

Article

# Reference Genomes from Distantly Related Species Can Be Used for Discovery of Single Nucleotide Polymorphisms to Inform Conservation Management

Stephanie J. Galla <sup>1,\*</sup>, Natalie J. Forsdick <sup>2,\*</sup> , Liz Brown <sup>3</sup>, Marc P. Hoepfner <sup>4</sup>, Michael Knapp <sup>2</sup>, Richard F. Maloney <sup>5</sup> , Roger Moraga <sup>6</sup>, Anna W. Santure <sup>7</sup>  and Tammy E. Steeves <sup>1</sup>

<sup>1</sup> School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand; tammy.steeves@canterbury.ac.nz

<sup>2</sup> Department of Anatomy, University of Otago, Dunedin 9054, New Zealand; michael.knapp@otago.ac.nz

<sup>3</sup> Te Manahuna, Department of Conservation, Twizel 7901, New Zealand; lbrown@doc.govt.nz

<sup>4</sup> Institute of Clinical Molecular Biology, Kiel University, 24105 Kiel, Germany; m.hoepfner@ikmb.uni-kiel.de

<sup>5</sup> Science and Policy Group, Department of Conservation, Christchurch 8011, New Zealand; rmaloney@doc.govt.nz

<sup>6</sup> Tea Break Bioinformatics Ltd., Palmerston North 4144, New Zealand; r.moraga@protonmail.com

<sup>7</sup> School of Biological Sciences, University of Auckland, Auckland 1142, New Zealand; a.santure@auckland.ac.nz

\* Correspondence: stephanie.galla@pg.canterbury.ac.nz (S.J.G.); natalie.forsdick@postgrad.otago.ac.nz (N.J.F.)

Received: 8 November 2018; Accepted: 19 December 2018; Published: 22 December 2018



**Abstract:** Threatened species recovery programmes benefit from incorporating genomic data into conservation management strategies to enhance species recovery. However, a lack of readily available genomic resources, including conspecific reference genomes, often limits the inclusion of genomic data. Here, we investigate the utility of closely related high-quality reference genomes for single nucleotide polymorphism (SNP) discovery using the critically endangered kakī/black stilt (*Himantopus novaezelandiae*) and four Charadriiform reference genomes as proof of concept. We compare diversity estimates (i.e., nucleotide diversity, individual heterozygosity, and relatedness) based on kakī SNPs discovered from genotyping-by-sequencing and whole genome resequencing reads mapped to conordinal (killdeer, *Charadrius vociferus*), confamilial (pied avocet, *Recurvirostra avocetta*), congeneric (pied stilt, *Himantopus himantopus*) and conspecific reference genomes. Results indicate that diversity estimates calculated from SNPs discovered using closely related reference genomes correlate significantly with estimates calculated from SNPs discovered using a conspecific genome. Congeneric and confamilial references provide higher correlations and more similar measures of nucleotide diversity, individual heterozygosity, and relatedness. While conspecific genomes may be necessary to address other questions in conservation, SNP discovery using high-quality reference genomes of closely related species is a cost-effective approach for estimating diversity measures in threatened species.

**Keywords:** conservation genomics; conservation genomics gap; SNP discovery; B10K; threatened species; birds

## 1. Introduction

The field of conservation genetics is in transition from using relatively few genetic markers (e.g., microsatellites, mitochondrial sequences) to using thousands of genome-wide single nucleotide polymorphisms (SNPs) discovered with high-throughput sequencing technologies (HTS) to inform conservation management of threatened species. In addition to providing greater resolution for

diversity estimates (e.g., nucleotide diversity, heterozygosity, relatedness [1]), these new genomic approaches provide an opportunity to tackle new questions regarding regions of the genome that underlie fitness-related traits (i.e., adaptive variation [2–4]). While the promise of a conservation genomic approach has been heralded for well over a decade [5], the uptake of these technologies by conservation management has been slow [6,7].

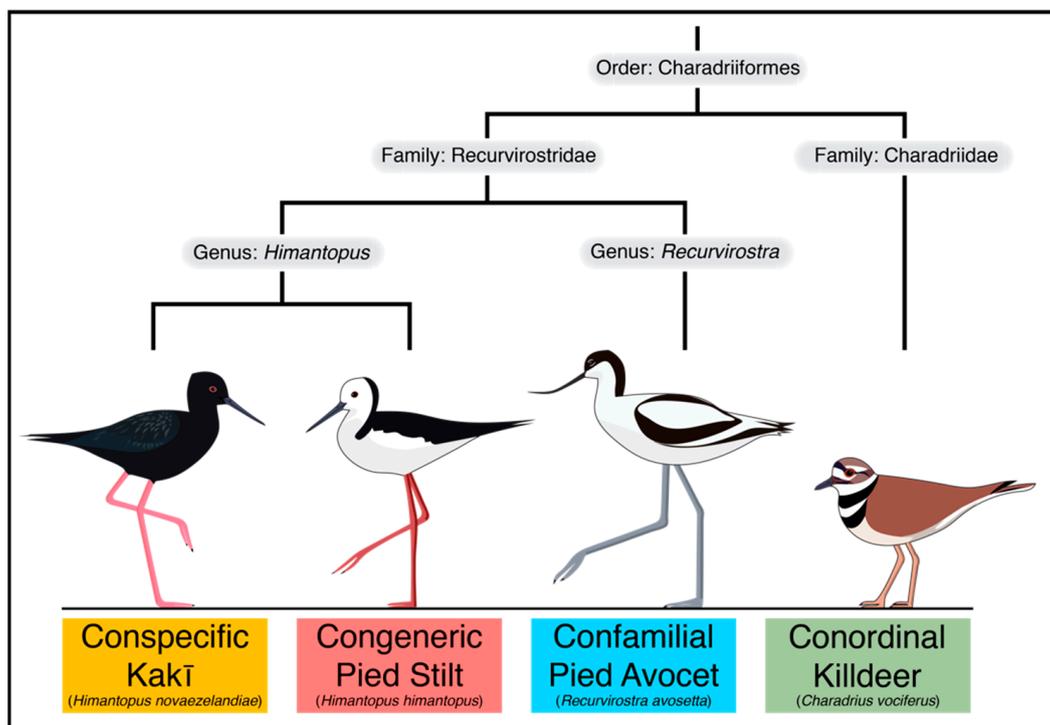
This time lag between technology availability and implementation (also termed the ‘conservation genomics gap’ [7]) may be caused by several interconnected issues, including a disconnect between conservation genetic researchers and practitioners [8,9], the time it takes for geneticists to upskill in bioinformatic expertise [6,7,10], and initial expense for HTS sequence production and generation of genomic resources (e.g., a high-quality reference genome). With that said, sequencing costs are dropping precipitously [11] (but see also [12]) and affordable reduced representation genomic approaches provide the ability to produce high-density marker sets, even in the absence of a reference genome (i.e., de novo marker discovery [13]). While it is possible to discover SNPs de novo, reference-guided approaches to SNP discovery offer many advantages, including enhanced computational efficiency, improved accuracy at low sequencing depth, higher confidence in identifying sequence contamination, greater ability to identify the location of SNPs, improved performance in determining linkage disequilibrium between SNPs, and greater ability to identify differences between paralogous and repetitive sequences from true SNP variants [14–17]. Reference genomes also allow for identifying variants in annotated gene regions, which is necessary for identifying adaptive variation [14]. While reference genomes are preferred for conservation genomic research, they are often unavailable for threatened species or out of reach for resource-constrained conservation projects (e.g., [18]).

There has been an exponential increase in the number of available eukaryotic genomes for non-model species that may be used as a reference [19], including the outputs from various genome consortiums (e.g., Genome 10K [20]; Bird 10,000 Genomes Project (B10K) [21]; 5000 Insect Genome Project (i5K) [22]; 1000 Plants Project (1KP) [23]; Oz Mammalian Genomics [24]; Earth BioGenome Project [25]). Readily available conspecific reference genomes for threatened species will likely enable faster uptake of a conservation genomics approach, for example, by avoiding the time and expenditure of sequencing and assembling a high-quality genome de novo. However, in many instances, the next best available resource may be a genome from a closely related species. There has been discussion on the utility of closely related reference genomes for reference-guided genome assembly (i.e., [26,27]). Additionally, there are many research studies to date that have used closely related reference genomes for SNP discovery using reduced-representation and whole genome resequencing (hereafter, resequencing) approaches (e.g., [28–31]).

Birds offer an exceptional opportunity to study the utility of SNP discovery using closely related reference genomes to inform conservation management. In comparison with other vertebrates, bird genomes are relatively small (~0.93–1.3 Gb), compact (i.e., low repetitive elements), and conserved between species [32,33]. Also, the availability of bird reference genomes has increased, due in part to the efforts of individual research groups that produce genomes to answer questions regarding primary production (e.g., chicken, *Gallus gallus* [34]; the turkey, *Meleagris gallopavo* [35]), evolution (e.g., zebra finch, *Taeniopygia guttata*, [36]; Galapagos cormorant, *Phalacrocorax harrisi* [37]), and conservation (e.g., ‘amakihi/Hawaiian honeycreeper, *Hemignathus virens* [38]; ‘alalā/Hawaiian crow, *Corvus hawaiiensis* [39]; kākāpō, *Strigops habroptilus* [40]; kakī/black stilt, *Himantopus novaezelandiae*, this study). A substantial increase in the number of reference genomes available for birds can also be attributed to the efforts of B10K [21,41], the international consortium whose goal is to produce a genome for every known species of bird. To date, B10K has published 38 de novo bird reference genomes [21]. These genomes, along with others that were available at the time of publication, make genomic resources available for at least one individual in almost every order of class Aves [42]. The next phase of B10K will include genomes representing one species from every bird family ( $n = 240$ , [42]), increasing the availability of conspecific or closely related reference genomes for conservation research.

Here, we explore the utility of closely related reference genomes for SNP discovery using a critically endangered wading bird, the kakī, as proof of concept. Once found on the North and South Islands of New Zealand, kakī experienced significant population decline throughout the 20th century due to habitat loss and degradation, and the introduction of mammalian predators. Today, there are approximately 132 kakī remaining (New Zealand Department of Conservation, *unpublished data*) and the population is contingent upon intensive management [43,44], including a captive breeding and rearing programme that uses genetic-based estimates of relatedness to pair distantly related individuals in captivity [45]. Beyond kakī, many programmes for threatened species incorporate neutral genetic measures (e.g., nucleotide diversity, individual heterozygosity or inbreeding, and relatedness) into management plans to minimise inbreeding [46] and loss of diversity [47,48] to reduce extinction risk [49,50].

To demonstrate that closely related reference genomes can yield sufficient SNPs to estimate diversity measures in threatened species, we map kakī genotyping-by-sequencing (GBS) and resequencing reads to genomes from members across the order Charadriiformes, including a conspecific reference genome (kakī, family: *Recurvirostridae*, *H. novaeseelandiae*), and members of the same genus (pied stilt, family: *Recurvirostridae*, *H. himantopus*), family (pied avocet, family: *Recurvirostridae*, *Recurvirostra avosetta*), and order (killdeer, family: *Charadriidae*, *Charadrius vociferus*) (Figure 1). Members from this comparison represent a wide evolutionary time scale: estimates based on traditional single-locus phylogenetic approaches suggest *Charadriidae* and *Recurvirostridae* diverged approximately 69 million years ago, avocets (genus: *Recurvirostra*) and stilts (genus: *Himantopus*) diverged approximately 36.9 million years ago, and kakī and pied stilt diverged approximately 1 million years ago [51,52] (but see [53]) (Figure 1). SNPs discovered from these reference-guided assemblies were then compared using estimates of diversity relevant to the conservation management of threatened species, including nucleotide diversity, individual heterozygosity, and relatedness.



**Figure 1.** Evolutionary relationships between species with reference genomes used in this proof of concept. The evolutionary tree indicates topology between taxa within the order Charadriiformes. Evolutionary tree is not to scale.

## 2. Materials and Methods

### 2.1. Tissue Sampling and DNA Extractions

Kakī blood samples were collected during routine health checks by the New Zealand Department of Conservation (DOC) at the captive breeding facilities in Twizel (DOC) and Christchurch (Isaac Conservation and Wildlife Trust), New Zealand, by approval of the DOC Animal Ethics Committee (AEC #283). These samples were stored in 95% molecular grade ethanol at  $-80\text{ }^{\circ}\text{C}$  at the University of Canterbury. Pied stilt blood samples were collected from one female and one male during routine health checks at Adelaide Zoo, with samples provided under the Royal Zoological Society of South Australia Specimen Licence Agreement (Import Permit: 2016061954). Pied stilt samples were stored in EDTA at  $-20\text{ }^{\circ}\text{C}$  at the University of Otago. The pied avocet blood sample was collected from a single individual from Hamburger Hallig, Germany, under a permit from the Ministry of Energy, Agriculture, the Environment, Nature and Digitization of the federal state of Schleswig-Holstein, Germany (Permit: V312-7224.121-37 [42-3/13]). Pied avocet samples were stored on filter paper at  $-20\text{ }^{\circ}\text{C}$  at the University of Kiel.

Genomic DNA for kakī and pied stilt reference genomes was extracted at the University of Otago using a Thermo Scientific™ MagJET™ Genomic DNA Kit (Waltham, USA) following manufacturer specifications. DNA was isolated for the pied avocet sample at the University of Kiel Institute for Clinical Molecular Biology (hereafter, IKMB) by adding 400  $\mu\text{L}$  of phosphorus buffered saline solution (PBS) to dried blood and using the Qiagen® QIAmp® DNA Blood Mini QIAcube® Kit (Hilden, Germany) following the manufacturer specifications. Genomic DNA for the kakī genotyping-by-sequencing (GBS) and resequencing approaches was extracted at the University of Canterbury using a lithium chloride chloroform extraction method (see Supplement 1 for details). Genomic DNA for all extractions were analysed for quality using a NanoDrop™ Spectrophotometer and for quantity using an Invitrogen™ Qubit™ Fluorometer.

### 2.2. Reference Genome Library Preparation and Sequencing

Paired-end libraries for the kakī and pied stilt were prepared at the University of Otago using the Illumina TruSeq® DNA PCR-free protocol according to manufacturer specifications, with genomic DNA fragmented to 350 bp. End repair and adapter ligation for sequence barcoding were carried out and libraries were indexed with unique 6 bp sequences. Sequencing of kakī and pied stilt libraries was completed by New Zealand Genomics Limited (NZGL), where sample libraries were pooled with three additional stilt samples and spread across five lanes of a flow cell for  $2 \times 125$  bp sequencing on an Illumina HiSeq 2500.

Paired-end libraries for the pied avocet were prepared using the TruSeq® DNA Nano Library Prep protocol according to manufacturer specifications, with genomic DNA fragmented to 350 bp. Library preparation and sequencing for the pied avocet was completed at IKMB using one lane of a flow cell on an Illumina HiSeq 4000 for  $2 \times 150$  bp sequencing.

### 2.3. Reference Genome Sequence Processing and Assembly

#### 2.3.1. Kakī and Australian Pied Stilt

Raw kakī and pied stilt sequence reads were evaluated for quality using FastQC v. 0.11.5 [54]. To test for exogenous contamination, the presence and abundance of non-avian reads was estimated by randomly subsampling 5000 reads from each library and searching these reads against the NCBI nucleotide database using BLAST [55].

Illumina adapters used for sequence barcoding were removed using Trimmomatic v. 0.35 [56]. Low quality bases were trimmed using ConDeTri v. 2.3 [57] with default settings. Read deduplication was carried out with ConDeTri, using the first 50 bp of both reads in a pair for comparisons. Raw reads were analysed using SGA-preqc v. 0.9.4 [58] to generate estimates of genome size and heterozygosity.

To determine the level of expected heterozygosity in the genome and assess potential signatures of contamination, paired-end reads were analysed using KmerGenie [59].

Trimmed sequences were assembled with SOAPdenovo2 [60] following initial testing of several assemblers and varying k-mer values. Draft assembly metrics were independently assessed with the assembly metrics script generated for Assemblathon [61]. BUSCO v. 3.0.1 [62,63] was used to determine completeness of the assembly outputs based on expected gene content using an avian ortholog set derived from OrthoDB v. 9 [64] and the chicken as reference. Both assembly metrics and BUSCO scores were used to determine the highest quality assemblies.

Trimmed sequence reads were used to close gaps between scaffolds in the highest quality assemblies for kakī and pied stilt with GapCloser v. 1.12 [60]. Scaffolds shorter than 5 kbp were removed, and genomes were syntenically aligned against the chicken reference genome (version 5.0, GenBank Assembly GCF\_000002315.5) using Chromosome in Satsuma v. 3.1.0 [65] to generate pseudochromosome-level assemblies by aligning the draft assembly scaffolds against the chicken genome, and retaining orthologous regions. Final drafts of kakī and pied stilt genomes are available (see Data Availability section).

### 2.3.2. Pied Avocet

Raw pied avocet sequence reads were evaluated for quality using FastQC v. 0.11.5 [54]. To remove low quality reads, paired-end data was trimmed for Illumina adapter contamination and low quality bases using Skewer v. 0.2.2 [66] with a mean Phred-score of 20, end-trim quality of 30, and a minimum length of 54 bp. Raw reads were analysed with SGA Preqc 0.10.15 [58] and KmerGenie [59] to estimate heterozygosity and potential signatures of contamination. These analyses indicated high expected heterozygosity (0.3%) compared to other birds. To eliminate highly abundant repeats and sequencing errors, a digital normalisation was conducted using Khmer 2.1.1 [67].

Pied avocet trimmed sequences were assembled using Velvet 1.2.10 [68] following initial testing with Meraculous-2D v. 2.2.5.1 [69], which failed to produce a high-quality assembly due to an overabundance of incorrectly merged diplotigs (i.e., contig pairs that share a unique k-mer at both ends [70]). To evaluate the misassemblies, a second assembly was done with Velvet using default parameters. All contigs were aligned against the assembly using LAST [71], with the -uNEAR seeding parameter. Alignments were filtered for trivial self-vs-self perfect alignments, with only single high-scoring pairs per sequence over 99% identical kept. These alignments revealed an unusual number of large and frequent indels (> 3 bp, higher than the default Velvet parameter for allowed gaps in graph bubbles) in extremely similar contigs, and therefore a final Velvet assembly was run with adjusted parameters (-ins\_length 410, -max\_branch\_length 50, -max\_divergence 0.1, -max\_gap\_count 10).

Assembled scaffolds were analysed with GapCloser v. 1.12 [60] to decrease gaps in the assembly. The gap-closed assembly was then aligned against the chicken genome using LAST [71] and the Chromosomer [72] toolkit was used to construct superscaffolds. The final draft of the pied avocet genome is available (see Data Availability section).

### 2.3.3. Killdeer

A killdeer genome was published in the ordinal phase of the B10K project [21]. To improve the assembly, a full de novo approach was used to construct a low-level base-accurate assembly. The data used in the original assembly of killdeer was downloaded from the GigaDB website [73]. This consisted of 12 libraries of Illumina sequence data, including five paired-end libraries with insert sizes ranging from 170 bp to 800 bp and seven mate-pair libraries of insert sizes ranging from 2000 bp to 20,000 bp.

FastQC v. 0.11.5 [54] was used to evaluate the quality of the Illumina data, as well as assess the contamination levels present in the samples. All paired-end libraries consisted of paired 100 bp reads, whereas mate-pair libraries were constructed of paired 50 bp reads. There was no evidence of any significant DNA contamination, but the per-base Phred-scores showed a consistently lower quality early in the reads. Due to the issues observed in the FastQC reports, reads were trimmed using Skewer

v 0.2.2 [66] to a minimum Phred-score of 30 and any read pair where at least one of the mates was trimmed to a length of < 32 bp was discarded.

Trimmed sequences were assembled using AllPaths-LG [74,75] following initial testing of several assemblers and varying k-mer values. The first run was made with the two 170 bp libraries and the complete collection of mate-pair libraries. As part of the AllPaths-LG pipeline, a set of diagnostic data was generated, including estimates of genome size, error rates, and SNP rates. Three of the mate-pair libraries were removed from subsequent analysis after low levels of utilisation were detected due to failed library construction.

The new draft assembly was aligned against the original killdeer reference genome produced by Zhang et al. [21] using the program LAST [71], which identified areas of conflict between the original and new draft killdeer genomes (e.g., short gaps, abundance of small indels, and poor resolution in heterozygous regions in the original genome). A custom set of scripts, 'SemHelpers' [76], was written to consolidate the changes detected via the genome-wide alignments into the existing reference genome. The resulting assembly has almost identical metrics when compared to the original assembly [21], given the method used. Post-correction alignments between the final assembly and the original reference genome show identities between 98 and 99%.

Quality of all final draft assemblies was assessed with the Assemblathon metrics script [61] and completeness assessed with BUSCO v. 3.0.1 [62,63] using the avian ortholog set and the chicken as reference. The final draft of the killdeer genome is available (see Data Availability section).

#### 2.4. Genotyping-by-Sequencing

Genotyping-by-sequencing (GBS), a reduced-representation genomics approach, was used to produce genome-wide SNPs for kakī. Briefly, GBS reduces genome complexity by sequencing regions that flank restriction enzyme cut sites [77]. The GBS data presented here were produced following the Elshire et al. [77] method, using 50 ng of genomic DNA with 0.72 ng of total adapters and the restriction enzyme ApeKI.

Because the kakī samples were collected during two different breeding seasons, library preparation and sequencing were completed in two separate batches. The first batch included captive parents and offspring from the 2015/2016 breeding season and other individuals sampled from 2014–2015 that represent diverse lineages based on the kakī pedigree ( $n = 52$ ; pedigree data not shown). This batch was sequenced with paired-end,  $2 \times 100$  bp reads on one lane of an Illumina HiSeq 2500 through NZGL. The second batch consisted of captive parents and offspring from the 2016/2017 breeding season plus one wild individual sampled in 2014 who represented a diverse lineage based on the pedigree ( $n = 47$ ). This batch was sequenced with paired-end,  $2 \times 150$  bp reads on one lane of an Illumina X Ten through CustomScience, Ltd. To assess the impact of batch effects (i.e., library and lane biases [78]), 10 individuals were represented in both batches to ensure similar genetic distance estimates were produced by each duplicated sample independently (see Table S1 for individual sample sequencing details).

FastQC v. 0.11.4 [54] was used to evaluate the quality of the raw Illumina data, as well as assess the contamination levels present in the samples. Paired-end reads were demultiplexed and barcodes were trimmed using Axe [79] with a maximum mismatch of 1. To minimise batch effects [78] and address sequence quality, reads from the 2016/2017 breeding season were trimmed to a maximum length of 100 bp using Skewer [66]. To remove low quality data, reads were filtered to discard short reads (< 32 bp) and reads with mean quality scores less than 30.

In order to be read by downstream pipelines, new single-end barcodes were generated for the ApeKI enzyme using the programme GBSX [80] and appended to the forward-end of reads through a custom Perl script, 'mux\_barcodes' [81]. For this study, the Tassel 5.0 [82] pipeline was used for SNP discovery and genotyping with GBS data. Due to the double-barcoding scheme of the GBS data generated here, a new class of enzymes was created specifically for Tassel 5 to add the enzyme cut site remnant, together with the reverse barcodes, as recognition sites for these datasets. The Tassel

5.0 GBSv2 pipeline was used with tag database and export plugins specifying a k-mer length of 64, a minimum k-mer length of 20, a minimum Phred-score of 30, and a minimum tag count of 10. Bowtie2 [83] was used to align tags to the each draft reference genome using the *-very-sensitive* presetting. The Tassel 5.0 GBSv2 discovery SNP caller plugin was run with a minimum minor allele frequency (-mnMAF) of 0.05 and a minimum locus coverage (-mnLCov) of 0.75. VCFtools v. 1.9 [84] was used to filter the dataset to a set of bi-allelic SNPs, with an average minimum SNP depth of 5, and 90% of all SNPs being shared amongst individuals. To minimise statistical bias of linkage disequilibrium, the data set was pruned for linkage disequilibrium using BCFtools v. 1.9 [85] with  $r^2$  set to 0.8 and a window size of 1000 sites. To ensure a more even spread of SNPs throughout the genome, VCFtools v. 1.9 [84] was used to reduce the number of SNPs to 1 SNP within 64 bp, which is the designated size of a GBS tag using Tassel 5.0. VCFs of the filtered data set are available (see Data Availability section).

In order to evaluate whether the same SNPs were likely to be mapped using different reference genomes, a custom script, 'pancompare' [86], was used to compare pairs of tags in SAM files that are unique or shared between Tassel 5.0 runs using different reference genomes. This method uses tag pair mapping as a proxy for SNP discovery, under the assumptions that tags all start at the restriction cut site and intersecting pairs of tags are likely to discover the same SNPs using different reference genomes.

### 2.5. Resequencing

In addition to a reduced representation approach, we also resequenced kakī genomes from 36 individuals for SNP discovery and genotyping. These individuals include parents and offspring from the 2015/2016 and 2016/2017 breeding seasons ( $n = 24$ ) and other individuals sampled between 2014–2017 that represent diverse lineages based on the pedigree ( $n = 12$ ). Libraries were prepared at IKMB using a TruSeq<sup>®</sup> Nano DNA Library Prep kit following the manufacturer's specifications. Libraries were sequenced across 34 lanes on a HiSeq 4000 at the IKMB.

FastQC v. 0.11.4 [54] was used to initially evaluate the quality of the raw Illumina data, as well as assess the contamination levels present in the samples. Reads were trimmed for the Illumina barcode and for a Phred-score of 20 using Trimmomatic [56]. Reads were mapped to each indexed genome using Bowtie2 [83] with the *-very-sensitive* presetting. Resulting SAM files were converted to BAM files and read coverage was analysed using mpileup with Samtools v. 1.9 [85]. To improve the computational efficiency of mpileup, a custom Perl script 'split\_bamfile\_tasks.pl' [87] was created to subdivide BAM files and run them in parallel. SNPs were detected, filtered, and reported using BCFtools v.1.9 [85]. Filtering settings included biallelic SNPs with a minor allele frequency >0.05, an average mean depth >10, and a Phred-score >20. BCFtools was used to filter for a maximum of 10% missing data per site. Resulting SNPs were pruned for linkage disequilibrium using BCFtools with  $r^2$  set to 0.8 and a window size of 1000 sites. To ensure a more even spread of SNPs throughout the genome, VCFtools v. 1.9 [84] was used to reduce the number of SNPs to 1 marker within 150 bp, which is the length of resequencing reads. VCFs of the filtered data set are available (see Data Availability section).

### 2.6. Diversity Estimates

Nucleotide diversity ( $\pi$ ) and individual heterozygosity ( $H_S$ ) were estimated using VCFtools v. 1.9 [84]. Pairwise relatedness ( $R$ ) matrices were produced using KGD [88], a programme that estimates relatedness while taking into account read depth of HTS data. Pairwise  $R$  values were scaled so that self-relatedness of all individuals was equal to 1 using the formula:

$$M_S = D \times M_O \times D$$

where  $M_S$  is the scaled matrix,  $M_O$  is the original matrix, and  $D$  is a diagonal matrix with elements:

$$D = 1/\sqrt{(\text{diag}(M_O))}$$

To compare  $H_S$  estimates generated from different reference genomes using GBS and resequencing data, analysis of variance (ANOVA) and Tukey multiple comparisons of means tests were performed using a linear mixed effects model with lme4 [89] to account for repeated measures (i.e., repeated individuals mapped to all four reference genomes). Mantel tests with 1000 iterations were used to test whether scaled pairwise  $R$  matrices using different reference genomes were significantly similar compared to a null distribution. Correlations were conducted between estimates of  $H_S$  and  $R$  (not including self-relatedness) using different reference genomes using Spearman's rank ( $r_S$ ), which accounts for the inherently non-normal distribution of the  $R$  estimates.

### 3. Results

#### 3.1. Reference Genome Sequencing and Assembly

Library sequencing produced 226–307 million paired-end sequences for each kakī, pied stilt, and avocet sample. Average sequencing depth was  $52\times$  for kakī,  $51\times$  for pied stilt, and  $70\times$  for avocet, based on an expected genome size of 1.2 Gb. Genomes produced were between 1.02–1.22 Gb in total length (Table 1), which is within the expected range for avian genomes [90]. Scaffold N50 sizes ranged from 3.66 to 105.71 Mb. The total number of scaffolds ranged from 67 to 15,167. BUSCO assessment indicated the presence of at least 82.4% of the orthologs from the avian database. Combined, these estimates indicate that the assembled genomes have high genome completeness.

**Table 1.** Genome assembly metrics for the genomes assembled in this study.

Species	Total Assembly Length (Gb)	Total Scaffolds	Scaffold N50 (bp)	Longest Scaffold (bp)	Average Scaffold Length (bp)	Complete Single-Copy BUSCOs (%)
Kakī	1.18	523	105,710,992	238,324,410	2,254,638	91.0
Pied Stilt	1.12	1443	99,457,149	221,521,436	773,955	85.9
Avocet	1.02	67	87,059,367	184,945,080	15,204,176	82.4
Killdeer	1.22	15,167	3,657,525	21,923,840	80,436	92.5

#### 3.2. SNP Discovery and Diversity Estimates—GBS

After demultiplexing and initial read filtering, kakī GBS sequencing resulted in a total of 802.4 million reads for 88 individuals (mean =  $9.1 \pm$  S.D. 4.9 million reads per individual). Five of these individuals were subsequently removed from the study after SNP filtering for having low average sample depths across sites ( $<4\times$  depth using conspecific reference genome). The resulting 82 individuals have an average depth of  $11.71\text{--}18.51\times$ , with average missingness of 2–4% depending on the reference genome used (Table 2).

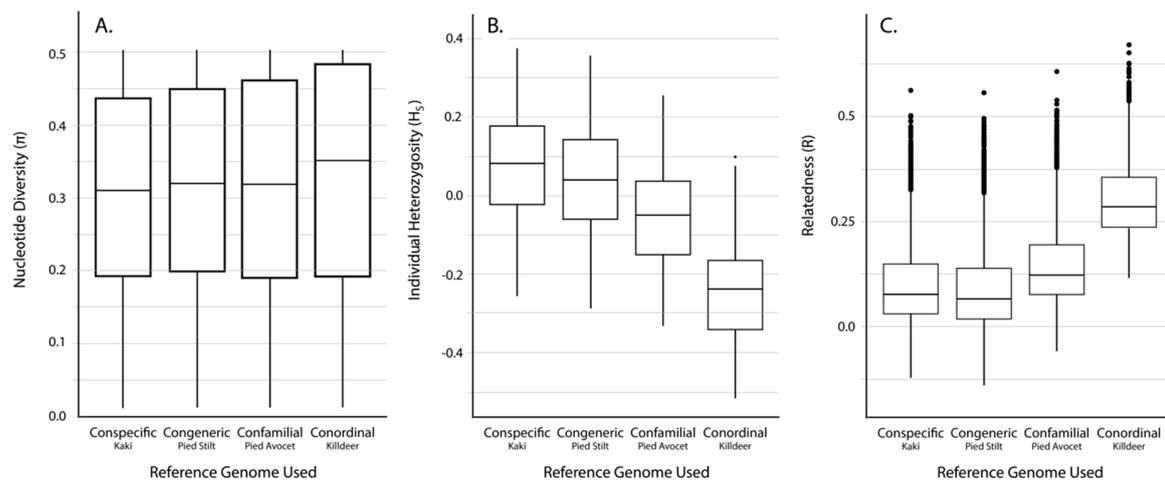
**Table 2.** Mapping statistics, single nucleotide polymorphisms (SNPs) discovered, SNP descriptive statistics, and average diversity statistics from genotyping-by-sequencing (GBS) reads mapped to different reference genomes.  $\pi$ : nucleotide diversity,  $H_S$ : individual heterozygosity,  $R$ : pairwise relatedness ( $\pm$  S.D. for each measure).

Reference Genome	No. of Mapped Tag Pairs	% Tags Shared with Kakī Mapping	No. Unfiltered SNPs	No. Filtered SNPs	Average Missingness	Average Depth	Average $\pi$	Average $H_S$	Average $R$
Kakī	392,652	100	634,695	19,396	$0.04 \pm 0.04$	$13.73 \pm 6.53$	$0.31 \pm 0.14$	$0.07 \pm 0.15$	$0.11 \pm 0.12$
Pied Stilt	372,906	91.04	604,573	18,625	$0.04 \pm 0.04$	$11.71 \pm 5.52$	$0.32 \pm 0.14$	$0.03 \pm 0.15$	$0.10 \pm 0.12$
Avocet	316,978	83.10	481,532	18,398	$0.03 \pm 0.04$	$13.90 \pm 6.58$	$0.31 \pm 0.15$	$-0.06 \pm 0.14$	$0.15 \pm 0.11$
Killdeer	151,546	72.42	242,493	10,440	$0.02 \pm 0.03$	$18.51 \pm 8.77$	$0.33 \pm 0.15$	$-0.25 \pm 0.14$	$0.30 \pm 0.09$

The number of GBS tag pairs mapped to each reference genome was greatest using a conspecific reference genome, with fewer tag pairs mapped the more phylogenetically distant the reference genome became (Table 2). Results from our analysis with 'pancompare' [86] indicate that more tags from the congeneric mapping were shared with those mapped to a conspecific reference genome

(91.04%) than more distantly related genomes (confamilial = 83.10% and conordinal = 72.42%; Table 2). Tag pairs always start at the GBS restriction enzyme cut site, making direct comparisons of tags mapped across different genomes possible. Because more mapped tags were shared between closely related genomes, these results suggest that SNPs discovered with the congeneric reference genome are more likely the same as those discovered with the conspecific reference genome than those discovered with the confamilial or conordinal references.

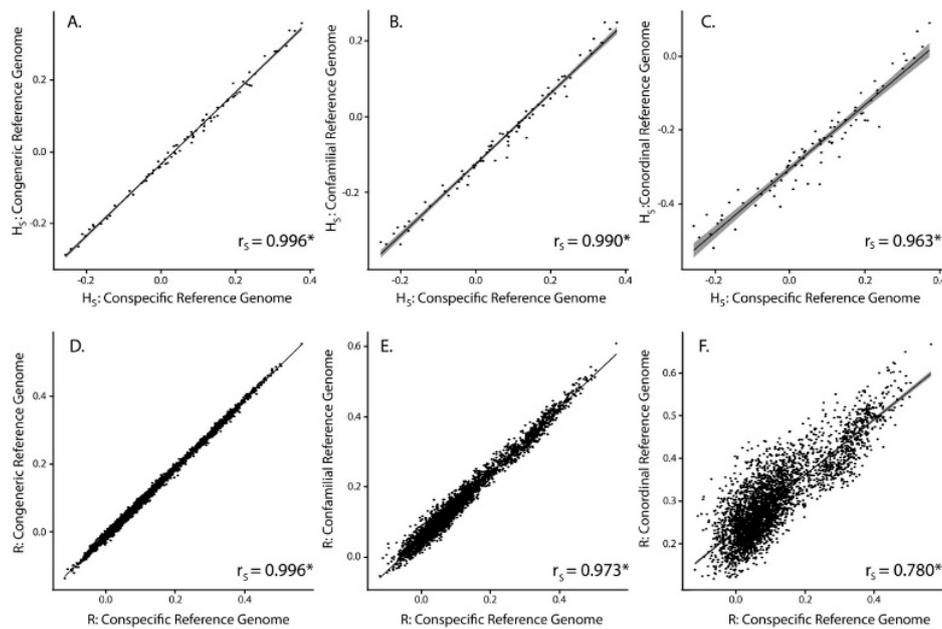
The number of unfiltered and filtered SNPs discovered was greatest when using a conspecific reference genome, with fewer SNPs discovered the more phylogenetically distant the reference genome became (Table 2). Despite the differences in number of SNPs discovered with each reference genome, average nucleotide diversity ( $\pi$ ) was similar across datasets (average  $\pi = 0.31$ – $0.33$ , Table 2, Figure 2A).



**Figure 2.** Distribution of different diversity estimates using SNPs discovered with GBS reads mapped against different reference genomes. (A) Nucleotide diversity ( $\pi$ ), (B) individual heterozygosity ( $H_S$ ), and (C) pairwise relatedness ( $R$ ) not including self-relatedness.

Average individual heterozygosity ( $H_S$ ) estimates differed depending on the reference genome used (Table 2, Figure 2B). Results show that using different reference genomes produced significantly different levels of  $H_S$  from one another (ANOVA,  $p < 0.001$ ; Tukey Contrasts,  $p < 0.001$ ). Using a congeneric reference genome resulted in  $H_S$  estimates that are on average 3.4% less than using a conspecific reference genome, with a confamilial being 12.9% less, and a conordinal being 31.6% less. Despite significant differences in  $H_S$  depending on the reference genome used, estimates of  $H_S$  using different reference genomes were significantly correlated (Spearman's correlation,  $p < 0.001$ ), with correlation coefficients between the conspecific and congeneric approaches ( $r_S = 0.996$ ) being higher than the conspecific and confamilial approaches ( $r_S = 0.990$ ) and conspecific and conordinal approaches ( $r_S = 0.963$ ; Figure 3A–C).

The range of scaled average pairwise estimates of relatedness ( $R$ ) shows a bimodal distribution, which reflects highly related individuals (siblings and parent-offspring relationships) along with more distantly related individuals that are captured in the study design. The range of scaled  $R$  values appeared different depending on the reference genome used, with average pairwise  $R$  in the conspecific and congeneric analyses being less than the confamilial and conordinal analyses (Table 2). Despite this pattern, the relationships between  $R$  using a conspecific reference genome and all other genomes were not significantly different (Mantel test,  $p < 0.001$ ). Estimates of pairwise  $R$  (not including self-relatedness) using different reference genomes were significantly correlated (Spearman's correlation,  $p < 0.001$ ), with correlation coefficients between the conspecific and congeneric approaches ( $r_S = 0.996$ ) being higher than the conspecific and confamilial approaches ( $r_S = 0.973$ ) and the conspecific and conordinal approaches ( $r_S = 0.780$ ; Figure 3D–F).



**Figure 3.** Scatterplots showing individual point estimates of  $H_5$  (A–C) and pairwise  $R$  estimates (D–F) using GBS reads mapped to different reference genomes. Self-relatedness estimates were not used in this analysis. Trend line in black, with 95% confidence intervals surrounding the trend line in gray. Spearman’s correlation coefficient ( $r_s$ ) provided in the lower right corner of each scatterplot. \* indicates significance  $p < 0.001$ .

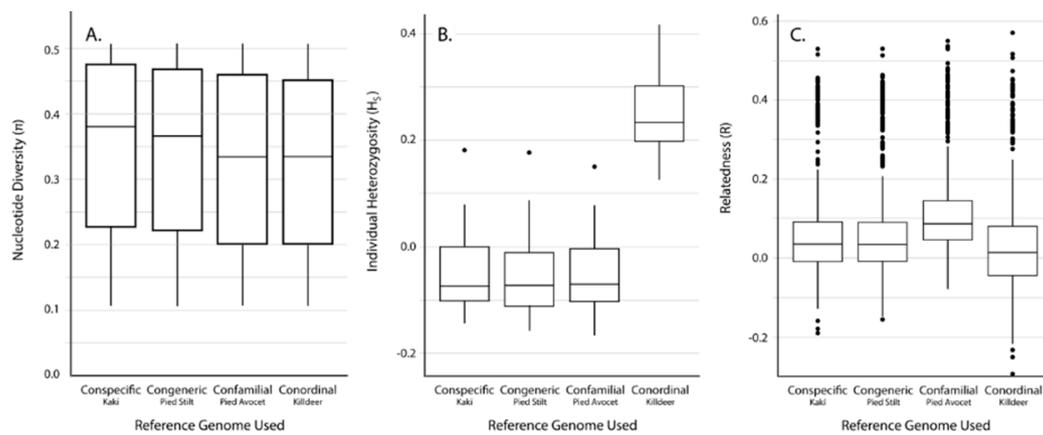
### 3.3. SNP Discovery and Diversity Estimates—Resequencing

After demultiplexing and initial read filtering, the kakī resequencing resulted in a total of 4.8 billion reads for 36 individuals (mean =  $135.8 \pm 54.1$  million reads per individual). After SNP filtering, these 36 individuals have an average depth of 13.95–17.44 $\times$  with average missingness of 0.2% across all reference genomes used (Table 3).

**Table 3.** Alignment rates, single nucleotide polymorphisms (SNPs) discovered, SNP descriptive statistics, and average diversity statistics from resequencing reads mapped to different reference genomes.  $\pi$ : nucleotide diversity,  $H_5$ : individual heterozygosity,  $R$ : pairwise relatedness. ( $\pm$  S.D. for each measure).

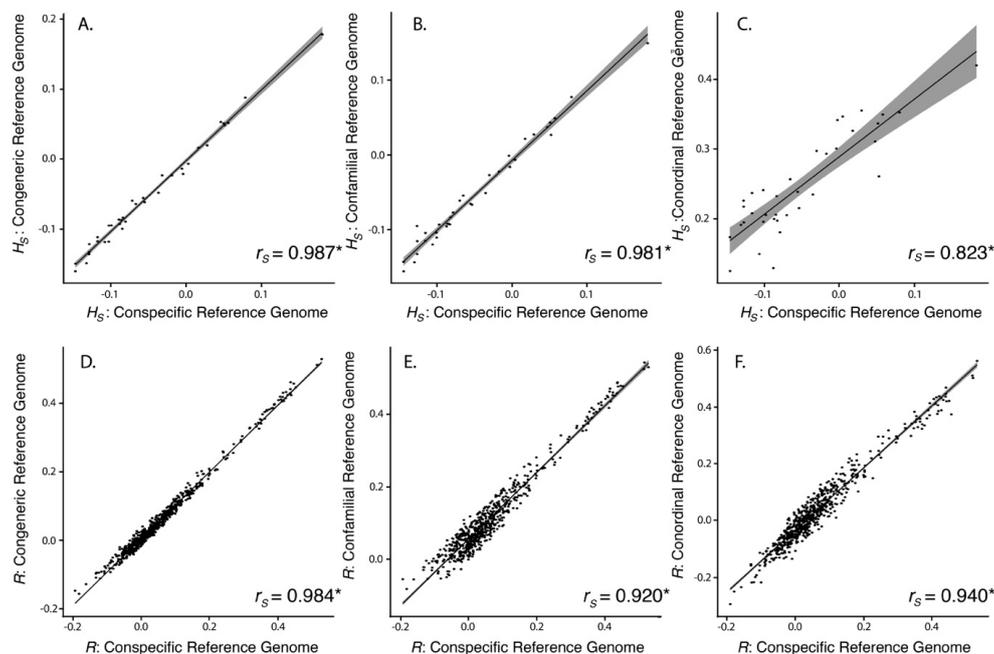
Reference Genome	Average Alignment Rate (%)	No. Unfiltered SNPs	No. Filtered SNPs	Average Missingness	Average Depth	Average $\pi$	Average $H_5$	Average $R$
Kaki	94.6 $\pm$ 0.50	4,246,100	91,854	0.002 $\pm$ 0.005	17.44 $\pm$ 6.79	0.35 $\pm$ 0.13	−0.05 $\pm$ 0.08	0.06 $\pm$ 0.11
Pied Stilt	88.1 $\pm$ 0.96	8,438,866	89,419	0.002 $\pm$ 0.005	14.99 $\pm$ 6.06	0.34 $\pm$ 0.13	−0.05 $\pm$ 0.08	0.06 $\pm$ 0.11
Avocet	78.5 $\pm$ 0.46	24,333,620	143,343	0.002 $\pm$ 0.004	16.02 $\pm$ 6.43	0.33 $\pm$ 0.14	−0.05 $\pm$ 0.07	0.11 $\pm$ 0.11
Killdeer	64.8 $\pm$ 4.89	62,888,931	89,145	0.002 $\pm$ 0.004	13.95 $\pm$ 5.54	0.32 $\pm$ 0.13	0.25 $\pm$ 0.07	0.03 $\pm$ 0.13

Average read alignment rates using Bowtie2 were highest when using a conspecific reference genome (94.6%), with fewer reads aligning with congeneric (88.1%), confamilial (78.5%), and conordinal reference genomes (64.8%, Table 3). In contrast to GBS, the number of unfiltered SNPs increased with phylogenetic distance of the reference genome, which is expected given resequencing SNPs are called by differences between reads and the reference. The number of SNPs discovered post filtering did not correspond with phylogenetic distance of the reference used, with the fewest filtered SNPs being discovered with the conordinal reference genome (89,145) and the most being discovered with the confamilial reference genome (143,343, Table 3). Similar to the GBS dataset, average  $\pi$  was similar across datasets generated using different reference genomes (average  $\pi = 0.32$ –0.35, Table 3, Figure 4A).



**Figure 4.** Distribution of different diversity estimates using SNPs discovered with resequencing reads mapped against different reference genomes. (A) Nucleotide diversity ( $\pi$ ), (B) individual heterozygosity ( $H_S$ ), and (C) pairwise relatedness ( $R$ ). Self-relatedness estimates were not used in this analysis.

Results show that using a conordinal reference genome produced significantly higher levels of  $H_S$  than the conspecific, congeneric, or confamilial approaches (ANOVA,  $p < 0.001$ ; Tukey contrasts,  $p < 0.001$ ; Table 2, Figure 4B). Using a congeneric reference genome resulted in  $H_S$  estimates that are on average 0.40% less than using a conspecific reference genome, with a confamilial being 0.31% less, and a conordinal being 29.9% greater. Despite significant differences in  $H_S$  depending on the reference genome used,  $H_S$  using different reference genomes is significantly correlated (Spearman's correlation,  $p < 0.001$ ), with correlation coefficients between the conspecific and congeneric approaches ( $r_S = 0.987$ ) being higher than congeneric and confamilial approaches ( $r_S = 0.981$ ) and congeneric and conordinal approaches ( $r_S = 0.823$ ; Figure 5A–C).



**Figure 5.** Scatterplots showing individual point estimates of  $H_S$  (A–C) and pairwise  $R$  estimates (D–F) using resequencing reads mapped to different reference genomes. Self-relatedness estimates were not used in this analysis. Trend line in black, with 95% confidence intervals surrounding the trend line in gray. Spearman's correlation coefficient ( $r_S$ ) provided in the lower right corner of each scatterplot. \* indicates significance  $p < 0.001$ .

Similar to the GBS approach, the range of scaled average pairwise estimates of relatedness ( $R$ ) based on resequencing also shows a bimodal distribution, which reflects the relationships of individuals captured in the study design. Average scaled pairwise estimates of  $R$  were similar across all reference genomes used (Table 2, Figure 4C). The relationship between  $R$  using a conspecific reference genome and all other genomes were not statistically different compared to the null distribution (Mantel test,  $p < 0.001$ ). Scaled pairwise  $R$  (not including self-relatedness) using different reference genomes is significantly correlated (Spearman's correlation,  $p < 0.001$ ), with correlation coefficients between the conspecific and congeneric approaches ( $r_S = 0.984$ ) being higher than conspecific and confamilial approaches ( $r_S = 0.920$ ) and conspecific and conordinal approaches ( $r_S = 0.940$ ; Figure 5D–F).

#### 4. Discussion

For species of conservation concern, limited conspecific genomic resources often impede inclusion of genomic data in conservation management strategies. Our proof of concept demonstrates that SNPs discovered using congeneric, confamilial, and even conordinal approaches yield diversity estimates that significantly correlate with estimates derived from SNPs discovered using a conspecific approach. Prior to this study, there was only one genome publicly available for the order Charadriiformes (i.e., the killdeer [21]). This study provides three additional high-quality de novo genome assemblies, all of which have practical applications for conservation.

The number of GBS tag pairs that aligned to each reference genome decreased the more phylogenetically distant the reference genome became. Because Tassel 5.0 calls SNPs based on differences among tag pairs [82]—as opposed to differences between tag pairs and the reference genome—the number of unfiltered SNPs discovered also decreased the more phylogenetically distant the reference genome became. The same pattern was observed for the number of filtered SNPs. The 'pancompare' analysis of GBS tag data suggests that SNP discovery using the conspecific and congeneric reference genomes are more likely to yield the same markers compared to SNPs discovered using the confamilial or conordinal reference genomes.

The number of resequencing reads that aligned to each reference genome also decreased the more phylogenetically distant the reference genome became. Unlike GBS, the number of unfiltered SNPs increased with phylogenetic distance. This is to be expected because the resequencing discovery pipeline calls SNPs based on differences between reads and the reference genome [85]. The number of SNPs discovered post-filtering was unexpected, however, as a similar number of SNPs were found in all but the confamilial reference approach, which resulted in  $\sim 1.5\times$  more SNPs than other reference-guided approaches. While the pied avocet genome shows signs of high completeness, complexities in the genome assembly due to high heterozygosity [69,91] may have resulted in less complete regions leading to higher false discovery rates [41].

Using GBS and resequencing data, the average and range of  $\pi$  estimates did not differ greatly based on reference genome used. Larger differences between reference genomes used were observed when estimating  $H_S$ . Using GBS data, mean estimates of  $H_S$  decreased significantly the more distant the reference genomes became, with the use of a conordinal reference genome producing a marked decrease in  $H_S$  estimates compared to the use of a conspecific reference. This decrease in  $H_S$  corresponds to an increase in  $R$ , although not significantly so. These combined results are consistent with expectations because SNPs called by Tassel 5.0 are based on identifying mapped tag pairs [82]; the more phylogenetically distant a reference genome is, the more conserved a region has to be to successfully map a pair of tags. Therefore, with GBS we expect  $H_S$  to be lower and  $R$  to be higher the more phylogenetically distant the reference used is, given that variants at these conserved regions are less frequent.

Using resequencing data, conspecific, congeneric, and confamilial approaches produced  $H_S$  that were not significantly different from one another, with the only significant difference seen with the conordinal approach, which resulted in a significant increase in  $H_S$  compared to other reference genomes. Unlike GBS tags, there is not an immediate explanation for this pattern. However, it may

be attributed to the fact that resequencing reads, which are longer and are more representative of the whole genome, can be mapped to more divergent regions than GBS tags.

While the range of  $H_S$  and pairwise  $R$  values may be different depending on the reference genome used, all estimates produced using different reference genomes correlate significantly with one another. Our results suggest that using a more closely related reference genome (e.g., congeneric) over a more distant reference genome (e.g., conordinal) will yield SNPs that have higher correlation coefficients with estimates generated using a conspecific, and therefore, are likely to result in similar conservation recommendations. Ongoing work incorporating genomic based estimates of relatedness into software that informs captive pairing recommendations (e.g., PMx [92]) will indicate whether more distantly related reference genomes indeed produce statistically similar pairing recommendations, as our correlation results suggest. In the meantime, we anticipate even small changes in  $H_S$  and pairwise  $R$  estimates will not greatly affect conservation recommendations, as diversity estimates are often used in relative terms. For example, pairing recommendations for intensively managed populations that lack reliable pedigrees are routinely informed by genetic- or genomic-based pairwise estimates of relatedness (e.g., [45,93–95]). In practice, pairing recommendations are made based on the relative ranking of these estimates and not the absolute values. Similarly, when investigating heterozygosity-fitness correlations (e.g., [96]), relative rankings of  $H_S$  among individuals are more informative than absolute values.

Still, there may be some instances where absolute diversity values may be of interest (e.g., parentage assignment, or management of individuals that exhibit  $H_S$  below a cutoff score [97]). SNPs derived using the conordinal reference genome provide markedly different ranges of  $H_S$  and pairwise  $R$  estimates and often the lowest correlation coefficients compared to SNPs derived from the conspecific reference genome. For birds, we recommend a confamilial reference genome as the most distant reference genome conservation researchers consider using for diversity estimates. However, this approach should be evaluated for use in other questions, such as the characterisation of adaptive variation [4,14].

The number of de novo bird genomes available to be used as reference is due to increase, especially as the next phase of B10K seeks to publish representative genomes for every recognised family of birds [42]. However, we recommend evaluating the quality of publicly available genomes prior to use, as lower quality genomes may produce lower SNP yield due to fewer alignable regions, or greater false discovery rate where there are assembly errors [98]. Here, we re-assembled the available killdeer reference genome for two reasons. First, the raw data available from the European Bioinformatics Institute European Nucleotide Archive (EBI ENA) showed poor sequencing quality and mapping of this raw data to the existing reference suggested inconsistencies where poor quality reads were more abundant. Second, mapping of the long-insert mate-pair data from the project showed little to no support for many of the scaffolding connections present in the published genome. Due to these factors, we reassembled the genome using much more stringent data curation and more cautious scaffolding. Given this, when using a genome “off the shelf”, we recommend careful assessment of the original genome publication, keeping in mind that genomes assembled from multiple libraries or data types, with greater depth of sequencing coverage, and a more complete and contiguous assembly, will be of higher quality [41]. When genomes with similar phylogenetic relationships are available, comparisons of synteny [65] and completeness [41] against the most closely related model genome may help identify which genome is most appropriate to use. Ultimately, the best way to assess existing genomic resources is to download the raw reads and evaluate them using tools such as FastQC [54] and SGA pre-QC [58], as we have done with the killdeer genome. Raw read quality may have the largest impact on final assembly quality, and initial quality checks will allow identification of any potential anomalies or limitations of the raw data that may have presented challenges to assembly, such as high heterozygosity [69,91,99]. If the raw data is of high quality, but there are inconsistencies between original reported statistics and those derived from raw reads, it may be worth investing in re-assembly to produce a genome of higher quality with greater confidence.

Indeed, re-assembly remains a more cost-effective option than starting a genome sequencing project from scratch. By our current (2018) estimates based on single libraries with paired-end reads, the use of a closely related high quality readily-available reference genome may save a conservation genomic project a minimum of EUR 6500 in library preparations, sequencing, computational power, and assembly time (Table S2, although prices subject to rapid change given new sequencing technologies). Among the 383 species in the order Charadriiformes, 51 are threatened with extinction [100]. The families Laridae (gulls, terns, and skimmers) and Scolopacidae (sandpipers) contain particularly high numbers of threatened species (14 and 13, respectively). Along with the genomes produced in this paper, there are now genomes available for four additional families within Charadriiformes (i.e., Alcidae [90], Charadriidae ([21], here), Recurvirostridae (here), and Scolopacidae [101]). Genome sequencing and assembly of one member of the Laridae family could benefit all 14 threatened species within this family, and combined with the existing genomes available as reference within Scolopacidae, could save conservation groups up to EUR 169,000 in sequencing and assembly costs. Using existing genomic resources will not only reduce these costs, but also the time needed to produce a high-quality reference genome, thereby allowing for a faster uptake of conservation genomics approaches to produce robust information for conservation management.

## 5. Conclusions

Many threatened species management programmes rely on measures of diversity, including nucleotide diversity, heterozygosity, and relatedness, in guiding management decisions [93,102]. While these measures have historically been calculated using small numbers of genetic markers, genomic markers offer the opportunity for increased resolution [1,6,103] and hence improved decision-making. Here, we have demonstrated that in the absence of a conspecific reference genome to map genomic sequence reads to, the availability of high-quality reference genome for a closely related species can provide highly correlated estimates for nucleotide diversity, individual heterozygosity, and relatedness. We anticipate the use of readily available reference genomes may provide resource-constrained conservation projects a way to minimise these costs and make a faster transition to using genomic data to improve conservation outcomes for threatened species.

**Data Availability:** The pied stilt Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession RSEF00000000. The version described in this paper is version RSEF01000000. The pied avocet genome raw reads have been deposited in Genbank under project number PRJNA508299. The reassembled killdeer genome is available at <http://www.uconsert.org/data/>. Kākī are taonga (treasured) to Māori (the indigenous people of Aotearoa New Zealand), and as such the genomes obtained from taonga species are taonga in their own right. Therefore, the genome for kākī and all VCFs for GBS and resequencing will be made available on recommendation of the iwi (tribes) that affiliate as kaitiaki (guardians) for kākī. A local genome browser is available to view the kākī genome and all VCFs presented here at <http://www.uconsert.org/data/>, along with details on how to request access.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4425/10/1/9/s1>, Table S1: Samples used in Genotyping-by-Sequencing and resequencing analyses; Table S2: Cost associated with genome sequencing and alignment; lithium chloride extraction protocol.

**Author Contributions:** All authors contributed to research conceptualization. Genomes presented here were prepared and assembled by N.J.F., R.M. and M.H., G.B.S. and resequencing data was prepared and mapped by S.J.G. with guidance from R.M. Marker discovery and subsequent diversity analyses were performed by S.J.G. Original draft preparation and writing completed by S.J.G. and N.J.F., with review and editing by all authors.

**Funding:** This research was funded by the Brian Mason Scientific & Technical Trust (SJC, TES), the Mohua Charitable Trust (TES), New Zealand Ministry of Business, Innovation and Employment Endeavour Fund (TES, AWS), and the Royal Society of New Zealand Rutherford Discovery Fellowship (MK).

**Acknowledgments:** We are grateful for the support of Te Rūnanga o Ngāi Tahu, Te Rūnanga o Arowhenua, Te Rūnanga o Waihao and Te Rūnanga o Moeraki. We thank Elshire Group Ltd. for their support in generating GBS data. We thank Sabine Müller of the Research and Technology Centre West Coast, Kiel University, for kindly providing access to the pied avocet blood sample, and the Royal Zoological Society of South Australia (Adelaide Zoo) for providing the pied stilt blood sample. Finally, we thank all members of the Kākī Recovery Programme, especially Simone Cleland and Cody Thyne, and the Isaac Conservation and Wildlife Trust, especially Anne

Richardson, for their assistance with sample collection and their ongoing support. We also thank three anonymous reviewers for their helpful comments on the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Allendorf, F.W.; Hohenlohe, P.A.; Luikart, G. Genomics and the future of conservation genetics. *Nat. Rev. Genet.* **2010**, *11*, 697. [[CrossRef](#)] [[PubMed](#)]
- Harrisson, K.A.; Pavlova, A.; Telonis-Scott, M.; Sunnucks, P. Using genomics to characterize evolutionary potential for conservation of wild populations. *Evol. Appl.* **2014**, *7*, 1008–1025. [[CrossRef](#)] [[PubMed](#)]
- Kohn, M.H.; Murphy, W.J.; Ostrander, E.A.; Wayne, R.K. Genomics and conservation genetics. *Trends Ecol. Evol.* **2006**, *21*, 629–637. [[CrossRef](#)]
- Mable, B.K. Conservation of adaptive potential and functional diversity: integrating old and new approaches. *Conserv. Genet.* **2018**, 1–12. [[CrossRef](#)]
- Luikart, G.; England, P.R.; Tallmon, D.; Jordan, S.; Taberlet, P. The power and promise of population genomics: From genotyping to genome typing. *Nat. Rev. Genet.* **2003**, *4*, 981. [[CrossRef](#)] [[PubMed](#)]
- 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. Building strong relationships between conservation genetics and primary industry leads to mutually beneficial genomic advances. *Mol. Ecol.* **2016**, *25*, 5267–5281. [[CrossRef](#)] [[PubMed](#)]
- Shafer, A.B.; Wolf, J.B.; Alves, P.C.; Bergström, L.; Bruford, M.W.; Brännström, I.; Colling, G.; Dalén, L.; De Meester, L.; Ekblom, R.; et al. Genomics and the challenging translation into conservation practice. *Trends Ecol. Evol.* **2015**, *30*, 78–87. [[CrossRef](#)] [[PubMed](#)]
- Knight, A.T.; Cowling, R.M.; Rouget, M.; Balmford, A.; Lombard, A.T.; Campbell, B.M. Knowing but not doing: Selecting priority conservation areas and the research-implementation gap. *Conserv. Biol.* **2008**, *22*, 610–617. [[CrossRef](#)] [[PubMed](#)]
- Taylor, H.R.; Dussex, N.; van Heezik, Y.J.G.E. Bridging the conservation genetics gap by identifying barriers to implementation for conservation practitioners. *Glob. Ecol. Conserv.* **2017**, *10*, 231–242. [[CrossRef](#)]
- McCormack, J.E.; Hird, S.M.; Zellmer, A.J.; Carstens, B.C.; Brumfield, R.T. Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol. Phylogenetics Evol.* **2013**, *66*, 526–538. [[CrossRef](#)]
- Hayden, E.C. The \$1,000 genome. *Nature* **2014**, *507*, 294–295. [[CrossRef](#)] [[PubMed](#)]
- Muir, P.; Li, S.; Lou, S.; Wang, D.; Spakowicz, D.J.; Salichos, L.; Zhang, J.; Weinstock, G.M.; Isaacs, F.; Rozowsky, J.; et al. The real cost of sequencing: Scaling computation to keep pace with data generation. *Genome Biol.* **2016**, *17*, 53. [[CrossRef](#)] [[PubMed](#)]
- Narum, S.R.; Buerkle, C.A.; Davey, J.W.; Miller, M.R.; Hohenlohe, P.A. Genotyping-by-sequencing in ecological and conservation genomics. *Mol. Ecol.* **2013**, *22*, 2841–2847. [[CrossRef](#)] [[PubMed](#)]
- Andrews, K.R.; Good, J.M.; Miller, M.R.; Luikart, G.; Hohenlohe, P.A. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* **2016**, *17*, 81. [[CrossRef](#)] [[PubMed](#)]
- Davey, J.W.; Hohenlohe, P.A.; Etter, P.D.; Boone, J.Q.; Catchen, J.M.; Blaxter, M.L. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* **2011**, *12*, 499. [[CrossRef](#)] [[PubMed](#)]
- Ilut, D.C.; Nydam, M.L.; Hare, M.P.J.B.R.I. Defining loci in restriction-based reduced representation genomic data from nonmodel species: Sources of bias and diagnostics for optimal clustering. *Biomed Res. Int.* **2014**, *2014*, 675158. [[CrossRef](#)] [[PubMed](#)]
- Oyler-McCance, S.J.; Oh, K.P.; Langin, K.M.; Aldridge, C.L. A field ornithologist's guide to genomics: Practical considerations for ecology and conservation. *Auk* **2016**, *133*, 626–648. [[CrossRef](#)]
- Waldron, A.; Mooers, A.O.; Miller, D.C.; Nibbelink, N.; Redding, D.; Kuhn, T.S.; Roberts, J.T.; Gittleman, J.L. Targeting global conservation funding to limit immediate biodiversity declines. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 12144–12148. [[CrossRef](#)]
- Ellegren, H. Genome sequencing and population genomics in non-model organisms. *Trends Ecol. Evol.* **2014**, *29*, 51–63. [[CrossRef](#)]
- Genome 10K Community of Scientists. Genome 10K: A proposal to obtain whole-genome sequence for 10 000 vertebrate species. *J. Hered.* **2009**, *100*, 659–674. [[CrossRef](#)]

21. Zhang, G.; Li, B.; Li, C.; Gilbert, M.T.P.; Jarvis, E.D.; Wang, J. Comparative genomic data of the Avian Phylogenomics Project. *GigaScience* **2014**, *3*, 26. [[CrossRef](#)] [[PubMed](#)]
22. Robinson, G.E.; Hackett, K.J.; Purcell-Miramontes, M.; Brown, S.J.; Evans, J.D.; Goldsmith, M.R.; Lawson, D.; Okamuro, J.; Robertson, H.M.; Schneider, D.J. Creating a buzz about insect genomes. *Science* **2011**, *331*, 1386. [[CrossRef](#)] [[PubMed](#)]
23. Matasci, N.; Hung, L.-H.; Yan, Z.; Carpenter, E.J.; Wickett, N.J.; Mirarab, S.; Nguyen, N.; Warnow, T.; Ayyampalayam, S.; Barker, M.; et al. Data access for the 1000 Plants (1KP) project. *GigaScience* **2014**, *3*, 17. [[CrossRef](#)] [[PubMed](#)]
24. Duchêne, D.A.; Bragg, J.G.; Duchêne, S.; Neaves, L.E.; Potter, S.; Moritz, C.; Johnson, R.N.; Ho, S.Y.W.; Eldridge, M.D.B. Analysis of phylogenomic tree space resolves relationships among marsupial Families. *Syst. Biol.* **2017**, *67*, 400–412. [[CrossRef](#)] [[PubMed](#)]
25. Lewin, H.A.; Robinson, G.E.; Kress, W.J.; Baker, W.J.; Coddington, J.; Crandall, K.A.; Durbin, R.; Edwards, S.V.; Forest, F.; Gilbert, M.T.P.; et al. Earth BioGenome Project: Sequencing life for the future of life. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 4325–4333. [[CrossRef](#)] [[PubMed](#)]
26. Card, D.C.; Schield, D.R.; Reyes-Velasco, J.; Fujita, M.K.; Andrew, A.L.; Oyler-McCance, S.J.; Fike, J.A.; Tomback, D.F.; Ruggiero, R.P.; Castoe, T.A. Two low coverage bird genomes and a comparison of reference-guided versus de novo genome assemblies. *PLoS ONE* **2014**, *9*, e106649. [[CrossRef](#)]
27. Lischer, H.E.; Shimizu, K.K. Reference-guided de novo assembly approach improves genome reconstruction for related species. *BMC Bioinform.* **2017**, *18*, 474. [[CrossRef](#)]
28. 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.; et al. Evolutionary genomics and conservation of the endangered Przewalski's horse. *Curr. Biol.* **2015**, *25*, 2577–2583. [[CrossRef](#)]
29. Ng, E.Y.; Garg, K.M.; Low, G.W.; Chattopadhyay, B.; Oh, R.R.; Lee, J.G.; Rheindt, F.E. Conservation genomics identifies impact of trade in a threatened songbird. *Biol. Conserv.* **2017**, *214*, 101–108. [[CrossRef](#)]
30. Nuijten, R.J.; Bosse, M.; Crooijmans, R.P.; Madsen, O.; Schaftenaar, W.; Ryder, O.A.; Groenen, M.A.; Megens, H.-J. The use of genomics in conservation management of the endangered visayan warty Pig (*Sus cebifrons*). *Int. J. Genom.* **2016**, *2016*, 5613862. [[CrossRef](#)]
31. Westbury, M.V.; Hartmann, S.; Barlow, A.; Wiesel, I.; Leo, V.; Welch, R.; Parker, D.M.; Sicks, F.; Ludwig, A.; Dalén, L.; et al. Extended and Continuous Decline in Effective Population Size Results in Low Genomic Diversity in the World's Rarest Hyena Species, the Brown Hyena. *Mol. Biol. Evol.* **2018**, *35*, 1225–1237. [[CrossRef](#)] [[PubMed](#)]
32. Organ, C.L.; Shedlock, A.M.; Meade, A.; Pagel, M.; Edwards, S.V. Origin of avian genome size and structure in non-avian dinosaurs. *Nature* **2007**, *446*, 180. [[CrossRef](#)] [[PubMed](#)]
33. Zhang, G.; Li, C.; Li, Q.; Li, B.; Larkin, D.M.; Lee, C.; Storz, J.F.; Antunes, A.; Greenwold, M.J.; Meredith, R.W.; et al. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **2014**, *346*, 1311–1320. [[CrossRef](#)] [[PubMed](#)]
34. Consortium, I.C.G.S. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **2004**, *432*, 695.
35. Dalloul, R.A.; Long, J.A.; Zimin, A.V.; Aslam, L.; Beal, K.; Blomberg, L.A.; Bouffard, P.; Burt, D.W.; Crasta, O.; Crooijmans, R.P.; et al. Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): Genome assembly and analysis. *PLoS Biol.* **2010**, *8*, e1000475. [[CrossRef](#)] [[PubMed](#)]
36. Warren, W.C.; Clayton, D.F.; Ellegren, H.; Arnold, A.P.; Hillier, L.W.; Kunstner, A.; Searle, S.; White, S.; Vilella, A.J.; Fairley, S.; et al. The genome of a songbird. *Nature* **2010**, *464*, 757. [[CrossRef](#)]
37. Burga, A.; Wang, W.; Ben-David, E.; Wolf, P.C.; Ramey, A.M.; Verdugo, C.; Lyons, K.; Parker, P.G.; Kruglyak, L. A genetic signature of the evolution of loss of flight in the Galapagos cormorant. *Science* **2017**, *356*, eaal3345. [[CrossRef](#)] [[PubMed](#)]
38. Callicrate, T.; Dikow, R.; Thomas, J.W.; Mullikin, J.C.; Jarvis, E.D.; Fleischer, R.C. Genomic resources for the endangered Hawaiian honeycreepers. *BMC Genom.* **2014**, *15*, 1098. [[CrossRef](#)]
39. Sutton, J.; Helmkampf, M.; Steiner, C.; Bellinger, M.R.; Korch, J.; Hall, R.; Baybayan, P.; Muehling, J.; Gu, J.; Kingan, S.; et al. A high-quality, long-read de novo genome assembly to aid conservation of Hawaii's last remaining crow species. *bioRxiv* **2018**, *2018*, 349035.
40. Strigops Habroptilus. 2018. Available online: [https://vgp.github.io/genomeark/Strigops\\_habroptilus](https://vgp.github.io/genomeark/Strigops_habroptilus). (accessed on 7 November 2018).

41. Peona, V.; Weissensteiner, M.H.; Suh, A. How complete are “complete” genome assemblies?—An avian perspective. *Mol. Ecol. Resour.* **2018**, *18*, 1188–1195. [[CrossRef](#)]
42. Zhang, G. Genomics: Bird sequencing project takes off. *Nature* **2015**, *522*, 34. [[CrossRef](#)] [[PubMed](#)]
43. Reed, C.E.M. *Management Plan for Captive Black Stilts*; Biodiversity Recovery Unit, Department of Conservation: Wellington, New Zealand, 1998.
44. Sanders, M.D.; Maloney, R.F. Causes of mortality at nests of ground-nesting birds in the Upper Waitaki Basin, South Island, New Zealand: A 5-year video study. *Biol. Conserv.* **2002**, *106*, 225–236. [[CrossRef](#)]
45. Hagen, E.N.; Hale, M.L.; Maloney, R.F.; Steeves, T.E. Conservation genetic management of a critically endangered New Zealand endemic bird: Minimizing inbreeding in the Black Stilt *Himantopus novaezelandiae*. *Ibis* **2011**, *153*, 556–561. [[CrossRef](#)]
46. Ford, M.J.; Parsons, K.; Ward, E.; Hempelmann, J.; Emmons, C.K.; Bradley Hanson, M.; Balcomb, K.C.; Park, L.K. Inbreeding in an endangered killer whale population. *Anim. Conserv.* **2018**. [[CrossRef](#)]
47. Jordan, S.; Giersch, J.J.; Muhlfeld, C.C.; Hotaling, S.; Fanning, L.; Tappenbeck, T.H.; Luikart, G. Loss of genetic diversity and increased subdivision in an endemic alpine stonefly threatened by climate change. *PLoS ONE* **2016**, *11*, e0157386.
48. Pacioni, C.; Hunt, H.; Allentoft, M.E.; Vaughan, T.G.; Wayne, A.F.; Baynes, A.; Haouchar, D.; Dortch, J.; Bunce, M. Genetic diversity loss in a biodiversity hotspot: Ancient DNA quantifies genetic decline and former connectivity in a critically endangered marsupial. *Mol. Ecol.* **2015**, *24*, 5813–5828. [[CrossRef](#)]
49. O’Grady, J.J.; Brook, B.W.; Reed, D.H.; Ballou, J.D.; Tonkyn, D.W.; Frankham, R. Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. *Biol. Conserv.* **2006**, *133*, 42–51. [[CrossRef](#)]
50. Spielman, D.; Brook, B.W.; Frankham, R. Most species are not driven to extinction before genetic factors impact them. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15261–15264. [[CrossRef](#)]
51. Baker, A.J.; Pereira, S.L.; Paton, T.A. Phylogenetic relationships and divergence times of Charadriiformes genera: Multigene evidence for the Cretaceous origin of at least 14 clades of shorebirds. *Biol. Lett.* **2007**, *3*, 205–210. [[CrossRef](#)]
52. Wallis, G. *Genetic Status of New Zealand Black Stilt (Himantopus novaezelandiae) and Impact of Hybridisation*; Department of Conservation Wellington: Wellington, New Zealand, 1999.
53. Jarvis, E.D.; Mirarab, S.; Aberer, A.J.; Li, B.; Houde, P.; Li, C.; Ho, S.Y.; Faircloth, B.C.; Nabholz, B.; Howard, J.T.; et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **2014**, *346*, 1320–1331. [[CrossRef](#)]
54. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. Available online: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 7 November 2018).
55. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
56. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
57. Smeds, L.; Künstner, A. ConDeTri—a content dependent read trimmer for Illumina data. *PLoS ONE* **2011**, *6*, e26314. [[CrossRef](#)] [[PubMed](#)]
58. Simpson, J.T. Exploring genome characteristics and sequence quality without a reference. *Bioinformatics* **2014**, *30*, 1228–1235. [[CrossRef](#)]
59. Chikhi, R.; Medvedev, P. Informed and automated k-mer size selection for genome assembly. *Bioinformatics* **2013**, *30*, 31–37. [[CrossRef](#)] [[PubMed](#)]
60. Luo, R.; Liu, B.; Xie, Y.; Li, Z.; Huang, W.; Yuan, J.; He, G.; Chen, Y.; Pan, Q.; Liu, Y.; et al. SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *GigaScience* **2012**, *1*, 18. [[CrossRef](#)] [[PubMed](#)]
61. Bradnam, K.R.; Fass, J.N.; Alexandrov, A.; Baranay, P.; Bechner, M.; Birol, I.; Boisvert, S.; Chapman, J.A.; Chapuis, G.; Chikhi, R.; et al. Assemblathon 2: Evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience* **2013**, *2*, 10. [[CrossRef](#)]
62. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)]

63. Waterhouse, R.M.; Seppey, M.; Simão, F.A.; Manni, M.; Ioannidis, P.; Klioutchnikov, G.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **2017**, *35*, 543–548. [[CrossRef](#)]
64. Zdobnov, E.M.; Tegenfeldt, F.; Kuznetsov, D.; Waterhouse, R.M.; Simao, F.A.; Ioannidis, P.; Seppey, M.; Loetscher, A.; Kriventseva, E.V. OrthoDB v9. 1: Cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* **2016**, *45*, D744–D749. [[CrossRef](#)]
65. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644. [[CrossRef](#)] [[PubMed](#)]
66. Jiang, H.; Lei, R.; Ding, S.-W.; Zhu, S. Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinf.* **2014**, *15*, 182. [[CrossRef](#)] [[PubMed](#)]
67. Crusoe, M.R.; Alameldin, H.F.; Awad, S.; Boucher, E.; Caldwell, A.; Cartwright, R.; Charbonneau, A.; Constantinides, B.; Edverson, G.; Fay, S.; et al. The khmer software package: Enabling efficient nucleotide sequence analysis. *F1000Research* **2015**, *4*. [[CrossRef](#)] [[PubMed](#)]
68. Zerbino, D.; Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **2008**, *18*, 821–829. [[CrossRef](#)] [[PubMed](#)]
69. Goltsman, E.; Ho, I.; Rokhsar, D. Meraculous-2D: Haplotype-sensitive Assembly of Highly Heterozygous genomes. *arXiv* **2017**, arXiv:1703.09852.
70. Chapman, J.A.; Ho, I.Y.; Goltsman, E.; Rokhsar, D.S. Meraculous2: Fast accurate short-read assembly of large polymorphic genomes. *arXiv* **2016**, arXiv:1608.01031.
71. Kielbasa, S.M.; Wan, R.; Sato, K.; Horton, P.; Frith, M. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **2011**, *21*, 487–493. [[CrossRef](#)]
72. Tamazian, G.; Dobrynin, P.; Krasheninnikova, K.; Komissarov, A.; Koepfli, K.-P.; O'brien, S.J. Chromosomer: A reference-based genome arrangement tool for producing draft chromosome sequences. *GigaScience* **2016**, *5*, 38. [[CrossRef](#)]
73. GigaDB. Genomic Data of the Killdeer (*Charadrius Vociferus*). 2014. Available online: <http://dx.doi.org/10.5524/101007> (accessed on 7 November 2018).
74. Gnerre, S.; MacCallum, I.; Przybylski, D.; Ribeiro, F.J.; Burton, J.N.; Walker, B.J.; Sharpe, T.; Hall, G.; Shea, T.P.; Sykes, S.; et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 1513–1518. [[CrossRef](#)]
75. Ribeiro, F.; Przybylski, D.; Yin, S.; Sharpe, T.; Gnerre, S.; Abouelleil, A.; Berlin, A.M.; Montmayeur, A.; Shea, T.P.; Walker, B.J.; et al. Finished bacterial genomes from shotgun sequence data. *Genome Res.* **2012**, *22*, 2270–2277. [[CrossRef](#)]
76. Moraga, R. SemHelpers [Custom Perl Script]. 2017. Available online: <https://github.com/Lanilen/SemHelpers> (accessed on 7 November 2018).
77. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **2011**, *6*, e19379. [[CrossRef](#)] [[PubMed](#)]
78. Leigh, D.; Lischer, H.; Grossen, C.; Keller, L. Batch effects in a multiyear sequencing study: False biological trends due to changes in read lengths. *Mol. Ecol. Resour.* **2018**. [[CrossRef](#)] [[PubMed](#)]
79. Murray, K.D.; Borevitz, J.O. Axe: Rapid, competitive sequence read demultiplexing using a trie. *Bioinformatics* **2017**. [[CrossRef](#)] [[PubMed](#)]
80. Herten, K.; Hestand, M.S.; Vermeesch, J.R.; Van Houdt, J.K. GBSX: A toolkit for experimental design and demultiplexing genotyping by sequencing experiments. *BMC Bioinform.* **2015**, *16*, 73. [[CrossRef](#)] [[PubMed](#)]
81. Moraga, R. Mux Barcodes [Custom Perl Script]. 2017. Available online: [https://github.com/sgalla32/mux\\_barcodes](https://github.com/sgalla32/mux_barcodes) (accessed on 7 November 2018).
82. Glaubitz, J.C.; Casstevens, T.M.; Lu, F.; Harriman, J.; Elshire, R.J.; Sun, Q.; Buckler, E.S. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* **2014**, *9*, e90346. [[CrossRef](#)] [[PubMed](#)]
83. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357. [[CrossRef](#)]
84. 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.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [[CrossRef](#)]

85. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
86. Moraga, R. Pancompare. 2018. Available online: <https://github.com/Lanilen/pancompare> (accessed on 7 November 2018).
87. Moraga, R. Split\_bamfile\_tasks.pl [Custom Perl Script]. 2018. Available online: <https://github.com/Lanilen/pancompare> (accessed on 7 November 2018).
88. Dodds, K.G.; McEwan, J.C.; Brauning, R.; Anderson, R.M.; Stijn, T.C.; Kristjánsson, T.; Clarke, S.M. Construction of relatedness matrices using genotyping-by-sequencing data. *BMC Genom.* **2015**, *16*, 1047. [[CrossRef](#)]
89. Bates, D.; Maechler, M.; Bolker, B.; Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **2015**, *67*, 1–48. [[CrossRef](#)]
90. Gregory, T.R. Animal Genome Size Database. 2001. Available online: <http://www.genomesize.com/> (accessed on 7 November 2018).
91. Tigano, A.; Sackton, T.B.; Friesen, V.L. Assembly and RNA-free annotation of highly heterozygous genomes: The case of the thick-billed murre (*Uria lomvia*). *Mol. Ecol. Res.* **2018**, *18*, 79–90. [[CrossRef](#)] [[PubMed](#)]
92. Lacy, R.C.; Ballou, J.D.; Pollak, J.P. PMx: Software package for demographic and genetic analysis and management of pedigreed populations. *Methods Ecol. Evol.* **2012**, *3*, 433–437. [[CrossRef](#)]
93. Willoughby, J.R.; Fernandez, N.B.; Lamb, M.C.; Ivy, J.A.; Lacy, R.C.; DeWoody, J.A. The impacts of inbreeding, drift and selection on genetic diversity in captive breeding populations. *Mol. Ecol.* **2015**, *24*, 98–110. [[CrossRef](#)]
94. Putnam, A.S.; Ivy, J.A. Kinship-based management strategies for captive breeding programs when pedigrees are unknown or uncertain. *J. Hered.* **2013**, *105*, 303–311. [[CrossRef](#)] [[PubMed](#)]
95. Hammerly, S.C.; de la Cerda, D.A.; Bailey, H.; Johnson, J.A. A pedigree gone bad: Increased offspring survival after using DNA-based relatedness to minimize inbreeding in a captive population. *Anim. Conserv.* **2016**, *19*, 296–303. [[CrossRef](#)]
96. Szulkin, M.; Bierne, N.; David, P. Heterozygosity-fitness correlations: A time for reappraisal. *Evolution* **2010**, *64*, 1202–1217. [[CrossRef](#)] [[PubMed](#)]
97. Sandoval-Castillo, J.; Attard, C.R.M.; Marri, S.; Brauer, C.J.; Möller, L.M.; Beheregaray, L.B. swinger: A user-friendly computer program to establish captive breeding groups that minimize relatedness without pedigree information. *Mol. Ecol. Res.* **2017**, *17*, 278–287. [[CrossRef](#)]
98. Trapnell, C.; Salzberg, S.L. How to map billions of short reads onto genomes. *Nat. Biotechnol.* **2009**, *27*, 455. [[CrossRef](#)]
99. Kajitani, R.; Toshimoto, K.; Noguchi, H.; Toyoda, A.; Ogura, Y.; Okuno, M.; Yabana, M.; Harada, M.; Nagayasu, E.; Maruyama, H.; et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* **2014**. [[CrossRef](#)]
100. IUCN. The IUCN Red List of Threatened Species. 2018. Available online: <http://www.iucnredlist.org> (accessed on 7 November 2018).
101. Küpper, C.; Stocks, M.; Risse, J.E.; dos Remedios, N.; Farrell, L.L.; McRae, S.B.; Morgan, T.C.; Karlionova, N.; Pinchuk, P.; Verkuil, Y.I.; et al. A supergene determines highly divergent male reproductive morphs in the ruff. *Nat. Genet.* **2016**, *48*, 79. [[CrossRef](#)]
102. Frankham, R. Challenges and opportunities of genetic approaches to biological conservation. *Biol. Conserv.* **2010**, *143*, 1919–1927. [[CrossRef](#)]
103. Taylor, H.R. The use and abuse of genetic marker-based estimates of relatedness and inbreeding. *Ecol. Evol.* **2015**, *5*, 3140–3150. [[CrossRef](#)] [[PubMed](#)]

