Original Research Article

Genomic sequencing confirms absence of introgression despite past hybridisation in a critically endangered bird

Natalie J. Forsdick[1*], Denise Martini[1], Liz Brown[2], Richard F. Maloney[3], Tammy E. Steeves[4], Michael Knapp[1]

[1.] Department of Anatomy, University of Otago, Dunedin, New Zealand.
[2.] Department of Conservation, Twizel, New Zealand.
[3.] Department of Conservation, Christchurch, New Zealand.
[4.] School of Biological Sciences, University of Canterbury, Christchurch, New Zealand.
* Author for correspondence (natalieforsdick@gmail.com)

Running head: Genomic data finds no introgression in kakī

1 Abstract

Genetic swamping resulting from interspecific hybridisation can increase extinction risk for threatened species. The development of high-throughput and reduced-representation genomic sequencing and analyses to generate large numbers of high resolution genomic markers has the potential to reveal introgression previously undetected using small numbers of genetic markers. However, few studies to date have implemented genomic tools to assess the impacts of interspecific hybridisation in threatened species. Here we investigate the utility of genome-wide single nucleotide polymorphisms (SNPs) to detect introgression resulting from past interspecific hybridisation in one of the world's rarest birds. Anthropogenic impacts have resulted in hybridisation and subsequent backcrossing of the critically endangered Aotearoa New Zealand endemic kakī (black stilts; *Himantopus novaezelandiae*) with the non-threatened self-introduced congeneric poaka (Aotearoa New Zealand population of pied stilts, *Himantopus himantopus leucocephalus*), yet genetic analyses with a limited set of microsatellite markers revealed no evidence of introgression of poaka genetic material in kakī, excluding one individual. We use genomic data for ~63% of the wild adult kakī population to reassess the extent of introgression resulting from hybridisation between kakī and poaka. Consistent with previous genetic analyses, we detected no introgression from

poaka into kakī. These collective results indicate that, for kakī, existing microsatellite markers provide a robust, cost-effective approach to detect cryptic hybrids. Further, for well-differentiated species, the use of genomic markers may not be required to detect admixed individuals.

2 Introduction

Growing genomic evidence indicates that interspecific hybridisation (hereafter, hybridisation) has been integral in the evolutionary history of many species (e.g., North American wolves, vonHoldt et al., 2011, 2016, but also see Hohenlohe et al., 2017; Rutledge et al., 2015), challenging existing perceptions of the intrinsic value of hybrids and hybrid species, and further highlighting the complexity of conservation policy relating to them (Haig & Allendorf, 2006; Jackiw et al., 2015; Wayne & Shaffer, 2016). Nevertheless, potential tradeoffs resulting from conservation management of hybrids and hybridisation – especially recent human-induced hybridisation between threatened endemic and non-threatened non-endemic congeners – warrants careful ethical and practical consideration, balancing conservation priorities alongside the ecological, social, and economic costs-benefits of hybrids (e.g., Hamilton & Miller, 2016; Estévez et al., 2015; Schlaepfer et al., 2011). From a conservation perspective, threatened endemic species are valued over non-threatened non-endemic species due to their rarity and their known (or perceived) ecological importance, which generally leads to ethical and moral obligations to conserve thesm, especially if they are also culturally significant species (Booth et al., 2011; Courchamp et al., 2006; Maguire & Justus, 2008; Richardson & Loomis, 2009). However, the potential conservation value of hybrids should not be viewed as static, especially as altered species ranges increase the prevalence of hybrids (Chunco, 2014), and the inclusion of genomic data continues to improve our understanding of the impacts of hybridisation (vonHoldt et al., 2018).

Regardless of the perceived value of hybrids, hybridisation negatively impacts threatened species recovery through the misplaced reproductive efforts of interspecific breeding (Allendorf et al., 2001). This can reduce the reproductive outputs of species of conservation concern by demographic swamping (Allendorf et al., 2001; Wolf et al., 2001). Another potential impact of hybridisation is introgression, where subsequent backcrossing to the parental species incorporates genetic material from one species into the genome of another (Rhymer & Simberloff, 1996). Negative impacts of introgression may include outbreeding

depression, where the breakdown of coadapted gene complexes or the introduction of maladaptive traits results in the decreased fitness of hybrid offspring (Arnold, 1997; Edmands, 2007; Lynch, 1991), and, at its most extreme, may result in extinction-by-hybridisation (Allendorf et al., 2001; Fitzpatrick et al., 2010; Quilodrán et al., 2018; Rhymer & Simberloff, 1996; Riley et al., 2003; Taylor et al., 2006; Todesco et al., 2016).

Genetic tools may be 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., 2006; Ma & Lambert, 1997; Milián-García et al., 2015; Pierpaoli et al., 2003). However, to date, most conservation-relevant 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, genomic sequencing technologies have progressed, and rapidly declining costs now enable the sequencing and assembly of complete genomes for threatened non-model organisms (e.g., Li et al., 2010; Sutton et al., 2018), or the generation of thousands of genomic markers (i.e., single-nucleotide polymorphisms (SNPs)) distributed throughout the genome via reduced-representation sequencing, 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 (K. R. Andrews et al., 2016; Seabury et al., 2011; Wright et al., 2019). Large genomic marker sets are expected to provide greater power to resolve questions relating to hybridisation and introgression (similar to that observed when estimating genetic diversity and differentiation (Fischer et al., 2017) and relatedness (Galla et al., 2020)), but studies comparing genomic data with previous genetic analyses related to hybridisation in a conservation context are thus far limited (Table 1). Here we demonstrate the utility of genomic tools for assessment of the impacts of hybridisation in a critically endangered wading bird.

One of the world's rarest bird species, the Aotearoa New Zealand kakī (black stilt, *Himantopus novaezelandiae*) provides a classic example of a threatened species affected by human-induced hybridisation (BirdLife International, 2018; Robertson et al., 2016). Anthropogenic impacts resulted in population decline during the 1900s, 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 of kakī, including predator control throughout Te Manahuna and a programme of captive breeding and rearing for translocation have been integral to increase the kakī population 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). Along with predation and altered habitat availability, interspecific hybridisation may pose a threat to species recovery (Steeves et al., 2010). The Aotearoa New Zealand population of congeneric Australian pied stilts (hereafter referred to as poaka; *Himantopus himantopus leucocephalus*) self-introduced from Australia at least 200 years ago, and anthropogenic impacts facilitated expansion of the species' range (Pierce, 1984b). Limited mate choice and a male sex bias among kakī at the peak of kakī decline promoted hybridisation between kakī and poaka, producing viable, fertile hybrid offspring (Pierce, 1984a; Steeves et al., 2010).

Current kakī conservation management policies reflect the conservation value of non-admixed kakī, which are individuals with pure-black plumage that genetically assign to kakī based on a small set of mitochondrial and microsatellite markers (Maloney & Murray, 2001; Reed et al., 1993; Steeves et al., 2010). Steeves et al. (2010) confirmed the genetic distinctiveness of kakī, and found no evidence of introgression from poaka except in a single individual. This finding was attributed to reduced fitness of female hybrid offspring, combined with an ephemeral skewed sex ratio and active management to exclude hybrids (Steeves et al., 2010). However, Steeves et al. (2010) acknowledged that the small marker set used may not be representative of genome-wide variation, and thus may lack the power to detect low levels of introgression resulting from past hybridisation (see also Brumfield, 2010).

Here, we use population-level GBS and a reference-guided approach to SNP discovery for kakī, known hybrids, poaka, and Australian pied stilts. We infer the genomic extent and pattern of introgression due to hybridisation between kakī and poaka in the contemporary kakī population, as compared with that of previous genetic analyses. Determining the utility

137 of genetic and genomic markers for detection of introgressive hybridisation is essential not
138 only for the conservation of kakī, but also for detecting hybridisation in threatened species
139 more broadly.
140
141 **3 Materials and Methods**
142
143 3.1 Sample collection and DNA extraction
144
145 Following Steeves et al., 2010, individuals sampled herein were grouped by plumage
146 morphology according to the nodes defined by (Pierce, 1984a) as used by the Department of
147 Conservation's Kakī Recovery Programme. Poaka and pied stilts (plumage nodes A–C2;
148 Pierce, 1984a) were labelled 'pied', completely black node J individuals were labelled 'kakī',
149 and individuals of intermediate plumage (nodes D1–I/J) or known hybrid parentage were
150 labelled 'hybrid' (Supplementary Table 1). Extracted genomic DNA (gDNA) was available
151 for 80 stilt samples including kakī, Australian pied stilts and poaka, and hybrids for previous
152 genetic analyses.
153
154 We extracted gDNA from an additional 155 feather samples collected as part of regular
155 handling practices for kakī under Aotearoa New Zealand's Department of Conservation
156 (DOC) ethics approvals (AEC #283) at the DOC's Kakī Recovery Programme, Twizel, and
157 the Isaac Conservation and Wildlife Trust kakī captive rearing facility, Christchurch,
158 Aotearoa New Zealand. Pedigree information is recorded for all kakī individuals as part of
159 routine Kakī Recovery Programme management, extending up to seven generations. In
160 addition, blood samples were collected as part of routine health checks from two Australian
161 pied stilts at Adelaide Zoo (provided under a Royal Zoological Society of South Australia
162 Specimen Licence Agreement; Import Permit #2016061954), two node B poaka from
163 Auckland Zoo (under Auckland Zoo animal ethics approval), along with a tissue sample from
164 one poaka from Hawke's Bay, North Island (Figure 1). Poaka from the North Island were
165 preferentially sampled due to a low likelihood of recent contact with kakī, minimising the
166 chance of these individuals having recent hybrid ancestry. We extracted gDNA from all
167 samples using a Thermo Scientific™ MagJET™ Genomic DNA kit, following Protocol E
168 (manual genomic DNA purification from up to 20 mg tissue). We quantified gDNA for all
169 samples using a NanoDrop™ 8000 Spectrophotometer to assess viability for GBS. GBS
170 guidelines from our sequencing provider (AgResearch Ltd., Mosgiel, Aotearoa New Zealand)

171 recommended a total of 1 µg of DNA per sample at a concentration of 80–150 ng/µL, and
172 260/280 and 260/230 ratios of 1.80–2.00 and > 1.0 respectively.
173
174 3.2 Genotyping-by-sequencing
175
176 Among all available DNA extracts, 145 samples (130 kakī, six pied stilts, and nine hybrids)
177 had DNA of sufficient quality and quantity as per GBS recommendations. This included 74
178 males and 62 females, with nine individuals of unknown sex (Supplementary Table 1). We
179 prepared two 96-well plates of samples for GBS, containing 145 samples, two negative
180 controls (DNA-free controls) per plate, and three positive controls (replicate samples) across
181 plates. We diluted samples with gDNA concentration > 150 ng/µL to 100 ng/µL, and
182 supplied ~1 µg of DNA for all samples. GBS optimisation and sequencing was conducted by
183 AgResearch Ltd.. The optimised GBS protocol used a double-digest with restriction enzymes
184 PstI-MspI. A single library was generated for all 145 samples and controls with a fragment
185 length filter of 193–500 bp including adapter sequences. This library was sequenced on one
186 lane of Illumina HiSeq 2500 v4 sequencing for 101 cycles.
187
188 3.3 Reference-guided variant discovery
189
190 We assessed sequence quality with FastQC v0.11.5 (Andrews, 2010), and confirmed absence
191 of contamination in negative controls through BLAST searches against the nucleotide
192 database (Altschul et al., 1990). To provide accurate SNP discovery, we mapped GBS data to
193 the kakī reference genome (Galla et al., 2019). We demultiplexed and filtered raw sequences
194 with Sabre v1.0 (Joshi, 2013) and adapter trimmed with Cutadapt v1.17 (Martin, 2011). We
195 indexed genomes with BWA v0.7.17 (Li & Durbin, 2009), and mapped the processed GBS
196 reads to the genomes with BWA-MEM. We pre-processed the mapped reads for variant
197 discovery by adding read-group information with SAMtools, marking duplicates with Picard
198 v2.18.0 (Picard Toolkit, 2019), and realigning indels with the Genome Analysis Toolkit
199 (GATK) v3.5 (McKenna et al., 2010).
200
201 We compared five independent pipelines for reference-guided variant discovery. The
202 'GATK' pipeline used GATK's HaplotypeCaller and GenotypeGVCFs to call variants. The
203 'Samtools' pipeline used SAMtools v1.7 mpileup and BCFtools v1.6 variant caller (H. Li,
204 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 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.

3.4 Variant processing

By using a reference-guided approach, genomic location data was available for all variants, and so we 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*. We visualised the intersections of common variants among pipelines with the package UpSetR (Conway et al., 2017; Lex et al., 2014) implemented in R v3.5.1 (R Core Team, 2018). To improve confidence that the SNPs discovered were true SNPs rather than the result of sequencing or mapping error, we produced a single variant set 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*. To produce a set of biallelic SNPs to investigate admixture between kakī and poaka, we removed indels and multiallelic SNPs from the composite variant set using VCFtools. To confirm absence of contamination and replicability of lab processes, preliminary filtering tests and downstream analyses had retained negative and positive controls, and once confirmed, these controls were removed. We then excluded sites with > 10% missing data, a minor allele frequency < 0.01, and a minimum quality score < 20. SNPs with mean depth over all individuals between 5× and 200× were retained, and individuals with > 50% missing data across all sites were excluded (final n = 140). Due to the putative nature of the sex chromosomes in the kakī reference genome, preliminary analyses were conducted both including and excluding SNPs located on putative sex chromosomes. No difference in the extent of introgression was observed between approaches (data not shown), so we conservatively excluded SNPs located on putative sex chromosomes.

3.5 Discriminant Analysis of Principal Components

In an exploratory multivariate approach to population clustering, we conducted Discriminant Analysis of Principal Components (DAPC) with adegenet v2.1.1 (Jombart, 2008; Jombart et al., 2010; Jombart & Ahmed, 2011; Jombart & Collins, 2015b) in R v3.5.1. 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 population structuring 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, we optimised DAPC parameters using the Bayesian Information Criterion (BIC) and *a*-scores, and performed cross-validation following (Jombart & Collins, 2015a). 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 confirmed 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. 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 analysis was visualised to infer species differentiation and individual clustering.

3.6 Analysis of introgression with ADMIXTURE

To estimate individual assignment to population clusters and to detect introgression, we analysed each SNP set with a maximum likelihood method implemented in ADMIXTURE v1.3.0 (Alexander et al., 2009). To minimise stochasticity across multiple runs, we conducted 100 iterations of ADMIXTURE analysis with each SNP set for $K = 1–6$, where K represents the hypothesised number of population clusters, 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 independently of the results of DAPC analysis, allowing for differentiation between the two species (kakī and pied stilts), along with potential population

273 structuring among kakī, or differentiation between Australian pied stilts and poaka. To
274 determine the most appropriate value of *K* for each SNP set, we averaged CV error across the
275 100 iterations and visualised the results, with the lowest CV error representing the most likely
276 *K*. We visualised mean assignment probabilities (*Q*-values) across all iterations with
277 pophelper v2.3.0 (Francis, 2017) in R v3.5.1. We used pophelper for file conversion for input
278 to CLUMPP to handle label switching. Consensus *Q*-values for each individual were
279 calculated with the Greedy algorithm over 100 iterations in CLUMPP vMacOSX 1.1.2
280 (Jakobsson & Rosenberg, 2007), and we visualised the results with pophelper in R v3.5.1.
281 We manually assessed the final *Q*-matrix for all individuals using the predefined assignment
282 threshold to assign individuals as kakī.
283
284 3.7 Combining pedigree data with genomic population assignment data
285
286 Following all admixture and population clustering analyses, only six node J kakī individuals
287 were identified with < 100% kakī assignment. A small number of hybrid individuals are
288 included among the founders in the kakī pedigree (Galla et al., 2020). To determine whether
289 these assignment probabilities could be attributed to known hybrid ancestry (< 7 generations
290 deep), we used the kakī pedigree (Galla et al., 2020) to assess the ancestry of these six
291 individuals.
292
293 **4 Results**
294
295 Following a reference-guided multi-pipeline approach to SNP discovery and filtering for 145
296 stilts, a total of 140,948 SNPs were used in downstream analyses which detected no evidence
297 of introgression from poaka into kakī. Of the 250 gDNA samples available, 145 extractions
298 contained DNA of the required quantity and quality for GBS, including 66 of the 106
299 (63.2%) adults alive in the wild kakī population when this study began in 2017. GBS of the
300 pooled set of kakī, Australian pied stilts and poaka, and interspecific hybrids produced a total
301 of 303,639,199 raw sequences with length 35–101 bp and high sequence quality.
302 Demultiplexing produced an average of 2,024,530 ± SD 1,031,208.21 reads per sample
303 (Supplementary Table 2), and no samples failed to sequence. Negative controls produced a
304 low number of reads (mean = 2,585 ± SD 1554.76 reads per negative). Contamination checks
305 of negative controls produced no matches to the BLAST nucleotide database.
306

307 4.1 Variant discovery and filtering

308

309 Mapping of trimmed, filtered reads for all samples to the reference kakī genome produced an
310 average of 1,138,306.05 ± SD 597,406.95 mapped reads per individual (Table 2). This
311 represents an average of 85.4% reads per sample successfully mapped to the reference
312 genome.

313

314 Variant calling with GATK produced the fewest variants (35,441), while SAMtools produced
315 the most (488,940, Fig. 2). There were 177,437 variants common to ≥ 3 pipelines (Fig. 3).
316 Despite GATK producing the fewest variants among the five pipelines, the majority
317 (92.68%) were retained in the common variant set. SAMtools had the lowest proportion
318 (35.14%) of discovered variants retained in the common set. A total of 15,851 SNPs
319 remained after filtering. The five individuals that had produced the fewest raw sequences
320 (26,725–313,884 reads) were subsequently excluded due to high (> 50%) levels of missing
321 data. The total genotyping rate was 96.91%.

322

323 Exploratory statistics indicated an even distribution of SNPs throughout the genome, with a
324 mean density of 0.014 ± SD 0.158 SNPs/kb (Table 3). Mean depth of coverage per individual
325 was 11.663×. Kakī were well-differentiated from pied stilts with mean F_{ST} = 0.622 (Table 3).
326 Per-population summary statistics identified pied stilts as having higher diversity in terms of
327 nucleotide divergence, a greater number of variant sites, and more private alleles than either
328 kakī or hybrids (Table 4). Kakī displayed the lowest nucleotide diversity and fewest
329 polymorphic sites among the predefined groups (Table 4).

330

331 4.2 Discriminant Analysis of Principal Components

332

333 DAPC analysis identified Australian pied stilts and poaka as clustering distinctly from kakī,
334 with hybrids intermediate to the two species, though grouping more closely with kakī than
335 poaka (Fig. 3). Two node I/J hybrid individuals (DNA777 and DNA779) were found to
336 cluster with kakī.

337

338 4.3 ADMIXTURE analysis of introgression

339

340 $K = 2$ was indicated as the most likely number of clusters for ADMIXTURE analysis based
341 on CV error values, consistent with previous genetic results confirming taxonomic
342 delimitation (Steeves et al., 2010), and concordant with the results of DAPC identifying two
343 distinct clusters, and therefore only results for $K = 2$ for are reproduced here. All individuals
344 categorised as kakī (node J individuals) had assignment probabilities to the kakī cluster above
345 the pre-defined 95% threshold, and only six individuals were assigned as kakī with
346 probability < 100% (kakī mean $Q = 0.9992 \pm$ SE 0.0009; Fig. 4, Table 5, Supplementary
347 Table 3). Both Australian pied stilts and the poaka individual Poaka1 were assigned with
348 100% probability to the pied stilt cluster. Assignment probabilities to the kakī cluster for
349 hybrids ranged from 23.01% (DNA2113) to 100% (DNA777 and DNA779; hybrid mean $Q =$
350 $0.6399 \pm$ SD 0.0175; Supplementary Table 3).
351
352 4.4 Incorporating pedigree data with results of genomic analysis of introgression for
353 individuals of interest
354
355 All node J individuals were assigned as kakī with probabilities above the 95% threshold.
356 Pedigree assessment revealed only 17 kakī individuals included in this study had no recorded
357 hybrid ancestry (i.e., all of their recorded ancestors were individuals with plumage node J
358 representing completely black birds). Among the node J kakī individuals identified as having
359 95.00–99.99% probability of assignment to the kakī cluster, all individuals had at least one
360 node I or I/J individual in their recorded ancestry. Individual DNA897 had the most recent
361 hybrid ancestry, with an I/J ancestor three generations deep. Individual DNA1252 had the
362 lowest probability of assignment to the kakī cluster among all kakī (Table 5), and while this
363 individual did not have more frequent or more recent recorded hybrid ancestry than other
364 kakī here, there is no recorded ancestry for the paternal lineage. The mother of individual
365 DNA1252 was included in sequencing and analyses, and was not identified as having any
366 hybrid ancestry, suggesting a paternal lineage origin for the relatively lower probability of
367 assignment as kakī. Individual DNA1694 has a relatively deep pedigree among kaki,
368 spanning seven generations. This individual also has the most frequent incidence of recorded
369 hybrid ancestry among these individuals, with a node I/J ancestor four generations deep, three
370 further I/J ancestors six generations deep, and a node I ancestor seven generations deep. The
371 two hybrid individuals that were consistently assigned to the kakī cluster were siblings
372 DNA777 and DNA779, with a node I/J mother, and a node I/J ancestor three generations
373 deep.

5 Discussion

With the development of genomic sequencing and associated methodologies, conservation management of threatened species can now incorporate high-resolution genomic tools to more robustly explore threats to species recovery, such as anthropogenic hybridisation (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 non-native species can result in the formation of a hybrid swarm, where widespread introgression results in the loss of genetic integrity 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 Press; Overbeek et al., 2017), such marker sets have proved less robust than genomic markers for analyses 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).

5.1 Impacts of hybridisation in stilts

All 130 kakī genotyped here—representing 63% of the contemporary adult population at the time of sampling—were assigned as kakī with > 95% probability. These data confirm that the genetic integrity of kakī has been maintained despite hybridisation with poaka, and are concordant with previous genetic analysis (Steeves et al., 2010).

No individuals with pure-black plumage (node J) were identified as cryptic hybrids, with all individuals assigned to the kakī cluster with probabilities above the predefined 95% threshold. 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 I/J). No kakī individuals included in this study had recorded hybrid ancestors with plumage lighter than node I. These results indicate that with every generation of kakī backcrossing 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

408 overwhelm introgression from poaka. The current kakī recovery strategy excludes node I/J
409 birds from active management (i.e., excluding any offspring of node I/J birds from the
410 captive rearing programme) based on plumage (Steeves et al., 2010).
411
412 The finding of no introgression from poaka to kakī likely results from a combination of
413 factors. First, the management strategy enacted by the DOC's Kakī Recovery Programme to
414 maintain kakī genetic integrity has successfully minimised opportunities for hybridisation
415 between these species. This management has been responsive to the results of genetic
416 analysis that led to the exclusion of all non-node J individuals from conservation
417 management (Steeves et al., 2010). Intensive population monitoring of breeding pairs and
418 assessment of putative hybrids using the microsatellite panel has enabled practitioners to
419 break up mixed pairs (allowing kakī to re-pair with kakī) and exclude hybrids from the
420 captive breeding programme (Maloney & Murray, 2001). Ongoing kakī recovery has resulted
421 in a relatively balanced sex ratio in the wild, and combined with the strong positive
422 assortative mating of kakī, has minimised the likelihood of kakī breeding with non-kakī
423 (Steeves et al., 2010). Moderate outbreeding depression and stochastic processes have also
424 contributed to reduce the reproductive success of hybrids (Steeves et al., 2010), further
425 limiting the likelihood of introgressed material to be maintained in the population.
426
427 5.2 Implications for kakī conservation
428
429 The results presented here provide evidence that active conservation management designed to
430 minimise hybridisation can be effective in maintaining species integrity, and support the
431 ongoing management strategy of poaka and hybrid exclusion from the captive management
432 programme based on the results of genetic analysis (Steeves et al., 2010). The goals of
433 current kakī management to maintain the genetic integrity of the species are appropriate due
434 to the strong differentiation between species observed here, reduced fitness of hybrid
435 offspring, and the previous observation of moderate levels of genetic diversity among kakī
436 compared with other threatened Aotearoa New Zealand endemic birds (Steeves et al., 2010).
437 Despite the increased resolution of genomic data, when individuals with anomalous plumage
438 are observed in the captive breeding and rearing facility, microsatellite genotyping remains
439 the most cost- and time-efficient, low-complexity method for confirming species status
440 (Table 6; Overbeek et al., In Press; Overbeek et al., 2017), and this may be the case for other
441 species threatened by hybridisation. Evidence of reduced fitness of hybrid offspring (Steeves

et al., 2010) suggests the genetic integrity of kakī will continue to be valued over any potential gain in genetic diversity that may be facilitated through managed introgression (e.g., the inclusion of cryptic hybrids). Further, the lack of cryptic hybrids in the contemporary kakī population indicates maintaining the genetic integrity of kakī is possible (Steeves et al., 2010; this study).

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, the success of the conservation breeding and rearing programme to date may mean that active management of the species could be scaled back. This may see a reduction in the number of adult kakī maintained in captivity for breeding, although management of wild nests, including egg-collection, artificial incubation and captive rearing are likely to continue to maximise population growth while kakī remain critically endangered. As kakī are one of the few threatened Aotearoa 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. 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 of high priority for conservation.

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 Aotearoa New Zealand, and subsequent hybridisation with kakī prior to species decline. Kakī only occur as vagrants in the North Island of Aotearoa New Zealand, observed in very low numbers since at least the 1950s (Pierce, 1984b). Given the limited contact between kakī and poaka in the North Island in recent years, we 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 were assigned to the pied stilt cluster with < 95% probability. The only North Island individual with an 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 a 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 among Australian pied stilts (Pierce, 1984a).

5.4 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 relatively cost-effective approach to population-level genomic sequencing of non-model species, producing a set of species-discriminating SNPs. Initial development of a GBS system is markedly less expensive than development of a microsatellite panel (3000 GBP for GBS development using 94 samples in this study in 2018 compared with 5000 GBP development and testing of a microsatellite panel of approximately ten loci using 94 samples based on the estimate of Galla et al. (2016; Table 6). However, ongoing costs of the microsatellite panel per sample remain considerably lower than that of GBS (10 GBP/sample for a microsatellite panel compared with 30 GBP/sample for GBS; Table 6), and the time required from individual sampling to completion of analysis is substantially reduced. 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 access to and experience with a variety of bioinformatic tools necessary to analyse GBS data, Table 6). While the ability to more readily characterise genome-wide variation will make a genomics approach desirable for many 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 expertise) will likely maintain the conservation genomics gap for at least some species for some time yet.

For kakī, the nature of GBS as a reduced-representation approach means that despite the large increase in data compared with the eight microsatellite loci used previously, this still only represents < 1% of the 1.1 Gb kakī genome (Galla et al., 2019). Declining costs associated with whole-genome resequencing have already overtaken reduced-representation sequencing. For example, combined with the kakī pedigree, whole-genome resequencing data is being used to inform the kakī conservation breeding programme (Galla et al., 2020). Thus, should hybridisation become a significant threat to kakī recovery in the future, or should the conservation value of cryptic hybrids be reconsidered, then a comparison of GBS with whole-genome resequencing data may be useful.

6 Conclusions

Studies comparing the utility of genetic and genomic approaches for generating estimates of population genetic diversity and differentiation indicate that large SNP sets generally outperform the small microsatellite sets typically used in conservation genetic studies (Hauser et al., 2011; Hohenlohe et al., 2013; Santure et al., 2010; Weinman et al., 2015). Thus, re-examining the extent of introgression between critically endangered kakī and non-threatened congeneric poaka using a genomic approach was essential to ascertain the efficacy of conservation management aimed at maintaining the genetic integrity of kakī. While results are concordant between genetic and genomic approaches for kakī and poaka, this may not be the case for other species, particularly when hybridisation may be widespread (e.g., hybridisation between koloa maoli/Hawaiian duck (*Anas wyvilliana*) and the invasive mallard (*A. platyrhynchos*; Wells et al., 2019). Thus, we recommend that when genetic assessment has not been conducted, or there is uncertainty as to whether genetic data have adequately captured the impact of hybridisation, a genomic approach should be used. Further, we suggest that when genetic and genomic results are concordant – which we anticipate will be more likely for well-differentiated species – conservation managers can confidently continue to use genetic tools, particularly when these remain more efficient and cost-effective.

7 Data accessibility

Kakī are a taonga (treasured) species for Māori (the Indigenous people of Aotearoa New Zealand) and as such, genomic data derived from kakī are also recognised as taonga in their own right. Due to the tapu (sacred) nature of these data, 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 (tribes) and hapū (subtribes) associated with kakī. These data include raw genotyping-by-sequencing data, and the VCF comprising the unfiltered SNP set.

8 Acknowledgements

We extend our thanks to the tangata whenua who are kaitiaki for 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 research was funded by a Royal Society of New Zealand Rutherford Discovery Fellowship to MK, and NJF was supported by a University of Otago Doctoral Scholarship. We are grateful to the Aotearoa New Zealand Department of Conservation's Kakī Recovery Team, Richard Jakob-Hoff at Auckland Zoo, David McLelland at Adelaide Zoo, and John Berry for assisting with provision of kakī, Australian pied stilt, poaka, and interspecific hybrid samples for use in this project. We also thank Dr Hugh Cross for additional bioinformatics assistance, and Jana Wold for helpful feedback on an early draft of the manuscript.

9 Declaration of interest:

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

10 Funding

This research was funded by a Royal Society of New Zealand Rutherford Discovery Fellowship to MK, and NJF was supported by a University of Otago Doctoral Scholarship.

11 Role of funding

The funding sources had no role in the study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

12 Credit authorship contribution statement

Natalie J. Forsdick: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Visualization, Writing - original draft, Writing - review & editing.
Denise Martini: Conceptualization, Methodology, Writing - review & editing.
Liz Brown: Conceptualization, Resources, Writing - review & editing.
Richard F. Maloney: Conceptualization, Writing - review & editing.
Tammy E. Steeves: Conceptualization, Supervision, Writing - review & editing.
Michael Knapp: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing - review & editing.

13 References

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

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

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

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

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

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

Booth, J. E., Gaston, K. J., Evans, K. L., & Armsworth, P. R. (2011). The value of species rarity in biodiversity recreation: A birdwatching example. *Biological Conservation*, *144*(11), 2728–2732. https://doi.org/10.1016/j.biocon.2011.02.018

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

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

Chan, C. H., Ballantyne, K. N., Aikman, H., Fastier, D., Daugherty, C. H., & Chambers, G. K. (2006). 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

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

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

Combosch, D. J., Lemer, S., Ward, P. D., Landman, N. H., & Giribet, G. (2017). Genomic signatures of evolution in Nautilus—An endangered living fossil. *Molecular Ecology*, *26*(21), 5923–5938. https://doi.org/10.1111/mec.14344

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

Courchamp, F., Angulo, E., Rivalan, P., Hall, R. J., Signoret, L., Bull, L., & Meinard, Y. (2006). Rarity value and species extinction: The anthropogenic Allee effect. *PLOS Biology*, *4*(12), e415. https://doi.org/10.1371/journal.pbio.0040415

Cunha, R. L., Forsman, Z. H., Belderok, R., Knapp, I. S. S., Castilho, R., & Toonen, R. J. (2019). Rare coral under the genomic microscope: Timing and relationships among Hawaiian Montipora. *BMC Evolutionary Biology*, *19*(1), 153. https://doi.org/10.1186/s12862-019-1476-2

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

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

Der Sarkissian, C., Ermini, L., Schubert, M., Yang, M. A., Librado, P., Fumagalli, M., Jónsson, H., Bar-Gal, G. K., Albrechtsen, A., Vieira, F. G., Petersen, B., Ginolhac, A., Seguin-Orlando, A., Magnussen, K., Fages, A., Gamba, C., Lorente-Galdos, B., Polani, S., Steiner, C., … 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

Dufresnes, C., & Dubey, S. (2020). Invasion genomics supports an old hybrid swarm of pool frogs in Western Europe. *Biological Invasions*, *22*(2), 205–210. https://doi.org/10.1007/s10530-019-02112-8

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., & 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

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

Estévez, R. A., Anderson, C. B., Pizarro, J. C., & Burgman, M. A. (2015). Clarifying values, risk perceptions, and attitudes to resolve or avoid social conflicts in invasive species management. *Conservation Biology*, *29*(1), 19–30. https://doi.org/10.1111/cobi.12359

Fischer, M. C., Rellstab, C., Leuzinger, M., Roumet, M., Gugerli, F., Shimizu, K. K., Holderegger, R., & Widmer, A. (2017). Estimating genomic diversity and population differentiation – an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. *BMC Genomics*, *18*(1), 69. https://doi.org/10.1186/s12864-016-3459-7

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

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

Galla, S. J., Buckley, T. R., Elshire, R., Hale, M. L., Knapp, M., Mccallum, J., Moraga, R., Santure, A. W., Wilcox, P., & 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., Moraga, R., Santure, A. W., & 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., Richardson, A., Slater, L., Santure, A. W., & Steeves, T. E. (2020). A comparison of pedigree, genetic and genomic estimates of relatedness for informing pairing decisions in two critically endangered birds: Implications for conservation breeding

programmes worldwide. *Evolutionary Applications*. https://doi.org/10.1111/eva.12916

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

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

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

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/

Hamilton, J. A., & Miller, J. M. (2016). Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conservation Biology*, *30*(1), 33–41. https://doi.org/10.1111/cobi.12574

Hanna, Z. R., Dumbacher, J. P., Bowie, R. C. K., Henderson, J. B., & Wall, J. D. (2018). Whole-genome analysis of introgression between the Spotted Owl and Barred Owl (*Strix occidentalis* and *Strix varia*, respectively; Aves: Strigidae) in Western North America. *G3: Genes, Genomes, Genetics*, *8*(12), 3945–3952. https://doi.org/10.1534/g3.118.200754

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

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

Hohenlohe, P. A., Day, M. D., Amish, S. J., Miller, M. R., Kamps-Hughes, N., Boyer, M. C., Muhlfeld, C. C., Allendorf, F. W., Johnson, E. A., & 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., Nowak, R. M., Patterson, B. R., Wydeven, A. P., Wilson, P. A., & 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

Jackiw, R. N., Mandil, G., & Hager, H. A. (2015). A framework to guide the conservation of species hybrids based on ethical and ecological considerations. *Conservation Biology*, *29*(4), 1040–1051. https://doi.org/10.1111/cobi.12526

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

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

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

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

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

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, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., Huang, Q., Cai, Q., Li, B., Bai, Y., Zhang, Z., Zhang, Y., Wang, W., Li, J., Wei, F., Li, H., Jian, M., Li, J., Zhang, Z., … 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

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

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

Maguire, L. A., & Justus, J. (2008). Why Intrinsic Value Is a Poor Basis for Conservation Decisions. *BioScience*, *58*(10), 910–911. https://doi.org/10.1641/B581002

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

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

Martin, C. H., Crawford, J. E., Turner, B. J., & Simons, L. H. (2016). Diabolical survival in Death Valley: Recent pupfish colonization, gene flow and genetic assimilation in the

smallest species range on earth. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1823), 20152334. https://doi.org/10.1098/rspb.2015.2334

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

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

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & 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

Milián-García, Y., Ramos-Targarona, R., Pérez-Fleitas, E., Sosa-Rodríguez, G., Guerra-Manchena, L., Alonso-Tabet, M., Espinosa-López, G., & 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

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

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

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

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

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.

Pierpaoli, M., Birò, Z. S., Herrmann, M., Hupe, K., Fernandes, M., Ragni, B., Szemethy, L., & 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

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

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

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

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

Rexer-Huber, K., Veale, A. J., Catry, P., Cherel, Y., Dutoit, L., Foster, Y., McEwan, J. C., Parker, G. C., Phillips, R. A., Ryan, P. G., Stanworth, A. J., Stijn, T. van, Thompson, D. R., Waters, J., & 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

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

Richardson, L., & Loomis, J. (2009). The total economic value of threatened, endangered and rare species: An updated meta-analysis. *Ecological Economics*, *68*(5), 1535–1548. https://doi.org/10.1016/j.ecolecon.2008.10.016

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 californiense*) 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., Wilkie, A. O. M., McVean, G., & 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., McArthur, N., O'Donnell, C. F. J., Sagar, P. M., Scofield, R. P., & Taylor, G. A. (2016). *Conservation status of New Zealand birds, 2016* (p. 27). Department of Conservation, Wellington, New Zealand.

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

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

Schlaepfer, M. A., Sax, D. F., & Olden, J. D. (2011). The potential conservation value of non-native species. *Conservation Biology*, *25*(3), 428–437. https://doi.org/10.1111/j.1523-1739.2010.01646.x

Seabury, C. M., Bhattarai, E. K., Taylor, J. F., Viswanathan, G. G., Cooper, S. M., Davis, D. S., Dowd, S. E., Lockwood, M. L., & 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

Sovic, M. G., Fries, A. C., & Gibbs, H. L. (2016). Origin of a cryptic lineage in a threatened reptile through isolation and historical hybridization. *Heredity*, *117*(5), 358–366. https://doi.org/10.1038/hdy.2016.56

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

Sutton, J. T., Helmkampf, M., Steiner, C. C., Bellinger, M. R., Korlach, J., Hall, R., Baybayan, P., Muehling, J., Gu, J., Kingan, S., Masuda, B. M., & 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

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., 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

Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., Heredia, S. M., Hahn, M. A., Caseys, C., Bock, D. G., & 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., Loeschcke, V., Gregersen, V. R., & 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

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

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., Shapiro, B., Wall, J., & Wayne, R. K. (2016). 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., Pollinger, J. P., Earl, D. A., Knowles, J. C., Boyko, A. R., Parker, H., Geffen, E., Pilot, M., Jedrzejewski, W., Jedrzejewska, B., Sidorovich, V., Greco, C., Randi, E., Musiani, M., Kays, R., Bustamante, C. D., Ostrander, E. A., Novembre, J., & 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

Wayne, R. K., & Shaffer, H. B. (2016). Hybridization and endangered species protection in the molecular era. *Molecular Ecology*, *25*(11), 2680–2689. https://doi.org/10.1111/mec.13642

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

Wells, C. P., Lavretsky, P., Sorenson, M. D., Peters, J. L., DaCosta, J. M., Turnbull, S., Uyehara, K. J., Malachowski, C. P., Dugger, B. D., Eadie, J. M., & 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

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

Wright, B. R., Grueber, C. E., Lott, M. J., Belov, K., Johnson, R. N., & Hogg, C. J. (2019). Impact of reduced-representation sequencing protocols on detecting population structure in a threatened marsupial. *Molecular Biology Reports*, 1–6. https://doi.org/10.1007/s11033-019-04966-6

14 Figures and Tables

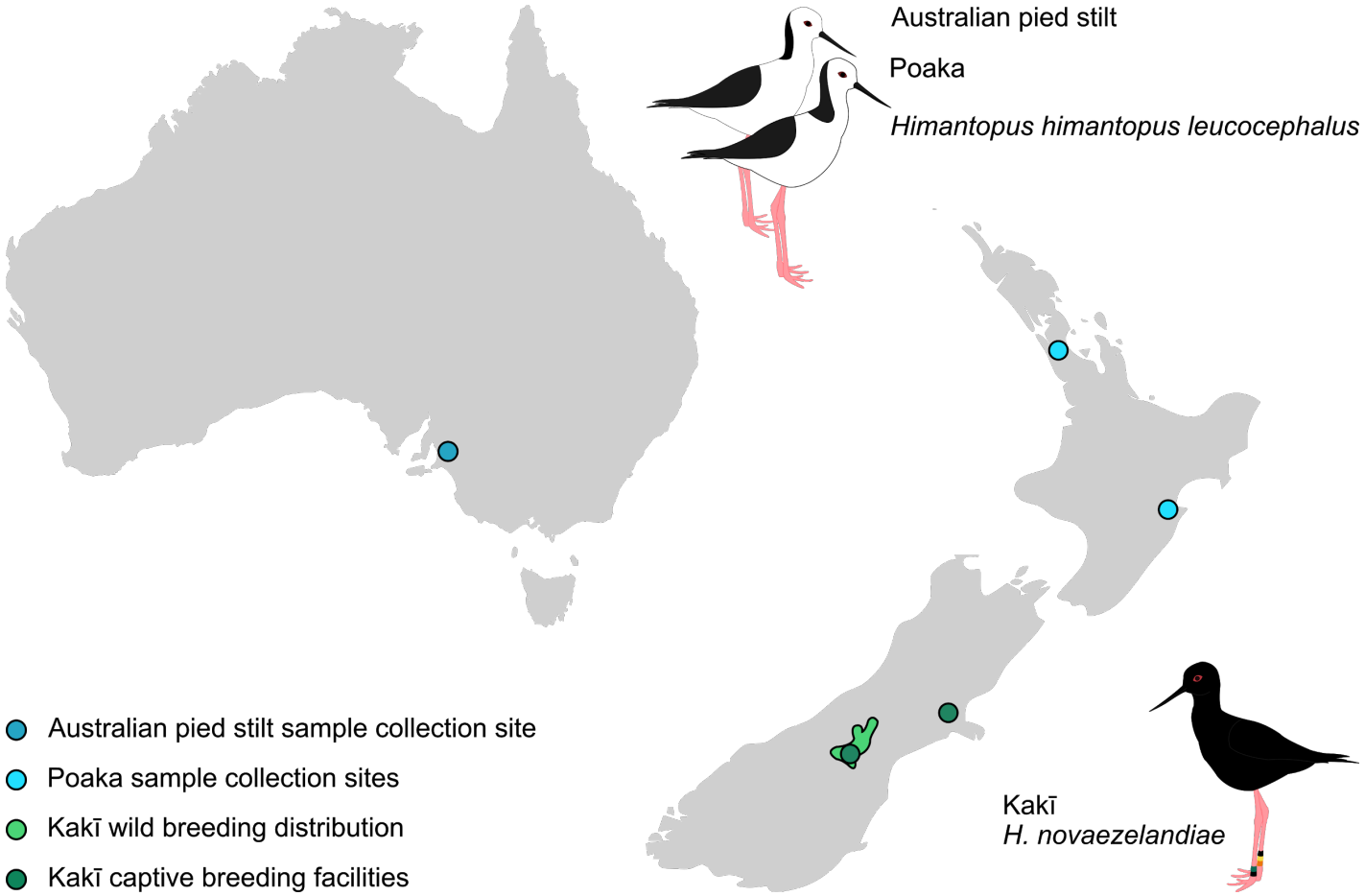


Figure 1 Locations of kakī captive breeding facilities and wild breeding distribution, and Australian pied stilt and poaka sampling sites in Australia and Aotearoa New Zealand (maps not to scale).

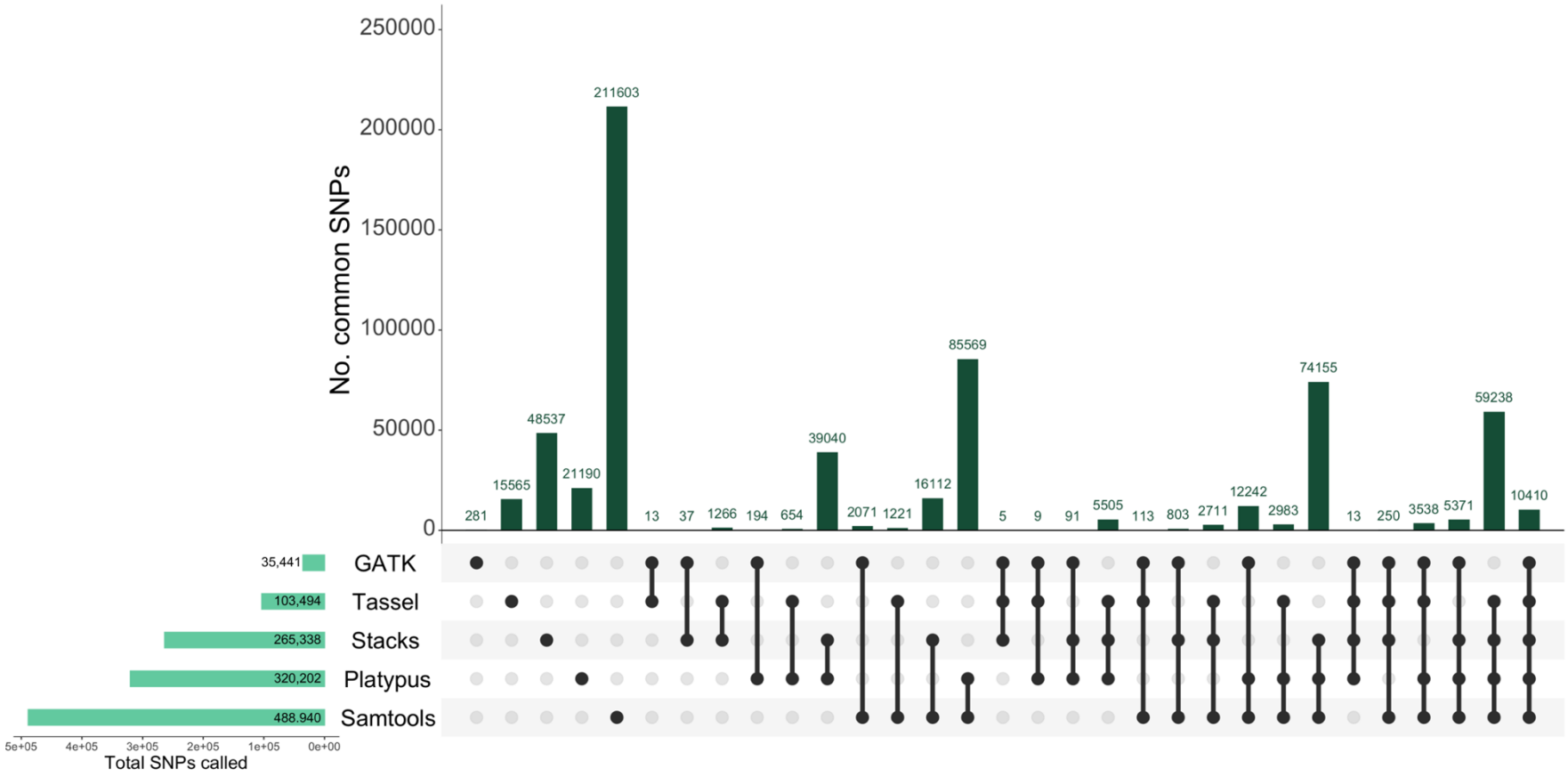


Figure 2 UpSetR plot of the intersections of the total variants discovered from GBS data for kakī, Australian pied stilts and poaka, and interspecific hybrids across five variant discovery pipelines: GATK, Platypus, Samtools, Stacks, and Tassel. Bottom left bars represent the total number of variants discovered with each pipeline, while the main bar plot represents the number of variants common to multiple pipelines as indicated by the filled circles below.

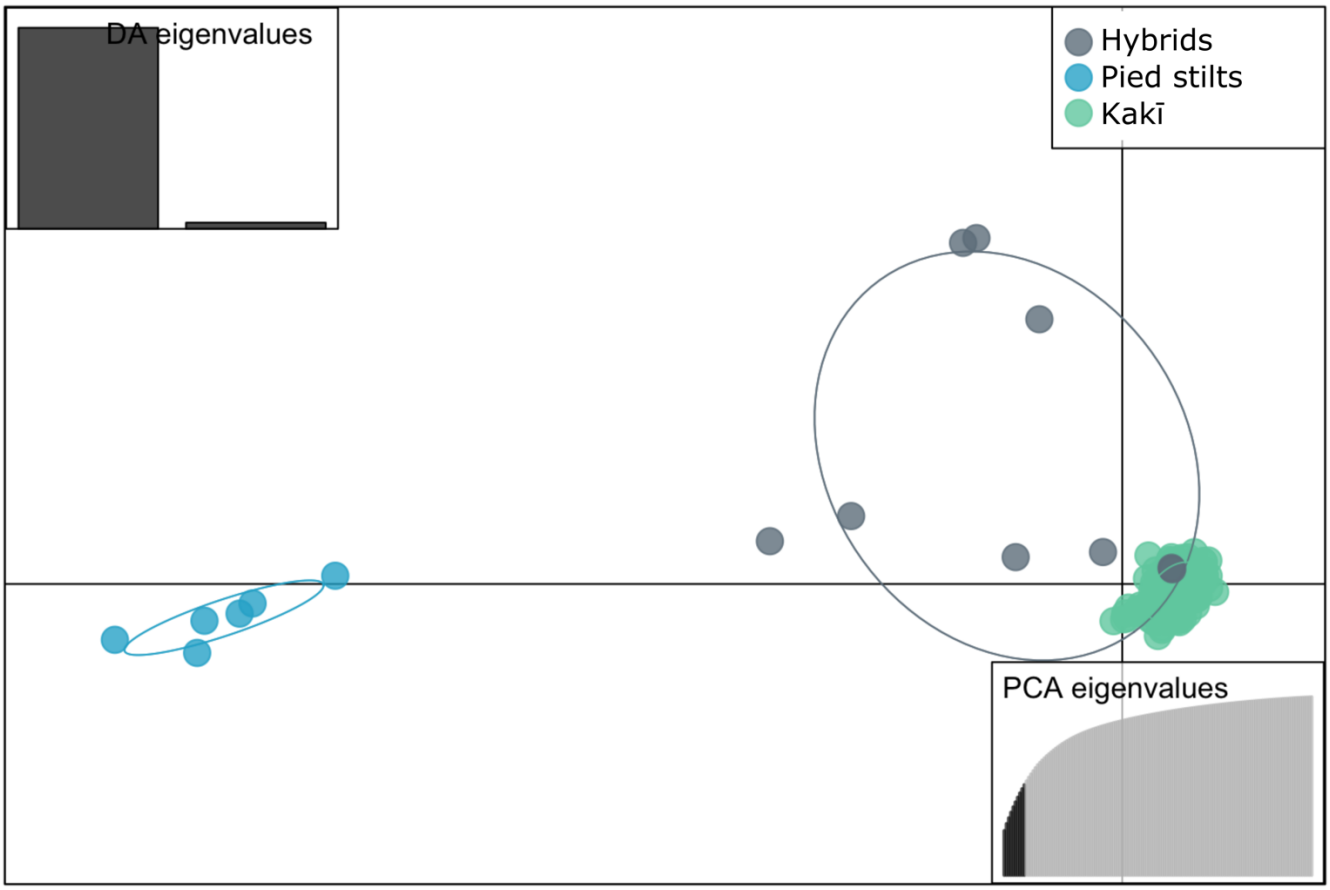


Figure 3 Scatterplot of a Discriminant Analysis of Principal Components (DAPC) produced from GBS data for kakī, Australian pied stilts and poaka, and interspecific hybrids conducted in adegenet. Individuals are coloured according to predefined population information as one of: kakī (*Himantopus novaezelandiae*), Australian pied stilts/poaka (*H. himantopus leucocephalus*), or interspecific hybrids. The closer individuals are to one another in the plot, 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 (2 discriminants, 10 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.

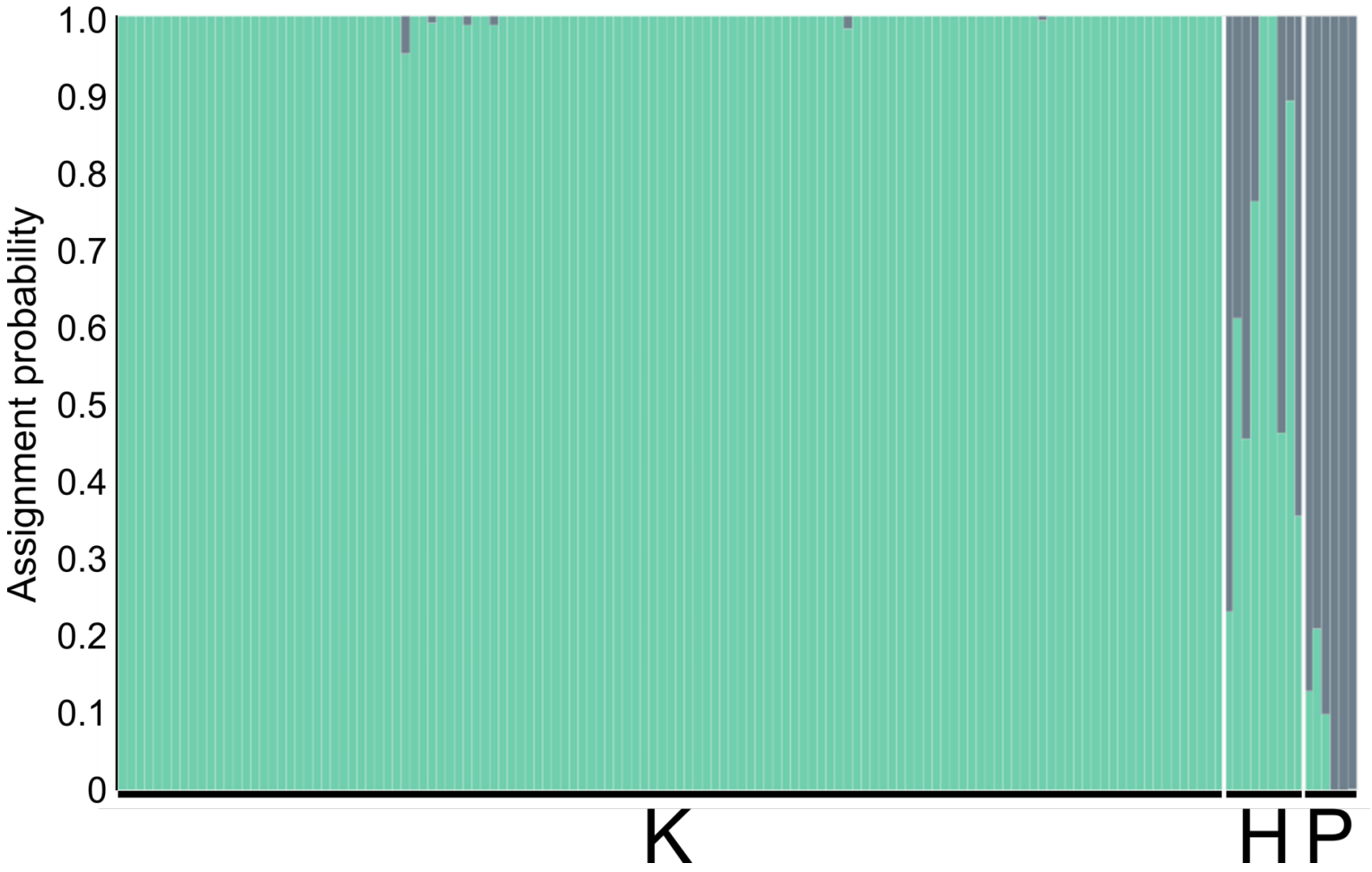


Figure 4 Assignment probabilities for kakī (K), Australian pied stilts and poaka (P), and interspecific hybrids (H) produced via pophelper visualisation of CLUMPP-permuted ADMIXTURE results when $K = 2$. Each individual is represented by a vertical bar, with colours indicating the assignment probability to the kakī (green) or Australian pied stilts/poaka (grey) cluster.

Table 1 Interspecific hybridisation research of conservation relevance where both genetic and genomic data have been generated and both have been used to assess hybrids or hybridisation. 'Conservation relevance' is used to mean that the hybridisation involves a threatened species, and the outcomes of the study have implications for conservation and/or present recommendations for conservation management. Web of Science search (19 July 2020): ALL FIELDS: (((genetic AND genomic) AND (hybridization OR introgression OR admixture) AND (conservation OR threatened OR endangered)) AND (fish OR bird OR mammal OR marsupial OR amphibian OR reptile OR plant OR invertebrate OR insect OR mollusc OR crustacean)). Of the 237 results, to capture empirical research that used high-throughput genomic sequencing techniques only, reviews and editorials, and articles published prior to 2014 were excluded. From the remaining 119 results, studies assessing intraspecific admixture or with no clear conservation relevance were excluded.

Taxa	Hybridisation relevance	Genetic data	Genomic data	Genetic and genomic data comparison	Conservation implications
Mottled duck (*Anas fulvigula*) and mallard (*A. platyrhynchos*)[1]	Species known to hybridise with subsequent backcrossing.	Microsatellite markers, introns, allozymes	ddRADseq	Genomic data provides more robust evidence of admixture with mallards.	Hybridisation creates potential for replacement of mottled ducks by mallards.
Golden-winged warbler (*Vermivora chrysoptera*) and blue-winged warbler (*V. cyanoptera*)[2]	Species morphologically differentiated, but genetically indistinguishable using genetic markers.	Allozymes, AFLPs, introns, microsatellite markers, mitochondrial markers	ddRADseq, whole-genome resequencing	Species indistinguishable based on genetic and ddRADseq data. Whole-genome resequencing data differentiates species and identifies both historical natural gene flow and recent human-mediated gene flow.	Taxonomic recommendation to retain species designations, and to use both genetic and phenotypic data to inform conservation.
Spotted owl (*Strix occidentalis*) and barred owl (*Strix varia*)[3]	Species known to hybridise with subsequent backcrossing.	Microsatellite markers, AFLPs	Whole-genome resequencing	Genetic data limited to identifying F1 and F2 hybrids. Genomic data provides no evidence of widespread introgression.	Hybridisation creates potential for replacement of the endangered spotted owl by the barred owl.
Przewalski's horse (*Equus ferus przewalskii*), and domesticated horse (*E. caballus*)[4]	Species known to hybridise with subsequent backcrossing.	Mitochondrial markers, sex chromosome markers	Whole genome resequencing	Genomic data provides more nuanced evidence of both historical natural gene flow and recent human-mediated gene flow in captivity.	Genomic data will be used to limit admixture in the Przewalski's horse conservation breeding program.

Eastern massasauga rattlesnake (*Sistrurus catenatus*) and Western massasauga rattlesnake (*S. tergeminus*)[5]	Western edge of Eastern massasauga range displays evidence of historic admixture with nearby non-threatened Western massasauga.	Nuclear markers, mitochondrial markers	RADseq	Limited population structure identified using genetic data. Genomic data provides evidence of additional population structure, including admixture attributed to historical natural gene flow.	Recommendation to categorise the admixed population as a 'Distinct Population Segment' under the Endangered Species Act (USA) due to its 'discreteness, significance, and population status'.
European pool frog (*Pelophylax lessonae*), edible frog (*P. esculentus*), Italian pool frog (*P. bergeri*)[6]	Species complex readily hybridises, including invasive *P. bergeri* with native *P. lessonae* and the hybridogenetic *P. esculentus*.	Mitochondrial markers, introns, microsatellite markers	RADseq	Genetic data identified admixed lineages but could not distinguish between hybridisation or incomplete lineage sorting. Genomic data confirms species complex as a hybrid swarm across most of the range, excluding one population of *P. lessonae*.	Genetically distinct population of *P. lessonae* may warrant conservation priotisation.
Devils Hole pupfish (*Cyprinodon diabolis*) and other *Cyprinodon* spp.[7]	Many pupfish species readily hybridise and produce viable offspring.	Mitochondrial markers, microsatellite markers	ddRADseq	Genomic data detects recent gene flow between *C. diabolis* and both *C. amargosae* and *C. pectoralis*.	Hybridisation creates potential for replacement of the Devils Hole pupfish, as has been common in the evolutionary history of the genus.
Montipora spp. (corals)[8]	Taxonomic uncertainty, including potential introgressive hybridisation.	Mitochondrial markers, nuclear markers	RADseq	Genetic data clarified taxonomy, but could not distinguish introgression, incomplete lineage sorting, or phenotypic plasticity. Genomic data further resolves taxonomy and confirms introgression, but timing difficult to resolve due to recent origin of the species complex.	Revised taxonomy will inform conservation.
Nautilus spp. (molluscs)[9]	Taxonomic uncertainty, including potential introgressive hybridisation.	Mitochondrial markers, microsatellite markers	ddRADseq	Genetic data insufficient for taxonomic delimitation. Genomic data resolves taxonomy, including evidence of recent admixture between species.	Revised taxonomy will inform conservation.

1. Peters et al., 2016; 2. Toews et al., 2016; 3. Hanna et al., 2018; 4. Der Sarkissian et al., 2015; 5. Sovic et al., 2016; 6. Dufresnes & Dubey, 2020; 7. Martin et al., 2016; 8. Cunha et al., 2019; 9. Combosch et al., 2017.

Table 2 Sequencing outputs and mapping success of GBS data from kakī, Australian pied stilts/poaka, and interspecific hybrids averaged by species. Overall includes all samples along with negative and positive controls.

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

Table 3 Summary statistics for single-nucleotide polymorphisms (SNPs) produced from GBS data for stilts. Mean ± standard deviation (SD). KB = kilobase, HWE = Hardy-Weinberg Equilibrium, FDR = False Discovery Rate, F_{ST} = measure of population differentiation, Ts/Tv = ratio of transitions to transversions.

	Stilt SNP metrics
Total SNPs	15,851
Total samples	140
Total Kakī / Pied / Hybrid individuals	125 / 6 / 9
Mean depth per SNP per individual	11.663 ± 6.107
Mean per SNP depth	1,632.77 ± 668.176
Mean SNP quality	1,248.19 ± 2447.43
Mean frequency of missing data per individual	0.031 ± 0.073
Mean SNPs/KB	0.014 ± 0.158
Total singletons/private doubletons	0
Mean nucleotide diversity; π	0.0912 ± 0.100
SNPs deviating from HWE (FDR-corrected $P \leq 0.05$)	758
Ts/Tv ratio	3.894
Weir & Cockerham mean F_{ST} (Kakī v Pied)	0.622
Weir & Cockerham weighted mean F_{ST} (Kakī v Pied)	0.637

Table 4 Population summary statistics from the three single-nucleotide polymorphism (SNP) sets produced from GBS data for kakī, Australian pied stilts/poaka, and interspecific hybrids, as calculated during format conversion from VCF to PLINK.

	Kakī	**Pied**	**Hybrid**
Mean samples per locus	121.990	5.542	8.141
Polymorphic sites	6,729	14,699	13,621
Private alleles	389	1,100	122
Mean nucleotide diversity (π)	0.057	0.396	0.235

Table 5 Mean individual assignment probabilities (*Q*-values) produced through 100 iterations of ADMIXTURE analysis for kakī that were assigned to the kakī cluster with probabilities above the 95% threshold, but below 100%.

DNA ID	**Mean assignment probability, *Q***
DNA1252	0.9514
DNA1429	0.9838
DNA1620	0.9882
DNA1694	0.9945
DNA897	0.9882
DNA932	0.9910

Table 6 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). With both platforms already established for kakī, future cost per sample is the primary deciding factor. All cost estimates are in British pounds (GBP) based on cost estimates in Aotearoa New Zealand.

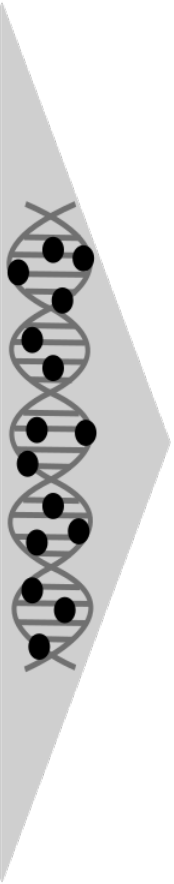
	Microsatellite panel	**Genotyping-by-sequencing**
Platform development cost	< 5,000 GBP. Includes development and screening of ~20 polymorphic loci, and genotyping of up to 94 individuals (based on a known decrease in costs since the estimate of Galla et al. (2016)).	~3,000 GBP. Includes testing of restriction 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 to SNP discovery.
Platform development time	3–4 months	3–4 months
Cost per sample for additional samples	10 GBP. Includes DNA extraction and quantification, and microsatellite genotyping at eight optimised loci. Excludes associated person-hours.	30 GBP. Includes DNA extraction and quantification, and GBS. Excludes associated person-hours.
Lab time for additional samples	1 week	1–3 months
Analysis time	1 week	4 weeks

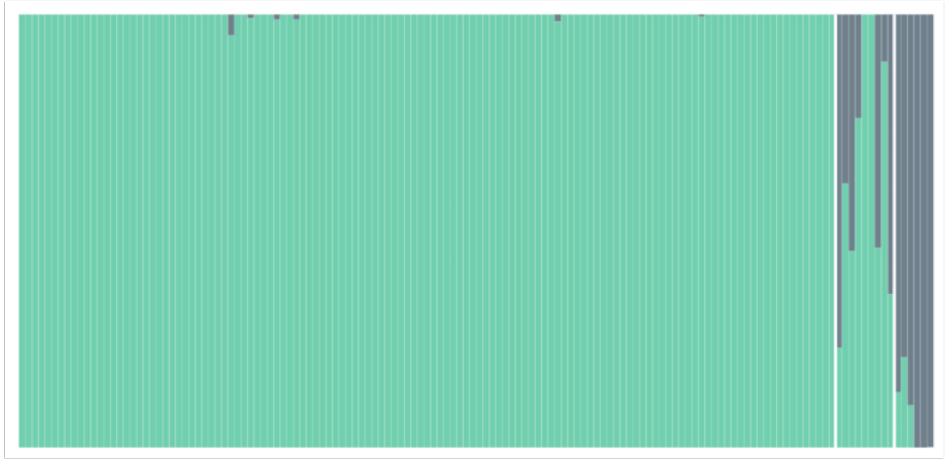
Analysis requirements	Access to a standard desktop computer. Access to allele-calling software (e.g., GeneMarker v2.2). Access to and experience with genetic population clustering tools optional.	Access to a high-capacity computing system or capable computing cluster. Access to and experience with a variety of population genomic and 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. Potential for these data to have additional application in the future with genomic analysis developments.
Limitations	Caveats associated with how the microsatellite loci are generated, and what they were 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. Potential for 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 different sequencing batches (Leigh et al., 2018). Analyses limited by software and models available in a developing field.

X

Are there cryptic hybrids among kakī?

?





We found no evidence of cryptic hybrids among kakī.