peters, J. L., Lavretsky, P., DaCosta, J. M., Bielefeld, R. R., Feddersen, J. C., & Sorenson, M. D. (2016). Population genomic data delineate conservation units in mottled ducks (*Anas fulvigula*). *Biological Conservation*, *203*, 272–281. https://doi.org/10.1016/j.biocon.2016.10.003

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, *7*(5), e37135–e37135. https://doi.org/10.1371/journal.pone.0037135

Picard Toolkit. (2019). *Picard Toolkit*. Broad Institute. http://broadinstitute.github.io/picard/

Pierce, R. J. (1984a). Plumage, morphology and hybridisation of New Zealand stilts *Himantopus* spp. *Notornis*, *31*(2), 106–130.

Pierce, R. J. (1984b). The changed distribution of stilts in New Zealand. *Notornis*, *31*(1), 7–18.

Pierce, R. J. (1986a). Foraging responses of stilts (*Himantopus* spp.: Aves) to changes in behaviour and abundance of their riverbed prey. *New Zealand Journal of Marine and Freshwater Research*, *20*(1), 17–28. https://doi.org/10.1080/00288330.1986.9516125

Pierce, R. J. (1986b). Differences in susceptibility to predation during nesting between pied and black stilts (*Himantopus* spp.). *The Auk*, *103*(2), 273–280. https://doi.org/10.1093/auk/103.2.273

Pierce, R. J. (1996). Ecology and management of the black stilt *Himantopus novaezelandiae*. *Bird Conservation International*, *6*(1), 81–88. https://doi.org/10.1017/S0959270900001325

Pierpaoli, M., Birò, Z. S., Herrmann, M., Hupe, K., Fernandes, M., Ragni, B., … Randi, E. (2003). Genetic distinction of wildcat (*Felis silvestris*) populations in Europe, and hybridization with domestic cats in Hungary. *Molecular Ecology*, *12*(10), 2585–2598. https://doi.org/10.1046/j.1365-294x.2003.01939.x

Pimm, S. L., Dollar, L., & Bass, O. L. (2006). The genetic rescue of the Florida panther. *Animal Conservation*, *9*(2), 115–122. https://doi.org/10.1111/j.1469-1795.2005.00010.x

Pop, M., Phillippy, A., Delcher, A. L., & Salzberg, S. L. (2004). Comparative genome assembly. *Briefings in Bioinformatics*, *5*(3), 237–248. https://doi.org/10.1093/bib/5.3.237

Primmer, C. R. (2009). From conservation genetics to conservation genomics. *Annals of the New York Academy of Sciences*, *1162*, 357–368. https://doi.org/10.1111/j.1749-6632.2009.04444.x

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

Prost, S., & Anderson, C. N. K. (2011). TempNet: A method to display statistical parsimony networks for heterochronous DNA sequence data. *Methods in Ecology and Evolution*, *2*(6), 663–667. https://doi.org/10.1111/j.2041-210X.2011.00129.x

Pryszcz, L. P., & Gabaldón, T. (2016). Redundans: An assembly pipeline for highly heterozygous genomes. *Nucleic Acids Research*, *44*(12), e113–e113. https://doi.org/10.1093/nar/gkw294

Puckett, E. E., & Eggert, L. S. (2016). Comparison of SNP and microsatellite genotyping panels for spatial assignment of individuals to natal range: A case study using the American black bear (*Ursus americanus*). *Biological Conservation*, *193*, 86–93. https://doi.org/10.1016/j.biocon.2015.11.020

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Puritz, J. B., Hollenbeck, C. M., & Gold, J. R. (2014). dDocent: A RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, *2*, e431. https://doi.org/10.7717/peerj.431

Qu, W.-M., Liang, N., Wu, Z.-K., Zhao, Y.-G., & Chu, D. (2019). Minimum sample sizes for invasion genomics: Empirical investigation in an invasive whitefly. *Ecology and Evolution*, *n/a*(n/a). https://doi.org/10.1002/ece3.5677

Quilodrán, C. S., Currat, M., & Montoya-Burgos, J. I. (2018). Effect of hybridization with genome exclusion on extinction risk. *Conservation Biology*, *32*(5), 1139–1149. https://doi.org/10.1111/cobi.13120

Quinzin, M. C., Sandoval-Castillo, J., Miller, J. M., Beheregaray, L. B., Russello, M. A., Hunter, E. A., … Caccone, A. (2019). Genetically informed captive breeding of hybrids of an extinct species of Galapagos tortoise. *Conservation Biology*, *33*(6), 1404–1414. https://doi.org/10.1111/cobi.13319

R Core Team. (2018). *R: A language and environment for statistical computing*. https://www.R-project.org/

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, *197*(2), 573–589. https://doi.org/10.1534/genetics.114.164350

Ralls, K., Ballou, J. D., Dudash, M. R., Eldridge, M. D. B., Fenster, C. B., Lacy, R. C., … Frankham, R. (2017). Call for a paradigm shift in the genetic management of fragmented populations. *Conservation Letters*, *11*(2), e12412. https://doi.org/10.1111/conl.12412

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Ramstad, K. M., & Dunning, L. T. (2020). *Population genomics advances and opportunities in conservation of kiwi (Apteryx spp.)*. 1–29. https://doi.org/10.1007/13836_2019_71

Randler, C. (2002). Avian hybridization, mixed pairing and female choice. *Animal Behaviour*, *63*(1), 103–119. https://doi.org/10.1006/anbe.2001.1884

Rat Genome Sequencing Project Consortium. (2004). Genome sequence of the brown Norway rat yields insights into mammalian evolution. *Nature*, *428*(6982), 493–521. https://doi.org/10.1038/nature02426

Reed, C. E. M., Butler, D., & Murray, D. P. (1993a). *Black Stilt Recovery Plan (Himantopus novaezealandiae)*. Department of Conservation, Wellington, New Zealand.

Reed, C. E. M., Nilsson, R. J., & Murray, D. P. (1993b). Cross-fostering New Zealand's black stilt. *The Journal of Wildlife Management*, *57*(3), 608–611. JSTOR. https://doi.org/10.2307/3809290

RepeatMasker v4.0.7. (1996). www.repeatmasker.org

Rexer-Huber, K., Veale, A. J., Catry, P., Cherel, Y., Dutoit, L., Foster, Y., … Robertson, B. C. (2019). Genomics detects population structure within and between ocean basins in a circumpolar seabird: The white-chinned petrel. *Molecular Ecology*, *28*(20), 4552–4572. https://doi.org/10.1111/mec.15248

Rhoads, A., & Au, K. F. (2015). PacBio sequencing and its applications. *Genomics, Proteomics & Bioinformatics*, *13*(5), 278–289. https://doi.org/10.1016/j.gpb.2015.08.002

Rhymer, J. M., & Simberloff, D. (1996). Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*, *27*(1), 83–109. https://doi.org/10.1146/annurev.ecolsys.27.1.83

Rhymer, J. M., Williams, M. J., & Braun, M. J. (1994). Mitochondrial analysis of gene flow between New Zealand mallards (*Anas platyrhynchos*) and grey ducks (*A. superciliosa*). *The Auk*, *111*(4), 970–978. JSTOR. https://doi.org/10.2307/4088829

Rick, K., Ottewell, K., Lohr, C., Thavornkanlapachai, R., Byrne, M., & Kennington, W. J. (2019). Population genomics of *Bettongia lesueur*: Admixing increases genetic diversity with no evidence of outbreeding depression. *Genes*, *10*(11), 851. https://doi.org/10.3390/genes10110851

Riley, S. P. D., Bradley Shaffer, H., Randal Voss, S., & Fitzpatrick, B. M. (2003). Hybridization between a rare, native tiger salamander (*Ambystoma californien*se) and its introduced congener. *Ecological Applications*, *13*(5), 1263–1275. https://doi.org/10.1890/02-5023

Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S. R. F., Wgs500 Consortium, … Lunter, G. (2014). Integrating mapping-, assembly-, and haplotype-based approaches for calling variants in clinical sequencing applications. *Nature Genetics*, *46*(8), 912–918. https://doi.org/10.1038/ng.3036

Robertson, H. A., Baird, K., Dowding, J. E., Elliott, G. P., Hitchmough, R. A., Miskelly, C. M., … Taylor, G. A. (2016). *Conservation status of New Zealand birds, 2016* (p. 27). Department of Conservation, Wellington, New Zealand.

Robertson, H. A., & Heather, B. D. (2015). *The hand guide to the birds of New Zealand* (3rd ed.). Penguin Random House.

Robinson, G. E., Hackett, K. J., Purcell-Miramontes, M., Brown, S. J., Evans, J. D., Goldsmith, M. R., … Schneider, D. J. (2011). Creating a buzz about insect genomes. *Science*, *331*(6023), 1386–1386. https://doi.org/10.1126/science.331.6023.1386

Rogers, R. L., Zhou, L., Chu, C., Márquez, R., Corl, A., Linderoth, T., … Nielsen, R. (2018). Genomic takeover by transposable elements in the strawberry poison frog. *Molecular Biology and Evolution*, *35*(12), 2913–2927. https://doi.org/10.1093/molbev/msy185

Rouche, G. P. L. (2003). *Birding in the United States: A demographic and economic analysis - addendum to the 2001 National survey of fishing, hunting and wildlife-associated recreation*. Division of Federal Aid, U.S. Fish and Wildlife Service.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, *34*(12), 3299–3302. https://doi.org/10.1093/molbev/msx248

Rudnick, J. A., Katzner, T. E., Bragin, E. A., Rhodes, O. E., & Dewoody, J. A. (2005). Using naturally shed feathers for individual identification, genetic parentage analyses, and population monitoring in an endangered Eastern imperial eagle (*Aquila heliaca*) population from Kazakhstan. *Molecular Ecology*, *14*(10), 2959–2967. https://doi.org/10.1111/j.1365-294X.2005.02641.x

Ruegg, K., Bay, R. A., Anderson, E. C., Saracco, J. F., Harrigan, R. J., Whitfield, M., … Smith, T. B. (2018). Ecological genomics predicts climate vulnerability in an endangered southwestern songbird. *Ecology Letters*, *21*(7), 1085–1096. https://doi.org/10.1111/ele.12977

Russell, J. C., Innes, J. G., Brown, P. H., & Byrom, A. E. (2015). Predator-Free New Zealand: Conservation country. *BioScience*, *65*(5), 520–525. https://doi.org/10.1093/biosci/biv012

Russell, T., Cullingham, C., Kommadath, A., Stothard, P., Herbst, A., & Coltman, D. (2019). Development of a novel mule deer genomic assembly and species-diagnostic SNP panel for assessing introgression in mule deer, white-tailed deer,

and their interspecific hybrids. *G3*, *9*(3), 911–919. https://doi.org/10.1534/g3.118.200838

Rutledge, L. Y., Devillard, S., Boone, J. Q., Hohenlohe, P. A., & White, B. N. (2015). RAD sequencing and genomic simulations resolve hybrid origins within North American Canis. *Biology Letters*, *11*(7), 20150303. https://doi.org/10.1098/rsbl.2015.0303

Sammler, S., Bleidorn, C., & Tiedemann, R. (2011). Full mitochondrial genome sequences of two endemic Philippine hornbill species (Aves: Bucerotidae) provide evidence for pervasive mitochondrial DNA recombination. *BMC Genomics*, *12*(1), 35. https://doi.org/10.1186/1471-2164-12-35

Sanders, M. D., & Maloney, R. F. (2002). Causes of mortality at nests of ground-nesting birds in the Upper Waitaki Basin, South Island, New Zealand: A 5-year video study. *Biological Conservation*, *106*(2), 225–236. https://doi.org/10.1016/S0006-3207(01)00248-8

Sant'Ana Sousa, L. M., Laganaro, N. M., de Camargo, C., Davanço, P. V., de Oliveira, P. R. R., Azeredo, R. M. A., … Francisco, M. R. (2013). Microsatellite markers for detecting hybrids between the extinct in the wild Alagoas Curassow (*Pauxi mitu*) and Razor-billed Curassow (*P. tuberosa*) (Aves, Galliformes). *Conservation Genetics Resources*, *5*(1), 181–183. https://doi.org/10.1007/s12686-012-9763-x

Santure, A. W., Stapley, J., Ball, A. D., Birkhead, T. R., Burke, T., & Slate, J. (2010). On the use of large marker panels to estimate inbreeding and relatedness: Empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. *Molecular Ecology*, *19*(7), 1439–1451. https://doi.org/10.1111/j.1365-294X.2010.04554.x

Saunders, A., & Norton, D. A. (2001). Ecological restoration at mainland islands in New Zealand. *Biological Conservation*, *99*(1), 109–119. https://doi.org/10.1016/S0006-3207(00)00192-0

Schaefer, N. K., Shapiro, B., & Green, R. E. (2016). Detecting hybridization using ancient DNA. *Molecular Ecology*, *25*(11), 2398–2412. https://doi.org/10.1111/mec.13556

Schibler, L., Roig, A., Mahé, M.-F., Save, J.-C., Gautier, M., Taourit, S., … Cribiu, E. P. (2004). A first generation bovine BAC-based physical map. *Genetics Selection Evolution*, *36*(1), 105. https://doi.org/10.1186/1297-9686-36-1-105

Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, *46*(8), 919–925. https://doi.org/10.1038/ng.3015

Schirtzinger, E. E., Tavares, E. S., Gonzales, L. A., Eberhard, J. R., Miyaki, C. Y., Sanchez, J. J., … Wright, T. F. (2012). Multiple independent origins of mitochondrial control region duplications in the order Psittaciformes. *Molecular Phylogenetics and Evolution*, *64*(2), 342–356. https://doi.org/10.1016/j.ympev.2012.04.009

Schubert, M., Ermini, L., Sarkissian, C. D., Jónsson, H., Ginolhac, A., Schaefer, R., … Orlando, L. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature Protocols*, *9*(5), 1056–1082. https://doi.org/10.1038/nprot.2014.063

Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Research Notes*, *9*. https://doi.org/10.1186/s13104-016-1900-2

Seabra, S. G., Silva, S. E., Nunes, V. L., Sousa, V. C., Martins, J., Marabuto, E., … Paulo, O. S. (2019). Genomic signatures of introgression between commercial and native bumblebees, *Bombus terrestris*, in western Iberian Peninsula—Implications for conservation and trade regulation. *Evolutionary Applications*, *12*(4), 679–691. https://doi.org/10.1111/eva.12732

Seabury, C. M., Bhattarai, E. K., Taylor, J. F., Viswanathan, G. G., Cooper, S. M., Davis, D. S., … Seabury, P. M. (2011). Genome-wide polymorphism and comparative analyses in the white-tailed deer (*Odocoileus virginianus*): A model for conservation genomics. *PloS One*, *6*(1), e15811. https://doi.org/10.1371/journal.pone.0015811

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, *19*(4), 198–207. https://doi.org/10.1016/j.tree.2004.01.003

Shafer, A. B. A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., & Wolf, J. B. W. (2018). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, 907–917. https://doi.org/10.1111/2041-210X.12700@10.1111/(ISSN)2041-210x.PracticalToolsFieldMethodsMEE32018

Shafer, A. B. A., Wolf, J. B. W., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., … Zieliński, P. (2015). Genomics and the challenging translation

into conservation practice. *Trends in Ecology & Evolution*, *30*(2), 78–87. https://doi.org/10.1016/j.tree.2014.11.009

Shringarpure, S. S., Bustamante, C. D., Lange, K., & Alexander, D. H. (2016). Efficient analysis of large datasets and sex bias with ADMIXTURE. *BMC Bioinformatics*, *17*, 218. https://doi.org/10.1186/s12859-016-1082-x

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Simpson, J. T. (2014). Exploring genome characteristics and sequence quality without a reference. *Bioinformatics*, *30*(9), 1228–1235. https://doi.org/10.1093/bioinformatics/btu023

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. M., & Birol, I. (2009). ABySS: A parallel assembler for short read sequence data. *Genome Research*, *19*(6), 1117–1123. https://doi.org/10.1101/gr.089532.108

Singh, T. R., Shneor, O., & Huchon, D. (2008). Bird mitochondrial gene order: Insight from three warbler mitochondrial genomes. *Molecular Biology and Evolution*, *25*(3), 475–477. https://doi.org/10.1093/molbev/msn003

Smeds, L., & Künstner, A. (2011). ConDeTri—A content dependent read trimmer for Illumina data. *PLoS ONE*, *6*(10), e26314–e26314. https://doi.org/10.1371/journal.pone.0026314

Smith, J. J., Timoshevskaya, N., Timoshevskiy, V. A., Keinath, M. C., Hardy, D., & Voss, S. R. (2019). A chromosome-scale assembly of the axolotl genome. *Genome Research*, *29*(2), 317–324. https://doi.org/10.1101/gr.241901.118

Smith, N. A. (2015). Sixteen vetted fossil calibrations for divergence dating of Charadriiformes (Aves, Neognathae). *Palaeontologia Electronica*, *18*(1), 1–18. https://doi.org/10.26879/410

Soares, A. E. R., Novak, B. J., Haile, J., Heupink, T. H., Fjeldså, J., Gilbert, M. T. P., … Shapiro, B. (2016). Complete mitochondrial genomes of living and extinct pigeons revise the timing of the columbiform radiation. *BMC Evolutionary Biology*, *16*(1), 230. https://doi.org/10.1186/s12862-016-0800-3

Steeves, T. E., Hale, M. L., & Gemmell, N. J. (2008). Development of polymorphic microsatellite markers for the New Zealand black stilt (*Himantopus novaezelandiae*) and cross-amplification in the pied stilt (*Himantopus*

himantopus leucocephalus). *Molecular Ecology Resources*, *8*(5), 1105–1107. https://doi.org/10.1111/j.1755-0998.2008.02170.x

Steeves, T. E., Maloney, R. F., Hale, M. L., Tylianakis, J. M., & Gemmell, N. J. (2010). Genetic analyses reveal hybridization but no hybrid swarm in one of the world's rarest birds. *Molecular Ecology*, *19*, 5090–5100. https://doi.org/10.1111/j.1365-294X.2010.04895.x

Stiller, M., Baryshnikov, G., Bocherens, H., Grandal d'Anglade, A., Hilpert, B., Münzel, S. C., … Knapp, M. (2010). Withering away—25,000 years of genetic decline preceded cave bear extinction. *Molecular Biology and Evolution*, *27*(5), 975–978. https://doi.org/10.1093/molbev/msq083

Sutton, J. T., Helmkampf, M., Steiner, C. C., Bellinger, M. R., Korlach, J., Hall, R., … Ryder, O. A. (2018). A high-quality, long-read de novo genome assembly to aid conservation of Hawaii's last remaining crow species. *Genes*, *9*(8), 393. https://doi.org/10.3390/genes9080393

Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, *10*(3), 512–526. https://doi.org/10.1093/oxfordjournals.molbev.a040023

Tan, M. H., Austin, C. M., Hammer, M. P., Lee, Y. P., Croft, L. J., & Gan, H. M. (2018). Finding Nemo: Hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (*Amphiprion ocellaris*) genome assembly. *GigaScience*, *7*(3). https://doi.org/10.1093/gigascience/gix137

Taylor, E. B., Boughman, J. W., Groenenboom, M., Sniatynski, M., Schluter, D., & Gow, J. L. (2006). Speciation in reverse: Morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. *Molecular Ecology*, *15*(2), 343–355. https://doi.org/10.1111/j.1365-294X.2005.02794.x

Taylor, H. R. (2015). The use and abuse of genetic marker-based estimates of relatedness and inbreeding. *Ecology and Evolution*, *5*(15), 3140–3150. https://doi.org/10.1002/ece3.1541

Taylor, H. R., Dussex, N., & van Heezik, Y. (2017). Bridging the conservation genetics gap by identifying barriers to implementation for conservation practitioners. *Global Ecology and Conservation*, *10*, 231–242. https://doi.org/10.1016/j.gecco.2017.04.001

Taylor, H. R., Kardos, M. D., Ramstad, K. M., & Allendorf, F. W. (2015). Valid estimates of individual inbreeding coefficients from marker-based pedigrees are not feasible in wild populations with low allelic diversity. *Conservation Genetics*, *16*(4), 901–913. https://doi.org/10.1007/s10592-015-0709-1

Taylor, R. H. (1975). Some ideas on speciation in New Zealand parakeets. *Notornis*, *22*(2), 110–121.

Te Manahuna Aoraki Annual Report. (2019). https://www.temanahunaaoraki.org/project/media/2019/11/Te-Manahuna-Aoraki-Annual-Report-2019.pdf

Thomas, J. E., Carvalho, G. R., Haile, J., Rawlence, N. J., Martin, M. D., Ho, S. Y., … Knapp, M. (2019). Demographic reconstruction from ancient DNA supports rapid extinction of the great auk. *ELife*, *8*, e47509. https://doi.org/10.7554/eLife.47509

Thompson, J. R., Marcelino, L. A., & Polz, M. F. (2002). Heteroduplexes in mixed-template amplifications: Formation, consequence and elimination by 'reconditioning PCR.' *Nucleic Acids Research*, *30*(9), 2083–2088. https://doi.org/10.1093/nar/30.9.2083

Thongda, W., Lewis, M., Zhao, H., Bowen, B., Lutz-Carrillo, D. J., Peoples, B. K., & Peatman, E. (2019). Species-diagnostic SNP markers for the black basses (*Micropterus* spp.): A new tool for black bass conservation and management. *Conservation Genetics Resources*. https://doi.org/10.1007/s12686-019-01109-8

Thrasher, D. J., Butcher, B. G., Campagna, L., Webster, M. S., & Lovette, I. J. (2018). Double-digest RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: A proof of concept in a highly promiscuous bird. *Molecular Ecology Resources*, *18*(5), 953–965. https://doi.org/10.1111/1755-0998.12771

Tigano, A., Sackton, T. B., & Friesen, V. L. (2018). Assembly and RNA-free annotation of highly heterozygous genomes: The case of the thick-billed murre (*Uria lomvia*). *Molecular Ecology Resources*, *18*(1), 79–90. https://doi.org/10.1111/1755-0998.12712

Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., … Rieseberg, L. H. (2016). Hybridization and extinction. *Evolutionary Applications*, *9*(7), 892–908. https://doi.org/10.1111/eva.12367

Toews, D. P. L., Taylor, S. A., Vallender, R., Brelsford, A., Butcher, B. G., Messer, P. W., & Lovette, I. J. (2016). Plumage genes and little else distinguish the

genomes of hybridizing warblers. *Current Biology*, *26*(17), 2313–2318. https://doi.org/10.1016/J.CUB.2016.06.034

Tokarska, M., Marshall, T., Kowalczyk, R., Wójcik, J. M., Pertoldi, C., Kristensen, T. N., … Bendixen, C. (2009). Effectiveness of microsatellite and SNP markers for parentage and identity analysis in species with low genetic diversity: The case of European bison. *Heredity*, *103*(4), 326–332. https://doi.org/10.1038/hdy.2009.73

Tollis, M., DeNardo, D. F., Cornelius, J. A., Dolby, G. A., Edwards, T., Henen, B. T., … Kusumi, K. (2017). The Agassiz's desert tortoise genome provides a resource for the conservation of a threatened species. *PLOS ONE*, *12*(5), e0177708. https://doi.org/10.1371/journal.pone.0177708

Toomey, A. H., Knight, A. T., & Barlow, J. (2017). *Navigating the Space between Research and Implementation in Conservation*. *10*(October), 619–625. https://doi.org/10.1111/conl.12315

Torkamaneh, D., Laroche, J., & Belzile, F. (2016). Genome-wide SNP calling from genotyping by sequencing (GBS) data: A comparison of seven pipelines and two sequencing technologies. *PLOS ONE*, *11*(8), e0161333. https://doi.org/10.1371/journal.pone.0161333

Torres, L., Welch, A. J., Zanchetta, C., Chesser, R. T., Manno, M., Donnadieu, C., … Pante, E. (2019). Evidence for a duplicated mitochondrial region in Audubon's shearwater based on MinION sequencing. *Mitochondrial DNA Part A*, *30*(2), 256–263. https://doi.org/10.1080/24701394.2018.1484116

Trewick, S. A., & Gibb, G. C. (2010). Vicars, tramps and assembly of the New Zealand avifauna: A review of molecular phylogenetic evidence. *Ibis*, *152*(2), 226–253. https://doi.org/10.1111/j.1474-919X.2010.01018.x

Vähä, J.-P., & Primmer, C. R. (2006). Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*, *15*(1), 63–72. https://doi.org/10.1111/j.1365-294X.2005.02773.x

Väli, Ü., Einarsson, A., Waits, L., & Ellegren, H. (2008). To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology*, *17*(17), 3808–3817. https://doi.org/10.1111/j.1365-294X.2008.03876.x

van Berkum, N. L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L. A., … Lander, E. S. (2010). Hi-C: A method to study the three-

dimensional architecture of genomes. *JoVE*, *39*, e1869. https://doi.org/10.3791/1869

van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics*, *30*(9), 418–426. https://doi.org/10.1016/j.tig.2014.07.001

van Heugten, R. A., Hale, R. J., Bowie, M. H., & Hale, M. L. (2017). Sleeping with the 'enemy': Hybridization of an endangered tree weta. *Conservation Genetics*, *18*(6), 1377–1387. https://doi.org/10.1007/s10592-017-0986-y

Verkuil, Y. I., Piersma, T., & Baker, A. J. (2010). A novel mitochondrial gene order in shorebirds (Scolopacidae, Charadriiformes). *Molecular Phylogenetics and Evolution*, *57*(1), 411–416. https://doi.org/10.1016/j.ympev.2010.06.010

Verry, A. J. F., Scarsbrook, L., Scofield, R. P., Tennyson, A. J. D., Weston, K. A., Robertson, B. C., & Rawlence, N. J. (2019). Who, where, what, wren? Using ancient DNA to examine the veracity of museum specimen data: A case study of the New Zealand rock wren (*Xenicus gilviventris*). *Frontiers in Ecology and Evolution*, *7*. https://doi.org/10.3389/fevo.2019.00496

Vezzi, F., Cattonaro, F., & Policriti, A. (2011). e-RGA: Enhanced Reference Guided Assembly of complex genomes. *EMBnet.Journal*, *17*(1), 46–54. https://doi.org/10.14806/ej.17.1.208

Vijay, N., Park, C., Oh, J., Jin, S., Kern, E., Kim, H. W., … Park, J.-K. (2018). Population genomic analysis reveals contrasting demographic changes of two closely related dolphin species in the last glacial. *Molecular Biology and Evolution*, *35*(8), 2026–2033. https://doi.org/10.1093/molbev/msy108

Vilà, M., Weber, E., & Antonio, C. M. D. (2000). Conservation implications of invasion by plant hybridization. *Biological Invasions*, *2*(3), 207–217. https://doi.org/10.1023/A:1010003603310

vonHoldt, B. M., Brzeski, K. E., Wilcove, D. S., & Rutledge, L. Y. (2018). Redefining the role of admixture and genomics in species conservation. *Conservation Letters*, *11*(2), e12371. https://doi.org/10.1111/conl.12371

vonHoldt, B. M., Cahill, J. A., Fan, Z., Gronau, I., Robinson, J., Pollinger, J. P., … Wayne, R. K. (2016a). Whole-genome sequence analysis shows that two endemic species of North American wolf are admixtures of the coyote and gray wolf. *Science Advances*, *2*(7), e1501714–e1501714. https://doi.org/10.1126/sciadv.1501714

vonHoldt, B. M., Kays, R., Pollinger, J. P., & Wayne, R. K. (2016b). Admixture mapping identifies introgressed genomic regions in North American canids. *Molecular Ecology*, *25*(11), 2443–2453. https://doi.org/10.1111/mec.13667

vonHoldt, B. M., Pollinger, J. P., Earl, D. A., Knowles, J. C., Boyko, A. R., Parker, H., … Wayne, R. K. (2011). A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. *Genome Research*, *21*(8), 1294–1305. https://doi.org/10.1101/gr.116301.110

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

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., … Earl, A. M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE*, *9*(11), e112963. https://doi.org/10.1371/journal.pone.0112963

Wallis, G. (1999). Genetic status of New Zealand black stilt (*Himantopus novaezelandiae*) and impact of hybridisation. *Conservation Advisory Science Notes*, *239*, 1–22.

Walther, B. A., & White, A. (2018). The emergence of birdwatching in China: History, demographics, activities, motivations, and environmental concerns of Chinese birdwatchers. *Bird Conservation International*, *28*(3), 337–349. https://doi.org/10.1017/S0959270917000557

Wandeler, P., Hoeck, P. E. A., & Keller, L. F. (2007). Back to the future: Museum specimens in population genetics. *Trends in Ecology & Evolution*, *22*(12), 634–642. https://doi.org/10.1016/j.tree.2007.08.017

Wang, B., Ekblom, R., Bunikis, I., Siitari, H., & Höglund, J. (2014). Whole genome sequencing of the black grouse (*Tetrao tetrix*): Reference guided assembly suggests faster-Z and MHC evolution. *BMC Genomics*, *15*(1), 180–180. https://doi.org/10.1186/1471-2164-15-180

Wang, B., Ekblom, R., Castoe, T. A., Jones, E. P., Kozma, R., Bongcam-Rudloff, E., … Höglund, J. (2012). Transcriptome sequencing of black grouse (*Tetrao tetrix*) for immune gene discovery and microsatellite development. *Open Biology*, *2*(4), 120054. https://doi.org/10.1098/rsob.120054

Wang, X., Maher, K. H., Zhang, N., Que, P., Zheng, C., Liu, S., … Liu, Y. (2019). Demographic histories and genome-wide patterns of divergence in incipient

species of shorebirds. *Frontiers in Genetics*, *10*. https://doi.org/10.3389/fgene.2019.00919

Warren, W. C., Hillier, L. W., Tomlinson, C., Minx, P., Kremitzki, M., Graves, T., … Cheng, H. H. (2017). A new chicken genome assembly provides insight into avian genome structure. *G3: Genes, Genomes, Genetics*, *7*(1), 109–117. https://doi.org/10.1534/g3.116.035923

Waterhouse, R. M., Seppey, M., Simao, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., … Zdobnov, E. M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, *35*(3), 543–548. https://doi.org/10.1093/molbev/msx319

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., … Lander, E. S. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, *420*(6915), 520–562. https://doi.org/10.1038/nature01262

Weinman, L. R., Solomon, J. W., & Rubenstein, D. R. (2015). A comparison of single nucleotide polymorphism and microsatellite markers for analysis of parentage and kinship in a cooperatively breeding bird. *Molecular Ecology Resources*, *15*(3), 502–511. https://doi.org/10.1111/1755-0998.12330

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370. JSTOR. https://doi.org/10.2307/2408641

Weisenfeld, N. I., Yin, S., Sharpe, T., Lau, B., Hegarty, R., Holmes, L., … Jaffe, D. B. (2014). Comprehensive variation discovery in single human genomes. *Nature Genetics*, *46*(12), 1350–1355. https://doi.org/10.1038/ng.3121

Wells, C. P., Lavretsky, P., Sorenson, M. D., Peters, J. L., DaCosta, J. M., Turnbull, S., … Engilis, A. (2019). Persistence of an endangered native duck, feral mallards, and multiple hybrid swarms across the main Hawaiian Islands. *Molecular Ecology*, *28*(24), 5203–5216. https://doi.org/10.1111/mec.15286

White, K. L., Eason, D. K., Jamieson, I. G., & Robertson, B. C. (2015). Evidence of inbreeding depression in the critically endangered parrot, the kakapo. *Animal Conservation*, *18*(4), 341–347. https://doi.org/10.1111/acv.12177

Whitehouse, A. M., & Harley, E. H. (2001). Post-bottleneck genetic diversity of elephant populations in South Africa, revealed using microsatellite analysis. *Molecular Ecology*, *10*(9), 2139–2149. https://doi.org/10.1046/j.0962-1083.2001.01356.x

Whiteley, A. R., Fitzpatrick, S. W., Funk, W. C., & Tallmon, D. A. (2015). Genetic rescue to the rescue. *Trends in Ecology & Evolution*, *30*(1), 42–49. https://doi.org/10.1016/j.tree.2014.10.009

Wick, R. R., Judd, L. M., & Holt, K. E. (2018). Deepbinner: Demultiplexing barcoded Oxford Nanopore reads with deep convolutional neural networks. *PLoS Computational Biology*, *14*(11). https://doi.org/10.1371/journal.pcbi.1006583

Wickes, C., Crouchley, D. P., & Maxwell, J. M. (2009). Takahe (*Porphyrio hochstetteri*) recovery plan 2007-2012. *Threatened Species Recovery Plan 61: Department of Conservation, Wellington, New Zealand*.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wilkinson, M., & Forsdick, N. (2017). A conglomeration of stilts: An artistic investigation of hybridity. *Junctures*, *2017*(18), 35–40.

Williams, C. L., Brust, R. C., Fendley, T. T., Tiller, G. R., & Rhodes, O. E. (2005). A comparison of hybridization between mottled ducks (*Anas fulvigula*) and mallards (*A. platyrhynchos*) in Florida and South Carolina using microsatellite DNA analysis. *Conservation Genetics*, *6*(3), 445–453. https://doi.org/10.1007/s10592-005-4978-y

Williams, M. (2017). The changing relative abundance of grey duck (*Anas superciliosa*) and mallard (*A. platyrhynchos*) in New Zealand. *Notornis*, *64*(4), 211–228.

Willing, E.-M., Dreyer, C., & Oosterhout, C. van. (2012). Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers. *PLOS ONE*, *7*(8), e42649. https://doi.org/10.1371/journal.pone.0042649

Wilmshurst, J. M., Moar, N. T., Wood, J. R., Bellingham, P. J., Findlater, A. M., Robinson, J. J., & Stone, C. (2014). Use of pollen and ancient DNA as conservation baselines for offshore islands in New Zealand. *Conservation Biology*, *28*(1), 202–212. https://doi.org/10.1111/cobi.12150

Wilson, G. H. (2001). *National distribution of braided rivers and extent of vegetation colonisation*. Landcare Research: Wellington, New Zealand.

Wolf, D. E., Takebayashi, N., & Rieseberg, L. H. (2001). Predicting the risk of extinction through hybridization. *Conservation Biology*, *15*(4), 1039–1053. https://doi.org/10.1046/j.1523-1739.2001.0150041039.x

Worthy, T. H. (1998a). A remarkable fossil and archaeological avifauna from Marfells Beach, Lake Grassmere, South Island, New Zealand. *Records of the Canterbury Museum*, *12*(1), 79–176.

Worthy, T. H. (1998b). Quaternary fossil faunas of Otago, South Island, New Zealand. *Journal of the Royal Society of New Zealand*, *28*(3), 421–521. https://doi.org/10.1080/03014223.1998.9517573

Worthy, T. H., & Holdaway, R. N. (2002). *The lost world of the moa: Prehistoric life of New Zealand*. Indiana University Press.

Wright, B., Farquharson, K. A., McLennan, E. A., Belov, K., Hogg, C. J., & Grueber, C. E. (2019). From reference genomes to population genomics: Comparing three reference-aligned reduced-representation sequencing pipelines in two wildlife species. *BMC Genomics*, *20*(1), 453–453. https://doi.org/10.1186/s12864-019-5806-y

Wright, L. I., Tregenza, T., & Hosken, D. J. (2007). Inbreeding, inbreeding depression and extinction. *Conservation Genetics*, *9*(4), 833. https://doi.org/10.1007/s10592-007-9405-0

Yan, F., Lü, J., Zhang, B., Yuan, Z., Zhao, H., Huang, S., … Che, J. (2018). The Chinese giant salamander exemplifies the hidden extinction of cryptic species. *Current Biology*, *28*(10), R590–R592. https://doi.org/10.1016/j.cub.2018.04.004

Yang, C., Wang, Q.-X., Li, X.-J., Yuan, H., Xiao, H., & Huang, Y. (2017). The mitogenomes of *Gelochelidon nilotica* and *Sterna hirundo* (Charadriiformes, Sternidae) and their phylogenetic implications. *Mitochondrial DNA Part B*, *2*(2), 601–603. https://doi.org/10.1080/23802359.2017.1372709

Yeates, D. K., Zwick, A., & Mikheyev, A. S. (2016). Museums are biobanks: Unlocking the genetic potential of the three billion specimens in the world's biological collections. *Current Opinion in Insect Science*, *18*, 83–88. https://doi.org/10.1016/j.cois.2016.09.009

Yu, L., Wang, G.-D., Ruan, J., Chen, Y.-B., Yang, C.-P., Cao, X., … Zhang, Y.-P. (2016). Genomic analysis of snub-nosed monkeys (Rhinopithecus) identifies genes and processes related to high-altitude adaptation. *Nature Genetics*, *48*(8), 947–952. https://doi.org/10.1038/ng.3615

Zdobnov, E. M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R. M., Simao, F. A., Ioannidis, P., … Kriventseva, E. V. (2017). OrthoDB v9.1: Cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal,

bacterial and viral orthologs. *Nucleic Acids Research*, *45*(D1), D744–D749. https://doi.org/10.1093/nar/gkw1119

Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, *18*(5), 821–829. https://doi.org/10.1101/gr.074492.107

Zhang, D., Song, G., Gao, B., Cheng, Y., Qu, Y., Wu, S., … Lei, F. (2017). Genomic differentiation and patterns of gene flow between two long-tailed tit species (Aegithalos). *Molecular Ecology*, *26*(23), 6654–6665. https://doi.org/10.1111/mec.14383

Zhang, G. (2015). Bird sequencing project takes off. *Nature*, *522*(7554), 34–34. https://doi.org/10.1038/522034d

Zhang, G., Li, B., Li, C., Gilbert, M. T. P., Jarvis, E. D., & Wang, J. (2014a). Comparative genomic data of the Avian Phylogenomics Project. *GigaScience*, *3*(1). https://doi.org/10.1186/2047-217X-3-26

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., … Wang, J. (2014b). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, *346*(6215), 1311–1320. https://doi.org/10.1126/science.1251385

Zhao, S., Zheng, P., Dong, S., Zhan, X., Wu, Q., Guo, X., … Wei, F. (2013). Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nature Genetics*, *45*(1), 67–71. https://doi.org/10.1038/ng.2494